Towards a more comprehensive monitoring of wildfire spread : contributions of model evaluation and data assimilation strategies

Mélanie, Catherine Rochoux

École Centrale Paris

THÈSE

présentée par

Mélanie ROCHOUX

pour l'obtention du

GRADE de DOCTEUR de l'ÉCOLE CENTRALE PARIS

Formation doctorale	▷ Sciences Pour l'Ingénieur
Laboratoires d'accueil	▷ Laboratoire d'Énergétique Moléculaire et Macroscopique, Combustion (EM2C) du CNRS et de l'ECP
	▷ Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS)

Vers une meilleure prévision de la propagation d'incendies de forêt : évaluation de modèles et assimilation de données

Présentée et soutenue publiquement le 21 Janvier 2014

JURY

A. Trouvé	Professeur - University of Maryland	Président
E. Blayo	Professeur - Université Joseph Fourier	Rapporteur
A. Simeoni	Professeur - University of Edinburgh	Rapporteur
J.-B. Filippi	Chargé de recherche CNRS - Université de Corse	Examinateur
D. Lucor	Chargé de recherche CNRS - UPMC	Examinateur
B. Cuenot	Chercheur Senior - CERFACS	Co-encadrante
S. Ricci	Chercheur Senior - CERFACS	Co-encadrante
D. Veynante	Directeur de recherche CNRS - ECP	Directeur de Thèse

École Centrale des Arts et Manufactures
Grand Établissement sous tutelle
du Ministère de l'Éducation Nationale
Grande Voie des Vignes
92295 Châtenay-Malabry Cedex
Tél : 33 (1) 41 13 10 00
Télex : 634 991 F EC PARIS

Laboratoire d'Énergétique Moléculaire et Macroscopique, Combustion (E.M2.C.)
UPR 288, CNRS et École Centrale Paris
Tél : 33 (1) 41 13 10 31
Fax : 33 (1) 47 02 80 35

2014ECAP0009

Remerciements

Il est temps d'écrire les derniers mots sur ces pages intransigeantes mais ô combien gratifiantes et captivantes, dont il m'a fallu plusieurs mois pour venir à bout. Telle une ascension dans les montagnes, le chemin a été long, sinueux, escarpé, mais la vue du sommet fait oublier tous les efforts endurés et ne fait perdurer que les instants magiques partagés sur le chemin. Qu'ils ont été précieux les conseils distillés au détour d'une discussion pour avancer dans mes recherches, qu'ils ont été encourageants les sourires confiants et petits mots bienvaillants dans les moments intenses de travail, qu'elles ont été salvatrices les petites pauses aux quatre coins de la France et au-delà pour repartir de plus bel. Je tiens ici à exprimer ma profonde reconnaissance à toutes les personnes qui, de près ou de loin, ont marqué de leur empreinte mes trois années de thèse et en ont fait une expérience inoubliable, d'une grande richesse à la fois scientifique et humaine.

Une heure avant le lever du soleil au milieu de l'Océan,
C'est un paradis qui semble hors du temps,
Posé sur cette masse liquide infinie,
Entouré par l'horizon,
Plongé dans ces nuances de bleu,
Je n'ai pas faim, pas soif, rien à penser, ni regarder,
Et dans le vertige de cette pureté,
Loin de tout, proche de l'essentiel,
Je déroule le fil de ma pensée,
L'esprit papillonne, butine quelques idées,
Prend de la hauteur,
Il voyage.

Corentin de Chatelperron, *L'aventure de Tara Tari*.

Il y a plus de quatre ans maintenant, je me présentais au CERFACS pour débuter mon stage de fin d'études sous la direction de *Sophie Ricci* et *Bénédicte Cuenot* côté CERFACS et *Arnaud Trouvé* côté Université du Maryland (UMD), impatiente d'en découdre avec les incendies de forêt. Des géosciences avec une collaboration franco-américaine, je ne pouvais rêver meilleur cocktail ! Cette thèse s'ensuivit, dans le cadre du projet IDEA financé par l'Agence Nationale de la Recherche, à l'interface entre le laboratoire EM2C de l'Ecole Centrale Paris, les équipes GLOBC/CFD du CERFACS et l'UMD. Je remercie chaleureusement l'ensemble des institutions et leurs acteurs qui se sont mobilisés pour m'accueillir dans les meilleures conditions possibles.

Sophie, Bénédicte, vous avez su me transmettre votre enthousiasme, votre curiosité pour un sujet encore inexploré, pour l'idée simple et audacieuse d'allier assimilation de données et combustion. Pourtant aucun doute sur votre visage, l'aventure allait être belle et palpitante. Nous n'aurions pas pu alors imaginer que ce projet sur les feux allait nous mener si loin. Un grand merci pour votre confiance des premiers instants, pour la grande liberté que vous m'avez octroyée, pour votre écoute et votre disponibilité sans égales. Travailler avec vous a été une aventure passionnante et ce ne sont que les prémisses ! *Bénédicte,* je tiens à souligner la pertinence de tes conseils et ta capacité remarquable à t'intéresser à des sujets toujours plus innovants et toujours plus transversaux. Merci pour toutes les discussions animées que nous avons pu partager. *Sophie,* tu as su m'initier et donner le goût de l'assimilation de données, tu as été d'un dynamisme sans limite, tant dans la relecture de mes travaux que dans la force de tes propositions. Merci pour ton engagement et pour toutes tes attentions. A bientôt à Montréal !

Arnaud, tu as su mobiliser autour de cette belle thématique qu'est l'assimilation de données pour la combustion et tu as su donner à ma thèse cette dimension internationale si précieuse et si enrichissante. Merci pour ton accueil au Maryland auquel j'associe sans hésitation tout le *Department of Fire Protection Engineering* et en particulier *Sharon Hodgson*, *Mary Lou Holt* et *Patricia Baker*. Merci également pour ton expertise de qualité sur les feux et pour ton implication continue dans ma thèse.

Je suis très reconnaissante envers mon directeur de thèse, *Denis Veynante*, de m'avoir laissée voler de mes propres ailes dans la conduite de cette recherche multi-institutionnelle. *Denis*, ta source de savoir intarissable et ta facilité à le partager ainsi que la finesse de tes remarques sont très appréciables. J'associe naturellement à ces remerciements *Nasser Darabiha*. *Nasser*, je ne pourrais oublier nos discussions sur la pyrolyse de la végétation et comment tu m'a aidée à appréhender la modélisation de ce problème.

Cette thèse, c'est donc un quintor de scientifiques de haut vol, *Sophie, Bénédicte, Arnaud, Denis* et *Nasser*. Je mesure la grande chance dont j'ai bénéficié en travaillant dans un environnement aussi dynamique, avec cet encadrement pluriel aux personnalités affirmées, aux approches multiples et aux compétences transversales.

Chacun à votre façon et avec vos propres éclairages scientifiques, vous avez su nourrir ma soif d'apprendre, d'ouvrir de nouveaux horizons et de mener ce travail de longue haleine. Cette thèse ne serait certainement pas ce qu'elle est devenue sans cette complicité tissée au fil du temps.

•

Je tiens naturellement à remercier mes deux rapporteurs, *Albert Simeoni* et *Eric Blayo*, qui m'ont fait l'honneur de relire mon travail avec grande attention et curiosité (malgré le marathon que représentait la lecture de mon manuscrit et malgré un délai un peu court), de formuler des remarques éclairées et constructives, et par ce biais de donner de la valeur à mon travail. *Albert,* merci pour nos discussions captivantes sur les modèles de propagation d'incendies et sur les mesures disponibles. *Eric,* merci pour ta lecture avec un œil d'expert en assimilation de données et d'avoir pris le temps de percer le mystère des feux. Un grand merci aussi aux autres membres du jury, *Jean-Baptiste Filippi* et *Didier Lucor*. *Batti,* merci pour ton enthousiasme né et pour avoir mené avec brio le projet IDEA. *Didier,* merci de m'avoir aidée à décoder les polynômes du chaos, merci aussi pour ton vif intérêt pour l'assimilation de données, pour la finesse et la richesse de tes questions qui ont été sources de nombreuses réflexions.

•

Incontestablement, le projet IDEA a été une grande réussite ponctuée en beauté par la conférence à Cargèse. Un grand merci à toute l'équipe, *Jean-Baptiste Filippi* et *Frédéric Bosseur* les chefs d'orchestre, *Christine Lac*, *Valéry Masson* et *Patrick Lemoigne* la team Méso-NH, *Céline Mari* et *Susanna Strada* les expertes en chimie atmosphérique, *Vivien Mallet* le maestro en incertitudes, *Dominique Morvan* pour la partie CFD, ou encore *Paul Santoni, Yolanda Pérez* et *Virginie Tihay* pour la partie expérimentale.

•

D'autres collaborations plus informelles ont pu voir le jour et ont ainsi pleinement contribué au développement de cette recherche. Aussi je voudrais remercier chaleureusement *Ronan Paugam* et *Martin Wooster* du King's College London pour m'avoir fourni des données de feux contrôlés. Merci également à *Pauline Crombette* et *Nicolas Merlet*, partenaires des pompiers des Pyrénées-Orientales, pour toutes les discussions intéressantes que nous avons pu partager sur les données aéroportées d'incendies, sans oublier *Hervé Yésou* du SERTIT pour ses réponses éclairées sur les données satellites. Je tiens également à saluer la disponibilité de *Philippe Moireau* de l'INRIA face à mes questions sur l'assimilation de données de front. Aussi surprenant soit-il, il n'y a qu'un pas entre le suivi des incendies et celui du rythme cardiaque. Je voudrais aussi mentionner les échanges que j'ai pu avoir avec *Maëlle Nodet* et *Arthur Vidard* de l'INRIA sur les calculs de distance entre fronts. Merci en particulier à *Maëlle* pour son aide dans la publication d'une brève dans le cadre de *L'année des Mathématiques pour la planète Terre*. Et merci également aux expérimentateurs du laboratoire EM2C, *Philippe Scouflaire* et *Laurent*

Zimmer, d'avoir lancé le défi de réussir des mesures PIV sur un feu.

●

Un grand merci aux stagiaires qui ont travaillé avec passion, avec moi sur les feux, *Blaise Delmotte*, *Charlotte Eméry* et *Clément Doche*. Je ne saurais taire votre implication, votre dynamisme et votre curiosité. Merci également à *Wellington da Silva* pour le travail commun que nous avons pu mener sur les filtres particulaires.

●

Le co-financement de cette thèse m'a permis de bénéficier de multiples bureaux et environnements de travail. De nombreuses personnes sont à remercier pour leur gentillesse, leur écoute, leur professionnalisme et leurs compétences. Merci à l'ensemble des permanents, administratifs et scientifiques, doctorants et post-doctorants que j'ai pu côtoyer ces trois années.

J'ai une pensée particulière pour *Eléonore Riber*, *Florent Duchaine* et *Olivier Vermorel* pour leur aide inestimable dans la mise en place des calculs multi-physiques. *Elé,* merci d'avoir toujours eu la porte ouverte à mes questions d'outsider et pour l'énergie que tu as su m'insuffler dans la dernière ligne droite. *Flo*, tu es incontestablement le roi du couplage, merci de ton aide précieuse et de ta patience face à toutes mes questions. *Olivier,* tu as toujours su me donner la réponse adéquate, alors que je débarquais de manière impromptue à ton bureau. Merci pour tes réponses précises et infaillibles sur le code AVBP, alors que j'étais noyée au milieu des conditions aux limites. J'associe également à ces remerciements *Laurent Selle* pour son regard extérieur toujours vif et percutant sur la combustion, *Matthieu Boileau* pour sa disponibilité et son aide technique lors de ma soutenance, ainsi que *Benedetta Franzelli* et *Olivier Cabrit*, anciens doctorants au CERFACS, qui ont été d'une grande réactivité face à mes questions sur la chimie ou encore les parois débitantes.

Un grand merci aussi à *Roberto Paoli* et *Sébastien Massart* d'avoir participé au projet d'assimilation de données et pour leurs multiples éclairages scientifiques. J'en profite également pour remercier *Anthony Weaver* pour sa relecture assidue de mon chapitre d'assimilation de données.

Merci aux équipes couplage et informatique au CERFACS, sans qui tout ce travail n'aurait pas été envisageable.

Merci aussi à *Sophie Valcke*, *Laurent Terray* et *Christophe Cassou* qui ne sont pas étrangers à l'obtention de mon post-doctorat canadien.

Je ne pourrais jamais assez remercier tous mes collègues, amis, rencontrés durant ces trois années, en particulier *Adrien Lemal* et *Carolyn Jacobs*, *Alessandro Scarpato*, *Aurélien Guy*, *Benedetta Franzelli*, *Benjamin Detraz* (et *Delphine*), *Clément Mirat* (et *Naïg*), *Da Xu*, *Diane Rusterholtz*, *Fabien Tholin*, *François Doisneau* et *Layal Hakim*, *François Péchereau*, *Jean-François Bourgouin* (et *Aude*), *Jean Caudal*, *Jean Lamouroux*, *Laurent Soucasse*, *Marien Simeni Simeni*, *Raphaël*

Baudoin, *Thibault Guiberti*, *Vincent Leroy*, *Yuxiang Ni*, *Elizabeth Harader*, *Elodie Jaumouillé*, *Elodie Fernandez*, *Emanuele Emili*, *Jean-Christophe Giret*, *Lola Corre*, *Margot Bador*, *Raphaël Mari*, *Yohan Ruprich* et tant d'autres. Merci pour les petites soirées sympathiques au bord de la Seine ou de la Garonne, pour les soirées crêpes, pour les joggings au Parc de Sceaux, pour les plongeons à la piscine, pour l'été à Palo Alto, pour toutes les discussions enflammées que nous avons pu partager...

•

Un petit clin d'oeil à mes amis éparpillés aux quatre coins du monde, qui m'ont fait sortir la tête de l'eau aux moments opportuns, que ce soit *Gabriel* et *Thérèse* (objectif Miami), *Frédéric* (pour ta capacité à relever les défis les plus fous, tu m'as bluffée avec l'Annapurna), *Joris* (pour ton énergie sans égale et ta culture illimitée), *Bérengère* et *Domi* (pour vos appels inopinés et vos éclats de rire), *Gauthier* (bravo Docteur !), *Eve* (objectif Singapour), *Emily* et *Damien* (pour les soirées de détente toulousaines), *Anaïs* et *Sean* (objectif Australie), *Guillaume* et *Claire* (pour les films et débats à l'Utopia), *Marine* et *Blaise* (pour le meilleur banofee cake ever), *Cédric* (Mister satellite), *Florian* (avec un petit clin d'oeil au projet sur les glaces), *Martin* (le plus sportif des sportifs, à quand le triathlon ?), *Cecilie* et *Aymeric* (objectif Norvège), *Mathieu* (pour les mois passés à McGill et la rando dans les Adirondacks), *Marjin* (pour ton message du Népal qui m'a fait rêver), *Bertrand* (pour tes traversées de l'Amérique à l'appel du voyage), *Saurya* (pour m'avoir montré une nouvelle image de l'Inde), *Robert*, *Mirka* et *Renate* (mes amis allemands de longue date), ou encore la team du *Festival de Cannes*, *Jessica*, *Pierre*, *Alexandre*, *Nicolas* et *Névénick*. Merci pour votre soutien, la Belle Province vous attend !

•

Mes mercis les plus forts vont naturellement à ma famille, être la 27ème a été un réel cadeau car j'ai pu apprendre de chacun de vous et continuer de grandir. J'ai une pensée émue pour mes deux grands-mères, qui auraient été fières et à qui je dédie cette thèse. Un merci incommensurable à mes parents que j'adore, pour leur bienvaillance, leur écoute et leur soutien infaillible, à *Alex* aussi (sans oublier *Mélanie*) qui depuis toujours m'accompagne avec cette affection bienvaillante de grand frère et est ma source d'inspiration, mon modèle de rider et mon conseiller comm'. Vous m'avez donné des ailes pour m'envoler au gré des vents, suivre mes idées et mes envies de voyages.

Un grand merci aussi à *Frédérique* et *Hervé* de m'avoir fait une si belle place dans leur famille. Et enfin, comment remercier suffisamment celui qui illumine ma vie par son sourire et qui a veillé sur moi pendant ces trois ans, alors que lui-même faisait face au marathon de la thèse ? *Flo*, tout a commencé avec un café de l'aventure et pour mon plus grand bonheur, le caravansérail continue sa route tel Marco Polo... en mettant le cap sur le Canada !

Bonne lecture !
Mélanie Rochoux, 6 avril 2014

Résumé

La prévision des incendies de forêt reste un défi puisque vitesse et direction de propagation dépendent des interactions multi-échelles entre la végétation, la topographie du terrain et les conditions météorologiques. Un modèle d'incendies de forêt à l'échelle régionale peut donc difficilement prendre en compte le détail des processus physiques mis en jeu. Toute modélisation est entachée de nombreuses incertitudes (modélisation incomplète, méconnaissance du terrain, de la végétation et des interactions flamme/atmosphère, etc.) qu'il est nécessaire de quantifier et de corriger afin de mieux comprendre la dynamique des incendies et de mieux prévoir leur progression en temps réel. Ces travaux de thèse proposent ainsi une modélisation régionale des incendies qui a des meilleures capacités de simulation et prévision, basée sur une évaluation des modèles et l'assimilation de données.

L'évaluation de modèles a consisté à développer des simulations multi-physiques détaillées à l'échelle de la flamme d'un feu de laboratoire afin de mieux comprendre les mécanismes physiques sous-jacents. Ces simulations multi-physiques impliquent la résolution des équations de Navier-Stokes réactives, l'évaluation du transfert radiatif vers la végétation, la construction d'un modèle de pyrolyse de la végétation ainsi que la formulation d'une interface flamme/végétation adéquate. La seconde approche a consisté à développer un prototype d'assimilation de données pour le suivi de la propagation du front de feu. L'idée est de rectifier la trajectoire simulée du front au fur et à mesure que de nouvelles observations sont mises à disposition, la différence entre les positions observées et simulées du front étant traduite en une correction des paramètres du modèle de vitesse de propagation ou directement de la position du front via le filtre de Kalman d'ensemble.

Ces approches, tenant compte des incertitudes à la fois sur la modélisation des incendies et sur les observations disponibles, permettent ainsi d'améliorer la prévision de la dynamique des feux ainsi que des émissions atmosphériques, ce qui constitue un enjeu de taille pour la protection civile et environnementale.[1]

[1] Publication dans le cadre de l'initiative *Un jour, une brève* de l'année des mathématiques pour la planète Terre, http://mpt2013.fr/lutter-contre-les-incendies-depuis-les-airs/.

Abstract

Because wildfires feature complex multi-physics occurring at multiple scales, our ability to accurately simulate their behavior at large regional scales remains limited. The mathematical models proposed to simulate wildfire spread are currently limited because of their inability to cover the entire range of relevant scales (i.e., from biomass pyrolysis to atmospheric dynamics), and also because of knowledge gaps and/or inaccuracies in the description of the physics as well as knowledge gaps and/or inaccuracies in the description of the controlling input parameters (i.e., the vegetation, topographical and meteorological properties). For this purpose, the uncertainty in regional-scale wildfire spread modeling must be quantified and reduced. In this context, the goals of this thesis are two-fold, model evaluation and data assimilation.

First, multi-physics detailed simulations of fire propagation, solving for the flame structure using Navier-Stokes equations for multi-species reacting flow and including radiation heat transfer, biomass pyrolysis as well as a flame/vegetation interface, were performed at the laboratory flame scale. These simulations were compared to measurements to provide a comprehensive understanding of the mechanisms underlying fire propagation and to examine the assumptions used to estimate fire spread-rates at regional scales. Second, the use of a data-driven simulator that sequentially integrates remote sensing (typically infrared imaging) measurements and that relies on an empirical spread-rate model was explored for regional-scale fire front tracking. The idea underlying this data assimilation strategy was to translate the differences in the observed and simulated fire front locations into a correction of the input parameters of the empirical spread-rate model or directly of the fire front location through the ensemble Kalman filter algorithm.

Since these two approaches account for uncertainties in fire spread modeling and measurements, they improve our general ability to forecast both wildfire dynamics and plume emissions. These challenges have been identified as a valuable research objective with direct applications in fire emergency response for civil defense and environmental protection.

Towards a more comprehensive monitoring of wildfire spread

Contributions of model evaluation and data assimilation strategies

Mélanie ROCHOUX

Contents

Nomenclature

Abbreviations

AMO	Analysis Minus Observation
AMT	Analysis Minus True
ASIR	Auxiliary Sequential Importance Resampling
BE	Burning Efficiency
BLUE	Best Linear Unbiased Estimator
CF	Combustion Factor
CFD	Computational Fluid Dynamics
CFL	Courant-Friedrichs-Lewy
CGS	Centimetre-Gram-Second
DOM	Discrete Ordinate Method
EF	Emission Factor
EFFIS	European Forest Fire Information System
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
FMO	Forecast Minus Observation
FMT	Forecast Minus True
FRP	Fire Radiative Power
FWI	Fire Weather Index
GRI-Mech	Gas Research Institute-Mechanism
HPC	High-Performance Computing
KF	Kalman Filter
LES	Large-Eddy Simulation
LSODE	Livermore Solver for Ordinary Differential Equations
MC	Monte Carlo
MEnKF	Morphing Ensemble Kalman Filter
MIR	Mid-InfraRed
MODIS	MODerate resolution Imaging Spectroradiometer
MPI	Message Passing Interface
NIR	Near-InfraRed

NSCBC	Navier-Stockes Characteristics Boundary Conditions
ODE	Ordinary Differential Equations
OI	Optimal Interpolation
OSSE	Observation System Simulation Experiments
PC	Polynomial Chaos
PDF	Probability Density Function
PEA	Pre-Exponential Adjustment
PF	Particle Filter
PIV	Particle Image Velocimetry
ROS	Rate Of Spread
RTE	Radiation Transfer Equation
SIR	Sequential Importance Resampling
SIS	Sequential Importance Sampling
STD	STandard Deviation
TFLES	Thickened-Flame model for Large Eddy Simulation
TGA	ThermoGravimetric Analysis
TIR	Thermal-InfraRed
TVD	Total Variation Diminishing
UAV	Unmanned Aerial Vehicles
UQ	Uncertainty Quantification

Roman characters

▷ Physical quantities

c	Reaction progress variable	[-]
c_p	Specific heat capacity at constant pressure	[J/kg]
$c_{p,v}$	Counterpart of c_p for biomass fuel	[J/kg]
D_k	Diffusivity of species k	[m^2/s]
D_{th}	Heat diffusivity	[m^2/s]
$E_{a,r}$	Activation energy of reaction r	[J/mol]
$\mathcal{E}$	Efficiency flame factor	[-]
$\mathcal{F}$	Thickening flame factor	[-]
F_i	Component i of the body force	[N/m^3]
$\mathbf{g}$	Acceleration induced by gravity	[m/s^2]
g_i	i-th component of the gravitational acceleration $\mathbf{g}$	[m/s^2]
h_k	Mass enthalpy of species k	[J/kg]
$h_{s,k}$	Mass sensible enthalpy of species k	[J/kg]
h_t	Total mass enthalpy	[J/kg]
I_{fr}	Fireline intensity	[W/m]
$I_{\nu_{rad}}$	Spectral radiance	[W/m^2/sr/Hz]
$\mathcal{J}_{j,k}$	Component j of the molecular diffusive flux of species k	[kg/m^2/s]

k_r	Pre-exponential factor of reaction r	[r-dependent]
K_r^{f}	Forward reaction constant for reaction r	[r-dependent]
K_r^{b}	Backward reaction constant for reaction r	[r-dependent]
m	Mass sample	[kg]
$m_v^{''}$	Biomass fuel loading	[kg/m^2]
$\dot{m}_v$	Total biomass fuel mass loss rate	[kg/m^3/s]
$\dot{m}_{v,char}$	Contribution of char oxidation to $\dot{m}_v$	[kg/m^3/s]
$\dot{m}_{v,pyr}$	Contribution of pyrolysis to $\dot{m}_v$	[kg/m^3/s]
$\dot{m}_{v,vap}$	Contribution of drying to $\dot{m}_v$	[kg/m^3/s]
M_v	Biomass fuel moisture content	[%]
$M_{v,ext}$	Biomass fuel moisture content at extinction	[%]
n_i^k	Number of atoms i in the k-th gas species	[-]
n_r	Temperature-exponent of reaction r	[-]
p	Pressure	[Pa]
p_k	Partial pressure of species k	[Pa]
p_∞	Atmospheric pressure	[Pa]
$\dot{q}$	Heat source term	[W/m^3]
q_j	Component j of energy flux	[J/m^2/s]
$\dot{q}_r$	Progress rate of reaction r	[mol/m^3/s]
$\dot{q}_{rad,fl}$	Flame-induced radiation in biomass fuel	[J/m^3/s]
$\dot{q}_{rad,l}$	Radiation heat losses in biomass fuel	[J/m^3/s]
$\dot{q}_v$	Heat exchanges in the biomass fuel	[J/m^3/s]
$\dot{q}_{v,chem}$	Contribution of chemical reactions to $\dot{q}_v$	[J/m^3/s]
$\dot{q}_{v,conv}$	Contribution of convection heat transfer to $\dot{q}_v$	[J/m^3/s]
$\dot{q}_{v,rad}$	Contribution of radiation heat transfer to $\dot{q}_v$	[J/m^3/s]
$\dot{q}_w$	Wall heat flux	[W/m^2]
R_g	Perfect gas constant	[J/mol/K]
s	Mass stoichiometric ratio	[-]
s_{mol}	Molar stoichiometric ratio	[-]
s_C	Consumption speed	[m/s]
s_L	Laminar flame speed	[m/s]
s_T	Turbulent flame speed	[m/s]
$\dot{s}_w$	Wall-normal fuel mass loss rate	[kg/m^2/s]
$\mathcal{S}_k$	Name of species k	[-]
t	Time	[s]
T	Gas mixture temperature	[K]
T_∞	Ambient temperature	[K]
T_g	Gas-phase temperature in biomass fuel	[K]
T_v	Solid-phase temperature in biomass fuel	[K]
T_{char}	Activation temperature of char oxidation	[K]
T_{pyr}	Activation temperature of pyrolysis	[K]
T_{vap}	Activation temperature of drying	[K]
T_w	Wall gas mixture temperature	[K]

$\mathbf{u}$	Flow velocity vector	[m/s]
u_i	Component i of the flow velocity vector	[m/s]
$\overline{\mathbf{u}}$	Mean flow velocity	[m/s]
$\mathbf{u}^{'}$	Root mean square of the flow velocity	[m/s]
$\mathbf{u}_w$	Wind velocity vector	[m/s]
u_w	Wind velocity magnitude	[m/s]
V	Gas mixture volume	[m^3]
$v_{inj,w}$	Wall-normal injection velocity	[m/s]
$V_{c,j}$	Correction velocity in direction j	[m/s]
$V_{k,j}$	Diffusion velocity in direction j for species k	[m/s]
$V_{k,n,w}$	Wall-normal diffusion velocity for species k	[m/s]
(x, y)	Two-dimensional spatial coordinates	[m]
X_k	Molar fraction of species k	[-]
$X_{w,k}$	Wall molar fraction of species k	[-]
$[X_k]$	Molar concentration of species k	[kg/m^3]
Y_k	Mass fraction of species k	[-]
$Y_{w,k}$	Wall mass fraction of species k	[-]
$Y_{air,k}$	Mass fraction of species k in air	[-]
$Y_{v,k}$	Mass fraction of species k in biomass fuel	[-]
$Y_{v,char}$	Mass fraction of biomass char material content	[-]
$Y_{v,dry}$	Mass fraction of biomass dry material content	[-]
$Y_{v,wat}$	Mass fraction of biomass moisture content	[-]
W	Mean molecular weight of the gas mixture	[kg/mol]
W_k	Atomic weight of species k	[kg/mol]
W_g	Mean molecular weight of the gas-phase in biomass fuel	[kg/mol]
Z_i	Mass fraction of atom i	[-]
z	Mixture fraction	[-]
z_{st}	Mixture fraction at stoichiometry	[-]

▷ Mathematical operators/quantities

$\mathbf{d}$	Innovation/residual vector
$\mathbf{e}$	Forecast or analysis errors
$\mathbb{E}$	Expectation operator
$\mathcal{F}$	Dynamic model of the state variables
$\mathcal{G}$	Generalized observation operator (possibly non-linear)
G	Linear generalized observation operator
$\mathbf{G}$	Tangent-linear of the observation operator
$\mathcal{G}_{pc}$	Surrogate model of the observation operator
$\mathcal{G}_{\Delta}$	LES filtering operator
$\mathcal{H}$	Observation selection operator (possibly non-linear)
$\mathbf{H}$	Tangent-linear of the observation selection operator
$\mathbf{I}$	Identity matrix

$\mathcal{J}$	Cost function
$\mathbf{K}$	Kalman gain matrix
$\mathcal{M}$	Dynamic model (possibly non-linear)
M	Linear dynamic model
$\mathbf{M}$	Tangent-linear of the dynamic model
n	Size of the control vector
n_s	Number of model state variables
$n_{\boldsymbol{\theta}}$	Number of model parameters
$\mathbf{n}$	Normal vector
$\mathbf{n}_b$	Normal vector to the computational domain boundaries
N_e	Number of members in Monte-Carlo sampling
N_{fr}	Number of markers along simulated front
N_{fr}^{o}	Number of markers along observed front
N_g	Number of gas compounds in the gas mixture
N_{pc}	Number of terms in the polynomial chaos expansion
N_{quad}	Number of quadrature roots
N_r	Number of reactions in the gas mixture
p	Size of the observation operator
p	Probability density function
$\mathcal{P}$	Dynamic model of the parameters
$\mathbf{P}$	Forecast/analysis error covariance matrix
Q_{po}	Maximum order of the polynomial basis
$\mathbf{R}$	Observation error covariance matrix
$\mathbb{R}$	Real numbers
$\mathbf{x}$	Control vector
$\hat{\mathbf{x}}_q$	q-th component of the polynomial chaos expansion
$\mathbf{y}^{\mathrm{o}}$	Observation vector
$\mathbf{y}$	Model counterparts of the observation vector
$\hat{\mathbf{y}}_q$	q-th component of polynomial chaos expansion

▷ Dimensionless numbers

Gr	Grashof number
Le_k	Lewis number of species k
Re	Reynolds number
Ri	Richardson number
Sc_k	Schmidt number of species k

Greek characters

▷ Physical quantities

α_{fr}	Fire propagation direction angle	[°]
α_{sl}	Terrain slope angle	[°]

α_w	Wind direction angle	[°]
β_{air}	Air dilution coefficient	[-]
β_v	Fuel packing ratio	[-]
Γ	Rate of spread (ROS)	[m/s]
δ_L	Thermal flame thickness	[m]
δ_R	Reaction zone thickness	[m]
δ_v	Biomass fuel layer depth	[m]
Δ	Filter size	[m]
Δt	Time step	[s]
Δt^A	Time step related to AVBP	[s]
Δt^{PY}	Time step related to PYROWO	[s]
Δh_c	Heat of combustion	[J/kg]
$\Delta h_{c,ef}$	Effective heat of combustion	[J/kg]
Δh_{char}	Heat of char oxidation	[J/kg]
$\Delta h^0_{f,k}$	Mass formation enthalpy of species k	[J/kg]
Δh_{pyr}	Heat of pyrolysis	[J/kg]
Δh_{vap}	Latent heat of evaporation	[J/kg]
ζ_{fr}	Fire emissivity	[-]
λ	Heat diffusion coefficient	[J/m/K/s]
μ	Dynamic flow viscosity	[kg/m/s]
ν	Kinematic flow viscosity	[m^2/s]
ν_{rad}	Wavenumber	[1/m]
$\nu_{r,k}$	Net molar stoichiometric coefficient of species k for reaction r	[-]
$\nu^{'}_{r,k}$	Molar stoichiometric coefficient of species k for forward reaction r	[-]
$\nu^{''}_{r,k}$	Molar stoichiometric coefficient of species k for backward reaction r	[-]
ρ	Gas mixture mass density	[kg/m^3]
ρ_k	Mass density of species k	[kg/m^3]
ρ_g	Mass density of the gas-phase in biomass fuel	[kg/m^3]
ρ_p	Biomass fuel particle mass density	[kg/m^3]
ρ_v	Biomass fuel bulk mass density	[kg/m^3]
σ_{sb}	Stefan-Boltzmann constant	[J/s/m^2/K^4]
Σ_v	Biomass fuel particle surface/volume	[1/m]
$\tau_{i,j}$	Component (i, j) of the stress tensor	[N/m^2]
Φ	Fuel/oxidizer equivalence ratio	[-]
χ	Scalar dissipation rate	[1/s]
χ_{char}	Fraction of pyrolysis rate due to char formation	[-]
χ_{conv}	Convection contribution to the heat release rate	[-]
χ_{ef}	Effective burning coefficient	[-]
χ_{O_2}	Stoichiometric coefficient related to char oxidation	[-]
χ_{rad}	Radiation fraction of the heat release rate	[-]

$\dot{\omega}_k$	Mass reaction rate of species k	[kg/m^3/s]
$\dot{\omega}_T$	Heat release rate due to combustion	[J/m^3/s]

▷ **Mathematical operators/quantities**

$\boldsymbol{\epsilon}$	Observation errors
$\boldsymbol{\varepsilon}$	Dynamic-model errors
ζ	Random event
$\boldsymbol{\theta}$	Input model parameters
σ^{a}	Analysis error standard deviation
σ^{f}	Forecast error standard deviation
σ^{o}	Observation error standard deviation
φ_q	q-th polynomial chaos basis function
φ_q^{1D}	q-th one-dimensional polynomial chaos basis function
$\omega^{(j)}$	Weight of the j-th quadrature root
$\omega^{(k)}$	Weight of the k-th sample particle
Ω	Computational domain
$\partial\Omega$	Boundary condition of the computational domain

Superscripts

0	Initial/reference	(k)	k-th ensemble member
a	Analysis	o	Observation
b	Background	t	True
e	Ensemble	T	Adjoint
f	Forecast		

Subscripts

0	Initial/reference	ign	Ignition
∞	Ambient conditions	r	Chemical reaction
b	Burnt	sl	Slope
f	Fresh	st	Stoichiometry
F	Fuel	t	Time step
fr	Front	v	Vegetation/biomass fuel
k	Gas compound	w	Wind
i	Front marker index		
O	Oxidizer		

Chapter 1

An eye onto wildfire spread modeling and monitoring

Wildfires, also referred to as wildland, forest or bush fires, constitute a global issue, affecting almost all climates, tropical belts as much as boreal ecosystems, as highlighted in Fig. 1.1. Real-time predictions of the behavior of a propagating wildfire have been identified by civil defense and forest agencies as a valuable research objective with direction applications in fire emergency management (Noonan-Wright et al., 2011). This chapter provides an introduction to the current operational monitoring of wildfire spread as well as to the academic research viewpoint. New mathematical modeling approaches and remote sensing technological breakthroughs are briefly described. In this context, we propose our novel strategies for increasing fire spread simulation reliability and delivering accurate forecasts, as these are highly needed but not yet available for wildfire applications.

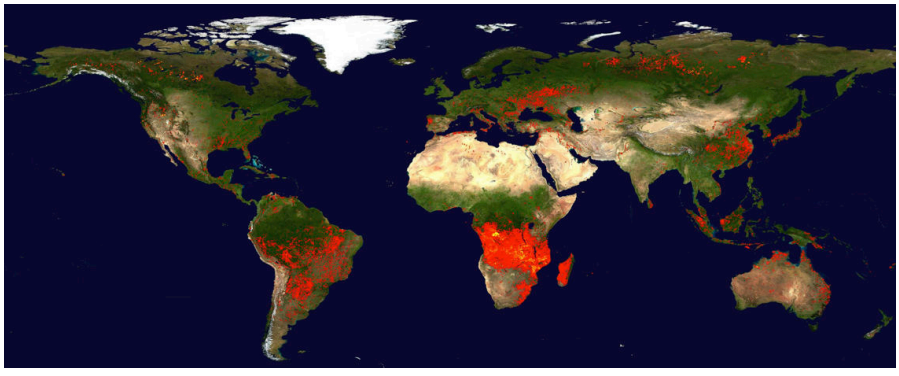


Figure 1.1: *Global active fire map delivered by the MODIS instrument (aboard the Terra and Aqua orbiting-satellites), representing accumulated locations of fires as colored dots over a 10-day period (9-18 August 2013). Credit: NASA, earthdata.nasa.gov/firms.*

Outline

1.1 Wildfires, worldwide natural disasters

During summer time, in the northern and southern hemispheres, there is a high fire danger severity for forest, grassland and savanna regions once vegetation reaches a high-level of drought. Based on satellite emissions of burned areas, Tansey et al. (2008) estimated that 350 to 440 million hectares were burned every year on average over the period 2000-2007, representing nearly 3.4 % of the Earth total vegetation area. As illustrated in Fig. 1.1, fire maps produced by MODIS[1] (Giglio et al., 2003; Davies et al., 2004) deliver global hot-spot locations (these hot-spots are identified as fire locations) over ten-day periods, and confirm this annual trend.

The northern shores of the Mediterranean region do not account for a large proportion of burnt areas at a global scale (only 0.1 %). However, they are still subject to an intense fire activity with the largest peak in summer, experienced in 2007 with the Greek forest fires. On average, nearly half a million hectares of wildland and forest areas burn in the European Union (EU) every year. Large fires (*large* referring to a final burnt area above 50 hectares) account for 75 % of this total burnt area, while they only account for 2.6 % of the total number of fires. Note that a burnt area of 50 hectares constitutes a major threat in Mediterranean ecosystems due to the presence of multiple wildland-urban interfaces. Figure 1.2 presents the EU fire annual statistics provided by the European Forest Institute (2009), based on the EFFIS[2] database. This figure shows in particular the spatial patterns

[1] MODerate resolution Imaging Spectroradiometer, modis-fire.umd.edu/index.html.

[2] European Forest Fire Information System, established in 2000 by the Joint Research Centre (JRC) and the Environment Directorate-General of the European Commission (EC), to support EU services in charge of forest protection against fires, forest.jrc.ec.europa.eu/effis/.

of Mediterranean areas subject to recurrent wildfires, mainly located in Portugal, Spain, Corsica, Sardinia, South of Italy and Greece.

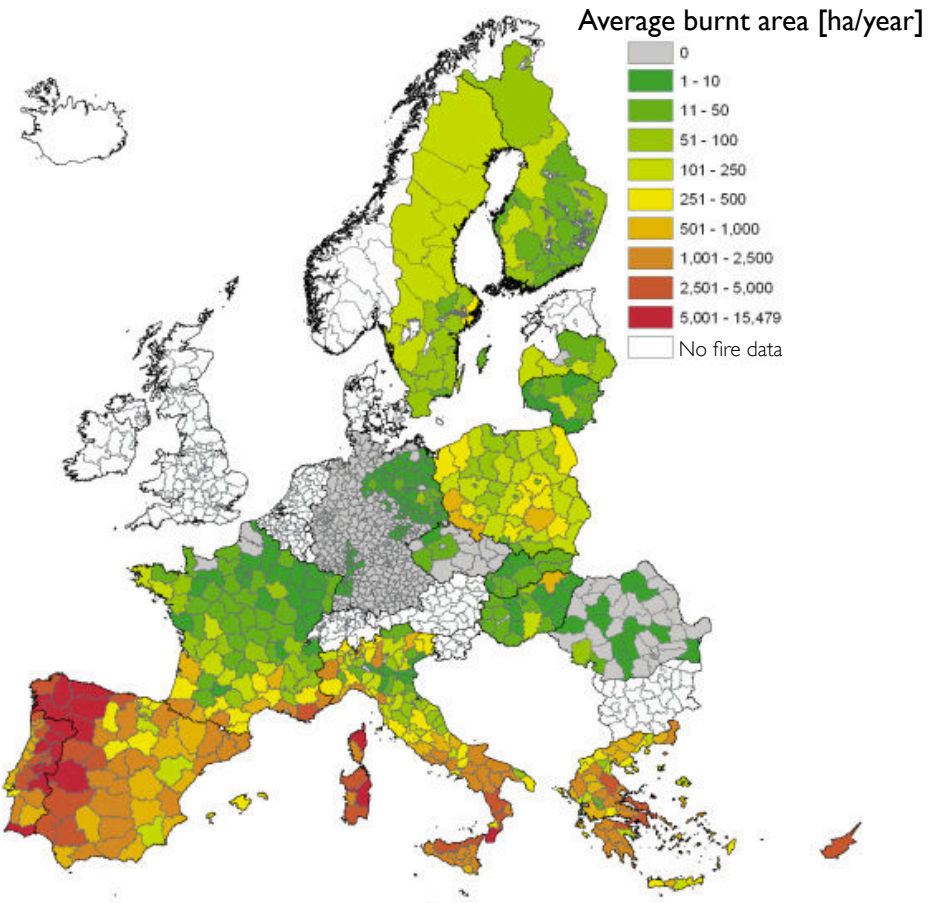


Figure 1.2: *Map of EU annual burnt areas. Credit: European Forest Institute (2009).*

Because wildfire behavior features complex multi-physical processes occurring at multiple scales (i.e., from local characteristics of vegetation to meso-scale atmospheric dynamics), which can also be affected by socio-economic factors (e.g., growing wildland-urban interfaces), wildfire risks can drastically change from one region to another. Wildfires can be particularly dangerous in canyons and valleys, where winds can strongly intensify wildfire spread (e.g., Santa Ana winds in California, Kochanski et al., 2013), and where terrain topography can significantly impact the directions in which wildfires propagate at the fastest rates.

As illustrated in Fig. 1.3, wildfires are also responsible for the release of significant amounts of CO_2 (equivalent to 25 % of the global annual industrial emissions, van der Werf et al., 2010), trace gases (e.g., CO, nitrogen oxides) and aerosols into the atmosphere, which can, locally, inhibit vegetation growth, reduce air quality and, globally, contribute to the greenhouse effect if vegetation does not recover (Miranda

et al., 2008; Strada et al., 2012). While fire intensity constitutes a local major threat, its effects can be felt hundreds of kilometers away due to the atmospheric convection of the smoke plume, as shown in Fig. 1.4 for the 1997 Indonesian wildfires that were mainly peat fires (Kunii et al., 2002).

Figure 1.3: *Snapshot of garrigue wildfires induced by tramontane winds (Pyrénées-Orientales, France, 9 August 2011). Credit: Pauline Crombette (CNES).*

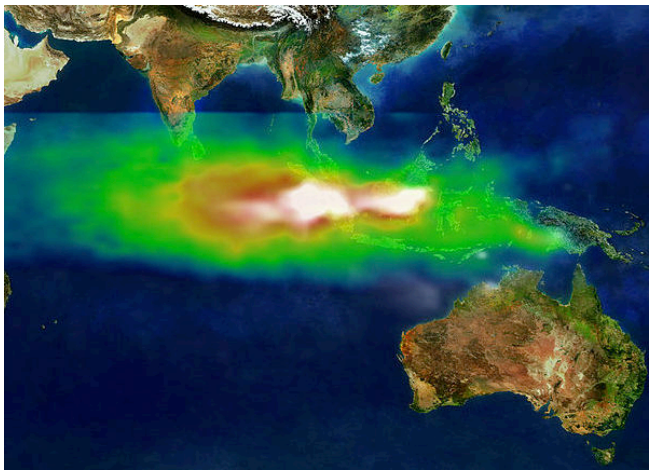


Figure 1.4: *Observation of the smoke plume over the Indian ocean due to 1997 Indonesian wildfires (22 October 1997). White colors represent aerosols in the vicinity of wildfires; green, yellow and red colors represent increasing amounts of tropospheric ozone (smoky fog) convected westward by high-altitude winds. Credit: NASA.*

Wildfires have recently shown the potential to reach colossal dimensions of highly-destructive power, beyond any currently-existing suppression capacity. These wildfires are commonly referred to as *megafires* or *firestorms* (Finney and Mcallister, 2011; Nijhuis, 2012). Typical examples are the Black Saturday wildfires in Australia in 2009 (Teague et al., 2010). These catastrophic bushfires were ignited under

extreme weather conditions, with daytime temperatures exceeding 45°C, strong changing surface winds and extremely dry conditions. They resulted in the burning of more than 450,000 hectares (i.e., about one-third of the annual burned area in Australia, reached only with 400 individual fires), along with the highest number of casualties on record (i.e., 173 people died, 414 were injured). Recent studies (Cruz et al., 2012; Engel et al., 2013) showed that meso-scale atmospheric features were partly responsible for the extreme behavior of these wildfires. In particular, a strong cooling-off in the late afternoon introduced variability in wind, temperature and humidity conditions at short temporal and spatial scales. These changes led, in turn, to high local variations in fire danger that were not predictable.[3] These extreme wildfire behaviors highlight the substantial interaction and feedback between a wildfire and the atmosphere (partly induced by the development of a thermal plume above active wildfire areas as shown in Fig. 1.5), which can, locally, modify environmental conditions and dramatically enhance wildfire spread. In this context, it is of primary importance to investigate physical and chemical processes underlying wildfires, which can be a threat not only to local ecosystems but also to public health.

Figure 1.5: *MODIS imaging of the 2009 Black Saturday bushfires (Victoria, Australia, 7 February 2009). Credit: NASA.*

Future perspectives of climate change and global warming tend to favor extreme drought events and alter precipitations (Milly et al., 2002; Palmer and Räisänen, 2002; Boé et al., 2009). These conditions dramatically increase the risk for the ignition and development of megafires (Lucas et al., 2007; Liu et al., 2010; Nijhuis, 2012). For instance, IPCC[4] expects 25 % rise by 2020 and 70 % rise by 2050 in the likelihood of megafires in Australia (Lucas et al., 2007). In this context, predictions of future extreme wildfire events cannot only rely on the analysis of past

[3] Following the carastrophic Black Saturday bushfires, an additional level of fire danger was incorporated in the fire danger rating system by the Australian Bureau of Meteorology.

[4] Intergovernmental Panel on Climate Change, i.e., leading international institution for climate change assessment, www.ipcc.ch/.

observed wildfire events. They inevitably have to rely on complex computer models to characterize future fire behaviors and intensities that are not observed to date. As computational capacity increases and more observations become available, these model simulation capabilities could be extensively evaluated and their uncertainties could be quantified for current wildfire behavior in a preliminary step. Similarly to climate change predictions, scenarios of future wildfire behavior could then be investigated using selected well-established model simulators. Real-time predictions of the direction and speed of a propagating wildfire have therefore been identified as a valuable research objective with direct applications in both fire risk management and fire emergency response for short- and long-term outlooks (Noonan-Wright et al., 2011).

1.2 Current status of potential fire danger evaluation

Wildfires generally feature a front-like geometry (see Fig. 1.6) and may be described at regional scales (i.e., at scales ranging from a few tens of meters up to several kilometers) as a thin flame zone that self-propagates (normal to itself) into unburnt vegetation. The local propagation speed is commonly referred to as the *rate of spread* (ROS) and is defined as the speed of the flame with respect to a fixed observer. Thus, the ROS can be regarded as the translation rate of the flame-ignition surface separating the burning zone and the unburnt vegetation; it directly results from multi-scale multi-physical interactions between vegetation, combustion and flow dynamics as well as atmospheric dynamics.

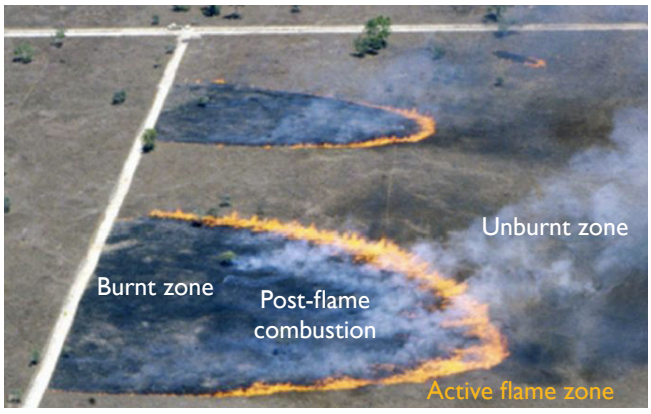


Figure 1.6: *Snapshot of Australian grassland controlled burns (200 m × 200 m domain), in which the orange zone corresponds to the flame zone. Credit: Cheney et al. (1993).*

For early warning of potential wildfire danger, operational systems have been designed worldwide by national civil defense authorities to identify geographical areas that are subject to possible extreme wildfire behavior in the next following days. *Fire danger* is a generic term referring to the assessment of both constant and variable fire precursor factors affecting the ignition, spread, intensity and suppression

capability of wildfires (Chandler et al., 1983). Current operational fire danger rating systems adopt a regional-scale viewpoint based on empirical and semi-empirical ROS modeling approaches and integrate remote sensing data (i.e., meteorological, terrain topography and biomass fuel information) into a reduced set of macroscopic qualitative and/or numerical indices. Note that due to the complexity of wildfire spread, a single index is not sufficient to provide a complete prediction of daily fire danger. An overview of wildfire danger indices follows.

- ▷ **Forest fire danger index (FFDI).** The McArthur forest and grassland fire danger indices (FFDI/GFDI), originally developed by McArthur (1966), is currently used by the Australian Bureau of Meteorology as a fire danger forecasting tool (Noble et al., 1980; Dowdy et al., 2009). While combining a record of drought, rainfall and weather variables such as wind magnitude, air temperature and humidity, the evaluation of the ROS is performed based on the grassland fire spread meter (GSFM), i.e., an operational system managed by the CSIRO[5] Australian national science agency.
- ▷ **National fire danger rating system (NFDRS).** Operational since 1972 in the United States (US), the national fire danger rating system (NFDRS) combines daily measurements of vegetation (e.g., living and dead fuel moistures[6]), terrain topography, weather and risk of ignition (i.e., human-caused and lightning) to provide local indices of fire occurrence or behavior (Deeming et al., 1978; Bradshaw et al., 1984; Burgan, 1988). In particular, the Keetch-Byram drought index represents the effects of evapo-transpiration and precipitation processes on soil moisture. Thus, it is useful to quantify drought in the deep soil/duff layer over seasonal times scales (Keetch and Byram, 1968). The NFDRS-based products are part of the WFAS[7] managed by the US Forest Service.
- ▷ **Fire weather index (FWI).** The Canadian counterpart of the NFDRS-based fire danger index is the fire weather index (FWI). The latter is part of CWFIS,[8] i.e., a service managed by the Canadian forest service that creates daily maps of fire weather and fire behavior across Canada (Turner and Lawson, 1978; Van Wagner, 1987; Hirsch, 1996). The FWI, presented in Fig. 1.7, includes the following five components:

  (i) Fine fuel moisture code (FFMC), which evaluates the moisture content of litter and other fine fuels at the top of the surface vegetation layer (this moisture content measures the ignition capacity of fine fuels, which are of primary importance in wildfire spread and ignition).

[5] Commonwealth Scientific and Industrial Research Organization, www.csiro.au/Outcomes/Safeguarding-Australia/GrassFireSpreadMeter.aspx.

[6] The description of living biomass fuels can be improved through the satellite-based normalized difference vegetation index (NDVI); this index provides with a spatial resolution on the order of 1 km, the departures from average greenness based on historical records (Burgan et al., 1996).

[7] Wildland Fire Assessment System, www.wfas.net/.

[8] Canadian Wildland Fire Information System, cwfis.cfs.nrcan.gc.ca/en_CA/.

(ii) Duff moisture code (DMC), which evaluates the moisture content of compacted organic layers at moderate depth and thereby, measures the potential fuel consumption in duff layers.

(iii) Drought code (DC), which evaluates the moisture content of deep compact organic layers and thereby, measures seasonal drought effects on wildland fuels.

(iv) Initial spread index (ISI), which evaluates the potential value for the wildfire ROS by combining the effects of wind and FFMC (without accounting for the influence of the wildland fuel spatial variability).

(v) Build-up index (BUI), which evaluates the total amount of wildland fuel available for combustion, by combining information coming from DMC and DC fuel moisture codes.

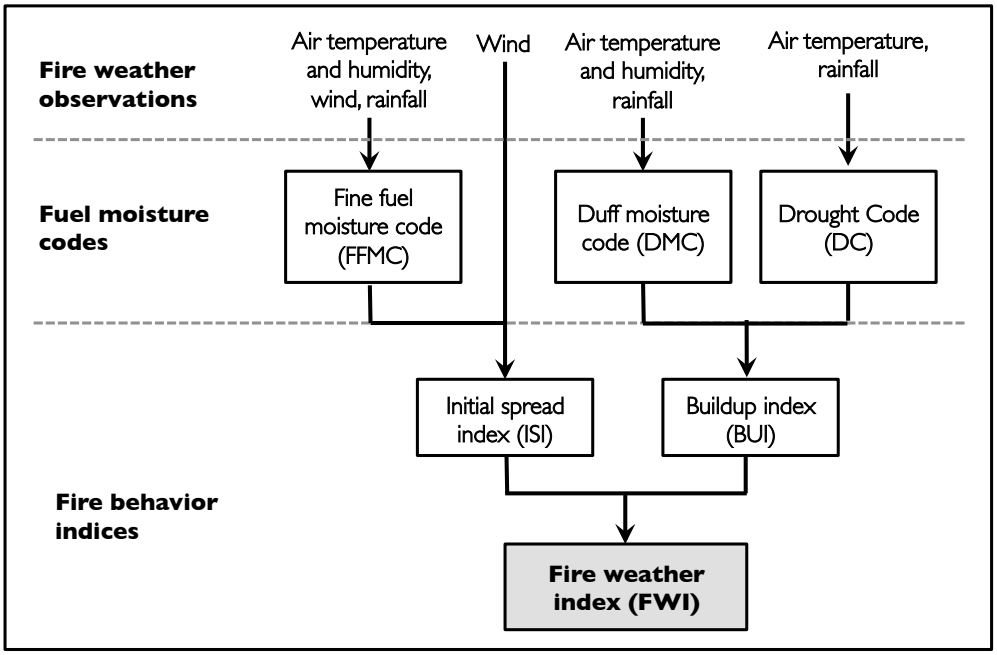


Figure 1.7: *Schematic of the components (e.g., remote sensing measurements, models) underlying the FWI-based fire danger rating. Credit: Canadian Forest Service, http://cwfis.cfs.nrcan.gc.ca/background/summary/ffws.*

Thus, the FWI-based fire danger rating provides an evaluation of the potential intensity of propagating fire fronts, based on daily measurements of air temperature and relative humidity, wind speed as well as 24-hour rainfall. Potential wildfire ROS values are included in the FWI evaluation through the ISI index. This FWI component is an empirical algebraic formulation of the ROS with respect to wind and fuel moisture content (Van Wagner, 1987; Camia and Bovio, 2000; Dowdy et al., 2009).

The performance of the Canadian FWI fire danger rating system has been demonstrated for a wide range of biomass fuels that cover forests and rural geographical areas, in particular throughout Europe (Viegas et al., 1999). Thus, this rating system has been adopted in the EU to produce twice daily, maps of fire danger using weather forecast data from French (Météo-France) and German (Deutsche Wetter Dienst) meteorological services. The FWI-based European system is based on six classes of fire danger (i.e., from very low to extreme), the highest class limit being based on the analysis of past fire danger conditions observed for 2,000 wildfires of more than 500 hectares. An example of EFFIS-based daily fire danger forecast map is provided in Fig. 1.8.

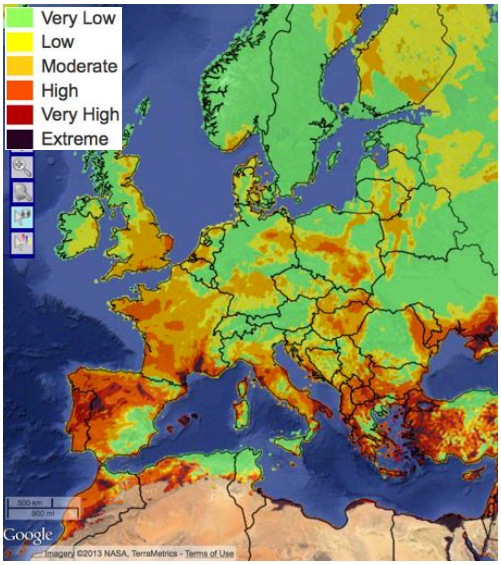


Figure 1.8: *Map of FWI-based fire danger forecast over Europe on September, 1st, 2013 (the fire suppression capacity limit is commonly assumed to be between the high and very-high classes, corresponding to a ROS threshold of about 1 km/h). Credit: EC Joint Research Centre,* `ec. europa. eu/ dgs/ jrc/ index. cfm`.

Note that the European system accounts for the effects of both biomass moisture and wind on fire behavior, but does not differentiate biomass fuel types. Thus, this fire danger rating mainly relies on meteorological information. In general, the evaluation of fire danger could be improved through a more detailed wildfire spread modeling and a more accurate forecast of the potential ROS, accounting for the spatial and temporal variability of environmental conditions, especially of biomass fuels and also of biomass fuel conditions for extreme meteorological events.

1.3 Overview of wildfire modeling research

The dynamics of wildfires are determined by multi-scale interactions between biomass dynamics and pyrolysis, combustion and flow dynamics as well as atmospheric dy-

namics and chemistry. As illustrated in Fig. 1.9, these interactions occur at: vegetation scales that characterize biomass fuels; flame scales that characterize combustion and heat transfer processes; topographical scales that characterize terrain and vegetation boundary layer; and meteorological micro-/meso-scales that characterize atmospheric conditions (Viegas, 2011). The magnitude of the wildfire ROS directly results from the interactions between the multi-physical processes over a wide range of temporal and spatial scales. These interactions make the modeling and prediction of wildfire behavior a challenging task.

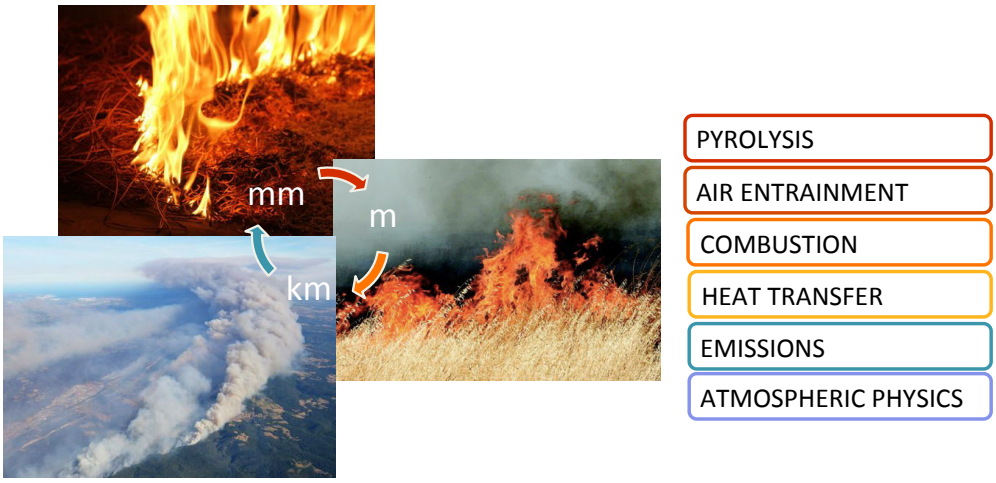


Figure 1.9: *Schematic of multi-scale multi-physical processes underlying wildfires.*

1.3.1 Physical features governing wildfire spread

To highlight important features governing wildfires, a qualitative description of the different modes of propagation is provided. An overview of the coupled physico-chemical processes that determine the rate and direction of wildfire spread is also presented.

↪ Modes of fire propagation

A fire requires an external heat source to start (e.g., human-induced ignition, thunderstorm lightning). However, once being ignited, the fire can self-sustain through a series of chemical reactions between oxygen and flammable gases that are released by the pre-heated vegetation (e.g., CH_4, CO, H_2). As suggested by Fons (1946), wildfire propagation can be regarded as a succession of ignitions inducing the displacement of the pyrolysis zone (and thereby of the flame zone) towards the unburnt region. Viegas (1998) distinguishes two main modes of fire propagation illustrated in Fig. 1.10, namely flaming and smoldering combustion. A brief description of these combustion modes follows. It is worth noting that this work exclusively focuses on the flaming mode (typical of surface fires), which is the main mode of wildfire spread and also the most studied one.

Figure 1.10: *Snapshot of undergrowth burning separating two modes of biomass combustion: (1) the flaming mode in grass (the flame is 10 mm tall); and (2) the smoldering mode in organic soils. Credit: Ashton et al. (2007).*

▷ **Flaming combustion** can occur when a large amount of pyrolysis gases is in contact with air, at high temperatures. In this case, combustion-related processes produce a flame, the visible part of the fire illustrated in Fig. 1.10, either for surface fires, crown fires or spot fires.

(i) **Surface fires** consume fine particles that are part of the vegetation at the ground surface (e.g., forest litter, herbaceous vegetation, shrubs), through the propagation of a flaming front towards unburnt vegetation areas. These particles mainly correspond to dead vegetation materials and are therefore characterized by a low moisture content (generally, in equilibrium with air humidity).

(ii) **Crown fires.** Surface fires propagate horizontally along the terrain surface, but can also spread vertically up to the top of the canopy and tree crowns. The resulting crown fires burn trees that drastically enhance the heat release rate and the size of the fire.

(iii) **Spot fires.** Crowning enhances the production of embers (i.e., glowing hot particles made of carbon-based materials), which are unexpectedly blown away by wind (beyond conventional heat transfer distances). These projected embers potentially initiate spot fires far ahead of the crown flaming front and thereby, drastically enhance fire spread.

▷ **Smoldering combustion (ground fires)** is a slow mode of combustion occurring through surface and sub-surface organic layers of the forest ground, at low temperatures and usually without any flame (Ashton et al., 2007; Hadden et al., 2013). Thus, smoldering is commonly opposed to the flaming mode of combus-

tion. The characteristic temperature and heat release of smoldering are significantly lower (i.e., a peak temperature of about 800 K, compared to 1200 K for the flaming mode), and the ROS within the fuel layer is at least reduced by a factor 10 compared to flaming combustion. Duff and peat lands (i.e., partially-decayed vegetation) are prone to sustain this mode of fire spread, as experienced in the 1997 Indonesian wildfires and 2010 Russian wildfires.[9] Smoldering is also the mode of combustion occurring behind the flaming front of wildfires (also referred to as *post-flame combustion*). This is in particular responsible for severe physical, chemical and biological soil damages as well as for large amounts of emissions into the atmosphere (Page et al., 2002; Rein et al., 2008).

↪ Fundamentals of fire dynamics

The combustion-related processes involved in wildfire spread are complex, due to the heterogeneous properties of vegetation (also referred to as *biomass fuel* or *wildland fuel*) and to the multiple underlying physico-chemical processes (e.g., chemical reactions in the gas and solid phases of the vegetation layer, radiation and convection heat transfer, buoyancy-driven flow).

In industrial applications, combustion systems are carefully controlled so that they involve a relatively limited set of fuels and the mass fraction of all these compounds is precisely controlled (except for impurities). However, this description does not apply to wildfires. Wildland fuels consist primarily of wood, grasses, shrubs, savannas, forests, etc., in different states depending on their age and level of biological decomposition. They cover therefore a wide range of physical structures and chemical compounds. Besides, these biomass components can absorb humidity from the atmosphere, a process that drastically modifies their physical and chemical properties over time. Thus, the temporal variability of vegetation is difficult to track in real-time, while significantly affecting the rate and direction of wildfire spread.

As shown in Fig. 1.11, the mechanism responsible for degrading biomass surface fuels and sustaining wildfire spread (once being ignited) can be decomposed into four main stages described in the following (Williams, 1982).

(1) **Flame-induced convection and radiation heat transfer.** The combustion zone, where combustion kinetic reactions occur, releases a large amount of heat through convection and radiation. In particular, the vegetation ahead of the combustion zone (in the pre-heated zone) receives a significant external heat flux from the flame and therefore, its temperature increases. The magnitude of this external heat flux decreases with distance from the flame.

(2) **Moisture evaporation.** The moisture contained in the porous vegetation of the pre-heated zone subsequently evaporates. This moisture evaporation is the primary stage of the vegetation thermal degradation, which breaks the chemical bonds within the porous organic material and modifies its composition.

[9] www.nasa.gov/topics/earth/features/asia-fire.html.

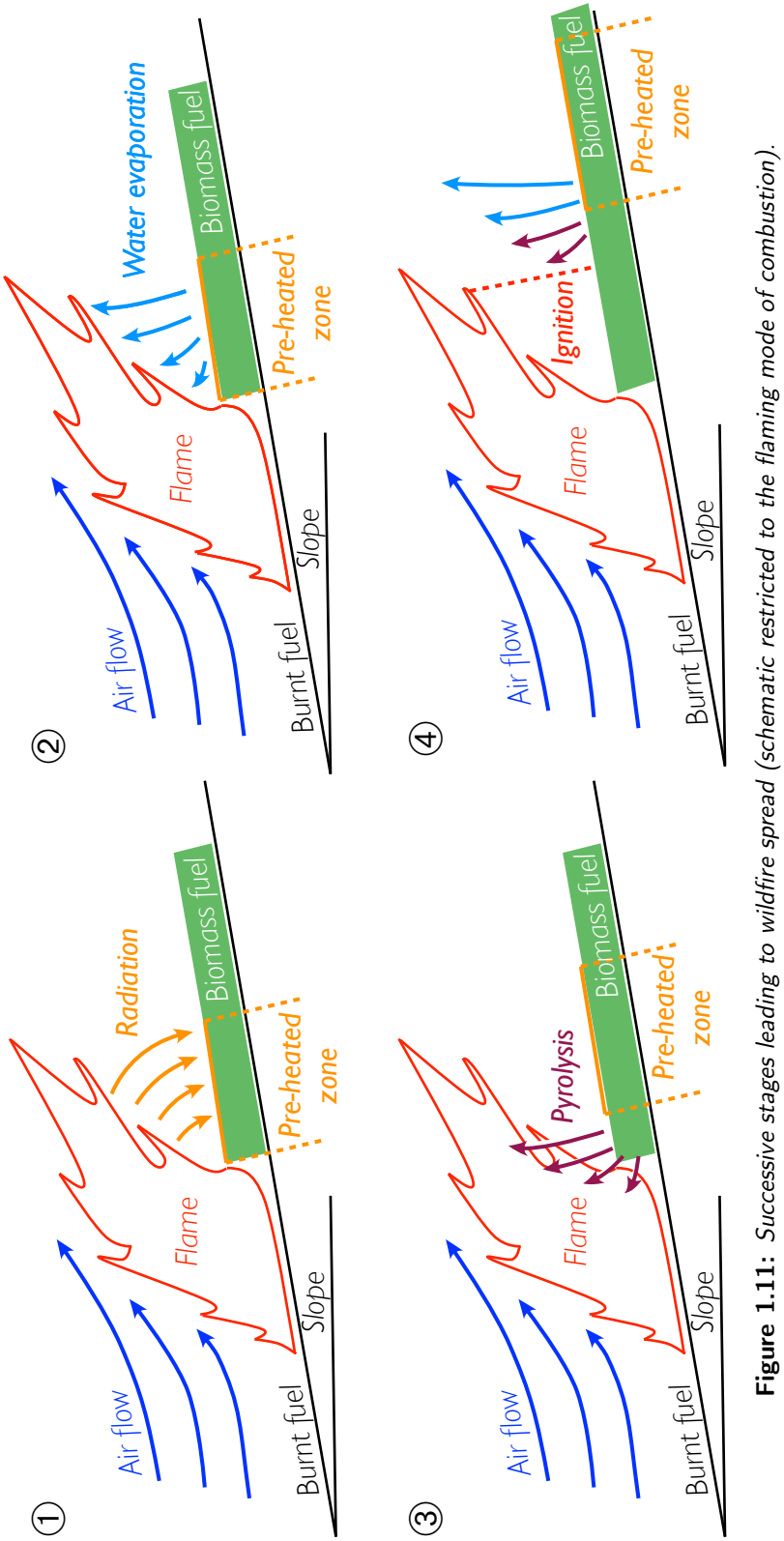


Figure 1.11: *Successive stages leading to wildfire spread (schematic restricted to the flaming mode of combustion).*

(3) **Pyrolysis gas release.** The temperature of the porous vegetation continues to rise and above a certain threshold temperature (typically, 450-650 K), the solid phase of the vegetation starts to release flammable gas compounds (e.g., CH_4, CO, H_2) that are convected through the vegetation layer towards the flame front. This entrainment is due to buoyancy effects. Since the burnt gases produced by the flame have a significantly lower density than ambient air (due to temperature discrepancies), they rise by convection and generate air streams (referred to as *air entrainment*) that push pyrolysis gas reactants towards the flame. This constitutes the pyrolysis stage, which can be regarded as a phase transformation (i.e., from solid-phase to gas-phase) within the porous vegetation.

(4) **Onset of combustion kinetic reactions.** Once the flammable gases released during the pyrolysis process are in contact with oxygen, oxidation reactions can proceed if the gas temperature is sufficiently high. A flame develops above the previously-mentioned pre-heated zone and in turn, releases heat towards the vegetation located ahead of the flaming front. This induces the displacement of the flame towards the unburnt vegetation. Note that the temperature at which pyrolysis gases are released (nearly 600 K) commonly defines the interface between the combustion zone and the pre-heated zone.

In this brief description, we highlight that a wildfire propagates due to a strong, non-linear coupling between heat transfer mechanisms, pyrolysis and combustion chemistry. Note that stages (2) and (3) can occur simultaneously. The vegetation is characterized by thermally-thick (e.g., tree trunks) and thermally-thin (e.g., leaves) solid particles, implying that the amount of heat absorbed by the solid vegetation is non-uniform through the vegetation layer. The composition of pyrolysis gases depends on the magnitude of the heat flux received by vegetation, since variations in heat transfer induce variations in the vegetation heating rates and thereby, variations in the pyrolysis rate (each flammable gas being released at specific temperatures). The presence of O_2 is also important. The features of the combustion kinetic reactions also strongly depend on the nature of these reactants (in terms of composition and temperature). Besides, the size of the pre-heated zone depends on the flame angle with respect to the ground surface: the larger the flame tilt angle, the wider the pre-heated zone and the higher the energy transferred, the faster flammable gases are emitted and thus, the faster the flame spreads. This implies that wildfires propagate faster in wind and upslope directions.

At ground level, conduction also plays a role in wildfire spread, independently of wind and slope. Also advection of pines and firebrands contributes to wildfire spread by spotting effects (as mentioned previously in the description of spot fires).

Furthermore, large amounts of heat and combustion products are released by the flame into the atmosphere. They modify the local state of the atmosphere, in particular local wind conditions and air humidity. These modifications affect, in turn, the state of the vegetation ahead of the flame zone and the local air flow. Thus,

they can potentially induce drastic changes in wildfire behavior over time. Wildfire is therefore a complex multi-scale natural hazard, governed by non-linear, scale-dependent, multi-physical processes (Viegas, 1998; Finney and Mcallister, 2011), which can produce unexpected and radical changes in its behavior (e.g., eruptive fires, Viegas and Simeoni, 2010).

1.3.2 A wide range of wildfire spread modeling approaches

Despite the devastating 1910 Great Fires in the Midwest (Pyne, 2001), interests in wildfire spread modeling only appeared at the end of the 1940s and quasi-exclusively in the US, due to the emerging Forest Service and the needs in wildland resource management. Curry and Fons (1938) and Fons (1946) laid the foundations of a rigorous physical approach to measure and mathematically model wildfire behavior, which served as a benchmark for wildfire research in the following decades (Emmons, 1964; Rothermel, 1972; Albini, 1985). However, computer-based wildfire spread modeling has only emerged during the past two decades, as a powerful tool for applications in both fire risk management and fire emergency response.

↪ Modeling aspects and issues

While length-scales of weather processes range over eight orders of magnitude, from hundred-kilometer-scale weather systems to millimeter-scale turbulent effects and flame-induced air entrainment, chemical reactions associated with vegetation thermal degradation and combustion occur at molecular scales to produce flame up to a few tens of meters height. Thus, the wildfire ROS depends on the ability of the flame and biomass burning region to supply a sufficient amount of heat to enhance fuel pyrolysis and ignite the mixture made of pyrolysis products and oxidizer ahead of the flame. The different length-scales involved in wildfire spread are illustrated in Fig. 1.12. Firelines travel at a ROS that results from complex interactions between pyrolysis, combustion and flow dynamics as well as atmospheric dynamics. Thus, our ability to accurately simulate the behavior of wildfires remains limited.

Due to its front-like topology at regional scales, similar to that of premixed flames,[10] a wildfire is generally considered as a propagating interface from the burnt area to the unburnt vegetation. This propagating interface (see Fig. 1.13) is referred to as the *fire front* or *fireline*. In this context, computer-based wildfire spread modeling aims at predicting the behavior of the fire front for a given set of environmental conditions and ignition location. In order to estimate the time-evolving location and perimeter of the fire front, the anisotropic ROS along the fireline must be determined at each time step.

[10] Flames in which the oxidizer is mixed with the fuel reactants before ignition, inducing the propagation of the flame towards the fresh mixture, in contrast to diffusion flames in which fuel and oxidizer are initially separated and in which the combustion process is confined at the interface between both compounds (see Chapter 2).

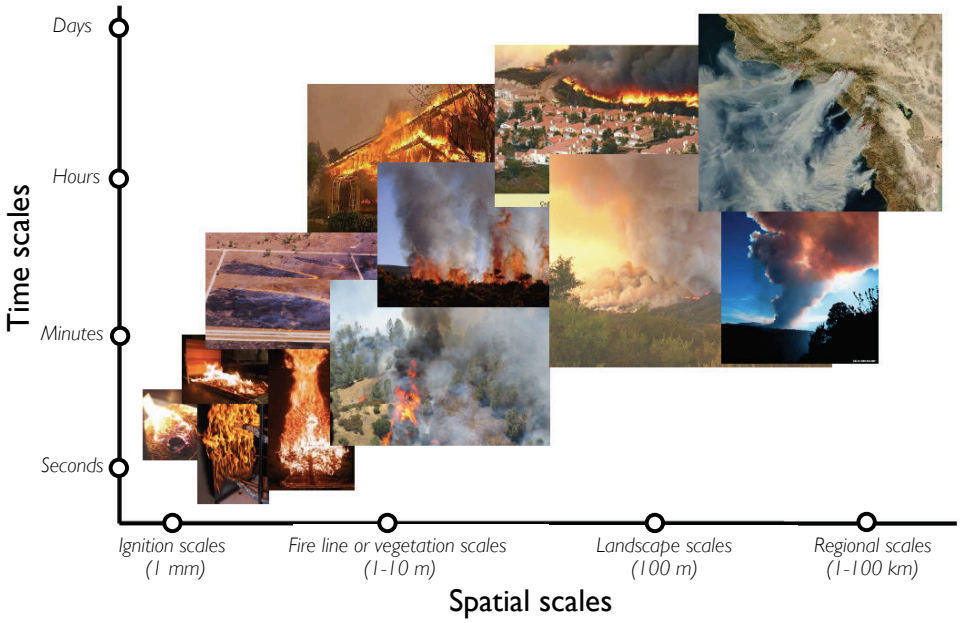


Figure 1.12: *Schematic of temporal and spatial (from mm-scale ignition to hundred-km-scale smoke transport) scales underlying wildfire spread. Credit: Rod Linn.*

The values for the ROS are determined by complex interactions between pyrolysis (e.g., fuel chemistry, solid-phase reactions, solid fuel transport), combustion and flow dynamics (e.g., gas-phase reactions, advection, buoyancy, radiation, topographic interactions) as well as atmospheric dynamics. From a modeling viewpoint, it is generally accepted that the ROS depends on the following set of factors.

- ▷ **Vegetation properties.** Vegetation can be regarded as porous organic materials, composed of a gas phase and of an ensemble of solid particles. While being non-uniformly distributed within the vegetation layer, these particles exhibit different sizes, chemical compositions, levels of biological decomposition and moisture contents. These properties are of primary importance in the pyrolysis process and significantly affect the rate at which the fire front can potentially propagate.
- ▷ **Weather conditions.** Wind velocity and direction are predominant factors in wildfire spread, since they are partly responsible for tilting the flame towards the unburnt vegetation. Air temperature and humidity, solar radiation, precipitations also play an important role.
- ▷ **Terrain topography.** Topography is associated with terrain configuration, altitude, slope, orientation parameters, which directly impact the directions in which a wildfire can propagate at the fastest rates.

It is worth mentioning that these different factors are not independent. For instance, terrain topography modifies the wind field near the ground.

Figure 1.13: *Snapshots of surface (left) and crown (right) wildfires, in which active fire areas are located at the interface between burnt and unburnt zones (referred to as* fireline*).*

Modeling-related issues are due in particular to the wide range of relevant length-scales, the complex set of coupled physical processes, the lack of knowledge in boundary and initial conditions (heterogeneous and poorly-defined vegetation, near-ground wind fluctuations) as well as difficult validation since it is usually limited to laboratory-scale and field-scale experiments, which are not fully representative of real-world wildfires.

To cope with the complexity of wildfire spread, different modeling approaches have been proposed in literature, from physics-based to empirical models as described in the following (Weber, 1991; Grishin, 1997; Perry, 1998; Pastor et al., 2003).

↪ Physics-based models

Relevant insight into wildfire dynamics has been obtained in recent years via detailed numerical simulations performed at flame scale, i.e., with a spatial resolution on the order of 1 m. These physics-based fire spread models are promising approaches to explicitly resolve interactions between the vegetation and the flame as well as between the flame and the atmospheric dynamics (Hanson et al., 2000). These models intend to simulate the fundamental chemical and physical processes within and above the vegetation (considered as a porous medium), by explicitly solving for mass, momentum and energy balance equations (Grishin, 1997; Larini et al., 1998; Linn et al., 2002; Morvan and Dupuy, 2004; Porterie et al., 2005). The porous medium includes the solid vegetation (branches, twigs, bark elements, etc., represented as solid fuel particles) and its surrounding gas (i.e., gas phase that incorporates the gas in-between the solid particles and above the vegetation layer or canopy).

The multiphase formulation implies that balance equations are applied to both solid and gas phases, which are coupled through non-linear heat and mass flux exchanges (e.g., vegetation mass loss, drag force of the fuel solid particles). Thus, physics-based models combine advanced physical modeling and classical methods of computational fluid dynamics (CFD) to accurately describe flame-scale processes,

while running at one or two orders of magnitude slower than real-time for a reasonable domain size (Hanson et al., 2000; Sullivan, 2009a). In particular, mean flow advection, large-scale effects of topography and buoyancy-induced flows are explicitly solved, while fine-scale turbulence, chemistry and combustion are subgrid-scale processes that require additional physical modeling.

▷ **Fine-scale two-dimensional models.** In FireStar for instance, following the choices made by Grishin (1997) and the mathematical formulation proposed by Larini et al. (1998), Morvan and Dupuy (2004) proposed a re-normalisation group (RNG) k-ϵ model combined with an eddy dissipation concept (EDC) to accurately account for the coexistence between regions of turbulent (e.g., thermal plume, burning zone) and laminar (e.g., near the ground) flows at flame scale. In this solver, the rate of combustion reactions is mainly limited by the turbulent mixing between the gaseous fuel and the oxidizer following EDC (Magnussen, 2005). Both radiation and convection are included in the energy balance equations of the gas and solid phases. In particular, the radiation heat transfer is quantified through the resolution of the radiation transfer equation (RTE) considering the vegetation as a black-body (i.e., with an emissivity equal to 1). Similarly to Séro-Guillaume and Margerit (2002) or Porterie et al. (2005), FireStar relies on an averaging formulation of the macroscopic vegetation to account for the porous structure of vegetation and in particular for the effects induced by the vegetation microstructures. Based on this formalism, the temporal evolution of the vegetation is described by mass balance equations for each component of the solid fuel (e.g., dry material, liquid water, carbon-based char material), in which each reaction source term is modeled as an Arrhenius-type law. As suggested by Grishin (1997), the soot volume fraction is imposed as a constant fraction of the pyrolysis rate (nearby 5 %). FireStar was evaluated against experimental fires through a homogeneous pine needle fuel bed (Morvan and Dupuy, 2001). The study presented in Morvan and Dupuy (2004) confirmed the existence of the two modes of wildfire spread proposed by Pagni and Peterson (1973):

(1) plume-dominated fires, in which the radiation heat transfer from the flame (due to soot particles and embers) is predominant (i.e., cases of surface fires on a flat terrain and without significant wind-induced convection velocities);

(2) wind-driven fires, in which the convection heat transfer is dominating, with still a significant radiation contribution (i.e., cases in which the flame is tilted towards unburnt vegetation due to wind effect).

The same approach has been used to study the propagation of surface fires through Mediterranean shrublands (Morvan, 2007) and crown fires (Dupuy and Morvan, 2005) in off-line mode and with simulations that are far from achieving real-time performance. For instance, Fig. 1.14 shows the importance of radiation heat losses within the flame zone and their effects on the flame topology.

Figures 1.15(a)-(b) illustrate a crown fire upslope propagation (that enchances air

entrainment and vegetation pre-heating zone). They confirm that the gradient of the fuel mass density (indicating a drastic vegetation mass loss) coincides with the temperature gradient at the flame front.

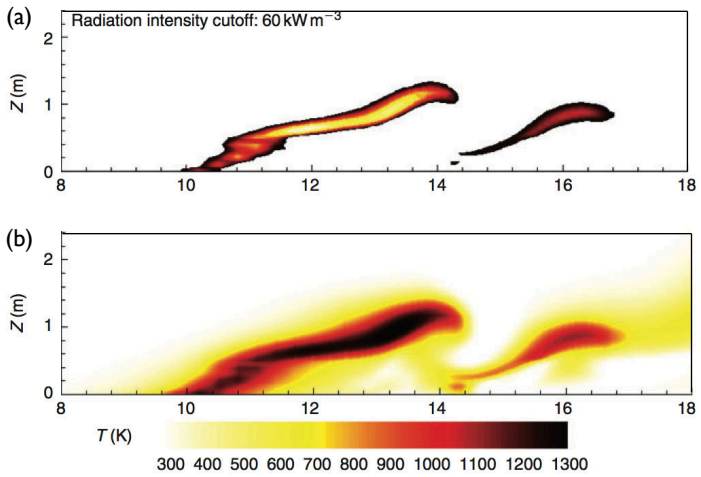


Figure 1.14: *Shrubland surface fire. (a) Reconstruction of the flame front from a radiation heat loss rate equal to* 60 *kW/m*^{3} *for a wind velocity* $u_w = 5$ *m/s (a pocket of burnt gases has ignited a secondary flame, propagating downwind of the flame front). (b) Associated gas temperature field. Credit: Morvan (2007).*

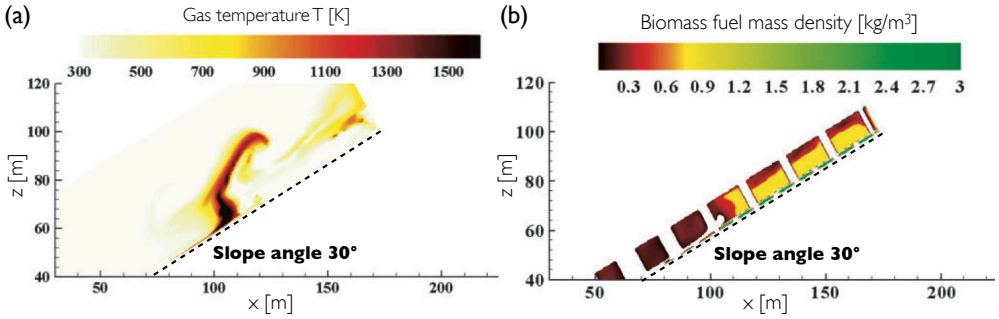


Figure 1.15: *Upslope crown fire for a terrain slope angle equal to* $30°$*. (a) Gas temperature field* T *[K]. (b) Associated biomass fuel mass density* ρ_v *[kg/m*^{3}*]. Credit: Dupuy and Morvan (2005).*

One of the objectives of these FIRESTAR studies (Morvan and Larini, 2001; Morvan and Dupuy, 2004; Morvan et al., 2009) is the improvement of knowledge on the fundamental physical and chemical mechanisms controlling wildfire spread. However, FIRESTAR is limited to two-dimensional (vertical) configurations and therefore cannot properly capture turbulent features. Note that a similar multiphase approach is proposed by Porterie et al. (2005), which showed promising results to simulate rapidly-propagating savanna fires and their emissions (in terms of gas compounds and soot particles). In contrast to Morvan and Dupuy (2004), Consalvi et al.

(2011) proposed a more physically-based char oxidation model due to Boonmee and Quintiere (2005), instead of an Arrhenius-type law.

▷ **Three-dimensional models.** Since wildfire spread exhibits fully three-dimensional features, two-dimensional solvers remain limited for the study of multi-physical multi-scale interactions. Thus, three-dimensional solvers such as FIRETEC and WFDS have been developed in the past years to simulate fire-related processes with a spatial resolution on the order of 1 to 2 m. FIRETEC[11] is a compressible multiphase LES solver, in which the pyrolysis of the solid phase and the combustion of the gas phase are treated together (Linn, 1997; Linn et al., 2002). As for WFDS[12], it is an extension of FDS for simulating biomass burning, which is based on a low-Mach number assumption and which handles gas and vegetation porous phases separately on different grids (Mell et al., 2007). In particular, WFDS adopts the pyrolysis model proposed by Morvan and Dupuy (2004), except that it does not consider char oxidation.

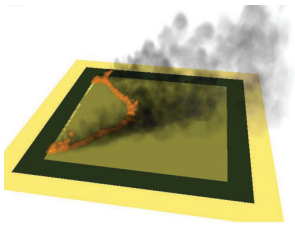
Mell et al. (2005) presented WFDS simulations for a surface fire through a uniform grassland fuel bed, for which in-situ and airborne measurements were available for a wide range of conditions (e.g., wind conditions, length of fireline ignition). Figures 1.16(a)-(b) compare the WFDS LES to the observed controlled burn and show consistent fire front location and shape (from a qualitative viewpoint). Note that since the vegetation is uniform, the shape of the fire front is relatively smooth and symmetric with a faster propagation at the head of the fire (aligned with the wind direction). However, this fire front does not remain straight, partly due to high heat losses at the front edges.

Figure 1.16(c) confirms this trend over time and the ability of WFDS to predict the propagation of the head of the fire as well as its flanks. Note that a wind shift occurred after 86 s that broke the front symmetry; this drastic change is not captured by WFDS since a constant wind direction was assumed over the simulation duration. Furthermore, Linn et al. (2005) highlighted how physics-based modeling approaches provide some important insights into the processes driving crown fires and how they are able to capture wind/fire interactions in discontinuous fuel beds. To illustrate how the canopy structure particularly affects wildfire spread, a simulation snapshot based on FIRETEC is shown in Fig. 1.17.

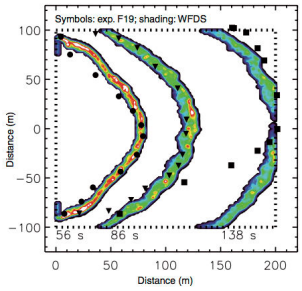
▷ **Successes and limitations.** Physics-based models have shown their overall ability to predict the macroscopic behavior of controlled burns, by explicitly solving for the multi-scale interactions between the vegetation, the flame and the atmosphere, and by accounting for each mode of heat transfer (i.e., conduction, radiation and convection). A review of physics-based models is presented in Sullivan (2009a).

[11] Los Alamos National Laboratory simulator, ees.lanl.gov/ees16/FIRETEC.shtml.
[12] NIST Wildland Fire Dynamics Simulator, www.openwfm.org/wiki/WFDS.

(a) Experimental fire (200 m $\times$ 200 m).



(b) Snapshot of WFDS simulation.



(c) Time-evolving fire front.

Figure 1.16: *Comparison of* WFDS *simulation and observation for a uniform grassland controlled burn. (a)-(b) Instantaneous snapshots at time 56 s. (c) Time-evolving location of the flame zone. Symbols correspond to experimental measurements and shaded contours correspond to* WFDS *simulations. Credit: Cheney et al. (1993), Mell et al. (2005).*

.

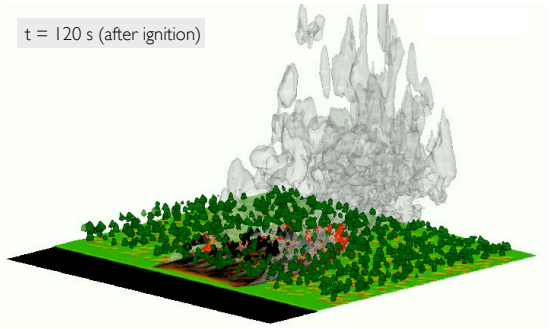


Figure 1.17: FIRETEC *simulation of* 20 *m* $\times$ 50 *m field-scale wildfire spread (Flagstaff, Arizona) in discontinuous fuel beds, 120 s after ignition. Colors on the horizontal plane represent the spatial variations in vegetation density, black indicating the absence of fuel and bright green indicating the* 1 *kg/m*^{3} *iso-contour (corresponding to a tall grass of 0.7-m depth). Dark green iso-surfaces indicate tree locations, while orange, red and grey iso-surfaces indicate regions of hot gases. Credit: Linn et al. (2005).*

Note that because of its high computational cost, flame-scale CFD is currently restricted to research projects and is not compatible with real-time forecast operational objectives. WFDS or FIRETEC are currently used in off-line mode to assess

the effects of prescribed burns on the vegetation, to evaluate the fire response to strategic fuel and firebreak policy as well a to study wildland-urban interfaces subject to fires.

Note also that flame-scale CFD still relies on several input parameters that are difficult to estimate (e.g., soot volume fraction, ash content, char fraction) and on modeling approximations (e.g., turbulent combustion model, pyrolysis model). Thus, their performance must be evaluated for more heterogeneous environmental conditions. Mell et al. (2005) stated that neither WFDS nor FIRETEC can explicitly solve for a detailed description of biomass fuels that differentiates a backing fire from a heading fire. The fuel bed is indeed unresolved on the computational grid (the horizontal and vertical spatial resolution being coarser than 1 m). From this viewpoint, FIRESTAR provides a finer representation of the flame/vegetation interactions and of the flame-scale combustion processes, even though it is limited to two-dimensional (vertical) configurations. This implies that further analysis is required to improve subgrid-scale models, for instance using CFD techniques primarily developed for gas engine applications. While industrial companies and policy-makers have been aware for a few decades that CFD could be helpful to optimize the design and emissions in aeronautical engines, wildfires have been identified only recently (in the EU at least) as a serious threat for public safety and ecosystem preservation. However, CFD applications to wildfire spread require significant modifications in comparison to gas engine applications (e.g., unconfined flow interacting with atmospheric dynamics, significant radiation heat transfer, partially known biomass fuel, buoyancy effects). These aspects are detailed further in the manuscript (see Chapters 2 and 3).

↪ Empirical models

An approach that is consistent with an operational framework relies on the evaluation of the ROS using statistical correlations of experimental data (e.g., wind-tunnel experiments, field-scale controlled burns). The resulting parameterization of the ROS depends on a reduced number of factors characterizing environmental conditions, namely the fuel moisture content M_v (defined as the amount of water contained in vegetation and expressed as a percentage of its dry mass) along with fuel intrinsic properties f_v (e.g., the vertical thickness of the fuel layer, the fuel loading, the fuel particle mass density), the wind velocity $\mathbf{u}_w$ at mid-flame height, and the terrain slope angle α_{sl}. Using this formalism, the ROS along the fireline noted Γ is of the following form:

$$\Gamma = \Gamma(\mathbf{u}_w, \alpha_{sl}, M_v, f_v)\,. \tag{1.1}$$

In a wildfire spread simulator, this ROS empirical model is combined with Lagrangian or level-set front-tracking techniques (Fendell and Wolff, 2001; Mallet et al., 2009; Rehm and McDermott, 2009) to simulate the two-dimensional propagation of the fireline. A review of empirical modeling approaches is provided by

Sullivan (2009b). These approaches are simple and computationally efficient; however, their domain of validity is limited to the conditions of the experiments used during their original development. This limitation implies that their extension from controlled burns to real-world regional-scale wildfires is not fully supported.

↪ Semi-empirical models

An intermediate approach between physics-based and empirical models consists in (1) formulating the ROS in Eq. (1.1) using the energy balance equation (applied to the unburnt vegetation located in the pre-heated zone); and (2) calibrating the resulting model parameters using experimental data. Resulting semi-empirical models do not distinguish between the different heat transfer modes. However, they exhibit the computational efficiency of empirical modeling approaches, while still including some relevant physical aspects of wildfire spread (through the energy balance equation). The most widely-used semi-empirical model is due to Rothermel (1972); its detailed formulation is provided in Appendix A. Further works (Weise and Biging, 1997) demonstrated that combining the original Rothermel's formulation (Rothermel, 1972) with Albini's description of radiation heat transfer (Albini, 1985, 1986) closely mimics the ROS response to multiple combinations of wind and slope conditions. For instance, Fig. 1.18 illustrates the consistency of Rothermel model predictions to observed ROS for different biomass fuels.

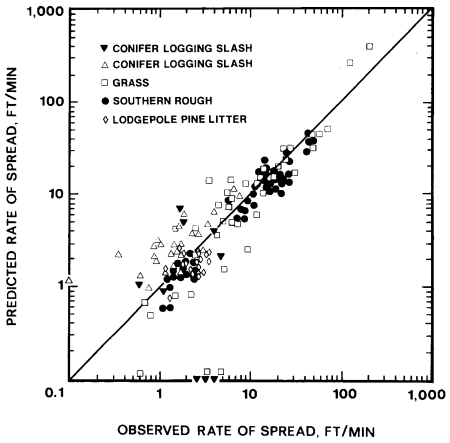


Figure 1.18: *Field verification of the consistency between Rothermel-based predictions and observed ROS for a wide range of biomass fuels (e.g., conifer logging slash, grass, pine litter). Credit: Rothermel (1983).*

Alternative ROS formulations exist in literature (Sullivan, 2009b; Cheney et al., 1998; Balbi et al., 2009). For instance, Cheney et al. (1998) provided a semi-empirical model that is specifically calibrated for Australian grassland fires; Balbi et al. (2009) described a semi-empirical approach that relies on mass, momentum and energy balance equations, while including geometrical simplifications and heat transfer assumptions (see Appendix A, in which a comparative study to the ROS

model due to Rothermel, 1972, is provided).

Since physics-based models exhibit a prohibitive computational cost, current operational wildfire spread simulators adopt a regional-scale viewpoint (i.e., a viewpoint that considers scales ranging from a few tens of meters up to several kilometers) based on empirical or semi-empirical ROS modeling approaches. They simulate a wildfire as a two-dimensional propagating front within the vegetation bed, using standard level-set or Lagrangian front-tracking techniques. For instance in the US, BehavePlus[13] (Andrews, 1986; Andrews et al., 2008) and Farsite[14] (Finney, 1998) developed at the Missoula fire sciences laboratory use the model due to Rothermel (1972) and provide guidance tools for fire management as part of WFDSS.[15] While BehavePlus considers uniform environmental conditions for the forecast period, Farsite relies on a perimeter expansion technique based on Huygens' principle (Richards, 1995) to propagate the fire front using anisotropic ROS values (due to spatially- and temporally-varying environmental conditions). However, the practitioner-oriented BehavePlus system includes predictions of fire behavior as well as fire effects (e.g., tree mortality) and fire environment (e.g., fuel moisture, wind adjustment factor). It is worth mentioning that in these operational simulators, surface winds are imposed as input parameters in the ROS models (Forthofer, 2007), while environmental conditions (e.g., vegetation properties, terrain topography) are integrated through in-situ and remote sensors (Lopez et al., 2002).

1.3.3 Limitations of wildfire spread modeling

While much progress has been achieved over the past decades in the basic understanding of wildfire dynamics, while also much progress has been achieved in the mathematical formulation and numerical simulation of wildfire spread, result accuracy remains limited for several reasons listed below (Viegas, 2011).

- ▷ **Absence of a model for fire/atmosphere interactions.** The mathematical models proposed to simulate regional-scale wildfire spread are unable to cover the entire range of relevant scales and to explicitly account for fire/atmosphere interactions (see Fig. 1.19). Their domain of validity is limited to the experimental conditions retained during their original development. Figure 1.20 shows that a particular form of semi-empirical modeling that is explicitly calibrated for Australian grasslands (Cheney et al., 1993, 1998) provides accurate predictions of ROS, consistently with WFDS simulations. However, Behave that relies on the formulation due to Rothermel (1972) is not able to track the wind-induced ROS variations, even though the vegetation conditions are homogeneous. These results confirm that the validity of a semi-empirical ROS model highly depends on the conditions for which the

[13] www.firemodels.org/index.php/national-systems/behaveplus.
[14] www.firemodels.org/index.php/national-systems/farsite.
[15] US Wildland Fire Decision Support System, wfdss.usgs.gov/wfdss/WFDSS_Home.shtml.

underlying parameters were calibrated and that predictions of physics-based models are valid over a much wider range of environmental conditions. Note that the WINDNINJA capability to simulate surface wind flows can improve FARSITE predictions (Forthofer, 2007).

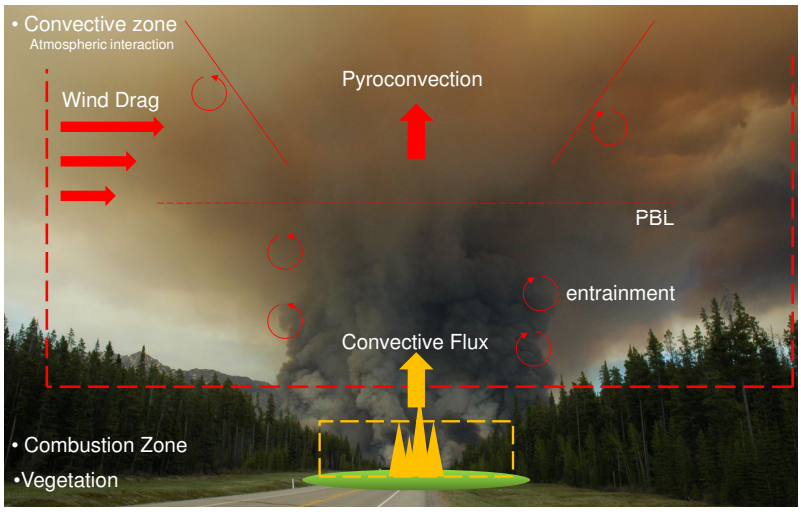


Figure 1.19: *Schematic of fire/atmosphere interactions (PBL standing for planetary boundary layer). Credit: Martin Wooster (private communication).*

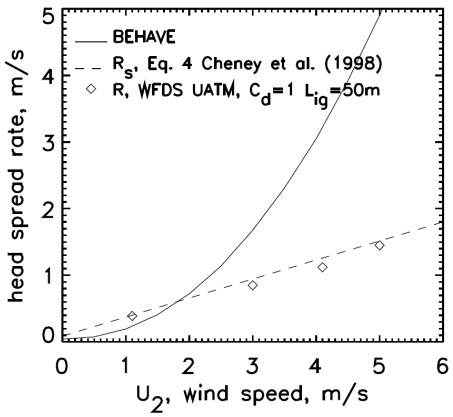


Figure 1.20: *ROS values with respect to the wind magnitude. Symbols correspond to* WFDS *predictions; the dashed line corresponds to the grassland-calibrated empirical model due to Cheney et al. (1998), and the solid line corresponds to* BEHAVE *predictions. Credit: Mell et al. (2005).*

- ▷ **Presence of physical modeling inaccuracies.** A second limitation is that semi-empirical modeling approaches exhibit knowledge gaps and/or inaccuracies in the description of the physics. While physics-based computational fire models are able to account for time-dependent wildfire behavior (Viegas, 2004), semi-empirical approaches are based on a steady-state assumption and

cannot properly capture acceleration processes such as in canyons (Viegas and Pita, 2004) among others. These accelerations in the upslope direction are illustrated in Fig. 1.21 for a laboratory-scale fire spread in a canyon experimental configuration.

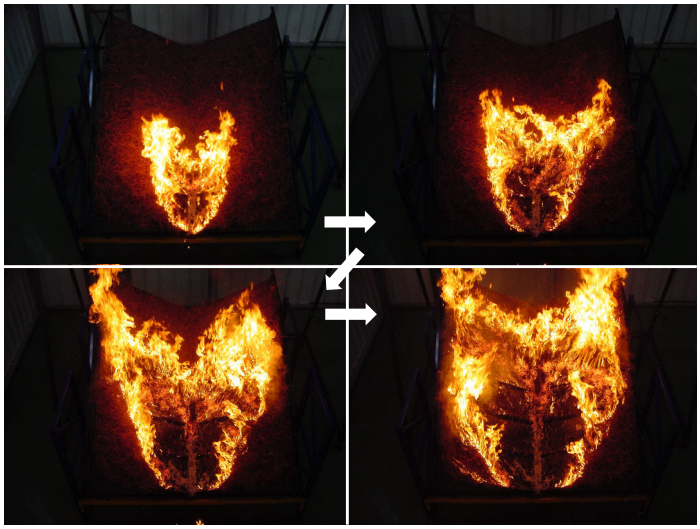


Figure 1.21: *Time-series of a laboratory-scale fire experiment in a canyon configuration. Credit: D.X. Viegas (private communication).*

- ▷ **Lack of high-resolution environmental conditions.** Another limitation shared by all fire spread models lies in the knowledge of the input variables that determine the ROS (i.e., vegetation properties, terrain topography, weather conditions), which are often unknown or are only known with limited accuracy.

Due to their computational requirements, physics-based modeling approaches cannot replace current operational wildfire spread simulators in the near-future. Still, they can provide reliable and detailed predictions of the behavior and effects of wildfires over a much wider range of conditions than operational simulators, but for very limited fire sizes. Thus, in spite of their uncertainties, wildfire spread simulators that adopt a regional-scale viewpoint (i.e., a front propagating approach) remain to date, the suitable operational tool for forecasting wildfire spread scenarios.

↪ Coupled fire/atmosphere simulation capabilities

One recent strategy to better account for time-varying weather conditions at regional scales consists in coupling a cost-effective wildfire spread model with a mesoscale CFD atmospheric model, see for instance WRF-Fire[16] (Clark et al., 2004; Mandel et al., 2011; Kochanski et al., 2013) that combines the weather research and forecasting (WRF) atmospheric model with the level-set-based surface fire

[16] www.openwfm.org/wiki/WRF-Fire.

behavior model SFIRE, or FOREFIRE/MÉSO-NH (Filippi et al., 2013) coupling the following components:

- the meso-scale atmospheric solver MÉSO-NH[17] that is a non-hydrostatic LES solver, which is able to describe kilometer-scale to meter-scale atmospheric dynamics along with chemical kinetic processes (Lafore et al., 1998).
- the front-tracking solver FOREFIRE[18] that is a Lagrangian front-tracking solver applied to wildfire spread, evolving the location and width of the flame front according to a semi-empirical ROS model (e.g., the formulation due to Balbi et al., 2009). This solver relies on a discrete-event numerical approach (Filippi et al., 2009, 2011). In contrast to conventional explicit or implicit schemes, this discrete-event approach performs time-integration in terms of increments of physical quantities (instead of time increments) and is therefore time-efficient (i.e., much faster than real-time).

A schematic of the coupled solver FOREFIRE/MÉSO-NH is presented in Fig 1.22. MÉSO-NH forces wildfire behavior through the surface wind field, while FOREFIRE imposes heat and vapor fluxes as surface boundary conditions to MÉSO-NH. An on-line chemistry module can also be activated in the coupling mode, in order to account for chemical kinetic processes of trace gases and aerosols emitted by wildfires (Filippi et al., 2011; Strada et al., 2012). Note that FOREFIRE/MÉSO-NH is the coupling approach developed in the project, to which this thesis contributes.

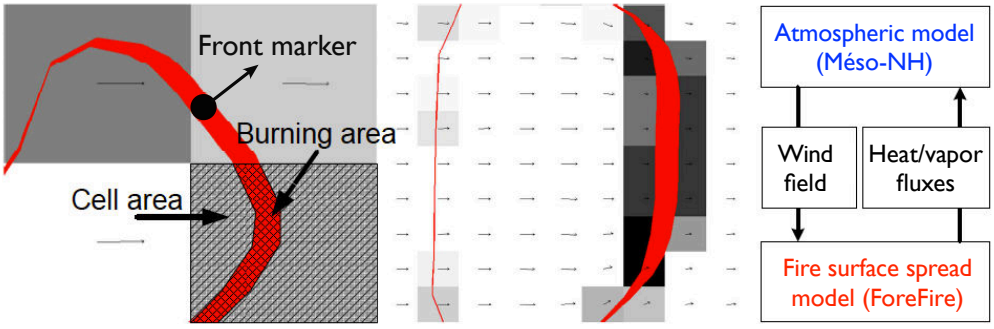


Figure 1.22: *Schematic of the FOREFIRE/MÉSO-NH coupling, in which the red zone indicates the flame zone discretized by front markers, gray cells correspond to the atmospheric computational grid and overlain vectors indicate the near-ground wind field. Credit: Filippi et al. (2011).*

Filippi et al. (2013) presented a validation study of the FOREFIRE/MÉSO-NH coupled simulator applied to the FireFlux grassland controlled burn (Clements, 2007). In addition, Santoni et al. (2011) demonstrated the improvements in the prediction of the fire front location due to the fire/atmosphere coupling (in comparison to the stand-alone fire spread simulator FOREFIRE). Figure 1.23 illustrates this

[17] www.aero.obs-mip.fr/.
[18] forefire.univ-corse.fr/.

improvement, since the simulated fire front in the coupling mode is much closer to the observation than the simulated fire front in the surface mode. Thus, the ForeFire/Méso-NH coupled simulator allows to investigate atmospheric feedback on wildfire behavior and predict fine-scale features of the wildfire behavior. Detailed simulations of the flow and wildfire patterns over a complex heterogeneous vegetation with Firetec or WFDS are currently beyond computational capacities. Thus, WRF-Fire and ForeFire/Méso-NH coupled approaches appear as a promising strategy to forecast the time-evolving location of the fire front (at a reduced-cost) as well as the atmospheric behavior (in terms of plume size, transport dispersion and smoke concentration).

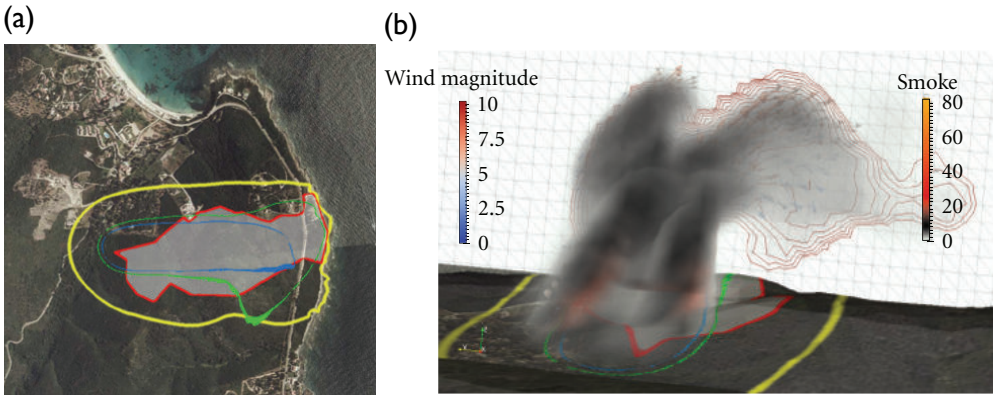


Figure 1.23: *Favone wildfire (8 July 2009, Corsica). (a) Comparison of fire front locations: the blue line represents the simulated fire front 50 min after ignition, the green line represents the simulated fire front 4 h after ignition (coupling mode), the yellow line is the equivalent for the non-coupled* ForeFire *simulation (surface mode), and the red line represents the final observed fire front. (b) 3-D view of the simulated plume nearly 50 min after ignition. Credit: Santoni et al. (2011).*

Still, many uncertainties remain due to simplifications in the description of the physics and to knowledge gaps in the description of environmental conditions. Further work aims at better representing biomass fuels and combustion, in order to improve surface fluxes models and to accurately simulate regional-scale wildfire spreads at a high spatial resolution. Even though these coupled fire/atmosphere models have already demonstrated their potential for forecasting real-world wildfires, they require further validation studies, similarly to the Aullène case study[19] presented in Fig. 1.24. Note that the French national database Prométhée[20] provides an extensive record of past wildfire events, which could be useful to evaluate uncertainties in coupled and non-coupled simulations against past observations (Filippi et al., 2013). Coupled fire/atmosphere models remain currently limited to research projects.

[19] www.cnrs.fr/insis/recherche/actualites/2013/incendie.htm.
[20] www.promethee.com/prom/home.do.

This overview of wildfire spread models (summarized in Table 1.1) highlights that the use of a regional-scale wildfire spread simulator that takes full advantage of the recent technological advances for geo-referenced front tracking is essential to improve fire front location predictions. This is motivated by the uncertainties inherent to regional-scale wildfire spread modeling and to the current impossibility of applying physics-based models to the operational framework of wildfire monitoring. Thus, data-driven wildfire spread simulators remain to date, the most promising strategy for wildfire spread forecast and thereby, for reliable fire danger evaluation.

It is worth noting that we use a semi-empirical ROS model due to Rothermel combined with a front-tracking simulator named Firefly to perform regional-scale wildfire spread simulations. This simulator has been developed in this work to allow for more flexibility in the implementation and evaluation of data assimilation techniques (see Chapter 6).

1.4 Remote sensing technologies for wildfire spread monitoring

Currently, observations of vegetation areas subject to wildfire spread are of primary importance for efficient detection and tracking of the location of flame front regions at large scales, while supporting both research and operational applications. Due to Earth observation national and international programs, available instruments for observations are of different nature, from in-situ sensors positioned on the ground to remote sensors aboard airborne platforms, unmanned aerial vehicles (UAV) or satellites, as reviewed by San-Miguel-Ayanz et al. (2005), Calle and Casanova (2012) as well as Wooster et al. (2013). We provide here a brief introduction to current remote sensing systems and emphasize their applicability for tracking regional-scale wildfire spread in real-time.

1.4.1 Detection of active fire areas

Detection refers to the determination of the location of hot-spots (i.e., areas of anomalous elevation of temperature above normal environmental temperatures), independently of their size. Early fire detection is essential for civil defense (due to the growing number of wildland-urban interfaces) and firefighting efficiency (a wildfire is easier to extinguish during its early stages than when it reaches a fairly large size, for which fire suppression capacities largely depend on meteorological conditions). By measuring the electromagnetic radiation that is emitted from burning biomass fuels at the Earth surface (i.e., about 10 to 30 % of the heat released by the flame zone, Byram, 1959), remote sensors are able to locate wildfires through the detection of either hot spots or smoke plumes formed by wildfire emissions into the atmosphere. It is worth mentioning that the majority of this emitted radiant energy can be regarded as a black-body-type radiation.

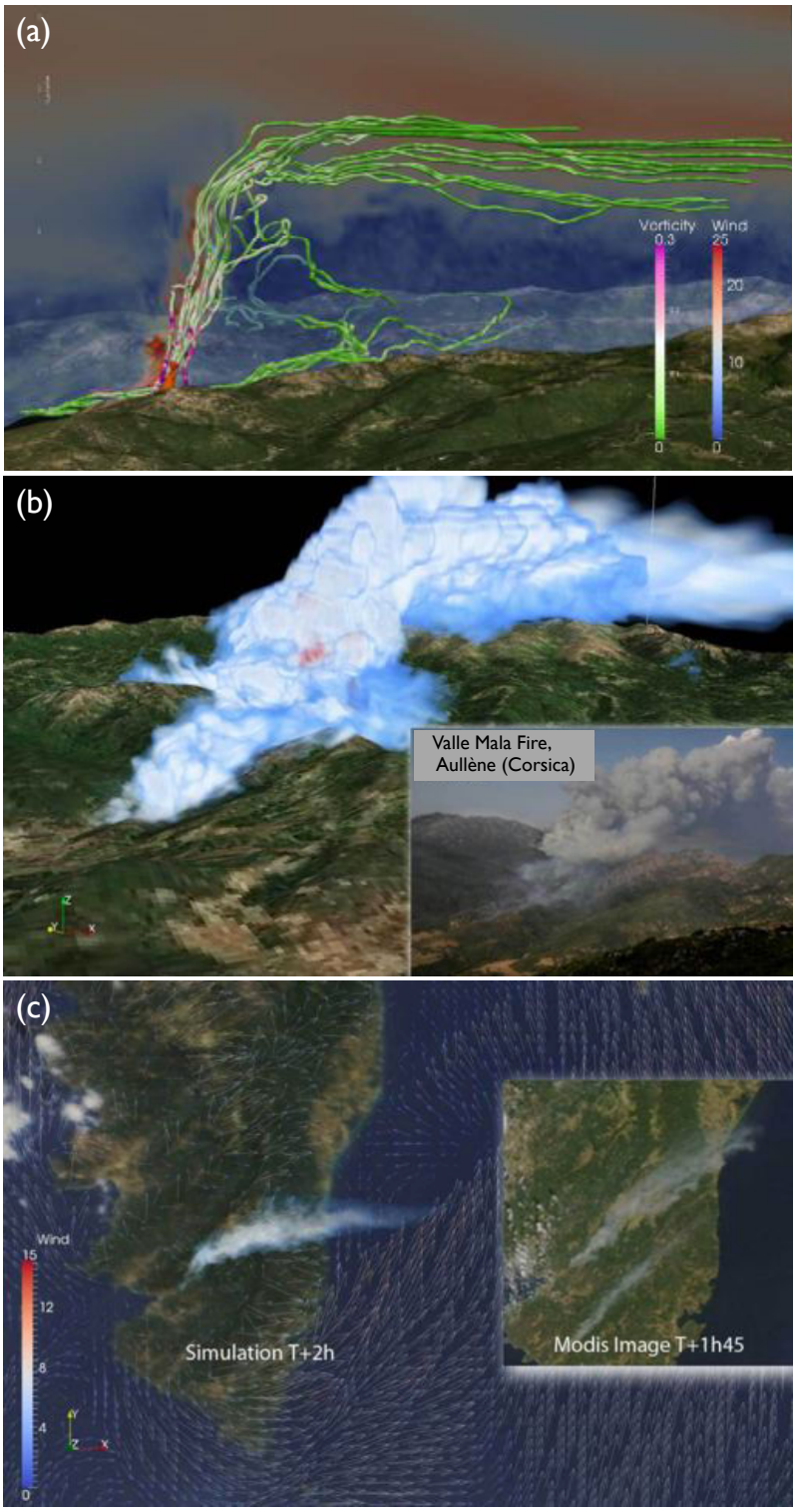


Figure 1.24: *Micro-scale (50 m resolution) to meso-scale (2400 m resolution)* ForeFire/Méso-NH *simulation of the Aullène wildfire (6000 ha, July 2009, Corsica). (a) Vertical wind profile at micro-scales. (b) Smoke plume at micro-scales compared to in-situ imaging. (c) Smoke plume at meso-scales compared to MODIS imaging (1 km resolution). Credit: CNRS-SPE,* anridea.univ-corse.fr.

Table 1.1: *Recap of physics-based and empirical wildfire spread simulators.*

Simulator name	Institution	Main characteristics	References
BehavePlus	US Forest Service	• Semi-empirical ROS modeling • Uniform conditions (0-D)	Andrews (1986)
Farsite	US Forest Service	• Semi-empirical ROS modeling • Ellipse fire growth simulator (2-D surface)	Finney (1998)
Firetec	Los Alamos national lab.	• Physics-based modeling • CFD multiphase solver (3-D)	Linn et al. (2002)
WFDS	NIST	• Physics-based modeling • CFD solver with wildland-fire module (3-D)	Mell et al. (2007)
FireStar	Univ. de la Méditerranée	• Physics-based modeling • CFD multiphase solver (2-D vertical)	Morvan and Dupuy (2004)
WRF-Fire	Univ. of Colorado, Univ. of Utah	• Coupled fire/atmosphere (3-D) • Level-set fire growth simulator (2-D surface) • Semi-empirical ROS modeling	Mandel et al. (2011), Kochanski et al. (2013)
ForeFire/Méso-NH	Météo-France and CNRS (LA and SPE national lab.)	• Coupled fire/atmosphere (3-D) • Lagrangian front-tracking simulator (2-D surface) • Semi-empirical ROS modeling	Filippi et al. (2011), Strada et al. (2012), Filippi et al. (2013)

Not all wavelengths in the electromagnetic spectrum are effective for remote sensing. However, a combination of spectral bands is generally required to identify a distinctive spectral response pattern of a particular emitting surface (referred to as the *spectral signature*). In practice, continental surfaces and vegetation are mainly observed within the mid-infrared (MIR) and thermal-infrared (TIR) regions of the electromagnetic spectrum (1.4 to 15 μm), without significant interference by the atmosphere. Table 1.2 indicates the common subdivisions of the infrared band in the geospatial remote sensing field.

Several wavelengths in the electromagnetic spectrum are able to distinguish a hot-spot fire from the surrounding ambient background. Planck's law calculates, for a wide range of wavelengths, the amount of electromagnetic energy radiated by a black-body in thermal equilibrium. The wavelength associated with the peak of the black-body radiation curve provides an indication of its temperature (through Wien's displacement law). In particular, the higher the temperature, the shorter the wavelength. In other words, when the temperature of the black-body increases, the peak of the radiation curve moves to shorter wavelengths as illustrated in Fig. 1.25. Thus, for the elevated temperatures encountered in wildfires, from some hundreds to more than 1000 K above the ambient background, the maximum radiant intensity generally occurs within the SWIR to MWIR regions. Since the SWIR region is also affected by very significant solar reflected radiation signals, the MWIR (between 3 and 5 μm) is generally the focus of active fire detection algorithms and is commonly referred to as the MIR region (Robinson, 1991; Kennedy et al., 1994; Arino and Melinotte, 1995).

Table 1.2: *Regions within the infrared electromagnetic spectrum.*

Band name	Wavelengths	Common appellation
Near-infrared (NIR)	0.75-1.4 μm	-
Short-wavelength infrared (SWIR)	1.4-3 μm	-
Mid-wavelength infrared (MWIR)	3-8 μm	MIR
Long-wavelength infrared (LWIR)	8-15 μm	TIR
Far-infrared (FIR)	15-1000 μm	-

Recent studies (Giglio et al., 1999; Justice et al., 2002; Wooster et al., 2005; Riggan and Robert, 2009) have confirmed that the MIR region centered on the 3.9-μm wavelength is suitable for wildfire detection using spaceborne sensors, since this wavelength is both sensitive to flaming and smoldering combustion modes and since this wavelength is not significantly affected by gaseous absorption and emission in the flame or in the atmosphere. Figure 1.26 shows indeed that contributions of both flames and hot ashes are significant at this particular wavelength. Note that wildfires can be detected even though they occupy a limited area of the sensor pixel (i.e., 0.1 to 1 % of the pixel area, while the spatial resolution of the MODIS sensor is 1 km for instance) or even though the Earth surface is covered by a smoke plume

(with some attenuation when the ground surface is covered by condensed-water clouds). However, the performance of spaceborne remote sensing also depends on the size of the fire, since the higher the fire temperature, the smaller the fire size required for its detection. This limitation can be overcome using airborne or UAV platforms (Ambrosia and Wegener, 2009).

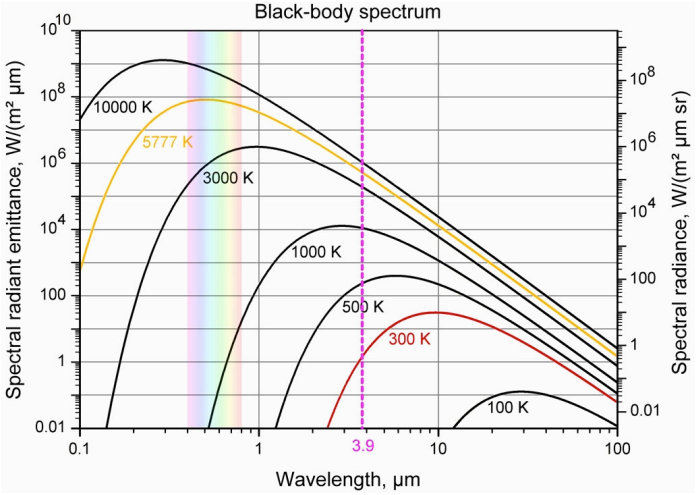


Figure 1.25: *Radiation emitted in the different bands of the electromagnetic spectrum by different temperatures of a black-body. The sun can be regarded as a 5777 K black-body, corresponding to a peak emission at 0.5 µm (yellow line), while the Earth surface can be regarded as a 300 K black-body, corresponding to a peak emission at 10.35 µm (red line). Credit:* www.eumetrain.org/data/3/30/.

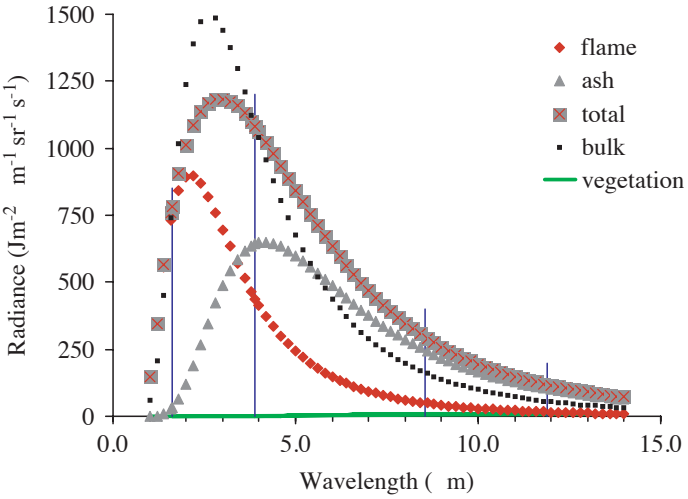


Figure 1.26: *Model fire radiances in tropical savanna showing flame (red diamonds) and hot-ash (grey triangles) contributions; the very low emitted radiance of vegetation (green line) is shown for comparison. Credit: Riggan and Robert (2009).*

The observation of a smoke plume is usually part of a fire detection system (especially in the case of extreme fire events) and in particular, it is used to eliminate false alarms produced by ground areas over-heating. However, this type of detection techniques presents one critical limitation since the time delay between wildfire ignition and detection is important (the formation of a large and easily detectable smoke plume requires a fairly large fire size). Besides ground over-heating, one of the main sources of false fire detection alarms is the presence of clouds. Sun-illuminated clouds typically appear as regions of elevated MIR values (due to their reflecting MIR radiation from the Sun) and thus, can be mistaken with wildfires (Giglio et al., 1999). Thus, complementary spectral channels (for instance, the MIR/TIR brightness temperature[21] difference between 3.9-μm and 11-μm channels) can be used to confirm whether each hot-spot pixel does indeed contain an actively-burning fire. This technique is known as the *contextual algorithm*, in which the thresholds for detection are obtained through a statistical analysis of the background environment (Li et al., 2002; Giglio et al., 2003; Wooster et al., 2005; Calle and Casanova, 2012). It is worth noting that the view zenith angle at which the ground surface is observed also affects fire detection capacities in TIR regions (Boles and Verbyla, 1999; Paugam et al., 2013).

1.4.2 Geo-location of time-evolving fire fronts

Beyond fire detection, remote sensing is regarded as a promising approach to provide a quantitative description of the fire radiation release to characterize sub-pixel fires and to estimate fuel consumption as well as smoke emissions (Kaufman et al., 1998; Wooster et al., 2003). These information are crucial to track the time-evolving fire front location and to quantify the impact of uncontrolled biomass burning on the Earth system that is recognized as major source of atmospheric gas pollutants and aerosol emissions.

↪ Fireline intensity

In addition to the ROS, a key parameter to characterize regional-scale wildfires is the fireline intensity I_{fr}, because the rapidity and completeness of combustion largely varies with intensity and because fires of different intensities release pollutant emissions at different rates. Defined by Byram (1959) as being the effective heat release rate per unit length of the fireline, the fireline intensity I_{fr} [W/m] reads:

$$I_{fr} = \Delta h_c \times m_v^{''} \times \Gamma, \tag{1.2}$$

with Δh_c [J/kg] the biomass fuel heat release (commonly assumed static with $\Delta h_c = 18.6 \times 10^6$ J/kg in Rothermel, 1972), $m_v^{''}$ [kg/m^2] the fuel mass consumed in the active flaming zone per unit area, and Γ [m/s] the ROS (Alexander, 1982; Whelan, 1995; Santoni et al., 2010). However, the fireline intensity is difficult to

[21] The brightness temperature can be easily calculated from the spectral radiance measurement (at different wavelengths) using the inverse Planck function (Wooster et al., 1995).

estimate because of uncertainties associated with providing accurate estimates of the biomass fuel load m_v'' in particular.

↪ Fire radiation power (FRP) measurements

Only the radiation fraction of the fire heat release (named fire radiation power and noted FRP [W]) can be detected using remote sensing technologies. Even though the flame front of the fire is mainly a sub-pixel phenomenon in spaceborne data, the MIR region of the electromagnetic spectrum ensures that FRP might be detected in active wildfire areas using airborne (Kaufman et al., 1996; Riggan et al., 2004) or spaceborne (Robinson, 1991; Wooster et al., 2003; Roberts and Wooster, 2008) platforms. In contrast, non-active areas remain blank. Using the Stefan-Boltzmann law, adapted to multi-thermal component situation,[22] the FRP can be theoretically expressed per unit area of the instantaneous ground-field of view, as follows:

$$\text{FRP} = \zeta_{fr}\,\sigma_{sb}\sum_{i=1}^{N_T} A_i T_i^4, \tag{1.3}$$

with N_T the number of separate thermal components in the fire at different temperatures, $\sigma_{sb} = 5.67 \times 10^{-8}$ [J/s/m^2/K^4] the Stefan-Boltzmann constant, ζ_{fr} the constant fire gray-body emissivity (i.e., lower than 1), A_i the fractional area of the i-th surface thermal component within the field of view, and T_i [K] the kinetic temperature of the i-th thermal component.

However, this FRP calculation technique is not reliable for sub-pixel fires observed with spaceborne sensors. As reviewed by Wooster et al. (2005), there exists different techniques to overcome these issues in practice, the bi-spectral approach combining MIR and TIR measurements (Dozier, 1981; Giglio and Kendall, 2001; Riggan et al., 2004), single waveband techniques corresponding to the MODIS approach (Kaufman et al., 1996, 1998; Justice et al., 2002; Giglio et al., 2003; Ichoku and Kaufman, 2005) and the MIR spectral radiance approach (Wooster et al., 2003, 2005), among others. The MODIS approach is based on an empirical relationship between the FRP and the brightness temperature of the fire pixel. Since it was calibrated specifically for the spectral and spatial characteristics of MODIS, this approach cannot be extended to lower or higher spatial resolution imagery without any modification. As discussed by Wooster et al. (2005), the MIR spectral radiance approach is more flexible and is for instance adapted to low spatial resolution imagery (e.g., geostationary satellites that exhibit a high revisit frequency compared to polar-orbiting satellites). This approach is therefore suitable for monitoring wildfire spread at high temporal resolution.

While showing promising results for controlled burns, further validation of MIR-based FRP measurements is required at larger scales. Still, FRP measurements have

[22] A single wildfire generally consists of multiple flaming and smoldering combustion zones, with a range of temperatures fluctuating at small spatial scales.

shown potential to offer new insight into wildfire dynamics (in terms of emissions and propagation), which provide information of primary importance for wildfire monitoring.

▷ **Retrieval of wildfire emissions.** The heat release measured by remote sensors directly results from the amount of energy stored in wildland biomass and released into the atmosphere through combustion processes. These processes also emit trace gases, soot and aerosols, which are essential to quantify at regional and global scales (Miranda et al., 2005, 2008; Strada et al., 2012; Urbanski, 2013). Typically, carbon-based emissions are estimated using the following equation (Seiler and Crutzen, 1980):

$$m_k = \int_{t_{ign}}^{t} \dot{\omega}_k \, \mathrm{d}t = \mathrm{EF}_k \left(\iint_{A_b} (\mathrm{BE} \times m_v'') \, \mathrm{d}x \, \mathrm{d}y \right), \tag{1.4}$$

with:

- m_k [kg] the total mass of species k released in the smoke plume;
- $\dot{\omega}_k$ [kg/s] the corresponding mass production rate of species k;
- t, t_{ign} [s] the current and ignition times of the fire, respectively;
- EF_k [kg (species k)/kg (fuel)] the emission factor for species k (the mass of k produced per unit mass of biomass fuel consumed);
- A_b [m^2] the burnt area;
- m_v'' [kg/m^2] the biomass fuel load (the mass of available fuel per unit wildland surface area);
- BE [kg (burnt fuel)/kg (available fuel)] the burning efficiency corresponding to the fraction of available biomass fuel actually pyrolyzed and burned over the fire duration.

The rate of biomass fuel consumption can be directly inferred through FRP measurements, since there is a semi-empirical linear correlation between the FRP and the biomass fuel consumption rate (Wooster et al., 2005; Freeborn et al., 2008). The coefficient of proportionality is referred to as the *combustion factor* (noted CF). In practice, the emission of a gas compound k defined in Eq. (1.4) is therefore estimated based on the measured FRP as follows:

$$m_k = \int_{t_{ign}}^{t} \dot{\omega}_k \, \mathrm{d}t = \mathrm{CF} \times \mathrm{EF}_k \int_{t_{ign}}^{t} \mathrm{FRP} \, \mathrm{d}t, \tag{1.5}$$

in which the combustion factor CF [kg/J] corresponds to the inverse of the radiation energy emitted by the fire per unit mass of fuel consumed. Thus, FRP measurements provide direct access to biomass fuel consumption and conveniently avoids uncertainties associated with the terms m_v'' and BE in Eq. (1.4). However, the FRP method relies on prior knowledge of the combustion factor CF and the

implicit assumption that this factor can be treated as a (calibrated) constant that is independent of fuel properties and fire conditions. To summarize, FRP measurements are essential to estimate carbon emissions from wildfires (Lentile et al., 2006; Roberts and Wooster, 2008).

▷ **Geo-location of active fire areas.** Since FRP measurements retrieve brightness temperatures to calculate wildfire emissions, they are also valuable to geo-locate spatio-temporal locations of the flame front and thus, derive the time-series of wildfire ROS. For instance, Paugam et al. (2013) used FRP measurements to track the temporal evolution of the flame front location on a thousand-meter-square controlled burn. This study showed in particular that FRP measurements based on the MIR radiance approach are able to properly capture the flaming zone (i.e., zone where the brightness temperature is above 700 K) and to dissociate it from burnt areas located immediately behind the fire front (i.e., zones of brightness temperatures between 500 and 600 K). Figure 1.27(a) presents the location of temperature iso-contours for one instantaneous snapshot of the controlled burn, while Fig. 1.27(b) presents the ROS values reconstructed along the flaming zone over the fire duration. Although further validation is required at larger scales, this study demonstrates the potential of FRP measurements for wildfire spread monitoring.

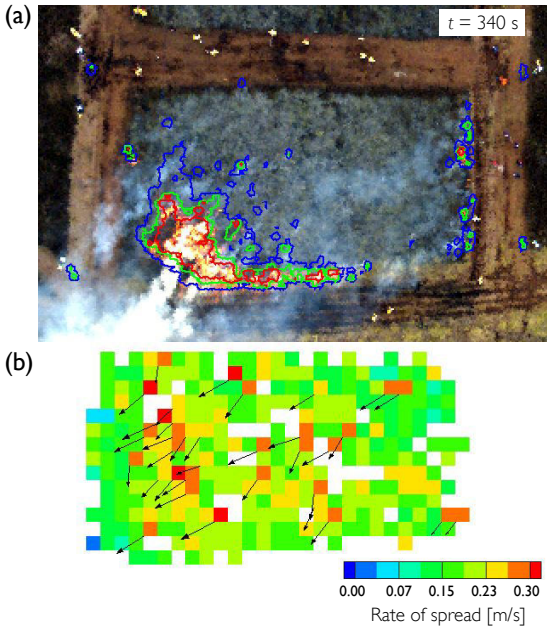


Figure 1.27: *Field-scale heather controlled burn in Northumberland (March 2010, UK). (a) Snapshot of the field-scale experiment ($45\ m \times 21\ m$) at time $t = 340$ s, where the blue line indicates the 500 K iso-temperature, the green line indicates the 600 K iso-temperature (assumed to represent the location of the fire front), and the red line indicates the 700 K iso-temperature. (b) Map of fire front ROS over the controlled burn (overlain vectors indicate local ROS above 0.25 m/s) reconstructed based on the displacement of the 600 K iso-contour. Credit: Paugam et al. (2013).*

1.4.3 Overview of remote sensing systems

The use of MIR and TIR imaging aboard airborne and spaceborne platforms allows for the detection and monitoring of active fire areas. A (non-exhaustive) list of current observation capabilities and emergency response services follows (San-Miguel-Ayanz et al., 2005; Doche et al., 2012; Wooster et al., 2013).

↪ Spaceborne systems

Spaceborne systems currently in orbit provide information on a wide range of wildfire features: the location of active fires, the mapping of burned areas, the tracking of trace gas and aerosol emissions, etc. A list of these spaceborne systems along with their related technical capabilities is presented in Table 1.3. These satellite systems exhibit different monitoring capabilities in terms of the following features:

- **spatial resolution**, corresponding to the pixel size of the instantaneous field of view of the sensor (i.e., the surface area being measured on the ground);
- **spectral resolution**, specifying the number and size of the wavelength intervals that the sensor detects;
- **temporal resolution**, referring to the amount of time between two successive observations of the same location of the Earth surface by the remote sensor;
- **swath**, corresponding to the spatial width of a single pass over the Earth surface.

▷ **Pioneer generation of sensors.** Operating aboard the NOAA (National Oceanographic and Atmospheric Administration) satellites and the EUMETSAT polar system,[23] the AVHRR (Advanced Very High Resolution Radiometer) sensor has been used as a research platform for the development of hot-spot detection algorithms (Arino and Mellinotte, 1998; Li et al., 2000, 2002) due to its high temporal resolution (among the polar-orbiting sensors). It is characterized by a 1.1 km spatial resolution in red, NIR and TIR spectral channels.

Based on the developments due to the AVHRR sensor, MODIS has significantly contributed to the emergence of an operational service for fire detection, in particular due to its high spectral resolution (Ichoku et al., 2003). This sensor operates aboard Terra and Aqua polar sun-synchronous orbiting satellites at 720 km altitude, with the capability of four daily revisits. It is composed of 36 spectral bands ranging from 0.4 to 14.4 μm (from visible through thermal-infrared imaging), with a spatial resolution varying between 250 m (red and NIR channels) and 1 km (MIR channels) at nadir.[24] While 1-km resolution channels operate to deliver the MODIS Active Fire Product[25] using the contextual algorithm due to Giglio et al. (2003), the 250-m resolution channel is used to reject false hot-spot alarms and to mask

[23] www.metops10.vito.be/index.html.
[24] Direction vertically downward from the observer.
[25] modis-fire.umd.edu/.

cloud. This product is made available through the fire information for resource management system (FIRMS)[26] within a few hours after data acquisition. As for the 500-m resolution channel, sensitive to char and ash deposits in the vegetation (by locating the occurrence of rapid changes in daily surface reflectance), it is used to deliver the MODIS Burnt Area Product and allows for the accurate mapping of fires of approximately 50 hectares or larger (within 3 or 4 days). These fire products provide an important contribution to the international GOFC project[27] and to EFFIS[28] through the Active Fire Detection and Rapid Damage Assessment modules.

▷ **High-spatial resolution platforms.** While MODIS is limited by its spatial resolution, a new generation of infrared sensors such as BIRD (Bi-Spectral InfraRed Detection) has been developed to explore the potential of high spatial resolution imagery for tracking hot-spots (e.g., wildfires, volcanoes). While equipped with a hot-spot recognition system based on a 370-m spatial resolution in MIR and TIR channels, BIRD has demonstrated its ability to detect small-scale fires down to 15 hectares (Briess et al., 2003), albeit at a much lower revisit frequency. This illustrates the difficulty to combine requirements of high spatial resolution and wide swath. The latter is essential for a high revisit frequency of the same location on the Earth surface. However, high spatial resolution systems exhibit in general a narrower swath. Still, such high spatial resolution systems play an important role in validating fire products of lower spatial resolution such as MODIS (Wooster et al., 2003) as well as in improving wildfire damage assessment over longer temporal scales. Typically, the additional Damage Assessment module in the EFFIS European fire system can be delivered with a higher level of details using AWiFS (Advanced Wide Field Sensor), a high spatial resolution system operating at a 56 m spatial resolution with four spectral bands (green, red, NIR and SWIR channels) aboard the polar-orbiting Indian Remote Sensing (IRS) satellite.[29] The performance of AWiFS-based burnt scar mapping has been demonstrated in Sedano et al. (2013) to map the impact of wildfires larger than 10 hectares over the fire season (corresponding to 90 % of wildfires in Europe) using the example of the 2007 Greece wildfires.

Alternative high spatial resolution systems are SPOT (Système Pour l'Observation de la Terre) satellites, RapidEye, DMC[30] or the Pléiades constellation. For instance, SPOT satellites operate along a polar sun-synchronous phased-orbit and exhibit a spatial resolution ranging between 1 and 20 m, depending on the satellite generation and on the spectral channel. As for the Pléiades constellation (i.e., Pléiades 1-A and 1-B), it offers a sub-metric spatial resolution (0.70 m at nadir), with a narrower field of view than SPOT satellites, but providing a detailed mapping of the Earth surface and possibly of wildfire spread.

[26] earthdata.nasa.gov/data/near-real-time-data/firms.
[27] Global Observation of Forest Cover, www.fao.org/gtos/gofc-gold/.
[28] forest.jrc.ec.europa.eu/effis/about-effis/technical-background/.
[29] www.isro.org/satellites/earthobservationsatellites.aspx.
[30] Disaster Monitoring Constellation, www.dmcii.com/.

▷ **Geostationary platforms.** Due to their high temporal resolution (in contrast to polar-orbiting satellites that exhibit a revisit frequency limited to a few times per day at most locations), geostationary satellites are useful for wildfire detection. For instance, GOES (Geostationary Operational Environmental Satellites) and the SEVIRI (Spinning Enhanced Visible and Infared Imager) instrument aboard MSG (Meteosat Second Generation) acquire images every 15 min, at up to 1 km resolution in the visible spectrum and 4 km resolution in the thermal-infrared spectrum (Prins and Menzel, 1994; Prins et al., 2004; Calle et al., 2006). They are part of the global geostationary system for fire monitoring (named GOFC/GOLD fire mapping and monitoring program) with the Japanese MTSAT (Multifunctional Transport Satellite). However, their spatial resolution remains limited and is not suitable for tracking the time-evolving location of fire flame fronts. Still, they are used for detecting wildfires and estimating emissions due to biomass burning (Calle and Casanova, 2012). Note that the IASI[31] instrument aboard EUMETSAT polar-orbiting satellites could be a valuable remote sensor for measuring amounts of trace gas compounds (e.g., ozone, CO) in the atmosphere due to its high spectral resolution (i.e., 0.25 cm^{-1}). However, its coarse spatial ground resolution (i.e., 12 km at nadir) makes IASI focus on atmospheric observations rather than surface observations (Coheur et al., 2009).

▷ **Operational services of fire monitoring.** One strategy to overcome the limited revisit frequency of high spatial resolution imagery and to allow for regional-scale wildfire spread monitoring with high spatial and temporal resolutions over the fire duration is to combine information coming from an ensemble of satellites. At European scale, a wide range of initiatives supported by the European Space Agency (ESA) and the European Commission (EC) has led to a strong background in emergency fire mapping.

- At the initiative of ESA and CNES (the French space agency), the International Charter *Space and Major Disasters*[32] constitutes a unique worldwide system for disaster response, operating since 2000 and relying on operational satellites and cooperation among space agencies. A 24-hour on-duty rapid mapping service (named *Emergency Mapping* and used as a decision support tool by civil defense agencies) is specifically dedicated to regional-scale wildfire monitoring and is notably supported by SERTIT[33] and ZKI.[34] This rapid mapping service consists in delivering geo-referenced maps of the specific natural hazard within six hours after reception of spaceborne data (e.g., SPOT1-3, DMC, Pléiades 1-A/1-B, BIRD, RapidEye, SPOT-4) as explained by Sarti et al. (2005) and Clandillon and Yesou (2011). An example of SERTIT rapid

[31] Infrared Atmospheric Spectrometer Interferometer, smsc.cnes.fr/IASI/.

[32] www.disasterscharter.org/.

[33] Service Régional de Traitement d'Image et de Télédétection, Univ. de Strasbourg (France), sertit.u-strasbg.fr/RMS/.

[34] Center for satellite-based crisis information at DLR, the German aerospace center, www.zki.dlr.de.

mapping for 2011 La Réunion wildfires is presented in Fig. 1.28.

- Within the framework of the European project Copernicus,[35] supported by EC and operated by a consortium led by e-GEOS, a network of real-time geo-location services to monitor natural hazards such as floods and forest fires has been developed. In this context, geo-referenced maps are delivered within 24 hours after reception of spaceborne data.

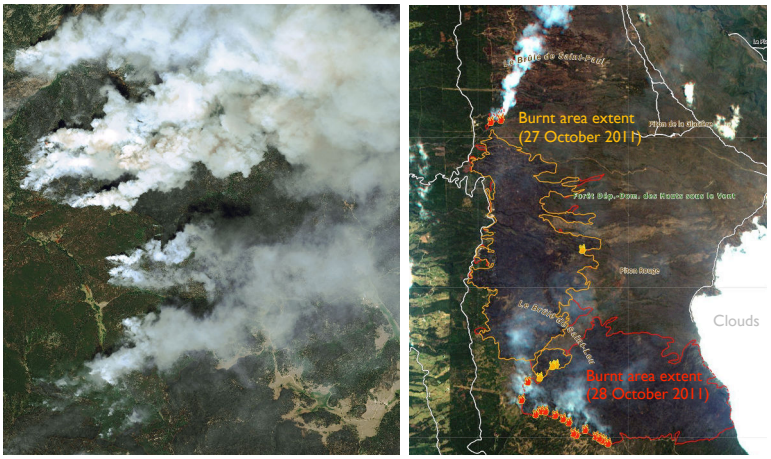


Figure 1.28: *Examples of wildfire spread monitoring using spaceborne data. Left: Pléiades 1-A optical imaging for 2012 Colorado wildfires. Credit: CNES. Right: RapidEye imaging with a 6.5 m spatial resolution for 2011 La Réunion wildfires. The orange line indicates the location of the fireline on October, 27th (710 ha), and the red line indicates the location of the fireline on October, 28th (1408 ha). Credit: SERTIT.*

↪ Airborne systems

Airborne fire detection usually relies on human surveillance from airplanes flowing at high altitudes. However, new airborne systems including onboard data processing and automated geo-location of active fire areas (using MIR channels along with global-positioning system) are currently investigated for operational applications with a spatial resolution finer than 50 meters and with a very high revisit frequency (Riggan and Hoffman, 2003; Riggan et al., 2004; Riggan and Robert, 2009). Airborne platforms can operate almost continuously over the battery life and map the fire propagation every few minutes, albeit at a logistical and financial cost. Note that issues concerning geo-location and calibration are more important than for spaceborne platforms. Furthermore, airborne platforms are useful to validate future spaceborne instruments and to develop detection/monitoring algorithms. Typical examples are the Parefeu program[36] supported by CNES (2003-2004) and the Livefire system (Merlet, 2008; Crombette, 2010), whose flowchart is presented in Fig. 1.29.

[35] www.emergencyresponse.eu/gmes/.
[36] www.pont-entente.org.

The US counterpart of the Livefire system is the FireMapper[37] LWIR sensor that has been operationally deployed since 2004 by the US Forest Service in collaboration with the US Department of the Interior Bureau of Land Management (Riggan and Robert, 2009). Figure 1.30 highlights the differences in spatial resolution between MODIS and FireMapper products: the former precisely locates the instantaneous flame zone, the latter provides a global viewpoint of the fire situation.

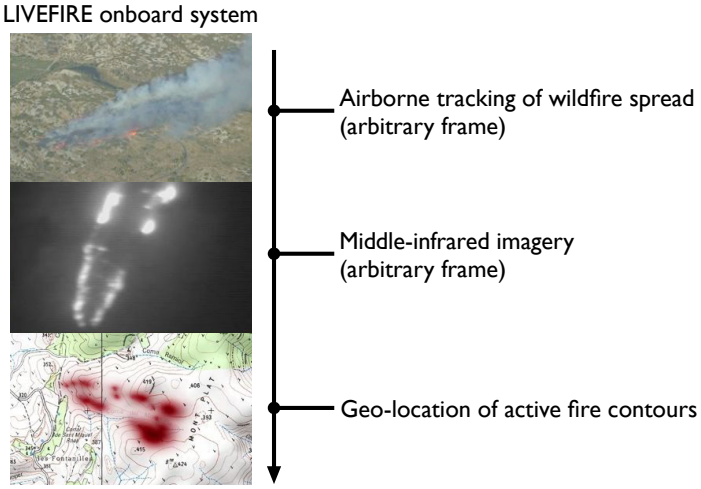


Figure 1.29: *Schematic of the* Livefire *system aboard the monitoring airplane Horus-66 that has a battery life of 6 hours (Pyrénées-Orientales, France). Credit: i-Tolosa,* www.itolosa.fr/.

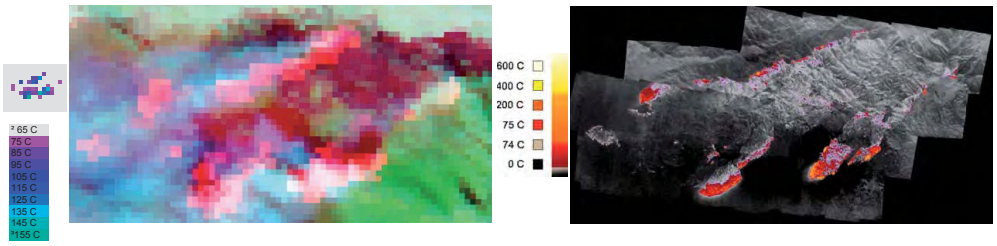


Figure 1.30: *Comparison of MODIS and FireMapper imaging for Esperanza fires (26 October 2006, California). Left: MODIS brightness temperature using blue/green and NIR channels (satellite). Right: FireMapper surface temperature (in Celsius) measured at 11.9 μm wavelength (airplane). Credit: Riggan and Robert (2009).*

This overview shows that recent progress made in airborne/spaceborne remote sensing provides new ways to examine wildfire behavior (in terms of radiant intensity, carbon mass flux and sensible heat flux, fuel consumption, etc.). In particular, high spatial resolution systems (at a scale on the order of a few meters) can locate the high spectral radiance typical of wildfires and contribute to monitor real-time locations of flame fronts.

[37] www.fireimaging.com.

Table 1.3: *Recap of spaceborne sensors aboard geostationary or polar-orbiting satellites.*

Name	Satellite type	Channels	Spatial resolution	Temporal resolution	Website
GOES	geostationary	visible	1 km	15 min	www.goes.noaa.gov/
SEVIRI	geostationary	• visible/NIR • MIR/TIR	• 1 km • 3 km	15 min	wdc.dlr.de/sensors/seviri/
MODIS	orbiting	• visible • MIR/TIR	• 250 m • 1 km	4 times a day	modis.gsfc.nasa.gov
SPOT					smsc.cnes.fr/SPOT/
1-3	orbiting	• yellow/green/red • NIR	• 10 m • 20 m	1-3 days	
4-5	orbiting	• yellow/green/red • NIR/SWIR	• 10 m • 2.5 m	1-3 days	
6-7	orbiting	• blue/green/red • NIR	• 6 m • 1.5 m	daily	
Pléiades 1A/1B	orbiting	blue	0.70 m	twice daily	/smsc.cnes.fr/PLEIADES/
BIRD	orbiting	NIR/MIR/TIR	370 m	experimental imaging	www.dlr.de/sc/en/desktopdefault.aspx/tabid-1262/1765_read-3202/
AWiFS	orbiting	green/red/NIR/SWIR	56 m	4 days	earth.esa.int/web/guest/-/awifs-data-products-1659
DIEMOS (DMC)	orbiting	green/red/NIR/SWIR	32 m	daily	www.dmcii.com/?page_id=6770
RapidEye	orbiting	blue/green/red/red-edge/NIR	6.5 m	daily (off-nadir), 5.5 days (nadir)	www.rapideye.com/products/
Sentinel-2	orbiting	blue/green/red/red-edge/NIR	10-25 m	future	earth.esa.int/web/guest/missions/esa-future-missions/sentinel-2

1.5 Thesis overview

The complexity of regional-scale wildfire spread requires a multi-disciplinary approach to address issues related to wildfire forecasting, in terms of burnt area extents and atmospheric emissions. Given the large number of processes and factors that occur in wildfires, covering several orders of magnitude in spatial and temporal scales, the collaboration of scientists with different specialized backgrounds is required to introduce a paradigm-shift in wildfire spread modeling. This could, in the long run, be part of the firefighter training as well as strengthen civil defense and environmental protection in fire prevention planning and emergency responses. Developing a novel approach of wildfire spread modeling is the purpose of the present Ph.D. thesis entitled *Towards a more comprehensive monitoring of wildfire spread: model evaluation and data assimilation strategies*.

This Ph.D. thesis was funded by the *Agence Nationale de la Recherche* (ANR) under the IDEA (*Incendies de forêts: simulation de la dynamique et des émissions atmosphériques par couplage de code*[38]) project grant ANR-09-COSI-006 (2010-2013). It was also supported by a LEFE-ASSIM grant (INSU-CNRS program named *Les Enveloppes Fluides de l'Environnement*, 2011-2013).

1.5.1 Scope of the thesis

The general objective of this Ph.D. thesis is to demonstrate the feasibility of using fire sensor technology combined with fire modeling software for real-time in-situ analysis, in order to improve real-time forecasts of wildfire evolution. From a technical perspective, it consists in what we consider a very innovative application of advanced methods for coupling observations and models that are developed in related scientific and engineering fields (e.g., numerical weather forecasting) to the area of combustion and fire science. These advanced methods are referred to as *data assimilation methodologies*. These methodologies are efficient to reduce uncertainties in the system predictions, especially if the underlying computer model properly captures fine-scale features of the system dynamics. Thus, the improvement of combustion models through the detailed analysis of flame-scale processes is also an important component towards an accurate data-driven wildfire spread simulator.

This Ph.D. thesis is part of an emerging collaborative, multi-discplinary, international research program between *Energétique Moléculaire et Macroscopique, Combustion*[39] (EM2C) CNRS laboratory at Ecole Centrale Paris, *Centre Européen de Recherche et Formation Avancées en Calcul Scientifique*[40] (CERFACS), and the Department of Fire Protection Engineering[41] at the University of Maryland (UMD). In

[38] anridea.univ-corse.fr/.
[39] www.em2c.ecp.fr/.
[40] www.cerfacs.fr/.
[41] www.fpe.umd.edu/.

this collaboration, EM2C brings extensive experience in combustion physics, combustion modeling and CFD; CERFACS brings experience in cutting-edge massively parallel scientific computing, CFD and data assimilation, while UMD brings experience in fire physics and fire modeling.

To overcome some of the limitations of wildfire spread modeling, this Ph.D. research includes the development of data-driven simulations of regional-scale wildfire spread through the application of a data assimilation methodology, and a critical analysis of available ROS models through LES of flame-scale fire spread configurations. The two main goals of this thesis can be summarized as follows.

▷ PART I - Insight from multi-physics flame-scale large-eddy simulations

Because wildfire dynamics features complex multi-physics occurring at multiple scales, our ability to accurately simulate the behavior of wildfires remains limited. The semi-empirical modeling approaches used at operational levels are limited, partly because of their inability to cover the entire range of relevant scales and their inaccuracies in the description of the physics. As for physical modeling approaches, they are too computationally-intensive to provide real-time predictions and still do not account accurately for all the relevant physical processes. As highlighted by Linn et al. (2005), only large-scale effects are explicitly solved in current wildfire LES (e.g., FIRETEC, WFDS), while fine-scale temperature distributions, mixing, turbulence, chemistry and combustion are considered as subgrid-scale processes and require additional models that must be further improved (Zhou et al., 2005, 2007).

To overcome these modeling issues, the ANR-IDEA project was aimed at developing a scientific computing platform for simulating regional-scale wildfire spread. This platform is based on the coupling of the wildfire spread model FOREFIRE and the meso-scale atmospheric solver MÉSO-NH. Its purpose is to address the wide range of scales involved in a wildfire, from the vegetation scales (less than 100 m) that are essential to describe fine-scale combustion processes, to the atmospheric scales (more than 100 km) that are of primary importance to account for the coupling between combustion, flow dynamics and atmospheric dynamics. Note that this project was piloted by the CNRS-SPE laboratory of the University of Corsica (J.-B. Filippi), with the following institutional partners: LA (CNRS/Paul Sabatier University, Toulouse), CERFACS, EM2C, CNRM (CNRS/Météo-France), INRIA and M2P2 (CNRS/University of Marseille). Within the ANR-IDEA project, the following issues have been addressed:

- **Vegetation scales.** Equation (1.4) is the classical way to estimate the amount of emissions of a specific gas compound k. However, this empirical calculation requires explicit knowledge of biomass fuels properties and consumption (e.g., flaming or smoldering combustion modes), through the burning efficiency BE and the emission factor EF_k. These parameters are subject to significant uncertainties; thus, Eq. (1.5) could be an alternative tech-

nique to estimate fire-induced emissions through FRP measurements. While megafires are expected to occur more frequently due to climate change and to induce large-scale air quality issues, there is a growing need to characterize gases that are released by the vegetation thermal degradation and that are partially consumed by flaming combustion (Pérez-Ramirez et al., 2012). These gases affect combustion dynamics and also determine the composition of the smoke plume (e.g., carbon monoxide CO, carbon dioxide CO_2, nitric oxide NO, sulphur dioxide SO_2, methane CH_4, hydrogen cyanide HCN, acetonitrile CH_3CN), which potentially extends over hundreds of kilometers in the atmosphere. Field-scale campaigns and measurements (Miranda et al., 2005) have been performed to determine emission factors corresponding to Mediterranean biomass fuels. However, it is evident that the list of emission factors still remains incomplete.

- **Flame scale.** The knowledge on flame-scale processes (resulting from multi-scale interactions between pyrolysis, combustion and flow dynamics) remains limited due to their high complexity as well as to their spatial and temporal variability. However, these dynamics are of primary importance to determine wildfire propagation and the resulting emissions into the atmosphere. While current operational fire spread simulators rely on empirical formulas for predicting the wildfire ROS or intensity, new physics-based approaches emerge and attempt to understand the controlling processes in a wildfire (e.g., wind, buoyancy-induced flow, combustion, thermal radiation and degradation of the vegetation). Currently, the limitations induced by insufficient computational capacities (for instance, resolving the processes of ignition would required a computational mesh resolution on the order of 1 mm) are addressed by subgrid-scale modeling as in industrial combustion applications. This subgrid-scale modeling needs to be improved through detailed analysis of wildfire spread, in order to build simulations that are more physically-consistent.
- **Surface/atmosphere interactions.** In a coupled system including surface wildfire spread and meso-scale atmospheric dynamics, interface conditions constitute a critical component. Surface forcings induced by FOREFIRE are prescribed as heat and vapor fluxes in the boundary layer of the meso-scale atmospheric model MÉSO-NH, through the surface modeling platform SURFEX[42] (see Fig. 1.31). Since FOREFIRE relies on a front-tracking approach and a semi-empirical ROS model, surface fluxes need to be represented using additional physical modeling. For this purpose, reduced models for combustion can be developed (Morvan, 2011).
- **Atmospheric scale.** The impact of a wildfire at atmospheric scales (over tens to hundreds of kilometers) is a direct result of the interaction between atmospheric dynamics and the ongoing chemical kinetics in the smoke plume (e.g., presence of CO, ozone titration and production). For this purpose, a

[42] www.cnrm.meteo.fr/surfex/.

specific chemical kinetic module is required in the meso-scale atmospheric model MÉSO-NH (Strada et al., 2012).

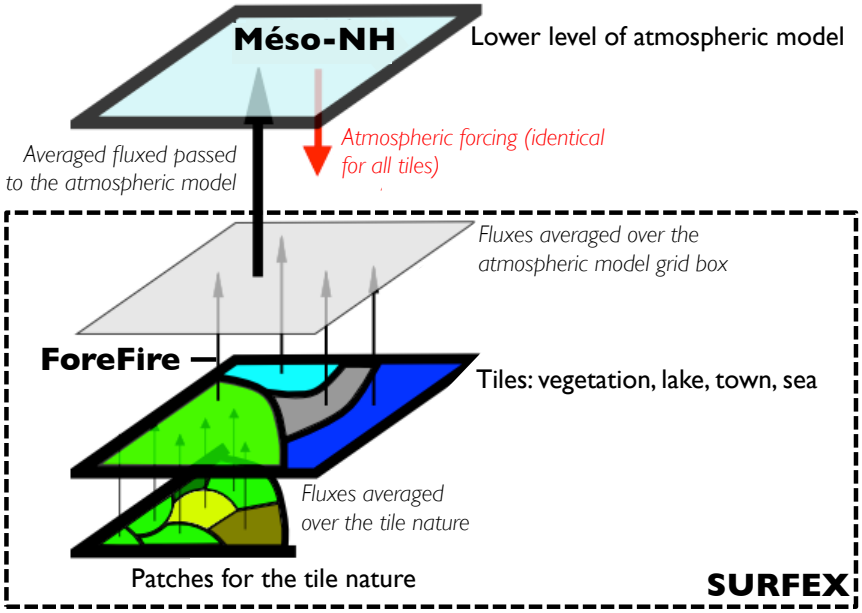


Figure 1.31: *Schematic of SURFEX tiling and coupling with* MÉSO-NH *meso-scale atmospheric model. Credit: Météo-France.*

The validation of this coupled fire/atmosphere platform has been performed through ensemble runs to characterize modeling errors over a large number of wildfire cases (Filippi et al., 2013). A typical example of coupled FOREFIRE/MÉSO-NH simulations is presented in Fig. 1.24.

Within the framework of the ANR-IDEA project, the first part of this Ph.D. thesis aims at bringing CERFACS/EM2C extensive expertise in combustion modeling (developed for industrial combustion applications) to the wildfire spread research field. As shown in Fig. 1.32, this work is aimed in particular at:

(1) investigating the thermo-chemical structure of typical wildfire flames to offer insight into the composition of the smoke plume (**vegetation-scale issue**);

(2) analyzing flame-scale interactions between pyrolysis, combustion and turbulence as well as the delicate balance between convection and radiation heat transfer (**flame-scale issue**).

For this purpose, multi-physics flame-scale LES of fire spread are performed at laboratory flame-scales and are compared to measurements to provide a comprehensive understanding of the mechanisms underlying fire spread. In particular, the assumptions used to estimate the ROS are examined in detail and some insights into the flame-induced air entrainment are provided through the analysis of particle image velocimetry (PIV) experimental measurements. To our knowledge, it is the

first time that a LES approach solving for the flame structure (at a spatial resolution on the order of 1 mm) and including radiation heat transfer and pyrolysis, is applied to natural fire propagation. This modular physics-based approach, allowing for continual improvement of process level models and relying on high-performance computing, constitutes a research tool for exploration of wildfire behavior and sensitivity to environmental conditions. In particular, this approach is a promising approach to better parameterize semi-empirical ROS formulations as a function of a reduced number of dominant factors (e.g., vegetation properties, weather conditions, terrain topography), which could then be used to regional-scale predictions of wildfire spread.

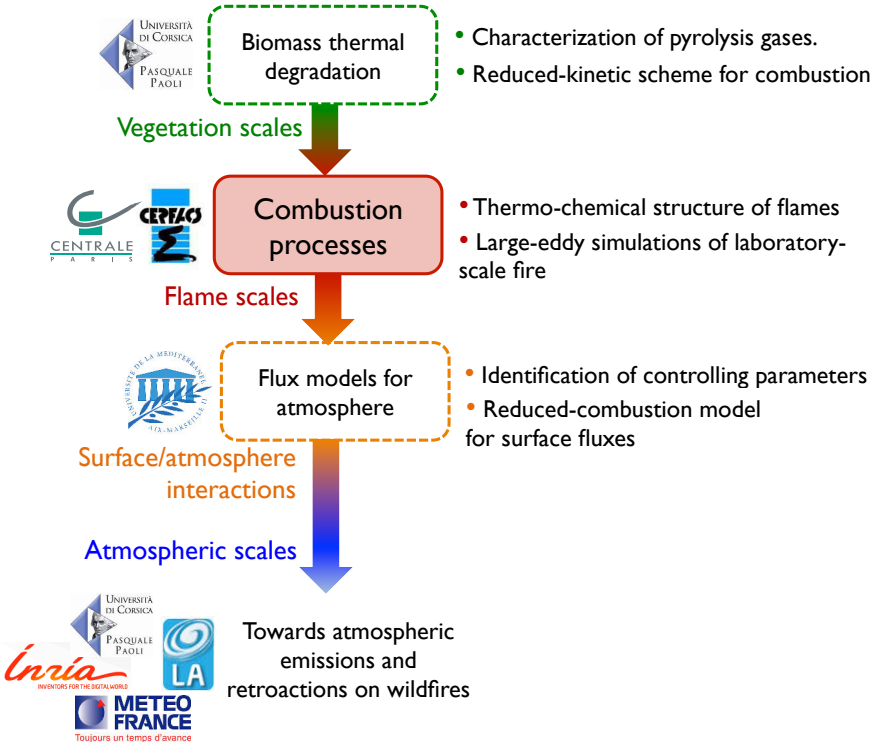


Figure 1.32: *Summary of the ANR-IDEA project with thesis objectives presented under the label Combustion processes.*

▷ PART II - Data assimilation for regional-scale wildfire spread forecast

A further strategy to overcome wildfire modeling limitations consists in coupling information coming from both measurements and computer model, taking into account that none of them, when used alone, provides a certain and complete description of the physical system. The idea is to use observations to improve the estimate of the set of parameters, initial/boundary conditions, or model state for the computer model, and improve its accuracy and high-fidelity. While still original in the field of fire and combustion, data assimilation provides a powerful framework

to perform such combination of remote sensing measurements and computer model, while explicitly accounting for the effects of both observation and modeling errors. The benefit of data assimilation has already been greatly demonstrated in numerical weather forecasting over the past decades. It fits into the wider domain of dynamic data-driven application systems, where data are used to formulate some feedback information on the physical system, leading to the reduction of uncertainties on the model and its predictions.

In the second part of this Ph.D. thesis, data-driven modeling is proposed as one of the two cornerstones of a wildfire spread forecasting capability, producing accurate predictions of the time-evolving location of the flame front with positive lead-times (consistent with operational applications) and without loss of accuracy. The other cornerstone corresponds to the integration of a variety of in-situ and remote sensors providing (real-time) information on fire location, vegetation, terrain topography and atmospheric conditions. The focus here is on the development and validation of data assimilation algorithms applied to a regional-scale wildfire spread simulator named FIREFLY. The problem of availability of the fire front observations and geographical-information-system-based information for the ROS model parameters is outside the scope of this research. In the continuation of Rochoux et al. (2010) that demonstrated the potential benefits of data assimilation for wildfire spread (albeit in much simplified configurations), the present work aims at extending this approach to more realistic configurations. Some critical questions must be addressed:

- Which type of observations is suitable for data assimilation (in terms of spatial and temporal resolutions for instance)?
- What are the main uncertainties in a wildfire spread simulator? To which parameters is wildfire spread sensitive?
- Which data assimilation algorithm is consistent with the features of the wildfire spread simulator (e.g., problems of non-linearities and non-Gaussian error statistics)?

In order to properly quantify modeling/observation uncertainties and to account for model non-linearities, ensemble-based data assimilation algorithms are explored in this work. This work has been developed within the framework of a collaboration between CERFACS and UMD. The main components of the proposed prototype data-driven simulator are: a regional-scale perspective in which the propagation of the fire front is described via a local description of the ROS; a semi-empirical model description of ROS (possibly modified through the analysis of flame-scale LES); a level-set-based solver for the fire propagation model (also referred to as the *forward model* and named FIREFLY); an assumed set of real-time observations of the flame front location; and ensemble-based data assimilation algorithms (also referred to as *inverse modeling procedure*). The innovation lies in the original application of these data assimilation algorithms to cases featuring variables conditions, including spatial variations in fuel properties and topography, as well as temporal variations in the

wind intensity. These experimental conditions are of great theoretical interest and are representative of the variability in vegetation, topographical and meteorological properties found in real-world applications.

The first objective is to evaluate the ability of the prototype data-driven wildfire spread simulator to correctly respond to temporal and/or spatial variations of the environmental conditions. In this case referred to as the *parameter estimation approach*, the estimation targets are the input parameters of the semi-empirical ROS model. The second objective is to evaluate the ability of data assimilation methodologies to reduce modeling uncertainties through the direct correction of the fire front locations. This approach is referred to as the *state estimation approach*. While the estimation targets differ, both approaches are expected to provide more accurate simulations and forecasts of the time-evolving location of the fire front. A validation study corresponding to a controlled burn across a small-scale flat open-field grassland lot is presented to evaluate and compare the performance of data assimilation algorithms. While the proposed fire spread forecasting capabilities are still at an early stage of development, it is envisioned that these future capabilities will be similar to current weather forecasting capabilities and that our general ability to predict the evolution of wildfires will rely on the continuous assimilation of observation data into a cost-effective wildfire simulator. The ultimate goal of this research (beyond this thesis) is to define a prototype data-driven wildfire simulator, able to produce real-time fire forecasts using thermal-infrared imaging data including a description of both wildfire dynamics and fire plume emissions.

1.5.2 Outline for the manuscript

▷ PART I - Insight from multi-physics flame-scale large-eddy simulations

- In **Chapter** 2, turbulent combustion is introduced. Balance equations are presented for reacting buoyancy-induced flows. The different approaches for chemistry description in turbulent combustion, combustion modeling as well as radiation heat transfer are also introduced along with the different CFD tools used in this work.
- We describe in **Chapter** 3 our strategy for building and simulating coupled multi-physics LES that are representative of laboratory-scale fire spread. In particular, we detail and validate the Pyrowo biomass thermal degradation model developed in this work to account for the pyrolysis of the biomass fuel. We also present the undertaken coupling strategy to combine Pyrowo and the Avbp LES solver for the reacting gas phase, including the methane/air flame description and the coupling with the Prissma radiation heat transfer solver. This coupling capability relies on the Open-Palm dynamic coupler.
- **Chapter** 4 presents the results of the multi-physics flame-scale LES for a laboratory-scale fire configuration. We compare these results to laboratory-scale measurements and in particular, we show the feasibility of PIV to mea-

sure flame-induced air entrainment. This study demonstrates the feasibility of this physics-based modeling approach to provide insight into the physical and chemical processes underlying fire spread and thereby, into the ROS formulation.

▷ PART II - Data assimilation for regional-scale wildfire spread forecast

- **Chapter 5** introduces general features on data assimilation, in terms of mathematical formalism and sequential methodologies for the numerical treatment of model non-linearities and of non-Gaussian modeling/observation error statistics. The Kalman filter and its extensions are presented along with an extensive comparison to variational approaches and particle filters.
- The aim of **Chapter 6** is to investigate the suitable strategy of data assimilation for tracking wildfire spread. We provide, in a first step, an overview of the recent efforts made at international level to apply inverse modeling procedures to fires. In a second step, we present our prototype data-driven wildfire simulator to correct inaccurate predictions of the fire front location and to subsequently, provide an optimized forecast of the wildfire behavior. This prototype simulator features a regional-scale wildfire spread model Firefly that deals with wind and complex terrain topography, coupled with a data assimilation methodology suitable for parameter or state estimation through OpenPalm.
- In **Chapter 7**, we conduct a comparative study between different data assimilation algorithms (extended Kalman filter, ensemble Kalman filter and particle filters) to highlight their respective benefits and disadvantages for tracking wildfire spread. Their performance is first evaluated in a series of verification tests using synthetically-generated observations and including configurations with spatially-varying vegetation properties and temporally-varying wind conditions. It is subsequently evaluated in a validation test corresponding to a controlled grassland burn. The data assimilation algorithm also features a choice between a parameter estimation approach in which the estimation targets are the input parameters of the ROS model and a state estimation approach in which the estimation targets are the fire front locations. We demonstrate in these tests the importance of assessing physically-consistent modeling errors to allow for an anisotropic correction of the fire front position, resulting in higher-fidelity data-driven simulations and optimized forecast of the wildfire behavior at positive lead-times.

▷ List of published and submitted articles

- Rochoux et al. (2014a, NHESS) - Rochoux, M.C., Ricci, S., Lucor, D., Cuenot, B., and Trouvé, A., *Towards predictive data-driven simulations of wildfire spread. Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation*, Natural Hazards

and Earth System Sciences, Special Issue: numerical wildland combustion, from the flame to the atmosphere, accepted for discussion.

- Rochoux et al. (2014b, NHESS) - Rochoux, M.C., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A., *Towards predictive data-driven simulations of wildfire spread. Part II: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread*, Natural Hazards and Earth System Sciences, Special Issue: numerical wildland combustion, from the flame to the atmosphere, submitted for publication.
- Rochoux et al. (2014, PROCI) - Rochoux, M.C., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A., *Ensemble-based data assimilation for regional-scale simulations of wildland fire spread*, in Proceedings of the Combustion Institute, submitted for publication.
- da Silva et al. (2014, HTHP) - da Silva, W.B., Rochoux, M.C., Orlande, H., Colaço, M., Fudym, O., El Hafi, M., Cuenot, B., and Ricci, S., *Application of particle filters to regional-scale wildfire spread*, High Temperatures-High Pressures, International Journal of Thermophysical Properties Research, under revision.
- Rochoux et al. (2014, IAFSS) - Rochoux, M.C., Emery, C., Ricci, S., Cuenot, B., and Trouvé, A. (2014) *Towards predictive simulation of wildfire spread at regional-scale using ensemble-based data assimilation to correct the fire front position*, in Fire Safety Science - Proceedings of the Eleventh International Symposium, International Association for Fire Safety Science.
- Rochoux et al. (2013, PROCI) - Rochoux, M.C., Delmotte, B., Cuenot, B., Ricci, S., and Trouvé, A. (2013) *Regional-scale simulations of wildland fire spread informed by real-time flame front observations*, in Proceedings of the Combustion Institute, vol. 34, pp. 2641-2647, doi:10.1016/j.proci.2012.06.090
- Rochoux et al. (2013, INCA) - Rochoux, M.C., Cuenot, B., Ricci, S., Trouvé, A., Delmotte, B., Massart, S., Paoli, R., and Paugam, R. (2013) *Data assimilation applied to combustion*, Compte-Rendus Mécanique de l'Académie des Sciences, vol. 341, pp. 266-276, doi:10.1016/j.crme.2012.10.011

The International Forum of Fire Research Directors designated the paper referred to as Rochoux et al. (2014, IAFSS) for the 2014 Sheldon Tieszen Student Award in the area of fire physics and fire modeling, at the 11th International Association of Fire Safety Science Symposium (University of Canterbury, Christchurch, New Zealand in February 2014).

Participation to the 2012 CTR Summer Program at Stanford University, USA, led to the following publication in the CTR Proceedings:

- Rochoux et al. (2012, CTR) - Rochoux, M.C., Ricci, S., Lucor, D., Cuenot, B., Trouvé, A., and Bart, J.-M., *Towards predictive simulation of wildfire spread using a reduced-cost ensemble Kalman filter based on polynomial chaos approximation*, in Proceedings of the Summer Program, Center for Turbulence Research, NASA AMES Stanford University, June/July 2012.

The collaboration between CERFACS (France), Mines Albi (France) and the Federal University of Rio de Janeiro (Brazil) conducted to the following proceeding:

- da Silva et al. (2013, IPDO) - da Silva, W.B., Rochoux, M.C., Orlande, H., Colaço, M., Fudym, O., El Hafi, M., Cuenot, B., and Ricci, S., *Application of particle filters in moving frontier problems*, in 4th Inverse Problems, Design and Optimization Symposium, June 26-28, Albi, France.

Part I

Insight from multi-physics flame-scale large-eddy simulations

Introduction

Given the large number of processes and factors that occur in wildfires, covering several orders of magnitude in spatial and temporal scales, semi-empirical modeling approaches used at operational levels for simulating regional-scale wildfire spread are limited, partly because of their inability to cover the entire range of relevant scales and their inaccuracies in the description of the physics. Real-time predictions of the direction and speed of a propagating wildfire have therefore been identified as a valuable research objective with direct applications in fire emergency response. Beyond the production of carbon dioxide $\mathrm{CO_2}$ and water vapor $\mathrm{H_2O}$, biomass burning constitutes a significant source of air pollutant emissions, among whom carbon monoxide CO, unburnt hydrocarbons (UHC), volatile organic compounds (VOC), nitrogen oxides $\mathrm{NO_x}$, sulfur oxides $\mathrm{SO_x}$ and aerosols (also known as *particulate matter*). These greenhouse and trace gases significantly affect the chemistry and radiation budget of the atmosphere; for instance, they alter the atmospheric composition and air quality by producing harmful ozone $\mathrm{O_3}$ and inducing acid rains. For this purpose, civil defense and environmental-related issues need to be addressed, in particular in terms of fire front tracking and air quality predictions.

To better track regional-scale wildfires and quantify their emissions, biomass combustion processes must be analyzed in details, within the flame zone as well as in the post-flame zone. Modeling fine-scale interactions between surface processes and atmospheric dynamics is also an important component to track the life-time of reacting species, from the combustion zone up to the troposphere, where airborne or spaceborne observations are available. Still, regional-scale wildfire spread models are currently limited to a semi-empirical parameterization of the rate of fire spread (ROS), see Eq. (1.1), and emission models rely on the specification of a burning efficiency coefficient and emission factors for the targeted chemical compounds, see Eq. (1.4), Chapter 1.

A possible approach to partly overcome the limitations found in wildfire modeling and to improve the knowledge on the multi-scale interactions between pyrolysis, combustion and flow dynamics is to develop flame-scale numerical simulations of fires (with a spatial resolution on the order of 1 mm). These simulations could provide insight into the ROS sensitivity to environmental conditions, but also into the fireline intensity, the vegetation/flame heat exchanges as well as on the composition of burnt products released by the flame into the atmospheric boundary layer. These information are essential for better characterizing wildfires, quantifying real-time emissions and assessing the long-term impacts of wildfires on the atmosphere. They are also of primary importance to evaluate the assumptions underlying semi-empirical ROS modeling since the Rothermel's model expresses the ROS as the ratio between the heat flux received by the unburnt vegetation and

the energy required to ignite the biomass fuel. The submodels for these quantities were calibrated using one-dimensional tunnel experiments, but detailed flame-scale simulations could directly solve for these quantities and thereby, evaluate the range of validity of these submodels. In particular, such detailed approaches could benefit to regional-scale wildfire spread models by improving the model parameterization of the fireline ROS. They can also be viewed as a promising approach to address modeling uncertainties in regional-scale wildfire spread models such as FIREFLY.

Numerical simulations of biomass combustion for turbulent flows belong to the wider area of computational fluid dynamics (CFD) and have recently demonstrated their potential to improve combustion-based technologies. This, in order to reduce fuel consumption in industrial applications (such as vehicles, industrial plants and furnaces, gas turbines, rocket propulsion, domestic boilers, etc.), consistently with European and international regulations. Similar CFD simulation techniques can be developed in wildfire research since they share common features such as phase-change, combustion/turbulence interactions as well as the importance of heat transfer. However, their application to wildfires is not straightforward for two main reasons illustrated in Fig. 1.33:

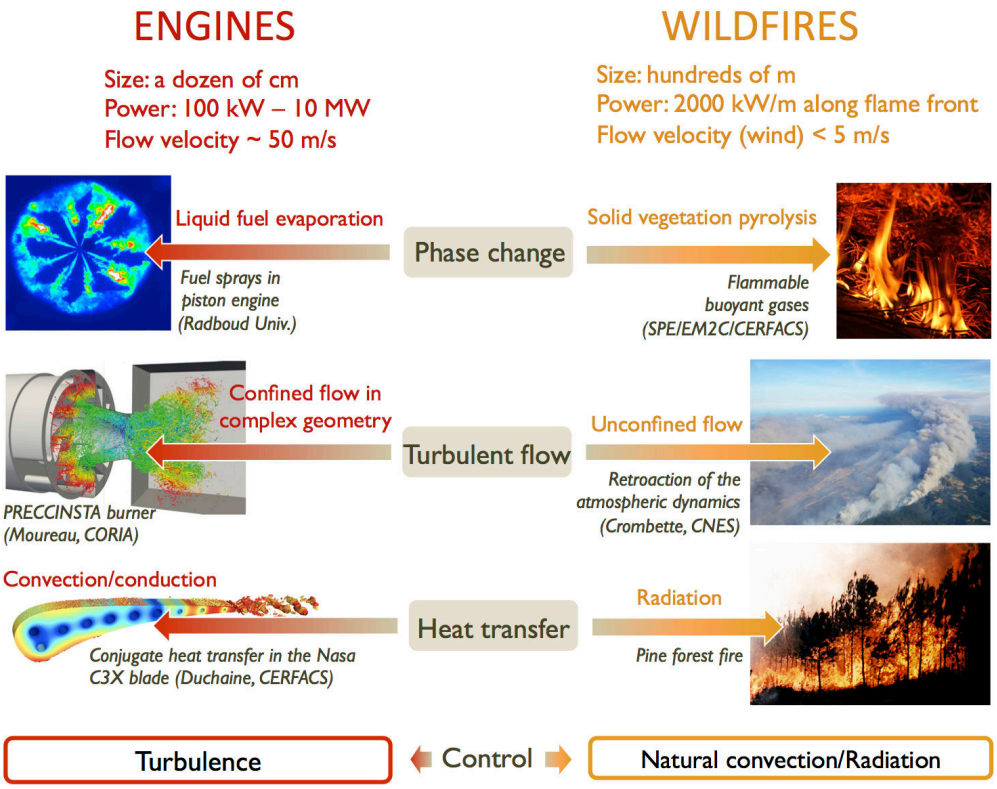


Figure 1.33: *Comparison of gas engines and wildfires (buoyancy-driven) combustion processes, sharing three common features: phase change, interactions between the turbulent flow and combustion as well as heat transfer.*

(i) The characteristic time- and length-scales of combustion processes in combustion chambers and wildfires are not of the same order of magnitude. In particular, wildfires propagate in an unconfined environment.

(ii) The controlling processes differ: while combustion in gas engines is driven by turbulent transport, wildfires are controlled by buoyancy (i.e., natural convection) and radiation. In wildfires, radiation is a key heat transfer mechanism that is largely responsible for the vegetation pyrolysis and that sustains flaming combustion, thus enhancing their propagation over large distances at the Earth's surface.

In spite of these discrepancies, this part of the thesis aims at demonstrating the feasibility of multi-physics large-eddy simulations (LES) performed at the laboratory flame-scale to provide a comprehensive understanding of the mechanisms underlying fire spread. This Ph.D.-level project is a collaboration between CERFACS (Drs. Bénédicte Cuenot, Florent Duchaine and Eléonore Riber) and the CNRS-EM2C laboratory at Ecole Centrale Paris (Drs. Nasser Darabiha and Denis Veynante), France. HPC resources from CERFACS and GENCI-CCRT[43] (Grant 2013-x20132b6074) were used.

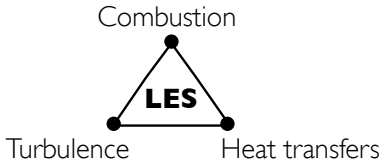
[43] Grand Equipement National de Calcul Intensif, www.genci.fr/en.

Chapter 2

General features of multi-physics turbulent combustion

Turbulent combustion is described in terms of Navier-Stokes equations corresponding to mass, momentum and energy balance equations, to which species transport equations are added to account for reacting flows. Modeling difficulties arise with the description of the exothermic fuel oxidation reactions producing the flame (that are highly non-linear and stiff reactions) and of its interactions with the turbulent flow. Based on high-performance computing (HPC), large-eddy simulation (LES) is the most recent and successful technique to properly account for unsteady flame/turbulence interactions as well as for the formation of pollutants in complex configurations. Models are used to account for the predominant physical features, in particular the fuel oxidation reactions are expressed through empirical production/consumption rates and small-scale turbulent features are parameterized using subgrid-scale modeling. Thus, by explicitly solving for the large-scale turbulent features present in the flow, LES is typically well-suited for simulating wildfires (large-scale buoyancy effects significantly affect the structure of the flame and of its surrounding flow).

In this chapter, the objective is to introduce the key ideas underlying turbulent combustion modeling. Balance equations governing reacting gas flows are presented along with filtering strategies and chemistry models.



Outline

2.1 Main challenges related to turbulent combustion

2.1.1 Oxidation pathways

A flame requires an external heat source to start, but once this preliminary ignition stage is reached, it can self-sustain (provided there is a continuous supply of gas reactants) through a series of chemical reactions between fuel and oxidizer. From a qualitative viewpoint, the resulting oxidation reactions can be represented as the following global reaction:

$$\text{Fuel (F)} + \text{Oxidizer } (O_2) \quad \longrightarrow \quad \text{Products (P)} + \text{Heat}, \tag{2.1}$$

where the fuel mainly consists of carbon (C), hydrogen (H) and oxygen (O) atoms, and where a large amount of heat is released and transferred to the surrounding environment (these oxidation reactions are therefore referred to as *exothermic*). As

part of the flaming combustion in wildfires, the global methane/air reaction can be expressed as:

$$\mathrm{CH_4} + 2\,\mathrm{O_2} \quad \longrightarrow \quad \mathrm{CO_2} + 2\,\mathrm{H_2O}, \tag{2.2}$$

in which $\mathrm{CO_2}$ and $\mathrm{H_2O}$ are the final products of the global reaction.

However, this global viewpoint is not representative of all the mechanisms occurring during methane oxidation. Generally, an oxidation reaction can be characterized by four main steps:

(1) **Initiation reactions**, corresponding to the production of first intermediate species (referred to as *radicals*) from the gas reactants;

(2) **Chain-branching reactions**, corresponding to a significant increase of the radical population within the gas mixture (made for instance of H, O and OH intermediate species);

(3) **Chain-carrying reactions** once the concentration of radicals is sufficiently high, corresponding to combustion of reactants and formation of products;

(4) **Termination reactions**, corresponding to the consumption of radicals (that can only be partial if the termination reactions are not completed).

This implies that multiple reaction pathways are followed during fuel oxidation; their detailed description in a chemical kinetic model is not systematically required to obtain accurate macroscopic quantities of interest such as the burning velocity. However, pollutant formation (e.g., nitrogen oxides $\mathrm{NO_x}$) and rich-mixture combustion (i.e., high fuel/air equivalence ratio) generally require the analysis of more complex chemical pathways. In particular, challenges found in pollutant predictions are due to their very low concentrations (i.e., a few tenths to a few hundreds parts-per-million) and to their very specific pathways of production or destruction, associated with intermediate radical species and slow reactions. Typically, biomass combustion processes are related to significant pollutant formation (e.g., $\mathrm{NO_x}$, $\mathrm{SO_x}$, CO, aerosols) and require specific chemical kinetic modeling associated with CO and $\mathrm{CH_4}$ oxidation (Dagaut and Lecomte, 2003; Pérez-Ramirez et al., 2012; Battin-Leclerc et al., 2013).

2.1.2 Premixed flames versus diffusion flames

As illustrated in Fig. 2.1, two types of ideal academic flame configurations are premixed flames and diffusion flames. Their main features are described below.

- ▷ **Premixed flames.** Fuel and oxidizer (referred to as *fresh gases*, with the index f standing for *fresh*) are mixed prior to reach the flame front. These fresh gases are separated from combustion products (referred to as *burnt gases*, with the index b standing for *burnt*) by the flame region. This thin flame region is characterized by a high temperature gradient and can be divided into three different layers:

(i) the pre-heating zone, in which fresh gas reactants are heated due to diffusion and radiation heat transfer (no chemical kinetic reactions occur);

(ii) the reaction zone, in which (fast) fuel oxidation occurs, leading to a significant heat release rate, associated with the production of intermediate species (e.g., CO) as well as radical compounds (e.g., OH);

(iii) the post-flame zone, in which slower recombination reactions release final combustion products and pollutants (e.g., $\mathrm{NO_x}$).

Thus, the flame region characterized by its thermal thickness δ_L (i.e., the thickness over which the temperature gradient is important) is propagating from burnt gases towards fresh gases at a finite velocity magnitude referred to as the *laminar flame speed* and noted s_L (the index L standing for *laminar*). The thermal thickness δ_L is commonly estimated from the inverse of the maximum temperature gradient ∇T as follows:

$$\delta_L = \frac{T_b - T_f}{\max\left(\nabla T\right)}, \tag{2.3}$$

with T_b and T_f the gas temperature in the burnt and fresh gases, respectively. The reaction zone of thickness δ_R is confined within the thin thermal region corresponding to the flame region. Within the fresh gases, the fuel/oxidizer proportion can be characterized using the equivalence ratio Φ expressed as follows:

$$\Phi = s\,\frac{Y_F}{Y_O} = \frac{\left(Y_F/Y_O\right)}{\left(Y_F/Y_O\right)_{st}}, \tag{2.4}$$

with Y_k the gas mass fractions (F standing for *fuel* and O standing for *oxidizer*) and s the mass stoichiometric ratio, i.e.,

$$s = \left(\frac{Y_O}{Y_F}\right)_{st} = \frac{\nu_O^{'} W_O}{\nu_F^{'} W_F},$$

for the one-step irreversible reaction scheme $[\nu_F^{'}\,\mathrm{F} + \nu_O^{'}\,\mathrm{O} \rightarrow \text{Products}]$. For instance, $s = 4$ for methane/oxygen reactions. Gas reactants and oxidizer are in stoichiometric conditions when $\Phi = 1$ (meaning that they are completely converted into final products), while $\Phi < 1$ for a lean mixture (i.e., fuel is the limiting reactant) and $\Phi > 1$ for a rich mixture (i.e., oxidizer is the limiting reactant). Classically in premixed combustion, a dimensionless reaction progress variable c (also called *reduced temperature*) is introduced as a flame marker such that:

$$c = \frac{T - T_f}{T_b - T_f}. \tag{2.5}$$

This implies that $c = 0$ in fresh gases ($T = T_f$), $c = 1$ in burnt gases ($T = T_b$) and the flame is the region where c takes values between 0 and 1 (describing the progression from fresh to burnt gases).

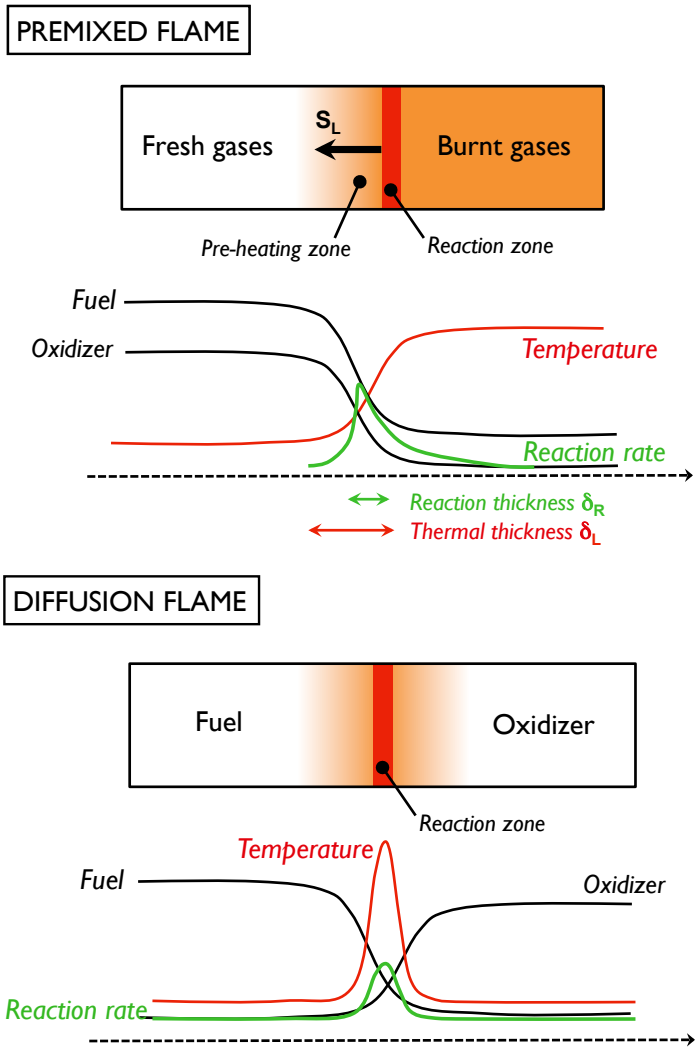


Figure 2.1: *Schematic of academic flame configurations: premixed flame (top) versus diffusion flame (bottom).*

▷ **Diffusion flames.** In contrast to premixed flames, diffusion flames correspond to a case in which fuel and oxidizer are not mixed prior to combustion and in which the reaction zone is located at their interface (fuel and oxidizer diffuse towards the reaction zone where they burn and release heat, see Fig. 2.1). Two boundary states must be defined, fuel (possibly diluted) on one side, and oxidizer (possibly diluted, for instance in nitrogen) on the other side. This implies that there is no flame propagation and no characteristic flame thickness such that the flame region drastically depends on the flow conditions (e.g., strain) and is more sensitive to velocity fluctuations than premixed flames (Peters, 1984; Bilger, 1989). Chemical reactions are confined in a restricted zone, in which the fuel/oxidizer equivalence ratio Φ is near

stoichiometry (the fuel/oxidizer interface is precisely located at stoichiometry for infinitely fast chemistry) and defined as follows:

$$\Phi = s \left(\frac{Y_F^0}{Y_O^0} \right), \tag{2.6}$$

with Y_F^0 and Y_O^0 the mass fractions of fuel and oxidizer in their respective boundary state. A typical example of diffusion flame is a candle.

↪ Flame structure underlying wildfire spread

The type of flame configuration associated with wildfires depends on the spatial scale at which their description is provided. At regional scales (i.e., from a few tens of meters up to several kilometers), wildfires can be regarded at first glance, as propagating premixed flames. In this configuration, the flame front separates the burnt zone (burnt gases in the standard terms of premixed flames) and the unburnt vegetation (fresh gases in the standard terms of premixed flames, corresponding to the gas reactants emitted by the vegetation when subject to thermal degradation and mixed with air that is entrained towards the flame).

However, when analyzing the fine-scale features of wildfires (i.e., from a millimeter up to a few meters), they are clearly of the diffusion flame type (Zhou and Mahalingam, 2002): the oxidizer is convected towards the flame through air entrainement, while combustion fuel reactants are released in the gas phase by the vegetation pyrolysis (corresponding to a phase-change process from solid-phase to gas-phase). Oxidation reactions can therefore proceed when these two reactants meet at the top of the vegetation bed or inside it for deep fuels at sufficiently high temperatures, a configuration that is typical of a diffusion flame. As illustrated in Fig. 2.2, this diffusion flame structure is due to the time delay of fuel injection introduced by the vegetation pyrolysis. Indeed, if the pyrolysis process were to occur instantaneously, the flammable gas compounds released by the pyrolysis would be mixed with air prior to flaming ignition and would thereby induce a premixed flame. Wildfire spread corresponds therefore to a diffusion flame that propagates with the biomass fuel source, a configuration that does not occur in industrial combustion and that makes the simulation of wildfires a challenging task.

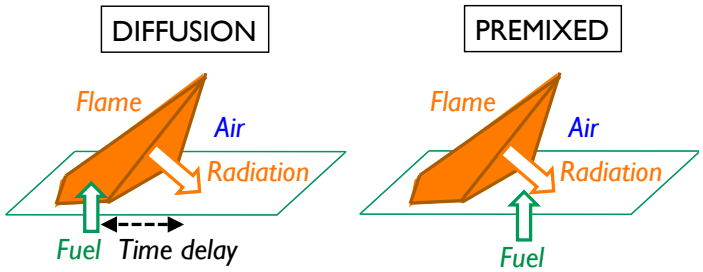


Figure 2.2: *Schematic of the flame structure of wildfire spread featuring a choice between a diffusion flame and a premixed flame: the time delay induced by biomass pyrolysis is in favor of a diffusion flame.*

Note however that fine-scale features of wildfires are more generally described as partially-premixed flames. Since the vegetation bed is a porous medium, air is also present. Furthermore, due to buoyancy effects, air entrainment induces a flow within the vegetation bed and enhances the convection of pyrolysis gases towards the flame zone. Thus, pyrolysis gases are most likely mixed with a certain proportion of air before ignition. However, this equivalence ratio is very difficult to quantify and depends on the (fine-scale) porosity of the vegetation.

2.1.3 Types of fluid flow

↪ Laminar flows versus turbulent flows

Whether a flow is laminar or turbulent depends on the relative importance of flow inertial forces compared to viscosity. Typically, flow conditions can be characterized by the Reynolds number (Re), a dimensionless quantity corresponding to the ratio of inertial to viscous forces such as:

$$\mathrm{Re} = \frac{|U|\,L}{\nu}, \tag{2.7}$$

with $|U|$ [m/s] the bulk flow velocity, L [m] a characteristic length-scale of the flow and ν [m^2/s] the fluid kinematic viscosity defined as $\nu = \mu/\rho$, μ [kg/m/s] representing the dynamic flow viscosity and ρ [kg/m^3] the gas mass density. The higher the Reynolds number, the more turbulent the flow, implying that turbulent flows are dominated by inertial forces, while viscous forces are predominant in laminar flows.

Common to both engineering and geosciences (e.g., boundary layer on an aircraft, atmospheric boundary layer, oceanic currents), turbulent flows are characterized by the development, motion and interaction of unsteady eddies (vortices) on a wide range of length-scales. Thus, turbulence dramatically affects the flow structure and mixing as illustrated by the tracer trajectory in Fig. 2.3. In the laminar region of the flow (the flow is uni-directional and does not exhibit any vorticity), the tracer particle strictly follows the mean free-stream velocity. In contrast, when the tracer enters the turbulent region of the flow, its trajectory becomes irregular and chaotic. Using Reynolds decomposition, the time-evolution of the flow velocity $\mathbf{u}$ can be decomposed into a mean component $\overline{\mathbf{u}}$ and a turbulent component $\mathbf{u}^{'}$ as follows:

$$\mathbf{u}(t) = \overline{\mathbf{u}} + \mathbf{u}^{'}(t), \tag{2.8}$$

where the overbar denotes a time average. This implies that turbulent flow eddies create temporal fluctuations in velocity that superimpose onto the mean flow. The trajectory of the tracer is therefore imposed by both the mean flow and the eddies in Fig. 2.3. In contrast, in steady laminar flows, $\mathbf{u}(t) = \overline{\mathbf{u}}$ for all time steps.

Due to the macroscopic mixing of fluid particles, turbulent flows are characterized by a fast rate of momentum and heat mixing. Thus, the energy is transferred from

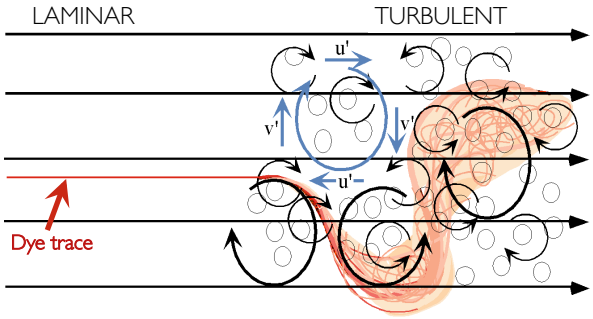


Figure 2.3: *Tracer transport (red lines) in laminar and turbulent flows. Credit: MIT.*

large-scale to small-scale eddies by non-linear interactions, until it is dissipated by viscous diffusion in the smallest eddies. Turbulent flows therefore require a continuous supply of energy to compensate for viscous losses. For more details, see, for example, Kundu et al. (2011).

↪ Natural convection versus forced convection

The structure of a flow can also be significantly affected by buoyancy effects. Density discrepancies $\Delta\rho$ in the flow (due to temperature and/or chemical spatial variations) induce the development of convective cells such as the Rayleigh-Bénard-type convection occurring in the mantle of Rocky planets (see Fig. 2.4). The importance of buoyancy effects on a flow can be characterized by the Richardson number Ri, a dimensionless number that represents the ratio of potential energy to kinetic energy such as:

$$\mathrm{Ri} = \frac{g\,(\Delta\rho/\rho)\,H}{U^2}, \tag{2.9}$$

with g [m/s^2] the acceleration induced by gravity, H [m] the vertical length-scale and U [m/s] the bulk flow velocity. The Richardson number typically represents the relative importance of natural convection to forced convection, the term $\sqrt{g\,(\Delta\rho/\rho)\,H}$ indicating the order of magnitude of the natural convection velocity. Thus, $\mathrm{Ri} < 0.1$ implies that natural convection is negligible, while for $\mathrm{Ri} > 10$ forced convection is in-turn negligible; in-between these threshold values, both effects co-exist.

In the context of wildfires, flow dynamics are buoyancy-driven such that the development of convective cells is important and induce air entrainment. However, this natural convection process combines to the wind velocity field (an external forcing of the flow that can be regarded as a forced convection process) to enhance wildfire spread. This implies that the fluid dynamics of wildfires is determined by mixed convection, where natural and forced convections occur simultaneously (Joulain, 1996; Tieszen, 2001; Finney and Mcallister, 2011). The dimensionless Grashof number (Gr) characterizes the ratio of buoyancy to viscous forces acting on the flow (i.e., how much of the convection is due to external forcing or to natural con-

vection) and can be expressed with respect to Reynolds and Richardson numbers as follows (Jaluria, 1980):

$$\mathrm{Gr} = \frac{g\,(\Delta\rho/\rho)\,H^3}{\nu^2} = \mathrm{Re}^2\,\mathrm{Ri}. \tag{2.10}$$

The Richardson number Ri is typically between 1 and 10 in fires, depending on the wind velocity, with Re in the range $[10^5, 10^6]$ and Gr in the range $[10^{12} - 5 \times 10^{12}]$.

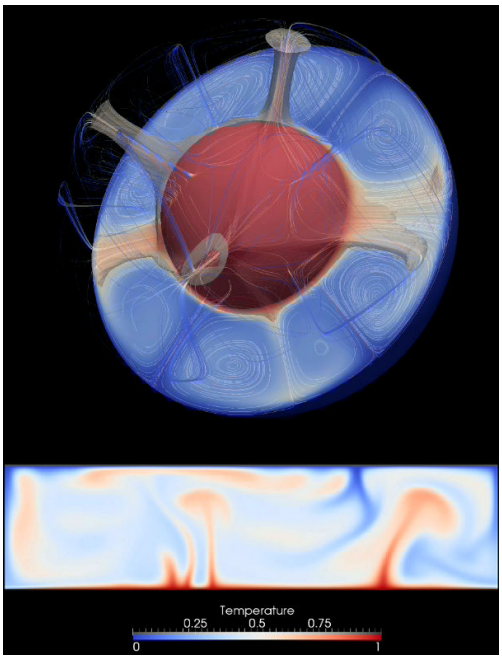


Figure 2.4: *Normalized temperature profile for simulations of terrestrial mantle convection (in spherical and cartesian geometry). Credit: DLR, http://www.dlr.de/pf/en/.*

2.1.4 Interactions between turbulence and combustion

Combustion can occur if fuel and oxidizer are mixed at molecular scales. However, in turbulent flows, combustion processes highly depend on turbulent mixing and especially, on velocity fluctuations. Interactions between turbulent shear flows and highly exothermic chemical reactions are difficult to account for, in particular due to the multiplicity of spatial and temporal scales involved (Veynante and Vervisch, 2002; Poinsot and Veynante, 2005). There is a wide range of temporal scales underlying detailed chemical kinetics: fuel oxidation is governed by fast reactions, while the production of pollutants (such as nitrogen oxides $\mathrm{NO_x}$) results from slower reaction processes. Turbulence also features unsteady eddies on multiple length-scales, from integral length-scales (i.e., the largest scales in the energy spectrum that exhibit the largest velocity fluctuations) to Kolmogorov length-scales (i.e., the smallest scales in the energy spectrum, where turbulence can be locally regarded as isotropic and homogeneous). Thus, interactions between turbulence and combustion

induce a competition between turbulent mixing and chemical reactions, resulting in a large variety of turbulent flame structures that can be described based on two qualitative regimes, flamelet and fast mixing (Damköhler, 1940).

▷ **Flamelet regime.** When the time-scale of chemical reactions is much shorter than the turbulent time-scale, the inner structure of the flame is not significantly affected by turbulence (although the flame front is wrinkled, stretched and convected by the turbulent flow). Thus, the overall structure of the turbulent flame is determined by large-scale eddy structures. This is referred to as the *flamelet regime*.

In the case of premixed flames, the flame front can be regarded as an ensemble of premixed laminar flame elements, which are stretched and wrinkled by the turbulent flow as represented in Fig. 2.5. This wrinkling leads to an increase of the effective flame surface area that enhances the reactant consumption rate and thereby, the propagating speed of the flame front. A model for the resulting turbulent flame speed noted s_T is therefore required. Since flamelet elements are supposed to propagate, locally, at the laminar speed s_L, the turbulent speed s_T is commonly modeled as the laminar flame speed s_L weighted by the ratio of the wrinkled instantaneous front surface A_T to the projected unwrinkled surface A_L. To incorporate the effects of flame strain and curvature due to turbulence, a stretch factor modeled as $(\overline{s}_c/s_L)$ is introduced in the generalized formulation of the turbulent flame speed (Law and Sung, 2000; Driscoll, 2008):

$$\frac{s_T}{s_L} = \left(\frac{A_T}{A_L}\right) \frac{\overline{s}_c}{s_L}, \tag{2.11}$$

with $\overline{s}_c$ [m/s] the mean consumption speed.

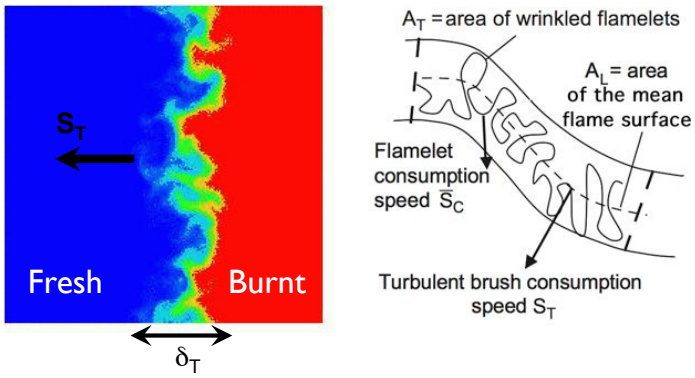


Figure 2.5: *Schematic of turbulent flame subject to wrinkling, with s_T the turbulent flame speed and δ_T the overall turbulent flame thickness. Credit: Driscoll (2008).*

▷ **Fast mixing regime.** In contrast, a short turbulence time-scale implies a fast mixing between oxidizer and fuel reactants, implying that small-scale eddies can interact with the transport mechanisms within the flame zone and modify the inner structure of the flame.

2.2 Equations for reacting flows

Combustion-related processes involve hundreds of species reacting through thousands of chemical reactions (the number of species in the gas mixture is denoted by N_g). The Navier-Stokes equations apply for such a multi-species multi-reacting gas; however, they require some additional terms and closure models as presented in the following. For more details, see Turns (2000), Poinsot and Veynante (2005).

2.2.1 Characterization of the gas phase

The k-th gas species (referred to as index k) is characterized by the following features:

- the **mass fraction** $Y_k = m_k/m$ $[-]$ defined as the ratio of the mass m_k of species k to the total mass m contained in a given homogeneous volume V, satisfying:
$$\sum_{k=1}^{N_g} Y_k = 1;$$
- the **mass density** $\rho_k = \rho\, Y_k$ [kg/m^3] corresponding to the mass m_k of the species k in the volume V, with ρ the mass density of the gas mixture;
- the **atomic weight** W_k [kg/mol];
- the **mass specific heat capacity at constant pressure** $c_{p,k}$ [J/kg]:
$$c_p = \sum_{k=1}^{N_g} c_{p,k}\, Y_k, \tag{2.12}$$
with c_p its counterpart for the gaseous mixture, which is a function both of temperature and composition;
- the **mass enthalpy** $h_k = h_{s,k} + \Delta h^0_{f,k}$ [J/kg] that is a composition of the sensible enthalpy $h_{s,k}$ satisfying:
$$h_{s,k} = \int_{T^0}^{T} c_{p,k}(T')\,\mathrm{d}T',$$
and the chemical enthalpy equal to the mass enthalpy of formation $\Delta h^0_{f,k}$ at the reference temperature T^0 (in principle, any value could be assigned to T^0 but for practical purposes, it is usually set to $T^0 = 298.15$ K with $\Delta h^0_{f,k} = -4675$ kJ/kg for $\mathrm{CH_4}$ and -13435 kJ/kg for $\mathrm{H_2O}$ for instance);
- the **partial pressure** p_k [Pa] that relates to the mass density ρ_k and atomic weight W_k through the ideal gas assumption:
$$p_k = \rho_k \frac{R_g}{W_k} T, \tag{2.13}$$

with R_g the perfect gas constant ($R_g = 8.314$ J/mol/K) and T [K] the temperature of the gaseous mixture;

- ▷ the **Lewis number** Le_k, a dimensionless number comparing heat and species diffusions that is expressed as:

$$\mathrm{Le}_k = \frac{\lambda}{\rho\, c_p\, D_k} = \frac{D_{th}}{D_k}, \tag{2.14}$$

with D_k [m^2/s] the diffusivity of species k in the rest of the mixture and $D_{th} = \lambda/(\rho\, c_p)$ [m^2/s] the heat diffusivity that partly depends on the heat diffusion coefficient λ [J/m/K/s].

Using this nomenclature, the equation of state for ideal gas reads:

$$p = \sum_{k=1}^{N_g} p_k = \sum_{k=1}^{N_g} \rho_k \frac{R_g}{W_k} T, \tag{2.15}$$

with p the overall pressure of the gaseous mixture. The atomic weight of the gaseous mixture denoted by W [kg/mol] can be derived from the atomic weight W_k and the mass fraction Y_k of each species k:

$$W = \left(\sum_{k=1}^{N_g} \frac{Y_k}{W_k} \right)^{-1}. \tag{2.16}$$

Thus, the equivalence between the mass fraction (Y_k) and the molar fraction ($X_k = \rho_k/\rho$) of each gas species k can be expressed as follows:

$$X_k = \left(\frac{W}{W_k} \right) Y_k. \tag{2.17}$$

2.2.2 Chemical kinetics

During combustion processes, fuel reactants are oxidized once a sufficiently high amount of energy is available in the gas mixture to activate reactions (in the zones where the fuel/oxidizer equivalence ratio Φ is within the flammability limits). Considering N_r reactions between the N_g species, these reactions can be schematized as follows (as generalization of Eq. 2.1):

$$\sum_{k=1}^{N_g} \nu'_{r,k}\, \mathcal{S}_k \quad \longleftrightarrow \quad \sum_{k=1}^{N_g} \nu''_{r,k}\, \mathcal{S}_k \quad (r = 1, \cdots, N_r), \tag{2.18}$$

with $\mathcal{S}_k$ the symbol of gas species k, $\nu'_{r,k}$ and $\nu''_{r,k}$ the respective forward and backward molar stoichiometric coefficients for species k in reaction r satisfying:

$$\sum_{k=1}^{N_g} \left(\nu''_{r,k} - \nu'_{r,k} \right) W_k = \sum_{k=1}^{N_g} \nu_{r,k}\, W_k = 0 \tag{2.19}$$

to enforce mass conservation. $\nu_{r,k}$ corresponds to the associated net stoichiometric coefficient. The chemical reaction rate of each gas species k noted $\dot{\omega}_k$ [kg/m^3/s] yields:

$$\dot{\omega}_k = W_k \sum_{r=1}^{N_r} \nu_{r,k}\, \dot{q}_r \quad (r = 1, \cdots, N_r). \tag{2.20}$$

This equation represents the contribution of each reaction r (through the progress rate of reaction $\dot{q}_r$ [mol/m^3/s] of the r-th reaction) to the production rate $\dot{\omega}_k$ (or consumption rate if negative) of species k. The following condition must be satisfied to ensure mass conservation:

$$\sum_{k=1}^{N_g} \dot{\omega}_k = 0. \tag{2.21}$$

Besides, the reaction progress rate $\dot{q}_r$ reads:

$$\dot{q}_r = K_r^{\mathrm{f}} \prod_{k=1}^{N_g} [X_k]^{\nu'_{r,k}} - K_r^{\mathrm{b}} \prod_{k=1}^{N_g} [X_k]^{\nu''_{r,k}}, \tag{2.22}$$

with $[X_k]$ [mol/m^3] the molar concentration of species k satisfying $[X_k] = \rho_k / W_k$, K_r^{f} the forward reaction rate of the r-th reaction and K_r^{b} its backward counterpart. It is worth noting that the forward reaction rate is usually expressed through an Arrhenius law of the following type:

$$K_r^{\mathrm{f}} = k_r\, T^{n_r}\, \exp\left[-\frac{E_{a,r}}{R_g\, T}\right], \tag{2.23}$$

where the Arrhenius-parameter triplet $(k_r, n_r, E_{a,r})$ represents the pre-exponential factor, the temperature exponent and the activation energy of reaction r. These parameters are generally calibrated for each reaction r against theoretical analysis or experimental data. From a molecular viewpoint, K_r^{f} represents the probability of occurrence of atomic exchanges through molecular collisions. It follows that the activation energy $E_{a,r}$ represents the minimal amount of collisional energy required to allow for reaction r to proceed, while the pre-exponential factor k_r is related to the collision frequency, the geometry as well as the orientation of the molecules during collisions. The backward reaction rate K_r^{b} is commonly specified as the ratio between the forward reaction rate K_r^{f} and the reaction equilibrium constant K_r^{eq} (that is not detailed here), since equilibrium induces the equivalence between forward and backward reaction rates (i.e., $\dot{q}_r = 0$ in Eq. 2.22). One additional quantity of interest in combustion is the heat release rate, noted $\dot{\omega}_T$ [J/m^3/s] and defined as follows:

$$\dot{\omega}_T = -\sum_{k=1}^{N_g} \Delta h_{f,k}^0\, \dot{\omega}_k. \tag{2.24}$$

Note that a thermodynamic database including the mass enthalpy of formation $\Delta h^0_{f,k}$ for each species k is required to calculate the heat release rate and thereby, the burnt gas temperature.

2.2.3 Balance equations for reacting flows

Reacting flows are described by Navier-Stokes equations to which reaction source terms are added. The following controlling variables are required: the mass density ρ, the velocity components u_i $(i = 1, 2, 3)$, the total enthalpy h_t, the species mass fractions Y_k $(k = 1, \cdots, N_g)$ and the pressure p. The resulting (5+N_g) balance equations can be expressed as follows.

▷ **Mass balance equation**

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho\, u_j)}{\partial x_j} = 0. \tag{2.25}$$

▷ **Momentum balance equations**

$$\frac{\partial (\rho\, u_i)}{\partial t} + \frac{\partial (\rho\, u_i u_j)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \rho\, g_i + \frac{\partial \tau_{ij}}{\partial x_j} + F_i \quad (i = 1, 2, 3), \tag{2.26}$$

with $\rho\, g_i$ the buoyancy force induced by gravity (g_i being the i-th component of gravitational acceleration), F_i [N/m^3] the (volume) body forces and τ_{ij} [N/m^2] the viscous stress tensor.

▷ **Species balance equations**

$$\frac{\partial (\rho\, Y_k)}{\partial t} + \frac{\partial (\rho\, u_j\, Y_k)}{\partial x_j} = -\frac{\partial \mathcal{J}_{j,k}}{\partial x_j} + \dot{\omega}_k \quad (k = 1, \cdots, N_g). \tag{2.27}$$

This equation includes convection and diffusion transports as well as the chemical source term associated with the k-th species. Note that $\mathcal{J}_{j,k}$ corresponds to the molecular diffusive flux of species k, which can be decomposed into the contributions of the species diffusion velocity $V_{k,j}$ and of the correction velocity $V_{c,j}$ as follows:

$$\mathcal{J}_{j,k} = \rho\, (Y_k\, V_{k,j} + Y_k\, V_{c,j}) = \rho\, \left(-D_k\, \frac{W_k}{W}\, \frac{\partial X_k}{\partial x_j} + Y_k\, V_{c,j} \right), \tag{2.28}$$

with D_k [m^2/s] the molecular diffusivity associated to species k in the gaseous mixture (usually characterized in terms of the dimensionless Schmidt number $Sc_k = \nu / D_k$, which compares the mixture kinematic viscosity ν to D_k and which can be regarded as the ratio of momentum diffusion to species diffusion) and $V_{c,j}$ [m/s] the correction velocity satisfying:

$$V_{c,j} = \sum_{k=1}^{N_g} D_k\, \left(\frac{W_k}{W} \right)\, \frac{\partial X_k}{\partial x_j}. \tag{2.29}$$

This expression for $V_{c,j}$ is derived through the approximation due to Hirschfelder et al. (1954), see Ern and Giovangigli (1995). While replacing the rigorous multi-species transport equation, this approximation is convenient in practice. In particular, it relates diffusion coefficients D_k to heat diffusivity D_{th} through the Lewis numbers of individual species Le_k. Then, Eq. (2.27) can be rewritten for $k = 1, \cdots, N_g$ as follows:

$$\frac{\partial(\rho\, Y_k)}{\partial t} + \frac{\partial\left(\rho\,(u_j + V_{c,j})\, Y_k\right)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\rho\, D_k\, \frac{W_k}{W}\, \frac{\partial X_k}{\partial x_j}\right) + \dot{\omega}_k. \quad (2.30)$$

▷ **Heat balance equation**

$$\begin{aligned}\frac{\partial(\rho\, h_t)}{\partial t} + \frac{\partial(\rho\, u_j\, h_t)}{\partial x_j} =& \frac{\partial p}{\partial t} - \frac{\partial q_j}{\partial x_j} + \frac{\partial(\tau_{i,j}\, u_j)}{\partial x_i} + \dot{q} \\ &+ \rho \sum_{k=1}^{N_g} Y_k\, F_{k,j}\,(u_j + V_{c,j}),\end{aligned} \quad (2.31)$$

in which:

- h_t [J/kg] corresponds to the total enthalpy of the gaseous mixture, formally defined as:

$$h_t = \sum_{k=1}^{N_g} h_k + \frac{1}{2}\, u_j u_j. \quad (2.32)$$

- q_j [W/m^2] corresponds to the energy flux composed by the heat diffusion term (following Fourier law) and by the diffusion between species with different enthalpies, such as:

$$q_j = -\lambda\, \frac{\partial T}{\partial x_j} + \rho \sum_{k=1}^{N_g} h_k\, Y_k\, V_{k,j}. \quad (2.33)$$

- $\tau_{i,j}\, u_i$ and $\rho \sum_{k=1}^{N_g} Y_k\, F_{k,j}(u_j + V_{c,j})$ correspond to the power due to viscous forces and volume forces $F_{k,j}$ applied on species k, respectively.
- $\dot{q}$ [W/m^3] corresponds to the heat source term (due, for instance, to an electric spark, a laser or a radiation flux). Note that the modeling of radiation is discussed in Section 2.5.

Note that in this thesis, the species diffusion under temperature gradients (i.e., the Soret effect), the molecular transport due to pressure gradients and the enthalpy diffusion due to mass fraction gradients (i.e., the Dufour effect) are neglected. Note also that there are three main differences with the Navier-Stokes for non-reacting flows:

(1) Since heat capacities change significantly with temperature and composition during combustion, the description of the thermodynamic state of a reacting gas requires individual tracking of each gas species.

(2) Specific modeling is required to describe the rate $\dot{\omega}_k$ at which each species k is produced or consumed.

(3) Attention must be paid to the transport coefficients (e.g., heat diffusivity, species diffusion, viscosity) in the multi-species gas mixture.

2.3 Turbulence considerations: large-eddy simulation and filtering

2.3.1 Overview of computational approaches for turbulence

A turbulent flow instantaneously satisfies Navier-Stokes Eqs. (2.25)-(2.26)-(2.30)-(2.31). However, due to the wide range of scales to be resolved, solving for all the involved fine scales present in complex configurations is not feasible.

- ▷ **Reynolds-averaged Navier-Stokes (RANS) equations.** Initial efforts have been devoted to design models for RANS equations (Launder and Spalding, 1974). As an approximate time-averaged solution to the Navier-Stokes equations, RANS equations only solve for the mean flow; all effects due to fluctuating motions (typically, the turbulent kinetic energy and the dissipation rate) must be modeled using a turbulence model. Thus, RANS equations show difficulties in predicting unsteady processes. Due to their relatively low computational cost, they are still routinely solved to design industrial devices (such as piston engines and aeronautical combustion chambers) using commercial codes nowadays.

- ▷ **Direct numerical simulations (DNS).** In contrast, DNS solve for the full instantaneous Navier-Stokes equations without any model for turbulent flows; all turbulent length-scales are explicitly resolved (Orszag, 1970). Thus, they can perfectly retrieve the fine-scale features of the flow and predict temporal variations of quantities of interest such as the temperature. From a historical perspective, they became feasible in the 1980s due to the development of HPC as well as to the progress in higher-order numerical schemes. Due to their high computational cost, they are still restricted to low Reynolds number and simple academic flow configurations (Trouvé et al., 1994; Poinsot, 1996; Vervisch and Poinsot, 1998; Poludnenko and Oran, 2010, 2011). Note that they often use simplified chemical kinetics to reduce their computational cost.

- ▷ **Large-eddy simulations (LES).** As an intermediate strategy, LES explicitly solve for the large-scale turbulent structures (i.e., at the integral length-scales, where the eddies containing most of the energy are present and mainly depend on the geometry configuration) and models the effects of the smallest

eddies (i.e., at the Kolmogorov length-scales involving the smallest length-scales in the energy spectrum that form the viscous sublayer range and that are characterized by local isotropic and homogeneous turbulence). Since LES can be regarded as a filtering procedure in which the smallest length-scales of the turbulent flow are filtered out, they are less computationally-intensive than DNS. They allow for a dynamic representation of the large-scale flow motions, which are essential to describe transient flows in complex configurations (Menon and Jou, 1991).

From a qualitative viewpoint, RANS, DNS and LES properties are summarized in terms of energy spectrum in Fig. 2.6. Consistently, Fig. 2.7 illustrates the effect of these different computational approaches for turbulence on the temporal evolution of the temperature at a specified sensor in a premixed flame.

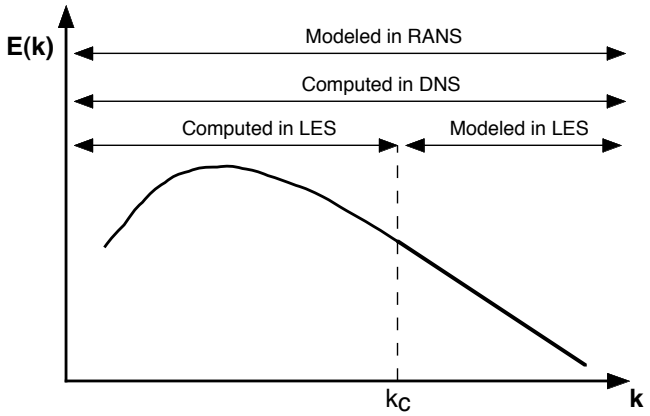


Figure 2.6: *Schematic of the energy spectrum with respect to the wave number k (inversely proportional to the eddy length-scale): DNS resolve all spatial frequencies in the spectrum, whereas only the largest scales (up to a cut-off wave number k_c) are computed in LES and no turbulent motion is explicitly captured in RANS. Credit: Poinsot and Veynante (2005).*

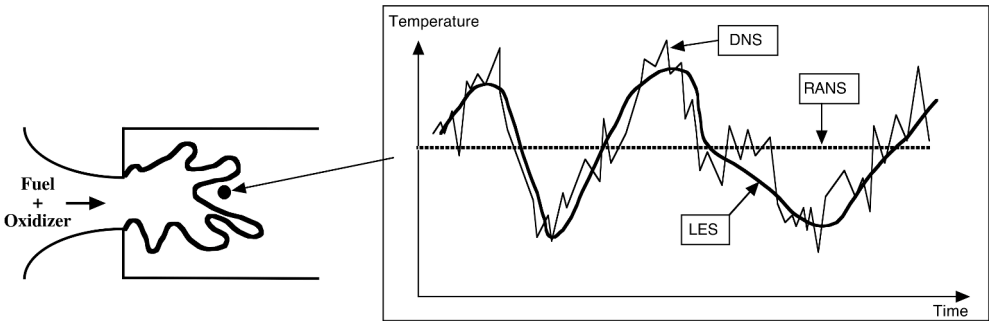


Figure 2.7: *Time-evolution of local temperature computed with DNS, LES and RANS in a turbulent premixed flame. Credit: Poinsot and Veynante (2005).*

↪ Large-eddy simulations and combustion applications

Many efforts have been recently devoted to the development of LES approaches for turbulent combustion, which currently appear as the most successful techniques to simulate turbulent combustion and capture unsteady large-scale gas flow features in industrial devices (Caraeni et al., 2000; Forkel and Janicka, 2000; Moureau et al., 2005; Boileau et al., 2008). The information gain obtained on the turbulent flow using LES instead of RANS is illustrated in Fig. 2.8.

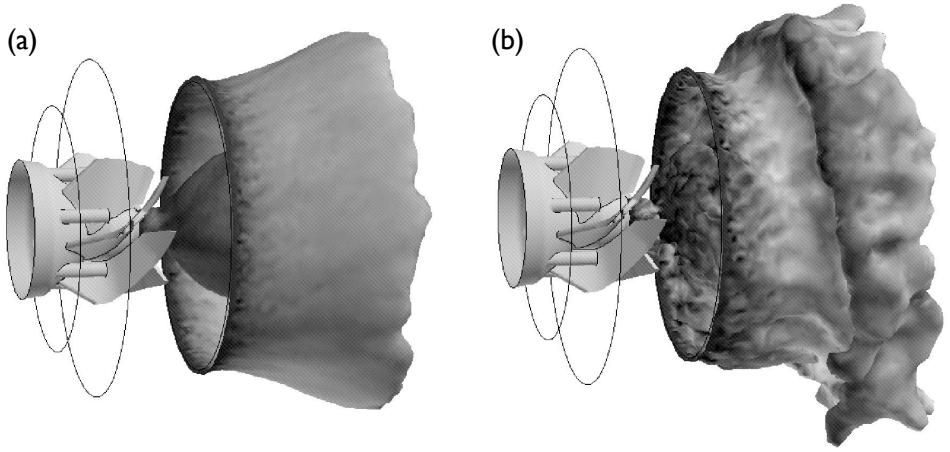


Figure 2.8: *Iso-surface of temperature (1100 K) in a turbulent swirled premixed flame. (a) RANS-based mean field. (b) LES-based instantaneous field. Credit: Selle et al. (2004).*

LES requires high-resolution computational grids to properly represent the flame and its interaction with turbulence as well as the formation of pollutants. Despite of its computational cost, LES is well-adapted for many combustion studies for the reasons listed below.

- ▷ While large-scale structures in turbulent flows highly depend on the configuration geometry, small-scale structures exhibit more universal features (i.e., they can be locally regarded as isotropic and homogeneous) and accordingly, they can be represented by models referred to as *subgrid-scale turbulence models*.
- ▷ The knowledge of large-scale turbulent motions may be used to infer the effects of unresolved small-scale motions, since the energy flows from large (resolved) structures to smaller (unresolved) scales. For instance, subgrid-scale models may be based on similarity assumptions between large and small turbulent scales. This knowledge is also useful to better characterize flow instabilities, since those are associated with large-scale eddy structures that are explicitly solved in LES.
- ▷ LES allows for a better description of combustion/turbulence interactions by explicitly identifying instantaneous fresh and burnt gases zones in premixed flames (in contrast, the RANS approach accounts for the local probability of

being in fresh or burnt gases). This clear identification of the flame front in LES is useful to quantify radiation heat transfer for instance.

The potential of LES approaches has been demonstrated for predicting combustion instabilities (Selle et al., 2006), pollutant emissions (Schmitt et al., 2007) or two-phase flows (Riber et al., 2009). Papers of Westbrook et al. (2005), Janicka and Sadiki (2005), Pitsch (2006), Bockhorn et al. (2009) and Gicquel et al. (2012) provide a very complete view of LES applications to turbulent combustion.

2.3.2 Filtering procedure

To separate large and small scales in LES, a low-pass spatial filter of given size Δ and noted $\mathcal{G}_\Delta$ is applied to the exact Navier-Stokes equations (see Poinsot and Veynante, 2005, for a description of usual formulations of $\mathcal{G}_\Delta$). From a mathematical viewpoint, this filter consists of a convolution product of any quantity of interest c with the filter function $\mathcal{G}_\Delta$:

$$\overline{c}(\mathbf{x},t) = \int_\Omega \mathcal{G}_\Delta\left(\mathbf{x}-\mathbf{x}^{'}\right)\, c(\mathbf{x}^{'},t)\,\mathrm{d}\mathbf{x}^{'}, \tag{2.34}$$

where $\overline{c}(\mathbf{x},t)$ is a spatially- and temporally-fluctuating quantity standing for the flow large-scale structures (the eddies of size smaller than the filter size Δ are removed), the bar referring to *Reynolds-filtering*. In contrast, RANS equations consider a statistically-averaged quantity (the mean component of the quantity). For variable density flows (e.g., combustion reacting flows), mass-weighted filtering (referred to as *Favre-filtering*) is introduced:

$$\overline{\rho}(\mathbf{x},t)\,\tilde{c}(\mathbf{x},t) = \overline{\rho\,c}(\mathbf{x},t) = \int_\Omega \rho(\mathbf{x}^{'},t)\,\mathcal{G}_\Delta\left(\mathbf{x}-\mathbf{x}^{'}\right)\, c(\mathbf{x}^{'},t)\,\mathrm{d}\mathbf{x}^{'}, \tag{2.35}$$

with $\tilde{c}(\mathbf{x},t) = \overline{\rho\,c}(\mathbf{x},t)/\overline{\rho}(\mathbf{x},t)$.

2.3.3 Filtered balance equations

The LES-filtered Navier-Stokes equations are obtained by Favre-filtering the instantaneous balance equations. Using this formalism, they can be summarized as follows.

▷ **Mass balance equation**

$$\frac{\partial\overline{\rho}}{\partial t} + \frac{\partial(\overline{\rho}\,\widetilde{u}_j)}{\partial x_j} = 0. \tag{2.36}$$

▷ **Momentum balance equations**

For $i = 1, 2, 3$,

$$\begin{aligned}\frac{\partial(\overline{\rho}\,\widetilde{u}_i)}{\partial t} + \frac{\partial(\overline{\rho}\,\widetilde{u}_i\,\widetilde{u}_j)}{\partial x_j} = &-\frac{\partial\overline{p}}{\partial x_i} + (\overline{\rho}\,g_i)\\ &+ \frac{\partial}{\partial x_j}\left(\overline{\tau}_{ij} - \overline{\rho}\,(\widetilde{u_i\,u_j} - \widetilde{u}_i\,\widetilde{u}_j)\right) + \overline{F}_i.\end{aligned} \tag{2.37}$$

▷ **Species balance equations**

For $k = 1, \cdots, N_g$,

$$\frac{\partial(\overline{\rho}\,\widetilde{Y}_k)}{\partial t} + \frac{\partial(\overline{\rho}\,\widetilde{u}_j\,\widetilde{Y}_k)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\overline{V_{k,j}\,Y_k} - \overline{\rho}\,(\widetilde{u_j\,Y_k} - \widetilde{u}_j\,\widetilde{Y}_k)\right) + \overline{\dot{\omega}}_k. \quad (2.38)$$

▷ **Heat balance equation**

$$\frac{\partial(\overline{\rho}\,\widetilde{h}_t)}{\partial t} + \frac{\partial(\overline{\rho}\,\widetilde{u}_j\,\widetilde{h}_t)}{\partial x_j} = -\frac{\partial}{\partial x_j}\left(\overline{\rho}\,(\widetilde{u_j\,h_t} - \widetilde{u}_j\,\widetilde{h}_t)\right) + \frac{\partial\overline{p}}{\partial t} - \frac{\partial\overline{q}_j}{\partial x_j} + \frac{\partial(\overline{\tau_{i,j}\,u_j})}{\partial x_i} + \dot{\overline{q}}. \quad (2.39)$$

Assumptions and closure models listed below are required for all right-hand side terms corresponding to unresolved transport and chemical terms.

▷ **Unresolved Reynolds stresses** $(\widetilde{u_i\,u_j} - \widetilde{u}_i\,\widetilde{u}_j)$. They require a subgrid-scale turbulence model that reproduces interactions between resolved and unresolved turbulent length-scales in terms of energy fluxes. This model aims at retrieving the fluid turbulent viscosity ν_T (usually based on turbulence modeling for non-reacting flows, using an eddy viscosity assumption). Reference subgrid-scale turbulence models are the Smagorinski model (Smagorinsky, 1963; Pope, 2000) and its dynamic counterpart (Germano et al., 1991), the WALE[1] model (Ducros et al., 1998) and the Sigma model (Nicoud et al., 2011). It is worth noting that the Smagorinski model is widely-used for its straighforward implementation. However, this model is too dissipative when approaching a solid boundary. To overcome this issue and to better capture scaling laws in wall modeling, the WALE model and more recently its upgrade, the SIGMA model, have been developed.

▷ **Unresolved species** $(\widetilde{u_j\,Y_k} - \widetilde{u}_j\,\widetilde{Y}_k)$ **and enthalpy** $(\widetilde{u_j\,h_t} - \widetilde{u}_j\,\widetilde{h}_t)$ **fluxes.** They are generally closed using a gradient assumption. For instance, the species fluxes $(\widetilde{u_j\,Y_k} - \widetilde{u}_j\,\widetilde{Y}_k)$ can be expressed as follows:

$$\widetilde{u_j\,Y_k} - \widetilde{u}_j\,\widetilde{Y}_k = -\frac{\nu_T}{Sc_k^{\mathrm{t}}}\left(\frac{\partial\widetilde{Y}_k}{\partial x_j}\right), \quad (2.40)$$

with ν_T the subgrid-scale turbulent viscosity (derived from a subgrid-scale turbulence model) and Sc_k^{t} the subgrid-scale turbulent Schmidt number.

▷ **Filtered laminar diffusion fluxes for species and enthalpy.** They might be neglected against turbulent transport for high Reynolds numbers, or modeled through a a classical gradient assumption such that:

$$\overline{V_{k,j}\,Y_k} \simeq -\overline{\rho}\,\overline{D}_k\,\frac{\partial\widetilde{Y}_k}{\partial x_j}, \quad \overline{\lambda\,\frac{\partial T}{\partial x_j}} \simeq \overline{\lambda}\,\frac{\partial\widetilde{T}}{\partial x_j}, \quad (2.41)$$

[1] Wall-Adapting Local Eddy-viscosity.

with $\overline{\lambda}$ and $\overline{D_k}$ the respective mean values of the heat diffusion coefficient and the diffusivity of species k (considering mean thermodynamic conditions).

- ▷ **Filtered chemical reaction rates $\overline{\dot{\omega}}_k$.** The modeling of these spatially-filtered reaction rates that represent the production or consumption rates of the k-th species in a turbulent gas flow is an essential step towards accurate LES for reacting flows. This issue is discussed in the next section of the manuscript.

2.4 Combustion modeling for large-eddy simulations

Modeling the chemical reactions rates $\overline{\dot{\omega}}_k$ in a LES capability relies on the two following cornerstones.

(1) Tracking the life-time of the gas compounds in a reacting mixture (e.g., fuel and oxidizer, radicals, combustion products) requires a chemical kinetic scheme, in which the production or consumption rate $\dot{\omega}_k$ of gas compounds (without filtering in a preliminary step) is represented through a set of elementary reactions. This chemical kinetic scheme must be consistent with the LES framework in terms of computational cost and stiffness, and its accuracy is commonly evaluated on simplified flame configurations (e.g. one-dimensional laminar premixed or counter-flow flames) for which experimental measurements are available. Note that this chemical kinetic scheme relevant for combustion does not intend to represent the biomass fuel thermal degradation leading to pyrolysis; this issue is discussed in Chapter 3.

(2) Accounting for the combustion/turbulence interactions at the LES subgrid-scale level requires the development of turbulent combustion models, which address for instance the loss of flame front wrinkling induced by the filtering size Δ.

2.4.1 Limitations of detailed chemical kinetics modeling

To describe chemical reaction rates $\dot{\omega}_k$ involved in the Navier-Stokes species balance Eq. (2.27), detailed chemical kinetic schemes including hundreds of gas species and thousands of reactions are available for a large number of light hydrocarbon-based fuels (Simmie, 2003). For instance, the detailed GRI-Mech3.0 scheme[2] (325 elementary reactions, 53 gas species) is used as a reference for (premixed) methane oxidation in the literature. GRI-Mech3.0 is optimized for an initial gas temperature varying from 1000 to 2500 K, a pressure varying from 0.01 to 10 atm, and an equivalence ratio Φ varying from 0.1 to 5. While these detailed kinetic schemes accurately predict flame dynamics for academic flame configurations (e.g., one-dimensional laminar premixed or counter-flow flames) and light hydrocarbon-based

[2] Gas Research Institute, http://www.me.berkeley.edu/gri_mech/.

fuels, their application to more complex turbulent combustion configurations remains prohibitive for two main reasons:

(1) The computational cost required to solve Navier-Stokes equations for reacting flows significantly increases with the number of gas species since (i) the spatial and temporal evolution of each extra gas species is described by an additional transport equation; and (ii) the determination of the thermo-chemistry of a reacting gas mixture induces an additional computational cost (chemical reaction rates and transport coefficients being complex functions of gas composition and temperature).

(2) The coupling between combustion and turbulence involves a wide range of temporal and spatial scales. In particular, pathways followed by detailed kinetic schemes for pollutant emissions (e.g., $\mathrm{NO_x}$) and radical predictions imply complex computational treatments.

It is worth noting that in the context of biomass burning, detailed chemical pathways of pyrolysis products are unknown (no detailed chemical kinetic scheme is available in the literature since the composition of biomass pyrolysis products is only partially known). Thus, alternative modeling strategies for the reaction rates $\dot{\omega}_k$ must be adopted to represent the combustion of pyrolysis products (e.g., $\mathrm{CH_4}$, CO) within a LES capability.

2.4.2 Effective strategies for chemical kinetics modeling

To overcome these issues, many efforts have been devoted to the development of simplified strategies for chemical kinetics modeling, partly through reduced and tabulated chemistry.

↪ Reduced chemistry

Global kinetic schemes aim at reproducing the main macroscopic flame features (e.g., burnt gas temperature, ignition delay, flame propagating speed for premixed flames) without significant loss of accuracy, but using only a few reactions (Westbrook and Dyer, 1981; Selle et al., 2004; Franzelli et al., 2010). While being computationally efficient, these global kinetic schemes are unable to capture information on intermediate radical compounds. These information are important when the objective is to quantify the formation of pollutants or when the fuel/air equivalence ratio Φ is high (air being the limiting reactant).

To overcome this issue, analytical mechanisms have been proposed in the literature (Peters, 1985; Jones and Lindstedt, 1988; Chen and Dibble, 1991; Chen et al., 1993; Boivin et al., 2011). This analytical approach consists in reducing the number of species and reactions involved in the chemical kinetic scheme by analyzing the time-scales of the different reactions. For instance, fast intermediate species or radicals that reach a quasi-steady state exhibit a negligible net reaction rate (i.e., $\dot{\omega}_k = 0$), in which case there is no need to solve for the transport equation

associated with the consumption or production of the associated species k. An additional analysis consists in determining which reactions are in equilibrium (at least partially), implying $\dot{q}_r = 0$ for the r-th reaction. Based on these assumptions, the number of species is decreased in the remaining chemical kinetic scheme. For instance, using this approach based on equilibrium approximations, Peters (1985) proposed a four-step analytical scheme for methane oxidation, including 7 species (i.e., CH_4, H_2O, CO, H_2, CO_2, H, O_2). In general, analytical mechanisms provide a physical insight into the flame structure and ignition delay as well as into intermediate species. However, they require a detailed understanding of the underlying chemistry to be able to select the relevant chemical steps. Besides, their implementation in CFD solvers is not straighforward due to the numerical treatments and the stiffness induced by their complex algebraic relations. For instance, Jones and Lindstedt (1988) proposed a four-step analytical mechanism for hydrocarbons of type C_nH_{2n+2} up to butane ($n = 4$); however, there is a negative water concentration exponent in the third reaction that induces numerical issues in the initiation of the simulation. An alternative formulation (of reduced accuracy in fuel lean regions, where there is a low fuel/air equivalence ratio) was proposed to avoid such a negative dependence.

Alternatively, Franzelli (2011) proposed a reduced chemical kinetic scheme (referred to as 2S-CH4-BFER) for premixed methane/air flames that exhibits a consistent behavior for a wide range of fuel/air equivalence ratios Φ and in particular, in rich mixtures. It is worth mentioning that this scheme results from pre-exponential adjustments (PEA) of the reaction rates as a function of the equivalence ratio Φ (Fernandez-Tarrazo et al., 2006) as detailed further in the manuscript.

Reduced (global and analytical) kinetic schemes provide a macroscopic description of the flame characteristics. The species remaining in these schemes incorporate the effects of a number of modeling choices. Thus, they cannot be directly compared to the physical species in detailed kinetic schemes (for instance, the H_2 species in Jones and Lindstedt, 1988).

↪ Tabulated chemistry

Tabulated chemistry represents the flame characteristics (i.e., the thermo-chemical variables such as the temperature T and the mass fractions Y_k) in a look-up table using a reduced number of (independent) controlling parameters such as the mixture fraction, the progress variable or the strain rate. As this number of controlling parameters is much lower than in the detailed chemistry formalism, tabulated chemistry can be regarded as a degraded representation of the detailed oxidation pathways. However, this degradation can be minimized by identifying the suitable set of controlling parameters and the appropriate flame reference configuration.

- ▷ A typical tabulation technique is ILDM[3] (Maas and Pope, 1992; van Oijen

[3] Intrinsic Low-Dimensional Manifold.

et al., 2001), based on mathematical arguments with a direct identification of the dynamic behavior of the non-linear response of the chemical system: using an eigenvalue decomposition, fast time-scales are identified and neglected in the resulting chemical system that is referred to as *manifold*.

- ▷ Gicquel et al. (2000) demonstrated that highly-reduced ILDM manifolds (with a maximum of two coordinates) cannot correctly capture molecular diffusion in mixing regions and proposed to include physical features into the look-up table through the FPI[4] technique. This approach primarily relies on the flamelet assumption, which states that a turbulent flame front can be regarded as a family of one-dimensional small laminar flame elements known as flamelets (Peters, 1984; Bradley et al., 1988).

The performance of tabulated chemistry highly depends on the amount of information stored in the look-up table and on the choice of the flame model (premixed or diffusion, adiabatic or with heat losses, etc.) that is adapted to a dedicated combustion regime. For instance, accounting for heat losses requires an additional controlling parameter as shown in Cavaliere and de Joannon (2004) and Lamouroux et al. (2013) for mild (flameless) combustion typical of industrial furnaces, or in Fiorina (2004) and Mercier et al. (2013) for premixed combustion (using RANS and LES, respectively). The formation of pollutants is also difficult to track with a low-dimensional look-up table. Since the formation of nitrogen oxides $\mathrm{NO_x}$ is governed by slow reaction processes, and since the classical definition of the progress variable cannot track the evolution of $\mathrm{NO_x}$, an additional progress variable specifically devoted to the description of $\mathrm{NO_x}$ is required (Ihme and Pitsch, 2008; Pecquery, 2013).

↪ Selected strategy for biomass combustion large-eddy simulation

Due to the characteristic time- and length-scales involved in wildfires, heat losses are important to account for in fire simulations, in particular near the interface between vegetation and flaming combustion. However, accounting for heat losses in turbulent combustion based on tabulated chemistry is still an ongoing research (Fiorina et al., 2003; Lamouroux et al., 2013; Mercier et al., 2013). One difficulty is that accounting for heat losses implies a new control parameter (related to enthalpy) in the look-up table, in particular for the case of diffusion flames (Lamouroux et al., 2013), and thereby, an additional computational cost. A second difficulty is that tabulation techniques are still difficult to extend to complex transient flame structures and fires precisely feature a transient behavior due to buoyancy effects. Since the combustion of biomass fuel generally occurs at high equivalence ratios ($\Phi > 1$) at the scales of fuel solid particles (where pyrolysis gases are released by the vegetation), reduced kinetic schemes able to represent the combustion in rich mixtures appear as a relevant strategy for representing the combustion of pyrolysis gases in this work.

[4] Flamelet Prolongation of ILDM.

2.4.3 Reduced kinetic schemes for methane/air flames

Since CFD simulations rely on a chemical kinetic scheme to represent the production or consumption rates of the gas compounds, the validation of this chemical kinetic scheme on simplified flame configurations is believed to be of primary importance in the development of a LES strategy. This validation step generally relies on the study of one-dimensional laminar premixed flames previously illustrated in Fig. 2.1 (since they are the most studied flame configuration from numerical and experimental viewpoints) and evaluates the ability of the chemical kinetic scheme to reproduce the flame macroscopic features (in terms of flame speed and thickness, burnt gas temperature, etc.).

Since this part of the thesis aims at applying turbulent combustion to fire configurations and since fuel reactants in fire applications can be considered (in a preliminary step) as a CO/CH_4 mixture, this discussion on reduced kinetic schemes primarily focuses on methane/air laminar flame properties (CO being an intermediate gas compound of combustion). In particular, the objective is to highlight the consistent behavior of the 2S-CH4-BFER reduced kinetic scheme for premixed methane/air flames with respect to experiments and detailed kinetic schemes (Franzelli, 2011).

↪ Sensitivity of the laminar flame speed to flow conditions

Since the turbulent flame speed s_T can be characterized using the laminar flame speed s_L in the flamelet regime (see Eq. 2.11), it is of primary importance to understand how this laminar flame speed s_L for methane/air combustion varies with respect to varying flow conditions in terms of fresh gas temperature T_f, pressure p and equivalence ratio Φ.

- ▷ The sensitivity of the flame speed s_L with respect to the equivalence ratio Φ is shown in Fig. 2.9(a) for ambient temperature ($T_f = 300$ K) and atmospheric pressure ($p = 1$ atm). Experimental measurements are compared to GRI-Mech3.0 predictions; results demonstrate the accuracy of GRI-Mech3.0 predictions over the range of measured equivalence ratios, even in rich-mixtures ($\Phi > 1.2$). The flame speed reaches its maximum value (0.4 m/s) near stoichiometry, while its value significantly decreases towards 0 when moving away to lean- or rich-mixture conditions.
- ▷ The effect of the unity Lewis number assumption on the flame speed s_L is also shown. While this assumption induces a simplified transport modeling (i.e., same diffusivity for heat and species), it leads to an underestimation of the flame speed over the whole range of equivalence ratios and in particular near stoichiometry. In contrast, GRI-Mech3.0 simulations with complex transport are able to reproduce experimental correlations.
- ▷ The initial temperature of the gas reactants T_f is also an important parameter that significantly modifies the flame speed s_L as highlighted in Fig. 2.9(b). The changes in s_L are described by experimental polynomial correlations (Gu

et al., 2000) such that at atmospheric pressure, the laminar flame speed s_L can be approximated as follows:

$$s_L(T_f) = s_L(T_f^0) \times \left(\frac{T_f}{T_f^0}\right)^{-\gamma_T}, \tag{2.42}$$

with for $\Phi = 1$, $s_L(T_f^0) = 0.360$ (for reference initial temperature $T_f^0 = 300$ K) and $\gamma_T = 1.612$ for methane/air premixed flames. Figure 2.9(b) shows indeed that the flame speed s_L drastically increases with the fresh gas temperature T_f. For instance, at $T_f = 600$ K, s_L is approximately equal to 1.4 m/s and is therefore multiplied by a factor higher than 3 compared to ambient temperature $T_f = 300$ K, where s_L takes a value of 0.4 m/s.

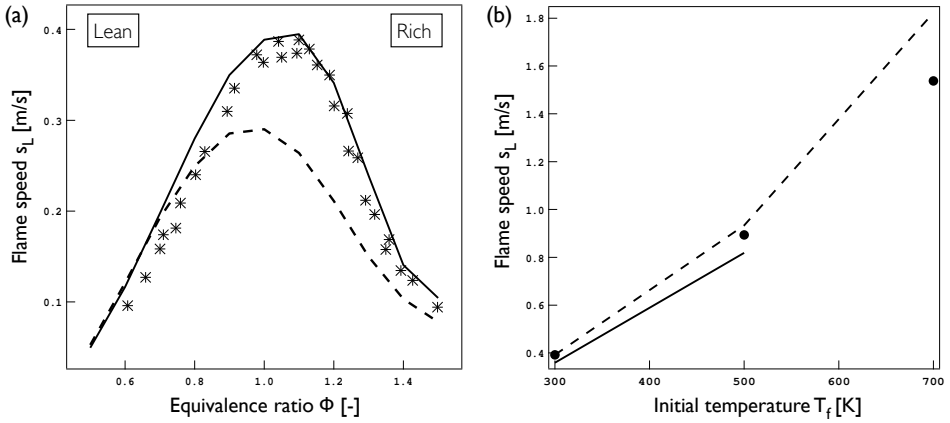


Figure 2.9: *Sensitivity of the flame speed s_L with respect to: (a) the equivalence ratio Φ (symbols correspond to experimental data, Vagelopoulos and Egolfopoulos, 1998; lines correspond to GRI-Mech3.0 predictions, the solid line being associated to complex transport and the dashed line to simplified transport); and (b) the fresh gas temperature T_f at atmospheric pressure and for $\Phi = 1$ (symbols correspond to experimental data, Gu et al., 2000; the solid line corresponds to experimental correlations, Gu et al., 2000; and the dashed line corresponds to GRI-Mech3.0 predictions). Credit: Franzelli (2011).*

↪ Evaluation of reduced kinetic schemes

In the perspective of CFD applications, the computational cost of the chemical kinetic scheme is of primary importance. For this purpose, Franzelli (2011) proposed the reduced chemical kinetic scheme 2S-CH4-BFER for methane/air flames,[5] which exhibits a consistent behavior for a wide range of equivalence ratios Φ, in particular in rich mixtures. This reduced kinetic scheme is a two-step mechanism accounting for 6 species (i.e., $CH4$, O_2, CO_2, CO, H_2O and N_2) as follows:

$$\begin{aligned} \mathrm{CH_4} + 1.5\,\mathrm{O_2} &\longrightarrow \mathrm{CO} + 2\,\mathrm{H_2O} \\ \mathrm{CO} + 0.5\,\mathrm{O_2} &\longleftrightarrow \mathrm{CO_2}. \end{aligned} \tag{2.43}$$

[5] 2S-CH4-BFER is proposed in this work for the combustion of pine needles, see Chapter 3.

In this scheme, the first reaction corresponds to methane oxidation into CO and H_2O, whereas the second reaction accounts for the recombination mechanism between CO and CO_2. The related kinetic parameters were calibrated to improve the description of the chemical flame structures for rich-mixtures and in particular, to correctly evaluate the burnt gas temperature. This calibration (based on laminar premixed flames) was performed for varying initial temperature T_f (300 K-800 K), pressure p (1-10 atm) and equivalence ratio Φ (0.6-1.6) using a PEA technique. This technique implies that a correction coefficient depending on the equivalence ratio Φ is introduced in the pre-exponential coefficient of each Arrhenius-based reaction rate. Formally, the reaction rates (noted $\dot{q}_1$ and $\dot{q}_2$ for reactions 1 and 2, respectively) are expressed as follows:

$$\begin{aligned} \dot{q}_1 &= A_1\, k_1(\Phi)\, \exp\left[-E_{a,1}/(R_g\, T)\right]\, [CH_4]^{n_{CH_4}}\, [O_2]^{n_{O_2,1}}, \\ \dot{q}_2 &= A_2\, k_2(\Phi)\, \exp\left[-E_{a,2}/(R_g\, T)\right]\, [CO]^{n_{CO}}\, [O_2]^{n_{O_2,2}}, \end{aligned} \tag{2.44}$$

where $k_1(\Phi)$ and $k_2(\Phi)$ correspond to the corrected pre-exponential coefficients associated with reactions 1 and 2, respectively. These PEA-based correction functions, represented in Fig. 2.10(b) with respect to the equivalence ratio Φ, are expressed as follows:

$$\begin{aligned} k_1(\Phi) = 2 \Bigg\{ &\left[1 + \tanh\left(\frac{\Phi_{0,1} - \Phi}{\sigma_{0,1}}\right)\right] \\ &+ B_1 \left[1 + \tanh\left(\frac{\Phi - \Phi_{1,1}}{\sigma_{1,1}}\right)\right] + C_1 \left[1 + \tanh\left(\frac{\Phi - \Phi_{2,1}}{\sigma_{2,1}}\right)\right] \Bigg\}^{-1}, \\ k_2(\Phi) = \frac{1}{2} \Bigg\{ &\left[1 + \tanh\left(\frac{\Phi_{0,2} - \Phi}{\sigma_{0,2}}\right)\right] + B_2 \left[1 + \tanh\left(\frac{\Phi - \Phi_{1,2}}{\sigma_{1,2}}\right)\right] \\ &+ C_2 \left[1 + \tanh\left(\frac{\Phi - \Phi_{2,2}}{\sigma_{2,2}}\right)\right] \left[1 + \tanh\left(\frac{\Phi_{3,2} - \Phi}{\sigma_{3,2}}\right)\right] \Bigg\}. \end{aligned}$$

The Arrhenius parameters as well as the coefficients for the correction functions $k_1(\Phi)$ and $k_2(\Phi)$ are provided in Table 2.1. Note that unity Lewis numbers were assumed for all species (consistently with the assumptions made in LES solvers).

Figure 2.10(a) shows the effect of the chemical kinetic scheme on the burnt gas temperature T_b by comparing predictions from GRI-Mech3.0 (53 species), 2S-CH4-BFER (6 species) and a one-step global kinetic scheme (with 5 species CH_4, CO_2, H_2O, O_2 and N_2). It is worth mentioning that this burnt gas temperature is controlled by the gas reactant composition since it directly results from species formation enthalpies $\Delta h^0_{f,k}$ and specific heat capacities $c_{p,k}$. As highlighted by Franzelli (2011), neglecting CO in the one-step global kinetic scheme leads to an overprediction of the burnt gas temperatures T_b for $\Phi > 1$, while 2S-CH4-BFER is able to retrieve accurate burnt gas temperature in rich mixtures. Consistently with the flame speed s_L, the maximum value for the burnt gas temperature T_b is reached near stoichiometry.

It was also found that 2S-CH4-BFER is able to retrieve correct laminar flame speeds s_L with respect to the equivalence ratio Φ, for a wide range of initial fresh gas temperatures T_f (e.g., 300 K in Fig. 2.11(a) and 700 K in Fig. 2.11(b)). Note that a similar scheme based on the PEA technique (referred to as 2S-CH4-BFER*) was developed for strained flames and could be useful to simulate partially-premixed turbulent flames, see Franzelli et al. (2010) for further details.

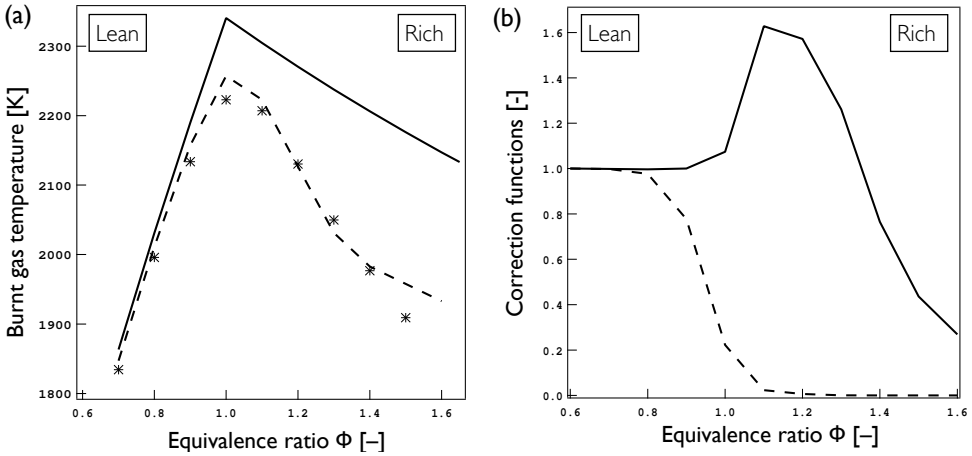


Figure 2.10: *(a) Sensitivity of the burnt gas temperature T_b with respect to the equivalence ratio Φ for a fresh gas temperature $T_f = 320$ K (symbols correspond to GRI-Mech3.0 predictions; the solid line corresponds to the one-step global scheme and the dashed line corresponds to 2S-CH4-BFER). (b) Correction functions of pre-exponential factors in 2S-CH4-BFER (the solid line corresponds to the first reaction, the dashed line corresponds to the second reaction). Credit: Franzelli (2011).*

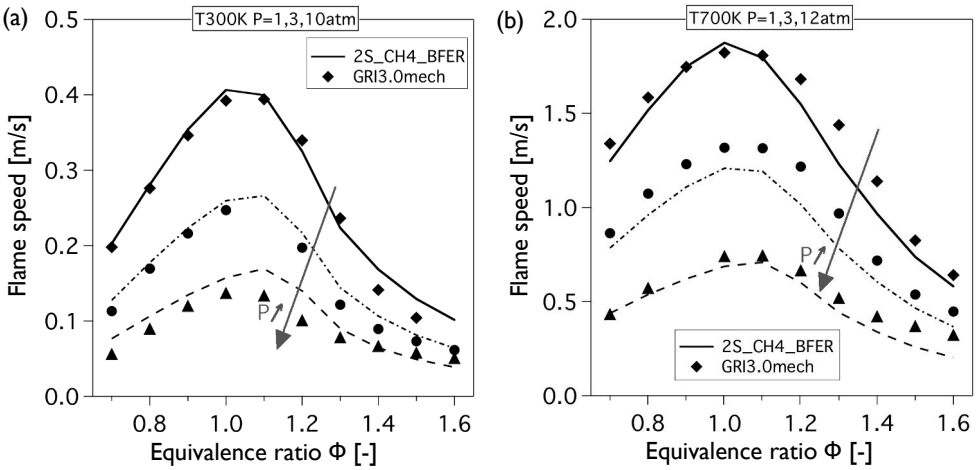


Figure 2.11: *Sensitivity of the laminar flame speed s_L with respect to the equivalence ratio Φ for varying pressures ($p = 1, 3, 10$ atm) and initial gas temperatures T_f. (a) $T_f = 300$ K. (b) $T_f = 700$ K. Symbols correspond to GRI-Mech3.0 predictions and lines correspond to 2S-CH4-BFER predictions. Credit: Franzelli (2011).*

Table 2.1: *Values for Arrhenius parameters and PEA-based correction parameters for the reduced chemical kinetic scheme 2S-CH4-BFER. Units are: mol, s, cm^3 and cal/mol.*

	CH_4 oxidation	CO-CO_2 equilibrium
Activation energy [cal/mol]	$E_{a,1} = 3.55 \times 10^4$	$E_{a,2} = 1.20 \times 10^4$
Pre-exponential factor [CGS]	$A_1 = 4.90 \times 10^9$	$A_2 = 2.00 \times 10^8$
Reaction exponents [-]	$n_{CH_4} = 0.50$ $n_{O_2,1} = 0.65$	$n_{CO} = 1.00$ $n_{O_2,2} = 0.50$
Correction parameters [-]	$\Phi_{0,1} = 1.1$ $\sigma_{0,1} = 0.09$ $B_1 = 0.37$ $\Phi_{1,1} = 1.13$ $\sigma_{1,1} = 0.03$ $C_1 = 6.7$ $\Phi_{2,1} = 1.60$ $\sigma_{2,1} = 0.22$ - -	$\Phi_{0,2} = 0.95$ $\sigma_{0,2} = 0.08$ $B_2 = 2.5 \times 10^{-5}$ $\Phi_{1,2} = 1.30$ $\sigma_{1,2} = 0.04$ $C_2 = 0.0087$ $\Phi_{2,2} = 1.20$ $\sigma_{2,2} = 0.04$ $\Phi_{3,2} = 1.2$ $\sigma_{3,2} = 0.05$

2.4.4 Overview of turbulent combustion models

Laminar flame configurations provide some important insights into the flame behavior for varying flow conditions, leading to the development of chemical kinetic schemes suitable for fuel oxidation over a wide range of equivalence ratios or gas mixture temperatures. For the purpose of LES, the next step is to incorporate the chemical kinetic scheme into a turbulent combustion framework, in order to model the filtered reaction rates $\overline{\dot{\omega}}_k$ for each gas species $k = 1, \cdots, N_g$ (see Eq. 2.38). As the most important contribution to reaction rates occurs in LES at the subgrid-scale level (Pope, 2000), flame/turbulence interactions also require modeling, either based on mixing, geometrical or statistical considerations (Fiorina et al., 2013). Note that the resolved flame thickness is generally controlled by numerical considerations in turbulent combustion models, since the flame front is generally too thin to be resolved on the LES computational grid.

↪ Mixing formalism

An early attempt to describe combustion/turbulence interactions consisted in assuming that combustion processes are essentially controlled by turbulent mixing (Spalding, 1971). Based on this assumption, the turbulent filtered reaction rates can be expressed as a function of the species mass fraction variance $(Y_k^{''})^2$. The filtered reaction rate of combustion products (referred to as index P) is for instance

expressed as follows:

$$\overline{\dot{\omega}_P} = \rho\, C_{EBU}\, \frac{\epsilon}{k} \left(\overline{(Y_P'')^2} \right)^{1/2}, \tag{2.45}$$

with k/ϵ the turbulent time scale (k being the turbulent kinetic energy and ϵ the rate of dissipation). This formulation is the foundation of the eddy break up (EBU) concept. The underlying constant C_{EBU} must be tuned in practice. The EBU model as well as the related eddy dissipation concept (in which the filtered mass fraction of combustion products is replaced by the filtered mass fraction of the deficient species, i.e., fuel in lean mixtures or oxygen in rich mixtures) eliminate the influence of chemical kinetics by representing the fast chemistry limit only (Magnussen, 2005). In spite of their limitations for temperature or pollutant predictions, they are widely spread (in particular in the fire research field) due to their steady convergence and straightforward implementation.

↪ Geometrical formalism

The geometrical formalism deals with the flame surface description and flame thickness issues using three different approaches listed below.

- ▷ **Level-set approach.** One approach consists in developing flame front-tracking techniques based on a level-set approach, also referred as the *G-equation* in the combustion research field (Kerstein et al., 1988; Pitsch, 2005, 2006; Moureau et al., 2009). This approach aims at tracking the location of the turbulent flame front (assumed infinitely thin) using a propagating transport equation that evolves the non-reacting, scalar G-field. The G-equation reads:

  $$\frac{\partial G}{\partial t} + \widetilde{\mathbf{u}} \cdot \nabla G = s_T \, |\nabla G|, \tag{2.46}$$

  with $\widetilde{\mathbf{u}}$ the filtered flow velocity vector and s_T the subgrid-scale turbulent burning velocity along the normal direction to the iso-contours of G. Based on the resolution of the Navier-Stokes equations, the flame front is then conveniently identified as an iso-contour of the G-field. However, s_T requires an additional model.

- ▷ **Filtering laminar flames.** The principle consists in filtering the laminar flame model with a filter size Δ larger than the LES computational grid size to be able to solve for the filtered flame. Thus, this approach solves for a transport equation that governs the filtered reaction progress variable c. While being similar to the G-equation, it includes physical, unsteady effects such as convection fluxes, which can be validated against DNS or experimental data (Boger et al., 1998; Duwig, 2007; Auzillon et al., 2011).

  One solution consists in tabulating the filtered flame in a look-up table with a modification of the equations, to allow for a physically-consistent flame

behavior, for instance using F-TACLES[6] (Auzillon et al., 2011, 2012). This model relies on tabulated chemistry, which has demonstrated its ability to represent thermo-chemical flame structures for LES of academic flames and laboratory-scale combustors. Note that a specific formalism named tabulated thermo-chemistry for compressible flows (TTC) is required to include tabulated chemistry in compressible LES solvers (Vicquelin, 2010; Vicquelin et al., 2011). However, the extension of tabulated chemistry to complex industrial configurations as well as to transient or non-adiabatic cases is a challenging task (Mercier et al., 2013; Lamouroux et al., 2013). Indeed, the dimension of the look-up table rapidly grows with the number of parameters required to characterize the flame structure, and the choice of the flame model used to generate the table is difficult when the combustion regime is unknown.

▷ **Artificially-thickened flames.** The principle is to operate a transformation of the spatial and temporal variables to thicken the flame, while conserving the laminar burning velocity s_L. Artificially-thickened flame models are applied when the flame (characterized by its thermal thickness δ_L) is much thinner than the LES filter size Δ. For instance, the TFLES[7] model artificially thickens the flame region by a factor $\mathcal{F}$ so that the flame is resolved on the LES computational grid (Colin et al., 2000; Angelberger et al., 2000; Légier et al., 2000). This implies that the filtered species and thermal reaction rates are divided by the factor $\mathcal{F}$ such that:

$$\overline{\dot{\omega}}_k = \frac{\mathcal{E}\,\dot{\omega}_k}{\mathcal{F}}, \quad \overline{\dot{\omega}}_T = \frac{\mathcal{E}\,\dot{\omega}_T}{\mathcal{F}}, \tag{2.47}$$

while the diffusivities are multiplied by $\mathcal{F}$ so that the flame speed s_L remains unchanged, i.e.,

$$s_L \propto \frac{\mathcal{E}\,(\mathcal{F}\,D_{th})}{\mathcal{F}\,\delta_L} = \frac{\mathcal{E}\,D_{th}}{\delta_L} = s_T. \tag{2.48}$$

Such artificially-thickened flame models have been successfully applied to LES complex configurations (Selle et al., 2004; Freitag et al., 2007; Schmitt et al., 2007; Boileau et al., 2008; Staffelbach et al., 2009); the main drawback of these models is that they change the combustion mode from transport-controlled to chemistry-controlled. Auzillon et al. (2011) compared F-TACLES and TFLES techniques and proposed a relation between the LES filter size Δ and the thickening factor $\mathcal{F}$ for a given numerical resolution.

It is worth noting that an efficiency coefficient noted $\mathcal{E}$ is introduced in Eq. (2.48) to overcome the loss in flame front wrinkling induced by artificial thickening (as illustrated in Fig. 2.12) and to therefore account for flame/turbulence interactions. The wrinkling factor is determined based on equilibrium considerations between

[6] Filtered-TAbulated Chemistry for LES.
[7] Thickened Flame model for LES.

flame wrinkling and turbulent fluctuations in TFLES (Colin et al., 2000; Charlette et al., 2002). Alternative formulations rely on similarity assumptions (Knikker et al., 2002) or dynamic adjustment following works from Germano et al. (1991), Knudsen and Pitsch (2010), Wang et al. (2011, 2012) or Veynante et al. (2012). In dynamic models, the thickening operation is only applied within the flame region (through a dynamic parameter adjustment), in order to preserve diffusion in non-reacting regions and to account for situations in which non-equilibrium flame/turbulence interactions occur. For instance, Wang et al. (2011, 2012) showed the ability of dynamic flame wrinkling factor models to reproduce the transient ignition of a flame kernel. These models are an important component of turbulent combustion models to describe unresolved flame/turbulence interactions, in level-set approaches (Pitsch, 2006), flame surface density modeling (Boger et al., 1998; Hawkes and Cant, 2000), F-TACLES (Auzillon et al., 2011) or TFLES (Colin et al., 2000).

Figure 2.12: *Impact of artificial thickening (TFLES) on the flame front wrinkling. The flame front represented in black solid lines is superimposed on the velocity field. Left:* $\mathcal{F} = 1$ *(no flame thickening). Right:* $\mathcal{F} = 5$*. Credit: Colin et al. (2000).*

↪ Statistical formalism

Combustion/turbulence interactions can also be formulated using probability density functions (PDF) to describe unresolved subgrid-scale distributions of thermochemical variables in space and time (Anand and Pope, 1987; Dopazo et al., 1997). The main issue is the determination of the filtered PDF and of the variable cross-correlations. Expectation values and correlations of the velocity field and scalar variables related to reacting flows are described based on presumed filtered PDF (FDF) using β-functions (Cook and Riley, 1998). A transport equation for the filtered PDF is derived and solved in this statistical formalism, in which the effect of chemical reactions is in a closed form (Gao and O'Brien, 1993; Colucci et al., 1998). This FDF approach is therefore suitable for LES of reacting flows. Conditional moment closure (CMC) is a more refined concept that relies on conditional moments (i.e., filtered mass fraction values for a given mixture fraction value) and a statistical closure (Bilger, 1993; Klimenko and Bilger, 1999; Garmory and Mastorakos, 2011).

2.4.5 AVBP, solver for large-eddy simulations

AVBP,[8] co-developed since 1993 by CERFACS and IFPEN, is a parallel explicit CFD code that solves the three-dimensional turbulent compressible Navier-Stokes equations for reacting flows on massive unstructured and hybrid grid meshes using DNS or LES approach (Schönfeld and Rudgyard, 1999; Colin et al., 2000; Selle et al., 2004; Moureau et al., 2005; Schmitt et al., 2007; Riber et al., 2009; Franzelli, 2011). It is dedicated to the prediction of unsteady reacting flows in combustor configurations. Employing a cell-vertex finite-volume approximation, the numerical methods underlying Avbp are explicit in time and based on a Lax-Wendroff or a finite-element type low-dissipation Taylor-Galerkin discretization in combination with a linear-preserving artificial viscosity model. Characteristics-based NSCBC[9] boundary conditions are used to handle acoustics properly (Poinsot and Lele, 1992). Avbp is highly portable to most standard computing platforms and has proven efficient on most parallel architectures. Tabulated chemistry based on F-TACLES formalism (Auzillon et al., 2011; Vicquelin et al., 2011) or Arrhenius-law-based reduced chemistry based on the TFLES formalism (Colin et al., 2000; Franzelli, 2011) allow to investigate combustion in complex configurations. It is worth mentioning that the methane/air reduced kinetic scheme 2S-CH4-BFER is available in Avbp.

2.5 Heat transfer considerations

One way to increase the physical consistency of solutions is to run multi-physics CFD simulations, coupling LES to heat transfer in solids, radiation, etc. In particular, radiation heat transfer must be accounted for in LES predictions of turbulent combustion, due to the importance of heat losses (e.g., wall heat fluxes) and to their impact on pollutant formation. Radiation heat transfer is involved in the heat balance equation through the heat source term $\dot{q}$ [W/m^3] (see Eq. 2.31 in Navier-Stokes equations) as well as in wall heat transfer. The description of radiation heat transfer is briefly described below. For a more detailed viewpoint, see Jensen et al. (2007), Coelho (2007), Poitou (2009) or Amaya et al. (2010).

2.5.1 Formulation of the radiation transfer equation

Radiation refers to the energy transfer in the form of electromagnetic radiation; this is a non-local, quasi-instantaneous, directional and spectral process (Goody and Yung, 1952; Viskanta, 1987; Taine and Petit, 1993; Modest, 2003). If not disturbed by external forcing, radiation propagates along a straight line-of-sight. However, when a beam of photons propagates across a non-transparent medium (typically, a reacting medium composed of CO_2, H_2O and soot particles), energy absorption, emission and/or scattering can occur. In the specific case of a black-body medium, all the energy carried by the beam is absorbed (there is no reflection).

[8] http://www.cerfacs.fr/4-26334-The-AVBP-code.php.
[9] Navier-Stokes Characteristics Boundary Conditions.

The radiation transfer equation (RTE) corresponds to the balance equation of the radiation energy (also referred to as *spectral radiance*) in a control volume of the crossing medium. More precisely, the spectral radiance in the direction $\mathbf{u}$ associated with the wavenumber ν_{rad} is noted $I_{\nu_{rad}}$ [W/m^2/sr/Hz] (sr standing for *steradian*) and corresponds to the monochromatic heat flux density $d\Phi_{\nu_{rad}}(\mathbf{u},t)$ that crosses the gas surface $\mathrm{d}s$ within a time delay $\mathrm{d}t$:

$$I_{\nu_{rad}}(\mathbf{x},\mathbf{u},t) = \frac{\mathrm{d}\Phi_{\nu_{rad}}(\mathbf{u},t)}{\mathbf{u}\cdot\mathbf{n}\,\mathrm{d}\Omega\,\mathrm{d}t\,\mathrm{d}s}, \tag{2.49}$$

with $\mathbf{n}$ the normal direction to the surface $\mathrm{d}s$ and $\mathrm{d}\Omega$ the solid angle around the direction $\mathbf{u}$. For clarity purposes, a schematic is presented in Fig. 2.13.

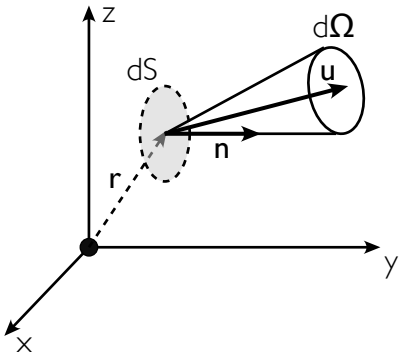


Figure 2.13: *Definition of the spectral radiance I_ν. Credit: Poitou (2009).*

For a given wavenumber ν_{rad}, the differential form of the RTE in a non-scattering medium can be formulated as follows:

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(I_{\nu_{rad}}(\mathbf{x},\mathbf{u})\Big) = \kappa_{\nu_{rad}}\Big(I^0_{\nu_{rad}}(\mathbf{x}) - I_{\nu_{rad}}(\mathbf{x},\mathbf{u})\Big), \tag{2.50}$$

with $\kappa_{\nu_{rad}}$ the absorption coefficient, $I^0_{\nu_{rad}}$ the equilibrium Planck radiance and $I_{\nu_{rad}}$ the incident intensity at the point $\mathbf{x}$ in the direction $\mathbf{u}$. Using this formalism, the macroscopic heat source term $\dot{q}$ involved in the fluid heat balance equation is retrieved by calculating a double integration of the RTE in physical and spectral spaces such that:

$$\dot{q} = \dot{q}(\mathbf{x}) = \int_0^\infty \kappa_{\nu_{rad}}\left(4\pi\, I^0_{\nu_{rad}}(\mathbf{x}) - \int_{4\pi} I_{\nu_{rad}}(\mathbf{x},\mathbf{u})\,\mathrm{d}\Omega\right)\mathrm{d}\nu_{rad}. \tag{2.51}$$

2.5.2 PRISSMA, radiation solver

PRISSMA[10] is a radiation solver specifically designed for combustion applications and relying on a discrete ordinate method (DOM) with different angular discretizations and spectral models (Poitou, 2009; Amaya et al., 2010; Poitou et al., 2011,

[10] http://www.cerfacs.fr/prissma/.

2012). The numerical strategy is based on the S4 quadrature (i.e., 24 directions) combined with the tabulated full spectrum SNBcK (FS-SNBcK) spectral model (Goutiere et al., 2000, 2002), allowing for a suitable balance between accuracy and computational cost.

Conclusion

This chapter provides an overview of the state-of-the-art modeling of turbulent combustion. Large-eddy simulation capabilities are able to capture flame-scale dynamic processes using subgrid-scale turbulence modeling to account for turbulent wrinkling effects and using chemical kinetic schemes to account for the production or consumption of chemical gas compounds.

In turbulent combustion research, AVBP is a reference software that has been validated in multiple configurations for premixed, partially-premixed or diffusion flames as well as for lean or rich multi-species mixtures. Since it features an explicit compressible solver with a small computational time step to solve for acoustics, the chemistry of combustion is accounted for by reduced or tabulated strategies, requiring evaluation and calibration in simplified flame configurations. Thus, AVBP coupled to the radiation solver PRISSMA provides a powerful tool to explore flame-scale processes underlying laboratory-scale fire propagation (if associated with a model for biomass fuel thermal degradation) and to assess/analyze the validity of empirical wildfire spread-rate modeling at regional scales.

Chapter 3

Strategy for multi-physics large-eddy simulations of fire spread

The direction and speed at which a wildfire propagates results from multi-scale interactions between multi-physical processes, namely the pyrolysis processes occurring at vegetation scale, the combustion and flow dynamics at flame scale, and the atmospheric dynamics and chemistry at large regional scale. Current wildfire spread simulators rely on a semi-empirical parameterization of the rate of spread (ROS) with respect to local environmental conditions (Rothermel, 1972; Balbi et al., 2009). This parameterization is based on a global energy balance in the unburnt vegetation subject to flame-induced pre-heating. Thus, empirical spread-rate models imply a rough heat transfer description and thereby, significant assumptions in the ROS behavior (e.g., dominating heat transfer, fuel bed configuration, homogeneity of local wind conditions). In contrast, relevant insight into wildfire dynamics has been recently obtained via detailed numerical simulations, which for instance differentiate the heat transfer mechanisms contributing to vegetation pre-heating (e.g., convection, radiation). Still at its early stages, this computational fluid dynamics (CFD) approach of wildfire spread could be improved by performing large-eddy simulations (LES) at fine flame scale (on the order of 1 mm), allowing for more accurate subgrid-scale combustion modeling.

This chapter presents the strategy developed in this thesis to simulate flame-scale interactions between combustion, turbulence and pyrolysis for a laboratory-scale fire configuration.

Outline

3.1 Strategy guidelines

This part of the thesis aims at building multi-physics flame-scale LES of a laboratory-scale fire based on a model coupling strategy and HPC, and then at validating them against experimental measurements. Typical laboratory-scale experiments are illustrated in Fig. 3.1 for flat or upslope propagation. Objectives are (1) to provide a more comprehensive understanding of the fine-scale mechanisms underlying fire propagation; and (2) to obtain more accurate correlations between quantities of interest that could be useful to:[1]

- ▷ improve the ROS parameterization in semi-empirical models and in current fire spread simulators such as Firefly or Forefire/Méso-NH;
- ▷ study wildfire emission factors (i.e., amount of chemical species released into the atmosphere) involved in current air quality estimations.

In particular, this study focuses on the interactions between combustion and turbulence by introducing fire physics into the existing compressible reactive LES solver AVBP, which has been actively applied in the field of industrial combustion in recent decades (see Chapter 2). Solving for laboratory-scale fires requires to generalize Navier-Stokes balance equations to buoyancy-induced flows and to introduce models for radiation heat transfer, biomass fuel pyrolysis as well as chemical kinetics of the combustion of pyrolysis gases. A multi-physics CFD coupling based on HPC is

[1] The objectives of the ANR-IDEA project are explained in detail in Chapter 1.

therefore necessary to describe interactions between porous biomass fuel thermal degradation, gas-phase convection, combustion and heat transfer processes.

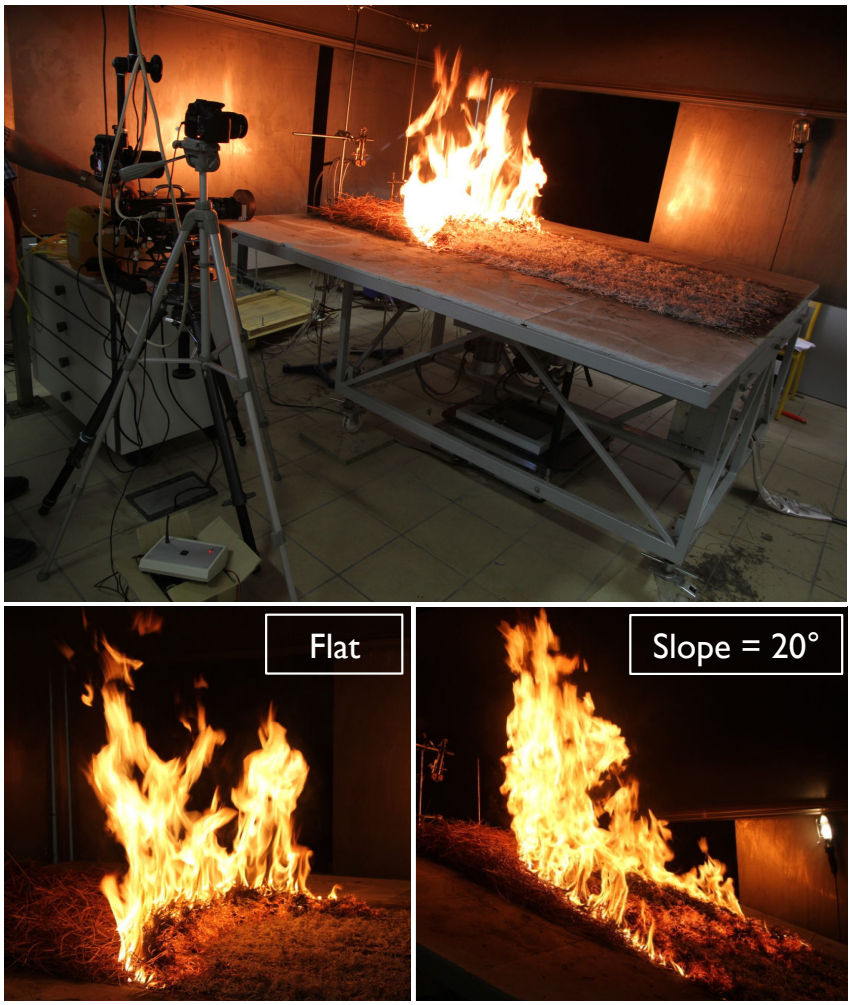


Figure 3.1: *Instantaneous snapshots of laboratory-scale fire experiments for flat and* $20°$*-upslope configurations. Fires propagate through a fuel bed of 2 m* $\times$ *1 m forming a pine needle layer. The mean height of the flame over the fire duration is 47 cm for flat propagation and 67 cm for upslope propagation. Credit: CNRS-SPE laboratory,* `http://spe.univ-corse.fr/`.

For this purpose, the following aspects are developed in this thesis and in particular in this chapter:

- ▷ **Vegetation representation.** Vegetation (i.e., layer of pine needles) is considered as a porous medium made of a solid phase (pine needles) and a gas phase (air in-between the solid fuel particles).
- ▷ **Vegetation thermal degradation model.** A model to account for the vegetation thermal degradation named Pyrowo is proposed and verified

against thermogravimetry experimental data (TGA): its objective is to provide a time-dependent gas flow from the vegetation solid-phase at the top of the pine needle layer (this gas flow is constituted by the gas compounds released during the pyrolysis step, e.g., CH_4, CO, CO_2, H_2O).

- ▷ **Combustion properties of pyrolysis gases.** Combustion properties of the gas flow from the vegetation solid-phase are studied with respect to its composition, temperature and flow-induced strain rate to gain insight into the related chemical time-scales: this study is performed in a simplified geometry (auto-ignition, one-dimensional premixed and counter-flow flames) and based on laminar flames using specific softwares (allowing for comparison with detailed kinetic schemes and complex transport properties). The objective is to validate the reduced chemical kinetic scheme 2S-CH4-BFER used in the AVBP fluid solver to simulate the combustion of pyrolysis gases in a laboratory-scale configuration. A discussion on the choice of the turbulent combustion model in AVBP is also provided.
- ▷ **Buoyancy-driven flow.** Since the fire plume is subject to buoyancy, the acceleration due to gravity is added in the AVBP momentum conservation equation to account for buoyancy-driven flows and induced non-stationary effects.
- ▷ **Radiation.** The radiation heat transfer from the flame to the vegetation that enhances its thermal degradation and the pyrolysis process is simulated using the PRISSMA radiation solver.
- ▷ **Coupling strategy for flame-scale simulations.** Simulating laboratory-scale fires requires a coupling strategy based on the most recent MPI[2] and HPC technologies to combine and synchronize AVBP, PRISSMA and PYROWO models. This coupling strategy relies on the OPENPALM dynamic code coupler (see Appendix B).

3.2 Formulation of the vegetation pyrolysis submodel

While much progress has been achieved in the recent decades in our basic understanding of biomass pyrolysis (*biomass* referring to all organic materials with wood as the main representative), its simulation remains a challenging task due to the unknown features of the vegetation and to the complex underlying physico-chemical mechanisms occurring at vegetation molecular scales. A wide range of modeling approaches has been proposed in literature, from the consideration of a single particle decomposition to the macroscopic representation of the vegetation pyrolysis, within the framework of biomass power generation or fire safety engineering. See Di Blasi (1993, 2008), Shafizadeh (1977, 1982) or Peters and Bruch (2003) for a detailed review of biomass pyrolysis modeling.

[2] Message Passing Interface.

3.2.1 Characterization of porous vegetation

↪ Representation of vegetation as a porous medium

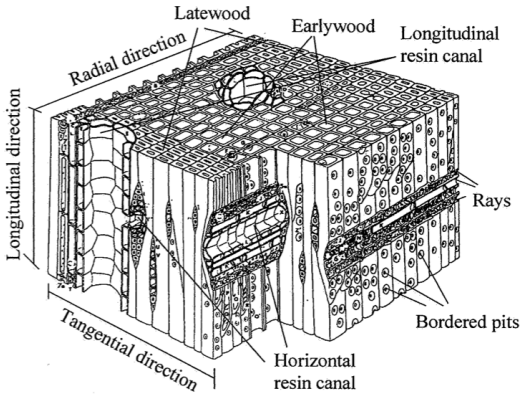
As illustrated in Fig. 3.2, the vegetation used in the present laboratory-scale experiments is a fuel bed made of maritime pine needles, typical of the Mediterranean ecosystem and referred to as *Pinus Pinaster*. From a macroscopic viewpoint, this vegetation bed can be regarded as a composition of a solid phase (pine needles can be assimilated to solid fuel particles with varying properties) and a gas phase (i.e., air in-between the solid fuel particles). The ratio of the volume occupied by the solid phase to the total volume of the vegetation bed is characterized by the packing ratio denoted by β_v; typically, β_v varies between 0.01 and 5 % for *Pinus Pinaster* (alternatively, the term $(1 - \beta_v)$ corresponds to the biomass fuel porosity). This porous structure is essential to account for, since combustion processes are highly dependent on the fuel/air equivalence ratio Φ and since the flammable compounds released during thermal degradation are in the gas phase of the biomass fuel bed before being entrained towards the flame. The macroscopic properties of the vegetation solid and gas phases (in terms of moisture content M_v and particle surface-area-to-volume ratio Σ_v for instance) must be regarded as equivalent properties to the biomass structure from microscopic to macroscopic scales. Thus, these macroscopic properties globally represent the mean effects of the underlying scales, where porous structures could also be identified (see Fig. C.1, Appendix C).

Figure 3.2: *Vegetation of thickness δ_v (5-10 cm) made of pine needles (solid fuel particles or solid phase) and a gas phase.*

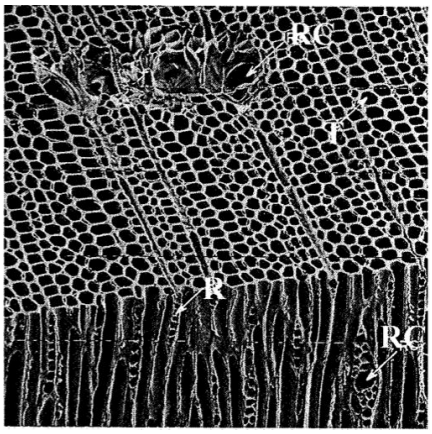
Biomass pyrolysis can be considered as the chemical degradation of an organic material through thermal decomposition. It is highly related to the material physico-chemical properties: physical properties characterize macroscopic structures that define the transport mechanisms as well as the moisture content distribution within the porous vegetation, while chemical properties refer to the biomass fuel composition at molecular scale.

↪ Chemical composition

Wood is composed of elongated cells, mainly oriented in the longitudinal direction of the stem (i.e., in the main growth direction) as shown in Fig. 3.3 for soft woods (e.g., conifers, pines). From a chemical perspective, these wood cells are made up of polymers (among whom cellulose, hemicellulose and lignin), which form a fibrous structural tissue with some amounts of extractives (e.g., resin) and inorganic compounds (e.g., potassium, sodium and magnesium mineral elements). It is worth noting that these inorganic compounds represent the non-combustible biomass content and are responsible for the formation of ashes during thermal degradation.



(a) Schematic of the inner structure.



(b) Transverse and longitudinal faces, T: Tracheid, R: Rays, RC: Resin canals.

Figure 3.3: *Typical structure of soft wood species (e.g., conifers, pines). Credit: Siau (1984), Butterfield (1980).*

As the main component of vegetation cell walls, cellulose is composed of long straight chains commonly represented with the structural formula $(C_6H_{10}O_5)_n$; n (typically 10,000 in unaltered wood) indicating the degree of polymerization. The different chains have a tendency to be tight together via hydrogen bonds that provide high tensile strength to vegetation. Hemicellulose surrounds the cellulose fibers of the cell walls and consists of shorter chains than cellulose (only 50 to 200 molecular units). While the cell walls of vegetation are formed by fibers made of cellulose and hemicellulose, lignin is located between the cells and thereby, serves the function of binding the fibers together. Thus, lignin is largely responsible for the rigidity of vegetation.

The proportion and structure of these chemical compounds depend on the biomass species and on its state of decomposition (living or dead cells). In particular, wood is known to be constituted by 40-60 % hemicellulose and cellulose, 25-40 % lignin, 4-15 % extractives and 0.5-5 % ash (in weight % on a dry basis). Table 3.1 compares this composition for different biomass types (i.e., soft wood, hard wood, bark, peat) and for *Pinus Pinaster*. Note that bark and peat generally contain

more extractives and less cell wall materials than woody tissues. Also, the lignin content is substantially higher for peat, indicating that more lignin is present when a biomass species is significantly decomposed and contains a large amount of dead cells (lignin could be an important factor enhancing smoldering combustion). A detailed description of wood composition is provided in Bronli (1996).

Table 3.1: *Chemical composition of biomass fuel types (weight % on a dry basis). PP corresponds to Pinus Pinaster. Credits:* `http://www.fao.org/docrep/T4470E/t4470e0a.htm` *and Leroy (2007) for Pinus Pinaster.*

Species	Soft wood	Hard wood	Pine bark	Peat	PP
Cellulose	41	39	34	10	38
Hemicellulose	24	35	16	32	5
Lignin	28	20	34	44	39
Extractives	2	3	14	11	13
Ash	0.4	0.3	2	6	5

↪ Anisotropic physical properties

At microscopic scale, the solid phase of the vegetation bed (also referred to as *p-phase*) can itself be considered as a porous medium that is made of a skeletal solid matrix (i.e., cell walls) and a pore network (i.e., voids filled with either liquid or vapor phase). The volume fraction occupied by the pores over the total solid vegetation volume is typically characterized by the p-phase porosity variable noted β_p; β_p varies from 0 (i.e., no pore within the material) to 1 (i.e., no solid matrix within the material). More details are provided in Appendix C.

Depending on the pore geometrical arrangement, vegetation exhibits highly anisotropic properties (in terms of thermal conductivity and permeability to gas flow for instance). It also contains a non-negligible moisture content M_v, either in the pores (i.e., as free water, liquid or vapor, held by capillary forces) or as bound water in the cell walls (i.e., water molecules attached to cellulose and hemicellulose via hydrogen bonds). From a modeling viewpoint, M_v is commonly expressed as the mass of moisture divided by the mass of dry vegetation (Rothermel, 1972). A description of some specific properties of vegetation moisture, which are useful to understand the parameters in semi-empirical ROS models, follows.

- ▷ **Bound water properties.** When organic materials undergo thermal degradation, moisture leaves first cell cavities and second cell walls. This bound water leaving cell walls transforms into free liquid water. Note that there always remains a low amount of moisture (a few % typically) in oven-dried vegetation, in the form of bound water.
- ▷ **Free water properties.** The energy required to evaporate the free liquid water can be considered to be equal to the latent heat of evaporation noted

Δh_{vap} ($\Delta h_{vap} = 2260$ kJ/kg).

- ▷ **Fiber saturation point.** The fiber saturation point (noted $M_{v,ext}$) is defined as the moisture content at which cell walls are saturated with bound water, while there is no free water within the pore network (i.e., in the cell cavities). An average value $M_{v,ext} = 30\ \%$ is typically assumed for biomass fuels, corresponding to the moisture of extinction introduced in the semi-empirical ROS model due to Rothermel (1972).
- ▷ **Equilibrium moisture content.** When only bound water is present in vegetation, an equilibrium exists between the biomass moisture content and the relative humidity of the surrounding air, called the equilibrium moisture content (EMC). Typical of dead cells in biomass fuels, this EMC is a dynamic equilibrium as it depends on air humidity and temperature.

↪ Energy content in biomass fuels

The thermal energy emitted by an active fire is a direct result of the energy stored in biomass, being released as heat when the biomass fuel undergoes combustion. It is therefore important to understand the factors affecting the energy content of vegetation (also called the *heat of combustion* and noted Δh_c [J/kg]). From a theoretical viewpoint, Δh_c is defined as the amount of heat released per unit fuel mass under complete combustion. Since the composition of biomass fuels (on a dry and ash-free basis) is relatively uniform in terms of carbon, oxygen and hydrogen atoms, the theoretical heat of combustion remains between 16 and 22 MJ/kg for most biomass species (Rothermel, 1972; Finney, 1998; Tihay et al., 2009). The value for Δh_c is set to 18.6 MJ/kg in Rothermel's ROS model. Note that this value is low compared to hydrocarbon-based industrial fuels typical of gas turbines, since the heat of combustion of kerosene (46 MJ/kg), methane (55.5 MJ/kg) and hydrogen (141.9 MJ/kg) is at least 2.5 times higher than for biomass fuels.

However, the actual amount of heat released during combustion is influenced to some degree by the presence of moisture, volatile resins or inorganic materials (Orfao et al., 1999). While the ash content ranges from about 1 % in wood to more than 5 % in grass, the moisture content M_v may vary from 2.5 % (e.g., in dead savanna grasslands) to 200 % (e.g., in fresh needles and leaves) of the dry vegetation weight. Since flaming combustion starts when the bound moisture is evaporated from vegetation, and since biomass moisture evaporation is a highly endothermic process associated with a high latent heat of evaporation Δh_{vap}, a high moisture content M_v has the capacity to stop a fire or to significantly slow it down, making combustion incomplete. That is why Rothermel's ROS model assumes that above a 30-% threshold value for moisture content (named the *moisture content at extinction* and noted $M_{v,ext}$), a fire cannot propagate. This moisture content at extinction corresponds to the fiber saturation point: dead fuel particles characterized by low bound moisture content and size (a few millimeters typically) significantly contribute to pyrolysis and sustain combustion. Note that the ignition and combus-

tion of living biomass fuels such as chaparral is still poorly understood (Mcallister and Finney, 2014).

3.2.2 PYROWO, model for biomass thermal degradation

↪ Key ideas underlying the PYROWO model

Since AVBP only solves for the gaseous phase, a model that accounts for the thermal degradation of the porous vegetation subject to strong radiation and convection heat fluxes is required. In general, physical submodels used to describe solid fuel sources in CFD solvers belong to two categories: (1) design fire models in which the fuel mass loss rate is prescribed using an empirical law; and (2) thermal feedback sensitive models in which the fuel mass loss rate is derived from a physics-based model as a function of the gas-to-fuel thermal loading (Hopkins and Quintiere, 1996; Novozhilov et al., 1996). In this work, the biomass fuel thermal degradation model named PYROWO belongs to the second category and was developed in the perspective of performing multi-physics flame-scale LES. The ultimate goal of PYROWO is to define the interface between the biomass fuel bed and the gaseous phase solved by AVBP and thereby, to describe properly the production of pyrolysis gases without considering a fully multiphase formulation retained, for example, in Linn et al. (2002), Morvan and Dupuy (2004) or Séro-Guillaume and Margerit (2002).

PYROWO requires a heat source, provided by the radiation heat flux received at the top of the vegetation bed. This radiation heat flux is a fraction of the heat released by the flame (noted χ_{rad}), which reaches the unburnt vegetation ahead of the fire front. It increases the temperature of the vegetation solid phase and leads to the release of pyrolysis gases. In-turn, these pyrolysis gases (i.e., the fuel reactants of the combustion) sustain flaming combustion and enhance fire spread, by providing a fuel-to-gas mass flux. PYROWO also provides the gas temperature and composition. The flowchart of PYROWO is provided in Fig. 3.4.

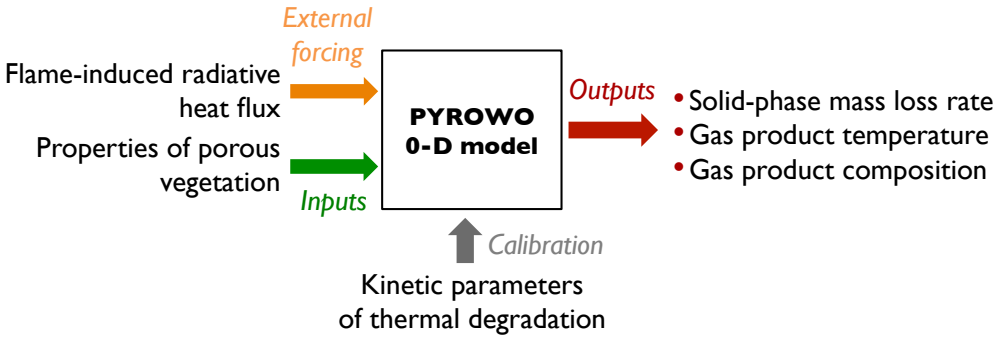


Figure 3.4: *Flowchart of* PYROWO*, the biomass thermal degradation model.*

Note that the properties of the solid particles of vegetation are difficult to estimate a priori, while being essential for the performance of a thermal degradation model.

For this purpose, they are calibrated against experimental data based on thermogravimetry and/or cone calorimeter analyses (Broido, 1969; Huggett, 1980), using optimization procedures. A set of vegetation properties and kinetic parameters that best reproduces experimental data can be retrieved and used as model parameters in PYROWO.

Note also that particular attention must be paid to the gas composition since it significantly affects the heat of combustion Δh_c associated with vegetation and thereby, its flammability potential capacity. Experimental measurements have shown that there is a wide range of gas compounds emitted by the pre-heated vegetation, among whom CH_4, CO, CO_2, H_2O and lower amounts of C_2 hydrocarbons (see Table 3.2). Since vegetation is a porous medium, air is also contained in its gas phase and therefore, some amounts of N_2 and O_2 must be accounted for in the gas mixture released by the vegetation. However, the ratio between air and pyrolysis gases (i.e., the fuel/air equivalence ratio Φ) in representative configurations of fire spread is difficult to estimate and thereby, subject to significant uncertainties.

Table 3.2: *Example of composition of pyrolysis gases $Y_{v,k}$ $[-]$ (mass fractions) released during Pinus Pinaster thermal degradation (experimental data). Credit: Tihay et al. (2009).*

Species	CO_2	H_2O	CO	CH_4	C_2H_4	C_2H_6
Mass fraction	0.640	0.089	0.171	0.029	0.007	0.011
Species	C_3H_6	C_3H_8	C_4H_6	C_4H_8	C_4H_{10}	H_2
Mass fraction	0.002	0.008	0.022	0.014	0.007	0.0

↪ Model assumptions

PYROWO is a 0-D model neglecting the transport within the biomass fuel bed and including a system of ordinary differential equations (ODE) that (1) accounts for the radiation source term emitted by the flame and absorbed by the biomass fuel bed (noted $\dot{q}_{rad,fl}$ [W/m^3] in the following[3]), and (2) reproduces the macroscopic behavior of vegetation thermal degradation. Inspired by the multiphase formulations due to Grishin (1997) as well as additional contributions of Larini et al. (1998), Morvan and Dupuy (2004) and Consalvi et al. (2011), PYROWO relies on the following assumptions.

- ▷ **Homogeneous vegetation.** Vegetation is considered as a porous medium made of a gas phase and a single family of solid particles sharing common properties (e.g., surface-area-to-volume ratio Σ_v, moisture content M_v). This assumption is relevant in the present study since *Pinus Pinaster* needles are uniformly-distributed in the target laboratory-scale fuel bed (see Fig. 3.2).

[3] Surface quantities at the top of the biomass fuel bed are involved in the coupling strategy presented in Section 3.4; the mathematical formulation underlying PYROWO considers physical quantities per unit volume of the fuel bed.

▷ **Reduced thermal degradation mechanism.** The solid phase of the vegetation is initially represented as an ensemble of dry wood and liquid moisture $(\mathrm{H_2O})_l$. It is supposed to undergo thermal degradation in three steps (which can overlap) illustrated in Fig. 3.5 and listed below (Bronli, 1996; Repellin, 2006).

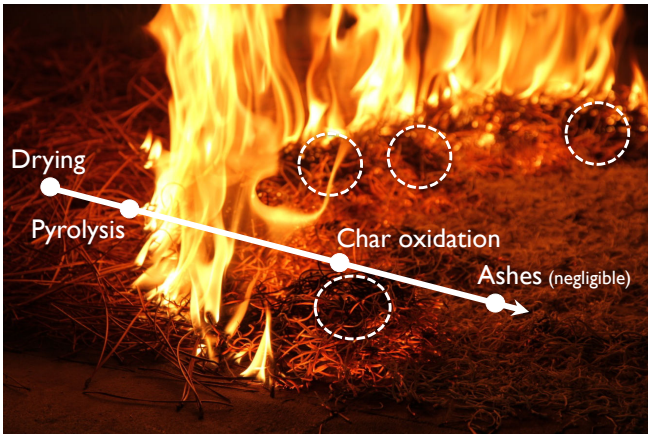


Figure 3.5: *Snapshot of the laboratory-scale fire with a schematic of the different steps underlying* Pinus Pinaster *thermal degradation (black zones surrounded by white dashed lines correspond to partially-burnt fuel where the combustion is momentarily stopped by oxygen defect).*

(i) **Drying** corresponds to the formation of water vapor $(\mathrm{H_2O})_v$ in replacement of fuel moisture $(\mathrm{H_2O})_l$ (i.e., free liquid and bound water) ahead of the flame zone. This step is represented as follows:

$$(\mathrm{H_2O})_\mathrm{l} \longrightarrow (\mathrm{H_2O})_\mathrm{v}.$$

(ii) **Pyrolysis** corresponds to the degradation of dry solid materials into (1) flammable gas compounds (e.g., $\mathrm{CH_4}$, CO), and (2) char materials in a certain proportion represented by the coefficient χ_{char} and that depends on the lignin content in vegetation (Di Blasi, 2008). This degradation of the dry mass content is represented as follows:

$$\text{dry} \longrightarrow \chi_{char}\,\text{char} + (1-\chi_{char})\,\text{volatiles}.$$

Note that *char* refers to a black residue mainly made of carbon, corresponding to the black zones in the post-flame region shown in Fig. 3.5. These zones are partially burnt and not supplied in oxygen such that combustion is momentarily stopped by oxygen defect.

(iii) **Char oxidation** corresponds to a heterogeneous mode of combustion at high temperatures, between the remaining char solid and the oxygen that is entrained towards the flame by buoyancy effects. This process typically occurs in the incandescent zones of the vegetation bed, behind

the flame region, in Fig. 3.5. This stage can be represented with the following complete oxidation reaction:

$$\text{char} + \chi_{O_2}\,O_2 \longrightarrow (1+\chi_{O_2})\,CO_2.$$

Once char oxidation is complete, only ashes remain; they are an indicator of the mineral content of vegetation (e.g., potassium, sodium, calcium). For the laboratory-scale fire experiments studied in this thesis, the mass of ashes only represents 2.8 % of the total mass of the initial fuel bed before ignition and is therefore assumed negligible. It is worth mentioning that the tar[4] intermediate products of vegetation thermal degradation is not considered here (Di Blasi, 2008). The priority in developing physically-consistent CFD simulations is to consider the important features of vegetation thermal degradation (i.e., pyrolysis and char reactions) and to fit the related-kinetic parameters against experimental data.

- ▷ **Instantaneous release to the gaseous phase.** To eliminate the complexity of the gas flow through the solid phase of the vegetation (i.e., p-phase), pyrolysis products are assumed to be removed out of the solid particles instantaneously upon their release. Thus, there is no accumulation of pyrolysis products and no further chemical reactions within the solid phase of the porous vegetation. Mass conservation therefore implies that the mass rate at which pyrolysis gases are released is equal to the negative counterpart of the vegetation solid-phase mass loss rate, or equivalently that the vegetation mass loss is totally gained by the gaseous phase (through the AVBP/PYROWO interface as detailed in Section 3.4).
- ▷ **Thermal equilibrium in the porous vegetation.** Thermal equilibrium between the gas and solid phases of the porous vegetation is assumed, meaning that the exit temperature of pyrolysis products (referred to as *gas temperature*) is assumed equal to the temperature of the vegetation solid phase (solid fuel particles are considered as thermally-thin).

↪ Characterization of the biomass solid phase

The solid phase of the biomass fuel bed in the PYROWO model is characterized by the following features (the subscript v referring to *vegetation*):

- ▷ the **temperature** of the gas and solid phases of the vegetation noted T_v [K], corresponding to the mean value through the vegetation layer thickness δ_v.
- ▷ the **moisture content** $Y_{v,wat}$ $[-]$, corresponding to the mass fraction of liquid water $(H_2O)_l$ contained in the solid phase of the vegetation and consumed

[4] Detailed models of biomass fuel thermal degradation generally represent a series of primary reactions where the vegetation transforms into char, flammable gases and tar (i.e., high-molecular weight products that vaporize at pyrolysis temperatures but condense at ambient temperatures), and a series of secondary reactions during which tar degrades into flammable gases and char.

during the drying phase (equivalent to M_v in empirical ROS models and typically between 1 and 10 % for pine needles). The drying mass rate is noted $\dot{m}_{v,vap}$ [kg/m^3/s].

- ▷ the **dry wood content** $Y_{v,dry}$ $[-]$, corresponding to the mass fraction of combustible dry materials contained in the solid phase of the vegetation and thereby, to the proportion of materials that can undergo pyrolysis and release flammable gases. The pyrolysis mass rate is noted $\dot{m}_{v,pyr}$ [kg/m^3/s]. By considering all hydrocarbons as CH_4 in Table 3.2 (Tihay et al., 2009; Consalvi et al., 2011), the mass fractions of pyrolysis gases $Y_{k,v}$ are taken as $Y_{v,CH_4} = 0.10$, $Y_{v,CO} = 0.171$, $Y_{v,CO_2} = 0.64$ and $Y_{v,H_2O} = 0.089$.
- ▷ the **char content** $Y_{v,char}$ $[-]$, corresponding to the mass fraction of char materials remaining in the biomass solid phase after drying and pyrolysis. Char formation is represented as a fraction of the pyrolysis mass rate $\dot{m}_{v,pyr}$ denoted by χ_{char}, typically between 0.30 and 0.40. Initially, $Y^0_{v,char} = 0$ and $Y^0_{v,wat} + Y^0_{v,dry} = 1$, the superscript 0 referring to the initial state. Char oxidation is represented through the char mass loss rate noted $\dot{m}_{v,char}$ [kg/m^3/s].
- ▷ the **mass density of the solid fuel particles** ρ_p [kg/m^3] corresponding to an intrinsic property of the biomass solid phase. Typically, ρ_p is between 500 and 800 kg/m^3 before thermal degradation and decreases when the vegetation undergoes drying and pyrolysis (i.e., solid-to-gas phase transformations without change in the volume of the fuel solid particles).
- ▷ the **fuel packing ratio** β_v $[-]$, corresponding to the volume fraction of the solid phase of the vegetation (typically, between 1 and 5 % for pine needles), which is only subject to change during char oxidation that induces volume variations of the fuel particles by degrading the remaining char-based solid materials. Note that in Rothermel's ROS model, the biomass packing ratio is assumed constant.
- ▷ the **specific heat of the vegetation** $c_{p,v}$ [J/K/kg], assumed to remain constant during thermal degradation (typically between 1,000 and 1,500 J/K/kg for dry and char materials).
- ▷ the **fuel particle surface-area-to-volume ratio** Σ_v [1/m], assumed to remain constant during thermal degradation (typically between 3,000 and 5,000 1/m for pine needles)

↪ Balance equations for the biomass solid phase

The governing equations of vegetation thermal degradation are defined in the following set of ODE. These 0-D equations (locally) apply to a control volume of vegetation, in which the properties of the solid particles are averaged and thereby, assumed uniform. The detailed derivation of this ODE system with respect to current multiphase formulation for a biomass porous medium (Larini et al., 1998;

Séro-Guillaume and Margerit, 2002; Morvan and Dupuy, 2004) is presented in Appendix C.

▷ **Energy balance in vegetation**

$$\beta_v \, \rho_p \, c_{p,v} \, \frac{\mathrm{d}T_v}{\mathrm{d}t} = \dot{q}_v = \dot{q}_{conv} + \dot{q}_{rad} + \dot{q}_{chem}. \tag{3.1}$$

▷ **Moisture content**

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\beta_v \rho_p \, Y_{v,wat}\Big) = -\dot{m}_{v,vap}. \tag{3.2}$$

▷ **Dry wood content**

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\beta_v \, \rho_p \, Y_{v,dry}\Big) = -\dot{m}_{v,pyr}. \tag{3.3}$$

▷ **Char content**

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\beta_v \, \rho_p \, Y_{v,char}\Big) = \chi_{char} \, \dot{m}_{v,pyr} - \dot{m}_{v,char}. \tag{3.4}$$

▷ **Mass density of the solid phase**

$$\beta_v \, \frac{\mathrm{d}\rho_p}{\mathrm{d}t} = -\dot{m}_{v,vap} - (1 - \chi_{char}) \, \dot{m}_{v,pyr}. \tag{3.5}$$

▷ **Vegetation packing ratio**

$$\rho_p \, \frac{\mathrm{d}\beta_v}{\mathrm{d}t} = -\dot{m}_{v,char}. \tag{3.6}$$

↪ **Mass loss rate model**

▷ The **drying mass rate** $\dot{m}_{v,vap}$ [kg/m^3/s] is represented as an endothermic reaction using the following Arrhenius-type formulation (Grishin, 1997; Morvan et al., 2000):

$$\dot{m}_{v,vap} = (\beta_v \, \rho_p \, Y_{v,wat}) \; T_v^{-\frac{1}{2}} \, k_{vap} \, \exp\left[\frac{-E_{vap}}{R_g \, T_v}\right], \tag{3.7}$$

with k_{vap} [K$^{1/2}$/s] and $T_{vap} = E_{vap}/R_g$ [K] the pre-exponential factor and the activation temperature related to drying, respectively. Note that moisture evaporation can be modeled in several ways, but preliminary tests have shown that a simple model based on an Arrhenius-type formulation provides macroscopic quantities of interest that are consistent with experiments: the activation temperature

T_{vap} triggers the moisture evaporation, the pre-exponential factor k_{vap} represents the evaporation characteristic time-scale with respect to the vegetation temperature T_v. This issue could be revisited in future work.

▷ The **pyrolysis mass rate** $\dot{m}_{v,pyr}$ [kg/m^3/s] is represented as an endothermic reaction (similarly to the drying process) and is expressed as follows:

$$\dot{m}_{v,pyr} = (\beta_v \, \rho_p \, Y_{v,dry}) \; k_{pyr} \; \exp\left[\frac{-E_{pyr}}{R_g \, T_v}\right], \tag{3.8}$$

with k_{pyr} [s^{-1}] the pre-exponential factor and $T_{pyr} = E_{pyr}/R_g$ [K] the corresponding activation temperature.

▷ The **char mass loss rate** $\dot{m}_{v,char}$ [kg/m^3/s] is determined assuming a complete char oxidation reaction in a single step:

$$\dot{m}_{v,char} = \left(\Sigma_v \, \beta_v \, (1-\beta_v) \, \rho_g \, \frac{Y^{\infty}_{v,O_2}}{\chi_{O_2}}\right) \; k_{char} \; \exp\left[\frac{-E_{char}}{R_g \, T_v}\right], \tag{3.9}$$

with ρ_g [kg/m^3] the density of the vegetation gas phase, Y^{∞}_{v,O_2} [$-$] the oxygen mass fraction in contact with remaining biomass (assumed constant and equal to 0.233), χ_{O_2} [$-$] the stoichiometric coefficient related to oxygen in the char oxidation reaction, and k_{char} [m/s] the pre-exponential factor. The corresponding activation temperature is noted $T_{char} = E_{char}/R_g$ [K]. The density of the gas phase ρ_g is derived from the equation of state for ideal gas:

$$\rho_g = \frac{p_{\infty} \, W_g}{R_g \, T_g}, \tag{3.10}$$

with $p_{\infty} = 1$ bar the atmospheric pressure, $R_g = 8.314$ J/K/mol the ideal gas constant, W_g the mean mass weight of the gas mixture and T_g the temperature of the gas phase in the vegetation layer. Using the previously-mentioned thermal equilibrium assumption, T_g is identical to the solid-phase temperature T_v given by Eq. (3.1). However, this simple global model for char oxidation could be revisited in future work following Boonmee (2004, 2005).

The total mass loss rate in vegetation noted $\dot{m}_v$ [kg/m^3/s] is directly linked to the time-evolution of the bulk mass density of the vegetation defined as $\rho_v = \beta_v \, \rho_p$. While ρ_p represents an intrinsic property of the solid particles, ρ_v represents the effective amount of biomass fuel available per unit volume in the vegetation bed. It is indeed defined as the ratio of the mass of solid particles to the total volume (i.e., the volume of the solid particles in addition to the gas volume in-between these particles). Summing Eqs. (3.2) to (3.4) leads to the following equation describing the temporal evolution of the mass density of the porous vegetation:

$$\frac{\mathrm{d}\rho_v}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \; (\beta_v \, \rho_p) = -\dot{m}_v = -\dot{m}_{v,vap} - (1-\chi_{char}) \; \dot{m}_{v,pyr} - \dot{m}_{v,char}. \tag{3.11}$$

The term $\dot{m}_v$ is an important diagnostic variable in Pyrowo since it directly provides the pyrolysis gas mass flux from the vegetation to the gas phase (assuming

instantaneous release to the gaseous phase). PYROWO is further validated against thermogravimetry experimental data for the variable $\dot{m}_v$.

↪ Heat exchanges in vegetation

The total energy balance $\dot{q}_v$ [W/m^3] exchanged by convection, radiation and mass transfer between the solid and gas phases of vegetation is given by the following equation:

$$\dot{q}_v = \dot{q}_{v,conv} + \dot{q}_{v,rad} + \dot{q}_{v,chem}, \tag{3.12}$$

with:

(i) $\dot{q}_{v,conv}$ [W/m^3] the convection (buoyant) heat exchange modeled as:

$$\dot{q}_{v,conv} = (\beta_v \, \Sigma_v) \; k_{conv} \, (T_\infty - T_v), \tag{3.13}$$

with T_∞ the ambient gas temperature and k_{conv} [W/K/m^2] the heat transfer coefficient (approximated using empirical correlations obtained for laminar or turbulent flows around assumed-circular solid particles, Incropera and DeWitt, 1996). Note that Σ_v is an important parameter in heat transfer: the higher Σ_v, the larger the available surface to absorb energy and release pyrolysis gases to the surrounding gas phase per unit time.

(ii) $\dot{q}_{v,rad}$ [W/m^3] the net radiation heat transfer to the vegetation due to the cumulative contribution of flame radiation $\dot{q}_{rad,fl}$ and radiation losses $\dot{q}_{rad,l}$ from the heated vegetation towards the surrounding environment such that:

$$\dot{q}_{v,rad} = \dot{q}_{rad,fl} - \underbrace{(\beta_v \, \Sigma_v) \, \sigma_{sb} \, T_v^4}_{\dot{q}_{rad,l}}, \tag{3.14}$$

with $\dot{q}_{rad,fl}$ the radiation source term from the flame that is received by the vegetation (calculated with PRISSMA, see Section 2.5 in Chapter 2, while estimated as a fraction of the total heat released by the flame in empirical ROS models) and $\sigma_{sb} = 5.67 \times 10^{-8}$ W/m^2/K^4 the Stefan-Boltzmann constant (the heated biomass fuel particles are treated as black-bodies). The term $(\beta_v \Sigma_v / 4)$ is assumed to represent the optical length-scale for the considered porous medium.

(iii) $\dot{q}_{v,chem}$ [W/m^3] the heat release due to the vegetation thermal degradation itself. While drying and pyrolysis processes are endothermic reactions, char oxidation is highly exothermic and provides an additional energy to the porous vegetation before flaming ignition. Thus,

$$\dot{q}_{v,chem} = -\dot{m}_{v,vap} \, \Delta h_{vap} - \dot{m}_{v,pyr} \, \Delta h_{pyr} + \dot{m}_{v,char} \, \Delta h_{char}. \tag{3.15}$$

with Δh_{vap} $[J/kg]$ the latent heat of evaporation, Δh_{pyr} $[J/kg]$ the heat of pyrolysis and Δh_{char} $[J/kg]$ the heat of char oxidation.

Typical values for the kinetic parameters underlying the thermal degradation of pine needle litters are given in Table 3.3.

Table 3.3: *Kinetic parameters for biomass fuel thermal degradation (Grishin, 1997; Morvan and Larini, 2001; Consalvi et al., 2011).*

Parameter	Value (literature)	Unit
T_{vap}	5500-6000	K
k_{vap}	6×10^5	$K^{1/2}/s$
Δh_{vap}	2.25×10^6	J/kg
T_{pyr}	7000-9500	K
k_{pyr}	3.64×10^4	1/s
Δh_{pyr}	4.18×10^5	J/kg
T_{char}	9000	K
k_{char}	4.3×10^2	m/s
Δh_{char}	3.2×10^7	J/kg
χ_{char}	0.30-0.40	-
χ_{O_2}	2.66	-

↪ Characterization of the biomass gas phase

Since an instantaneous release from the biomass solid phase to the biomass gas phase is assumed in PYROWO, the mass gain of the gas phase is represented by the term $(-\dot{m}_v)$, i.e., the vegetation mass loss is totally gained by the biomass gas phase. The pyrolysis gases that are injected in this gas phase are characterized by their temperature T_g taken as equal to T_v (due to the equilibrium assumption within the biomass fuel layer) and by their mass fractions $Y_{v,k}$ (specified based on experimental data, since PYROWO is not detailed enough to characterize the composition of released gas compounds, see Di Blasi, 2008).

↪ Numerical resolution

The ODE system made of Eqs. (3.1) to (3.6) is numerically resolved using the ODEPACK[5] Fortran77 library. This library implements a wide range of ODE solvers (e.g., explicit, implicit, with Krylov-based preconditioning, treatment of sparse matrix). Here the time-integration of PYROWO relies on the basic explicit solver of the library named LSODE.[6] This LSODE solver is suitable for non-stiff and stiff systems of the form $\mathrm{d}\mathbf{y}/\mathrm{d}t = f(\mathbf{y}, t)$.

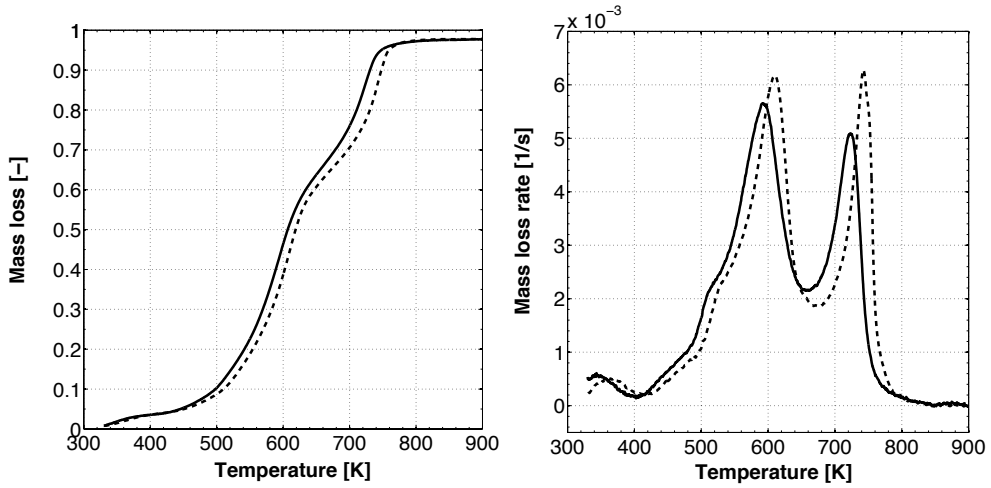
[5] http://people.sc.fsu.edu/~jburkardt/f77_src/odepack/odepack.html.
[6] Livermore Solver for ODE.

3.2.3 Calibration of the pyrolysis kinetic parameters

The production rate of pyrolysis gases is studied with respect to the vegetation intrinsic properties within the framework of the Pyrowo 0-D model. This means that a single vegetation cell (corresponding to a control volume) is considered, in which uniform properties of the porous vegetation are defined. The objective is to track the time-evolving behavior of the vegetation cell when subject to a constant radiation source term $\dot{q}_{rad,fl}$, which increases the temperature of the solid fuel particles T_v.

↪ Thermogravimetry experiments

Thermogravimetric analysis (TGA) monitors the weight m of a material sample subject to a constant thermal heating, with respect to the elapsed time or to the increasing material temperature in a controlled atmosphere (Tang, 1967; Shafizadeh and McGinnis, 1971; Broido, 1991; Bronli, 1996). The initial sample mass m_0 is typically on the order of a few milligrams. Thus, TGA is a fundamental tool to characterize thermal degradation mechanisms and reaction kinetics for a given type of vegetation in a laboratory-scale configuration. It also allows to determine the ash content in the sample as the remaining mass at the end of biomass thermal degradation. TGA data at different heating rates (from 1 to 30 K/min) are available in literature to calibrate Pyrowo model parameters and thereby, accurately reproduce the weight loss related to a type of biomass fuel. Figure 3.6(a) provides an example of weight loss curve with respect to the solid-phase temperature T_v for oven-dried *Pinus Pinaster* for two different heating rates, 10 K/min and 30 K/min.



(a) Mass loss $(1 - m/m_0)$. (b) Mass loss rate $^{\mathrm{d}}/_{\mathrm{d}t}(m/m_0)$.

Figure 3.6: *TGA measurements of Pinus Pinaster for different heating rates: the solid (dashed) line corresponds to 10 K/min (30 K/min). Credit: Leroy (2007).*

Analyzing the derivative of the TGA weight loss curve presented in Fig. 3.6(b) reveals three different peaks; each region can be described in terms of the dominant active reaction: the first peak between 300 and 400 K corresponds to vegetation moisture evaporation (the moisture content M_v is released from the vegetation solid phase), the second peak in the temperature region between 550 and 650 K is attributed to the degradation of the wood solid particles into flammable gas compounds and char (pyrolysis step), and the third peak between 650 and 800 K is due to char oxidation. The presence of the first peak below 400 K means that a low amount of moisture remains in oven-dried dead vegetation cells due to their equilibrium with the surrounding air. Figures 3.6(a)-(b) indicate a self-similar behavior of thermal degradation for the studied range of heating rates (10 to 30 K/min), implying that a unique calibration of the Pyrowo model parameters is sufficient to capture the three peaks and to retrieve a valid weight loss curve for this range of heating rates.

For higher heating rates than 30 K/min encountered in real-world fires (about 100 K/s), the three peaks of the solid-phase mass loss rate may significantly overlap and induce changes in the material response to thermal heating. However, the present work assumes a self-similar behavior of the biomass fuel for a wide range of heating rates since this is a common assumption in fire research. Besides, TGA measurements are limited by the assumption of thermally-thin materials (the temperature is assumed to be uniform within the solid particles). To evaluate the assumptions underlying TGA in the context of biomass burning, cone calorimeter measurements (using Fire Propagation Apparatus - FPA) could be useful to infer temperature gradients within fuel solid particles (Bartoli, 2011). This aspect is out of the scope of this work since TGA provides a global viewpoint of vegetation thermal degradation, which is sufficient for the preliminary developments of a multi-physics flame-scale CFD capability.

↪ Calibrated model behavior

The kinetic parameters in Pyrowo (i.e., pre-exponential factor, activation temperature and heat yield related to moisture evaporation, pyrolysis and char oxidation) are calibrated against TGA experimental data based on the Friedmann kinetic analysis method (Friedman, 1964; Trick et al., 1997). The proposed calibration procedure referred to as Calwo separates the three peaks shown in the derivative TGA curve and (separately) calibrates the Arrhenius-type kinetic parameters associated with each peak. The criterion for each peak calibration is based on the minimization of the distance between TGA mass loss measurements (symbols in Fig. 3.7(a)) and simulated mass loss (over the temperature range related to each reaction involved in biomass thermal degradation), see Appendix C.

The simulation of pine needles thermal degradation with Pyrowo corresponds to a case where the moisture content is taken initially as $Y^0_{v,wat} = 0.10$ and the wood dry content as $Y^0_{v,pyr} = 0.90$ (the superscript 0 referring to the initial state).

This case assumes that no char material is initially present in pine needles such that $Y^0_{v,char} = 0$. The mass sample is initially at ambient temperature ($T^0_v = 300$ K); pine needle properties are approximately known, i.e., $\rho^0_p = 500$ kg/m^3, $\beta^0_v = 0.05$, $c_{p,v} = 1,500$ J/K/kg and $\Sigma_v = 4,000$ m^{-1}. The convection heat transfer coefficient is specified as $k_{conv} = 10$ W/K/m^2. Over the duration of the TGA experiment, the mass sample is subject to a constant radiation source term, equivalent to a heating rate of 1.6 K/min, with $\dot{q}_{rad,fl} = 1.5 \times 10^6$ W/m^3 (Morvan and Dupuy, 2004). Note that nearly 60 s (physical time) are required to simulate the whole thermal degradation of the mass sample (from 0 to 100 % mass-loss, i.e., from drying to char oxidation) with PYROWO. The numerical integration is achieved for a time step Δt^{PY} that takes values on the order of $10^{-2}/10^{-3}$ s. Results of the calibration procedure are presented in Fig. 3.7(a); the associated calibrated parameters are indicated in Table 3.4. Their values can be compared to literature presented in Table 3.3; the calibration procedure mainly modifies the pre-exponential factors k_{vap}, k_{pyr} and k_{char} of the three reactions underlying biomass fuel thermal degradation. Note that these values are not due to physics-based arguments, they are derived from an optimization procedure, in which several sets of parameters can lead to the same mass loss curve.

Table 3.4: *TGA-based calibrated kinetic parameters with* PYROWO.

Parameter	Value (calibration)	Unit
T_{vap}	4800	K
k_{vap}	6×10^6	K$^{1/2}$/s
Δh_{vap}	2.25×10^6	J/kg
T_{pyr}	7100	K
k_{pyr}	3.64×10^5	1/s
Δh_{pyr}	4.18×10^5	J/kg
T_{char}	9200	K
k_{char}	1.0×10^5	m/s
Δh_{char}	1.2×10^4	J/kg
χ_{char}	0.48	-
χ_{O_2}	2.66	-

Results show that the observed behavior of the vegetation weight loss with respect to its increasing temperature is well captured by the PYROWO calibrated simulation. The three main reaction mechanisms (i.e., moisture evaporation, pyrolysis and char oxidation) are reproduced; they correspond to the different accelerations observed in the biomass mass loss curve in Fig. 3.7(b), which represents the total mass loss rate $\dot{m}_v$ in the vegetation. This vegetation behavior is consistent with the mass fractions presented in Fig. 3.7(c), the evaporation of the moisture content $Y_{v,wat}$ is associated with the first peak of the mass loss rate curve, while the consumption of

dry materials $Y_{v,dry}$ and the production of char materials $Y_{v,char}$ is closely related to the second peak associated with pyrolysis. This process leads to the decrease of the solid particle mass density ρ_p represented in Fig. 3.7(d). The subsequent char oxidation enhances the drop in the biomass fuel packing ratio β_v, also shown in Fig. 3.7(d). The proportion of the fuel bed occupied by the solid phase decreases (the gas volume increases). Thus, the remaining solid particles are exclusively made of char, $Y_{v,char} = 1$ above 550 K in Fig. 3.7(c).

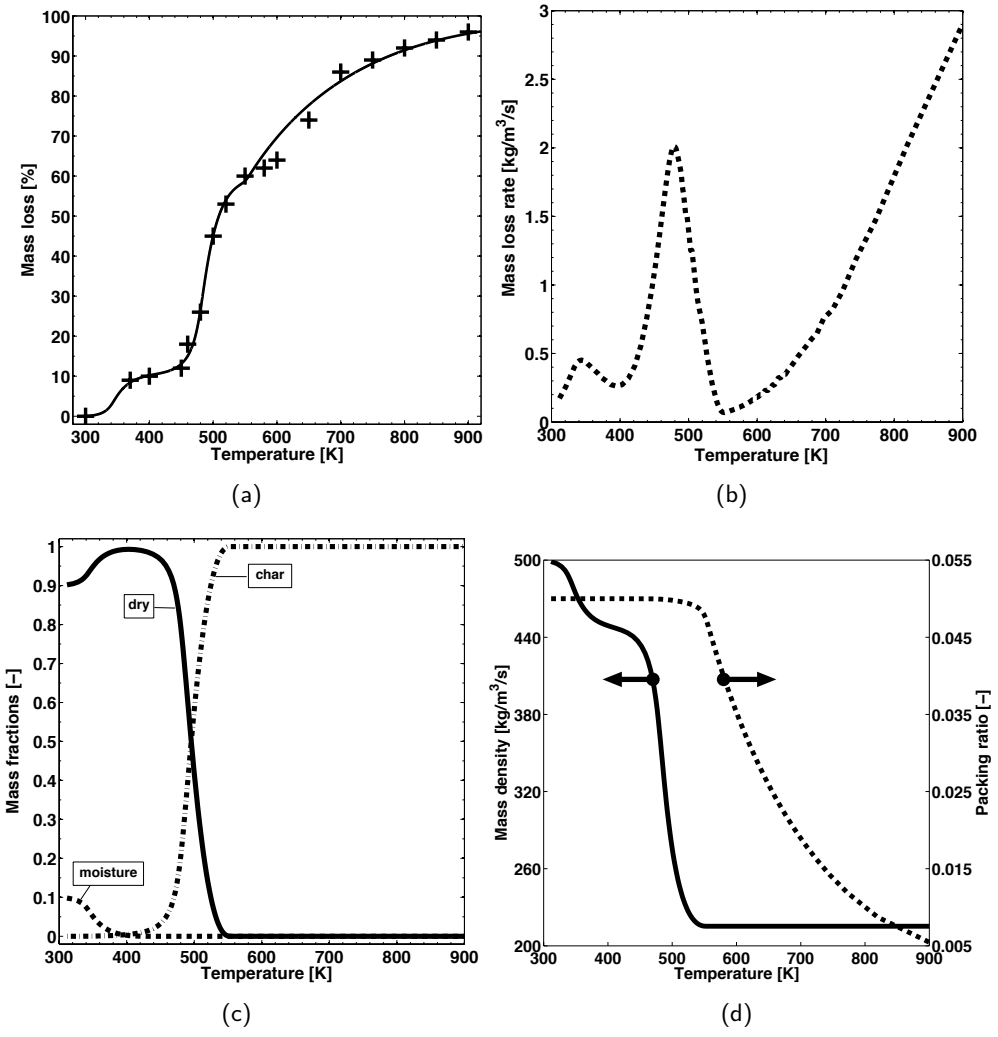


Figure 3.7: *Simulated behavior of pine needles with respect to the vegetation temperature T_v using* Pyrowo *with calibrated kinetic parameters. (a) Simulated mass loss $(1-m/m_0)$ in solid line; black crosses corresponds to TGA experimental measurements (INRA, private communication). (b) Simulated mass loss rate $\dot{m}_v$ (dashed line). (c) Mass fractions: moisture content $Y_{v,wat}$ (dashed line), dry material content $Y_{v,dry}$ (solid line), and char content $Y_{v,char}$ (dashed-dotted line). (d) Solid particle mass density ρ_p (solid line) and biomass fuel packing ratio β_v (dashed line).*

The Pyrowo-based simulated trend is consistent with TGA data and provides a good first-order approximation of the observed behavior. The char oxidation mechanism is the most difficult to calibrate: even though a wide range of values for the kinetic parameters related to char oxidation was tested, the slope of the mass loss between 600 and 700 K is not perfectly recovered. However, since this thesis focuses on the flaming zone, this approximation remains relevant for the preliminary developments of a multi-physics flame-scale CFD capability.

3.3 Analysis of the flame thermo-chemical properties

During the pyrolysis of wildland vegetation, large amounts of flammable compounds (referred to as *pyrolysis gases*) are released and convected towards the flame by buoyancy effects. Provided their temperature and the amount of air in contact with pyrolysis gases are large enough, flaming combustion can self-sustain and thereby, enhance fire propagation.

- ▷ Combustion between pyrolysis gases and oxidizer requires a chemical kinetic scheme adapted to compressible LES capability such as Avbp, in terms of computational cost and chemical stiffness. It is therefore important to validate the reduced kinetic scheme 2S-CH4-BFER used in this work, against reference detailed predictions for realistic biomass combustion conditions.
- ▷ Since the detailed flame structure is partially known for fires (diffusion or partially-premixed flame, see discussion in Section 2.1, Chapter 2), it is relevant to perform a parameter study to analyze the impact of pyrolysis gases on the flame thermo-chemical features (in terms of temperature and fuel/oxidizer equivalence ratio for instance) and thereby, gain insight into the actual combustion occurring in wildfires. While studies proposed by Grishin (1997) and Morvan and Dupuy (2004) are limited to the oxidation of CO, the combined oxidation of CH_4/CO is considered here; CH_4 and CO are indeed the two main flammable gases in pyrolysis gases (see Table 3.2). The objective is therefore to highlight the impact of chemical kinetics on the flame structure.

3.3.1 Characterization of the multi-species reacting gas mixture

↪ Definition of combustion fuel reactants

During vegetation thermal degradation, a multi-species gas mixture (the pyrolysis gases) is released when dry wood materials are consumed. In addition to CH_4 and CO, recent studies have shown the importance to include H_2O in pyrolysis gases (Tihay and Gillard, 2010). Using a perfectly-stirred reactor[7] analysis, Pérez-Ramirez et al. (2010) showed that oxidation reactions start at a lower temperature

[7] Ideal reactor in which fuel reactants and oxidizer are mixed at a certain temperature and in which the time-evolution of the system is studied for a given residence time (when this residence time tends to infinity, the chemical system tends towards its equilibrium state).

in presence of H_2O. Indeed, the presence of hydrogen-based compounds such as H_2 and H_2O enhances the formation of OH radicals that largely consume CO (Kee et al., 2003). Thus, H_2O plays an important role in the initiation of the combined oxidation of CH_4/CO. With regards to emissions, H_2O tends to enhance the formation of CO_2, C_2H_6, C_2H_4 and to decrease the concentration of CO, nitric oxide NO and hydrogen cyanide HCN, among others (Pérez-Ramirez et al., 2010).

In this work, pyrolysis gases are taken as a multi-species mixture made of CH_4, CO, CO_2 and H_2O. Based on experimental measurements (see Table 3.2), this mixture is described with the following mass fractions: $Y_{v,CH_4} = 0.10$, $Y_{v,CO} = 0.171$, $Y_{v,CO_2} = 0.64$ and $Y_{v,H_2O} = 0.089$. While transported through the vegetation layer towards the flame, they are mixed with atmospheric air ($Y_{air,O_2} = 0.233$, $Y_{air,N_2} = 0.767$) in a certain proportion as illustrated in Fig. 3.8. Thus, a 6-species mixture (CO, CO_2, CH_4, H_2O, O_2, N_2) is adopted to describe the gas phase solved by Avbp; the air proportion is defined by the mass dilution coefficient β_{air} such that the composition of the combustion fuel reactants satisfies:

$$\begin{cases} Y_{CH_4} = (1-\beta_{air})\, Y_{v,CH_4} \\ Y_{CO} = (1-\beta_{air})\, Y_{v,CO} \\ Y_{CO_2} = (1-\beta_{air})\, Y_{v,CO_2} \\ Y_{H_2O} = (1-\beta_{air})\, Y_{v,H_2O} \\ Y_{O_2} = \beta_{air}\, Y_{air,O_2} \\ Y_{N_2} = \beta_{air}\, Y_{air,N_2} \end{cases} \tag{3.16}$$

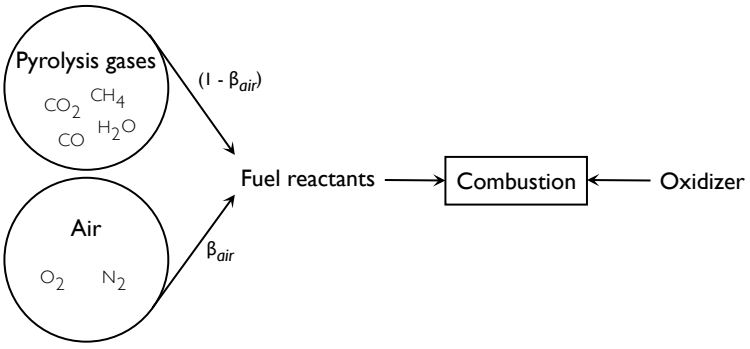


Figure 3.8: *Composition of combustion reactants. Pyrolysis gases are mixed with air in a certain mass proportion β_{air}.*

↪ Definition of equivalence ratio

An essential parameter in combustion applications is the fuel/air equivalence ratio Φ (see Chapter 2); Φ is calculated from the mass dilution coefficient β_{air}. The individual global oxidation reactions related to CH_4 and CO are defined as follows:

$$\begin{aligned} \mathrm{CH_4} + 2\,\mathrm{O_2} \quad &\longrightarrow \quad \mathrm{CO_2} + 2\,\mathrm{H_2O} \\ \mathrm{CO} + \frac{1}{2}\,\mathrm{O_2} \quad &\longleftrightarrow \quad \mathrm{CO_2} \end{aligned} \tag{3.17}$$

The associated mass stoichiometric coefficient ratios (noted s_{CH_4} and s_{CO}) are expressed with respect to the molar stoichiometric coefficient ratios (noted s_{mol,CH_4} and $s_{mol,CO}$, respectively):

$$s_{CH_4} = s_{mol,CH_4}\left(\frac{W_{O_2}}{W_{CH_4}}\right) = 4, \quad s_{CO} = s_{mol,CO}\left(\frac{W_{O_2}}{W_{CO}}\right) = \frac{4}{7}, \tag{3.18}$$

with $s_{mol,CH_4} = {\nu'_{O_2}}/{\nu'_{CH_4}} = 2$ and $s_{mol,CO} = {\nu'_{O_2}}/{\nu'_{CO}} = 0.5$ based on previously-mentioned global reactions. The equivalence ratio Φ for the combined oxidation of $\mathrm{CH_4}/\mathrm{CO}$ can be expressed as follows:

$$\begin{aligned} \Phi &= \frac{1}{Y_{O_2}}\left(s_{CH_4}\,Y_{CH_4} + s_{CO}\,Y_{CO}\right), \\ \Longrightarrow \Phi &= \left(\frac{1-\beta_{air}}{\beta_{air}}\right)\left(\frac{s_{CH_4}\,Y_{v,CH_4} + s_{CO}\,Y_{v,CO}}{Y_{air,O_2}}\right). \end{aligned} \tag{3.19}$$

For the pyrolysis-gas composition adopted in this work, Eq. (3.19) reduces to:

$$\Phi = 2.14\left(\frac{1-\beta_{air}}{\beta_{air}}\right), \tag{3.20}$$

so that a stoichiometric mixture ($\Phi = 1$) is obtained for $\beta_{air} = 0.68$. The equivalence ratio Φ and species mass fractions ($\mathrm{CH_4}$, CO and $\mathrm{O_2}$) are represented with respect to β_{air} in Fig. 3.9.

↪ Tools to study chemical kinetics of pyrolysis gases

Softwares such as Cantera[8] and Regath[9] may be used to simulate 0-D/1-D academic flame configurations in order to study detailed chemical kinetics, thermodynamics and complex transport processes. For instance, they can be used to study equilibrium properties (in terms of burnt gas temperature and gas mixture composition), to evaluate the impact of thermodynamics and transport properties on the flame speed/thickness of one-dimensional premixed flames, or to analyze the effect of the flow-induced strain rate on counter-flow diffusion flames. Both solvers include an adaptive mesh-refining algorithm to properly solve the reaction zone (in which the temperature gradients are typically steep) at a low computational cost. Both Cantera and Regath softwares are used in this work to study the thermo-chemical properties of pyrolysis gases.

[8] Object-oriented open-source code, https://code.google.com/p/cantera/.

[9] REal GAs THermodynamics, consisting in a Fortran90 library developed at EM2C laboratory (Candel et al., 2011; Caudal, 2013) and similar to the Chemkin-II formalism (Kee et al., 1993).

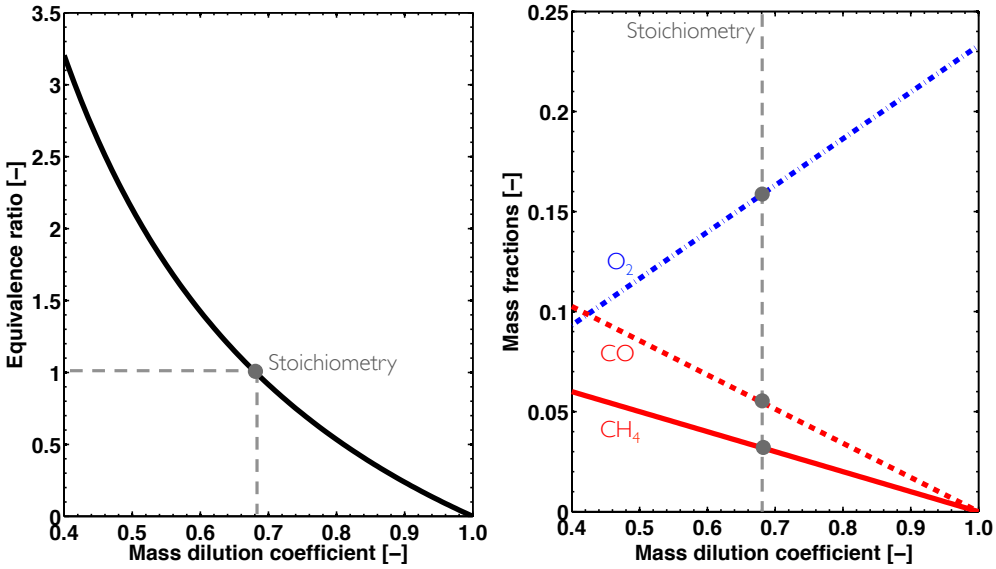


Figure 3.9: *Left: Equivalence ratio Φ with respect to the mass dilution coefficient β_{air}. Right: Mass fractions of CH_4 (red solid line), CO (red dashed line) and O_2 (blue-dashed-dotted line) with respect to the mass dilution coefficient β_{air}. Vertical gray-dashed lines indicate stoichiometry.*

- The detailed GRI-Mech3.0 scheme (see Section 2.4, Chapter 2) is adopted as reference to study the combined oxidation of CH_4/CO. Pérez-Ramirez et al. (2012) showed indeed that, while being optimized for the stand-alone oxidation of CH_4 (referred to as *pure-methane oxidation* in the following), GRI-Mech3.0 is able to retrieve a burnt gas composition that is consistent with experimental data for the oxidation of pyrolysis gases, for lean and rich mixtures (i.e., for Φ varying between 0.6 and 1.4) as well as for a wide range of initial gas temperatures (i.e., from 773 to 1273 K).
- Since the compressible LES solver AVBP cannot integrate detailed kinetic schemes due to computational cost issues, current strategies rely on reduced schemes such as 2S-CH4-BFER to model chemical source terms $\dot{\omega}_k$ (without considering filtering in a preliminary step) for each gas species k. These strategies simplify the description of combustion kinetics but also biomass fuel chemistry, and a large amount of chemical compounds released by the vegetation thermal degradation is not accounted for. In the following, reference results from GRI-Mech3.0 are compared to the PEA-based reduced 2-step scheme 2S-CH4-BFER retained in this work (see Section 2.4.3, Chapter 2, for the presentation of 2S-CH4-BFER) as well as to a global 5-step scheme specifically dedicated to pyrolysis gases of Mediterranean biomass fuels (Pérez-Ramirez et al., 2012), referred to as 5S-GLO-pyr (see Appendix D for the presentation and calibration of 5S-GLO-pyr).

In summary, the impact of pyrolysis gases on the flame thermo-chemical features is studied in the following with respect to the stand-alone oxidation of CH_4 (that is a reference in literature) using the detailed GRI-Mech3.0 scheme for different flame configurations: auto-ignition, one-dimensional premixed flames and (diffusion) counter-flow flames. Reduced kinetic schemes (2S-CH4-BFER and 5S-GLO-pyr) are also evaluated against GRI-Mech3.0 predictions.

3.3.2 Thermo-chemical flame characteristics

↪ Introduction to auto-ignition calculations

Auto-ignition calculations in adiabatic conditions consist in studying the time-evolution of a gas mixture based on 0-D unsteady simulations. The induction period (referred to as *auto-ignition delay time*) is followed by a rapid heat release rate, during which the gas temperature rises abruptly and intermediate species are produced and consumed rapidly. Combustion products are formed; the gas mixture can therefore reach a steady-state or equilibrium state. Thus, auto-ignition calculations provide insight into the composition of burnt gases and thereby, into the thermodynamical properties of the gas mixture given its initial temperature and composition if the residence time of the gas mixture is large enough. Figure 3.10 compares the temperature evolution of pure-methane and pyrolysis-gas stoichiometric mixtures, initially at temperature 1000 K, using the detailed kinetic scheme GRI-Mech3.0. The corresponding evolutions of the mass fraction of main gas species and radicals (e.g., H_2, OH, H) are presented in Fig. 3.11. Temperature profiles indicate that the heat released by combustion ignition occurs very rapidly (the temperature gradient following auto-ignition is very steep).The auto-ignition delay time is more than twice longer for pyrolysis gases than for pure methane: since flammable gas compounds are diluted in pyrolysis gases, they require more time to produce radical compounds in a sufficiently high concentration to activate ignition.

These results also indicate that the burnt gas temperature for a pyrolysis-gas mixture is lower than for a pure-methane mixture (2303 K and 2541 K, respectively). This temperature is directly determined by the initial composition of the gaseous mixture and depends on the species considered in the problem. Since CO exhibits a significantly lower heat of reaction than CH_4 (10 MJ/kg versus 50 MJ/kg), the heat of reaction of the mixed composition is also lower than for a pure-methane mixture and therefore, the burnt gas temperature is reduced. The variations of burnt gas temperature and auto-ignition delay time with respect to the initial gas temperature and composition as well as with respect to the equivalence ratio Φ are studied in the following.

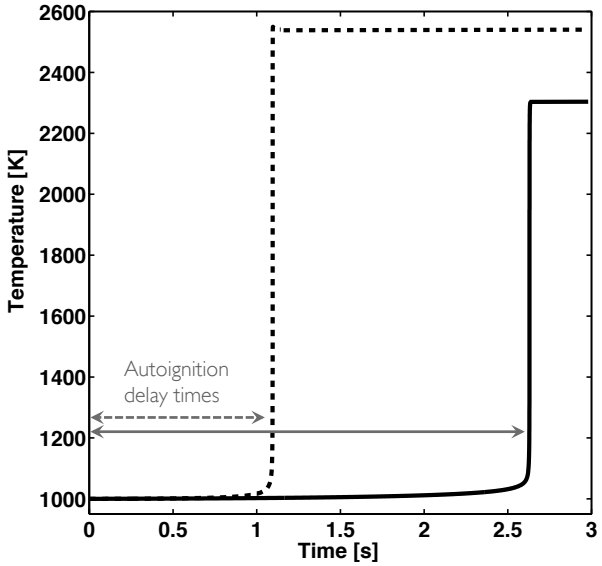


Figure 3.10: *Temporal evolution of the temperature corresponding to a gas mixture made of pure methane (dashed line) or pyrolysis gases (solid line) based on GRI-Mech3.0 predictions; both mixtures are stoichiometric and initially at 1000 K.*

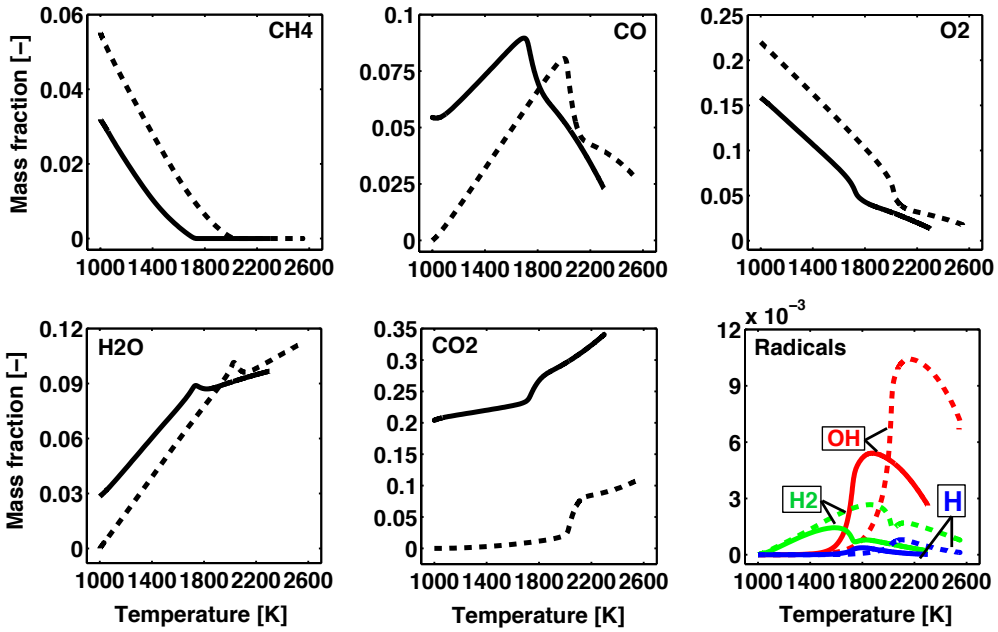
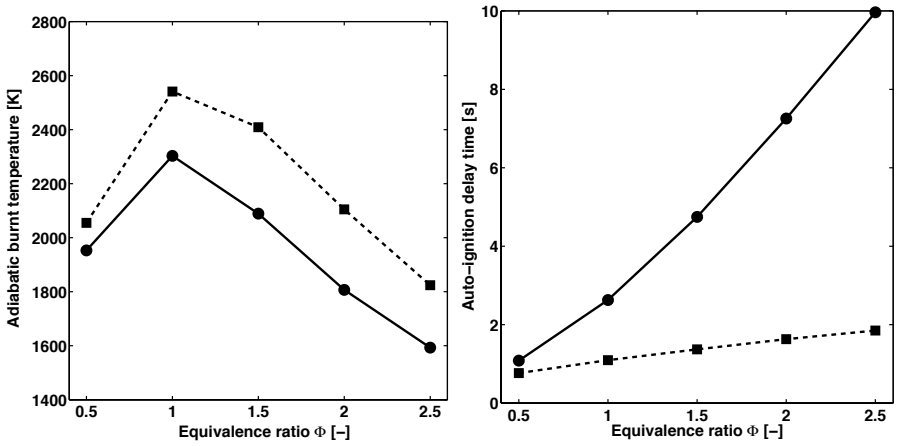


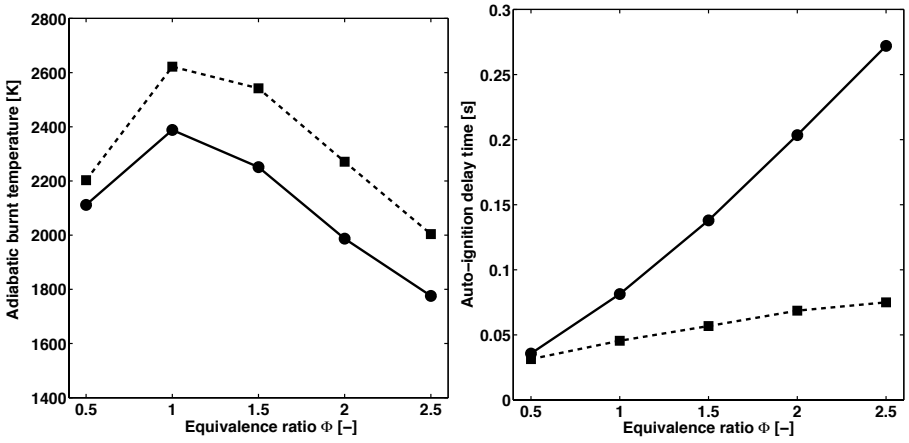
Figure 3.11: *Mass fractions Y_k with respect to the temperature of a gas mixture made of pure methane (dashed lines) or pyrolysis gases (solid lines) based on GRI-Mech3.0 predictions; both mixtures are stoichiometric and initially at 1000 K.*

↪ Insight from detailed chemical kinetics

In a first step, the adiabatic burnt gas temperature and the auto-ignition delay are studied based on GRI-Mech3.0 predictions for a wide range of equivalence ratios (i.e., from 0.5 to 2.5). Figure 3.12 illustrates the results obtained for an initial gas temperature of 1000 and 1200 K.



(a) Initial gas temperature: $1000\ K$. (b) Initial gas temperature: $1000\ K$.



(c) Initial gas temperature: $1200\ K$. (d) Initial gas temperature: $1200\ K$.

Figure 3.12: *Auto-ignition simulations using GRI-Mech3.0 for 1000 K (top) and 1200 K (bottom) initial gas temperatures. The squared-dashed line correspond to predictions for pure methane and the circled-solid line corresponds to predictions for pyrolysis gases. (a)-(c) Adiabatic burnt gas temperature with respect to Φ. (b)-(d) Auto-ignition delay time with respect to Φ.*

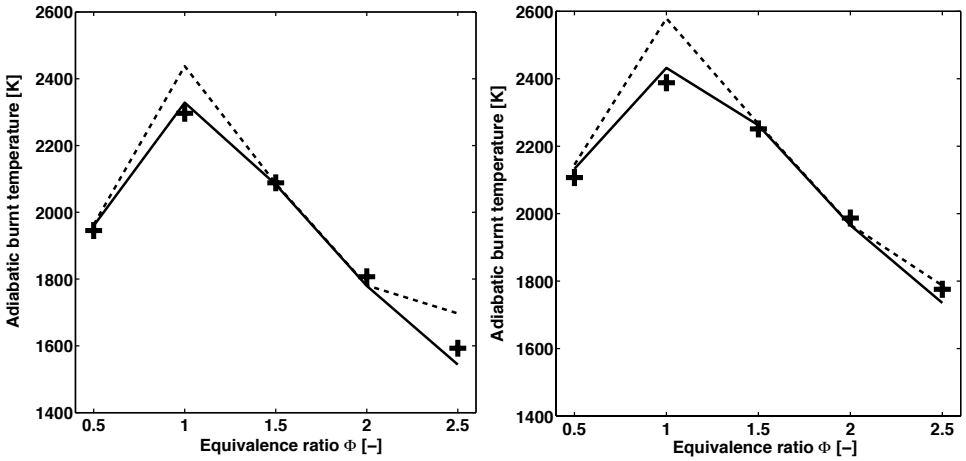
Results show that the pure-methane mixture exhibits the highest burnt gas temperature for all equivalence ratios and initial gas temperatures, partly due to the low heat of combustion associated with CO. For a given initial gas composition, the highest burnt gas temperature occurs near stoichiometry. Note that the decrease

in the burnt gas temperature on both sides of stoichiometry occurs at a similar rate for all initial gas compositions. Auto-ignition occurs much faster (lower ignition delays) for pure methane than for a pyrolysis-gas mixture because of dilution. The differences between these two types of gas mixture composition drastically increase, either when moving towards richer mixtures or when decreasing the initial gas temperature. These results highlight that the combustion of biomass pyrolysis gases is a slow burning process (especially at high fuel/oxidizer equivalence ratios Φ) with burnt gas temperatures much lower than in industrial combustors, typically.

↪ Predictions from global schemes

In a second step, GRI-Mech3.0 predictions in terms of auto-ignition are used as reference to analyze the results provided by 2S-CH4-BFER and 5S-GLO-pyr reduced kinetic schemes. Figure 3.13 displays auto-ignition simulation results with GRI-Mech3.0 and 2S-CH4-BFER for a wide range of equivalence ratios (i.e., Φ varying between 0 and 2.5) for the pure-methane and pyrolysis-gas mixtures as well as for 1000 K and 1200 K initial gas temperatures.

Results show that the global scheme 2S-CH4-BFER provides a good approximation of GRI-Mech3.0 predictions over the considered range of equivalence ratios Φ and initial gas temperatures. The global scheme 5S-GLO-pyr gives larger errors than 2S-CH4-BFER, especially near stoichiometry and for rich mixtures above $\Phi = 2.0$ with high initial gas temperatures. Thus, Appendix D proposes a calibration of 5S-GLO-pyr to retrieve more physically-consistent burnt gas temperatures and auto-ignition time delays. The resulting calibrated scheme is named 5S-GLO-pyr*.



(a) Initial gas temperature: $1000\ K$. (b) Initial gas temperature: $1200\ K$.

Figure 3.13: *Comparison of auto-ignition simulations based on GRI-Mech3.0 (crosses), 2S-CH4-BFER (solid line) and 5S-GLO-pyr (dashed line) for the pyrolysis-gas mixture; two initial gas temperatures are considered, (a) 1000 K, and (b) 1200 K.*

This parameter study related to auto-ignition shows that reduced kinetic schemes such as 2S-CH4-BFER and 5S-GLO-pyr* are able to provide accurate macroscopic features of the combustion of pyrolysis gases, typical of pine needle pyrolysis. Since they involve a limited number of gas compounds and thereby features a low computational cost, they are possible candidates to model combustion chemistry in AVBP.

It is worth mentioning that 5S-GLO-pyr* induces numerical difficulties due to its stiffness and to its formulation with a negative exponent factor in the first methane dissociation reaction (see Table D.1 in Appendix D). Thus, a modification of the global scheme formulation (out of the scope of this work) is necessary to preserve the equilibrium state and to avoid numerical instabilities when calculating 1-D laminar premixed flames (Jones and Lindstedt, 1988; Franzelli, 2011).

3.3.3 Analysis of laminar flame structure and strain effects

One-dimensional (premixed and diffusion) laminar flames are a canonical configuration in combustion, especially to study the effect of multi-species transport properties neglected in 0-D auto-ignition calculations. This configuration allows extensive comparison between experiments, theory and simulations; it is also regarded as the elementary component of turbulent combustion modeling through the flamelet assumption, in which the turbulent flame front is represented as a collection of laminar flame elements (Poinsot and Veynante, 2005). From this viewpoint, the study of 1-D laminar flames constitutes a preliminary and unavoidable step towards more complex flame configurations and allows to investigate the flame response to a wide range of conditions, in terms of initial gas temperature/composition and flow strain rate.

↪ Flame speed and thickness of one-dimensional premixed flames

In the context of pine needle litter fires, 1-D laminar premixed flames are defined as planar flames propagating into a premixed mixture made of pyrolysis gases and air (referred to as *fresh gases*), which is characterized by the equivalence ratio Φ given in Eq. (3.19) and an initial gas temperature T_f (see Fig. 2.1, Chapter 2).

▷ **Reference solutions.** The computation of 1-D laminar premixed flames in CANTERA or REGATH relies on a Newton-based method that requires an initial guess of the flame solution and an iterative algorithm with some convergence criterion to ensure the accurate determination of the solution. This initial guess is obtained in reference conditions, at ambient temperature, atmospheric pressure and stoichiometry.

Figure 3.14 illustrates a typical solution for the pyrolysis-gas mixture and highlights the differences with a pure-methane mixture for these reference conditions using the detailed kinetic scheme GRI-Mech3.0 (i.e., with full chemistry and complex transport, implying non-unity Lewis numbers for the considered gas species).

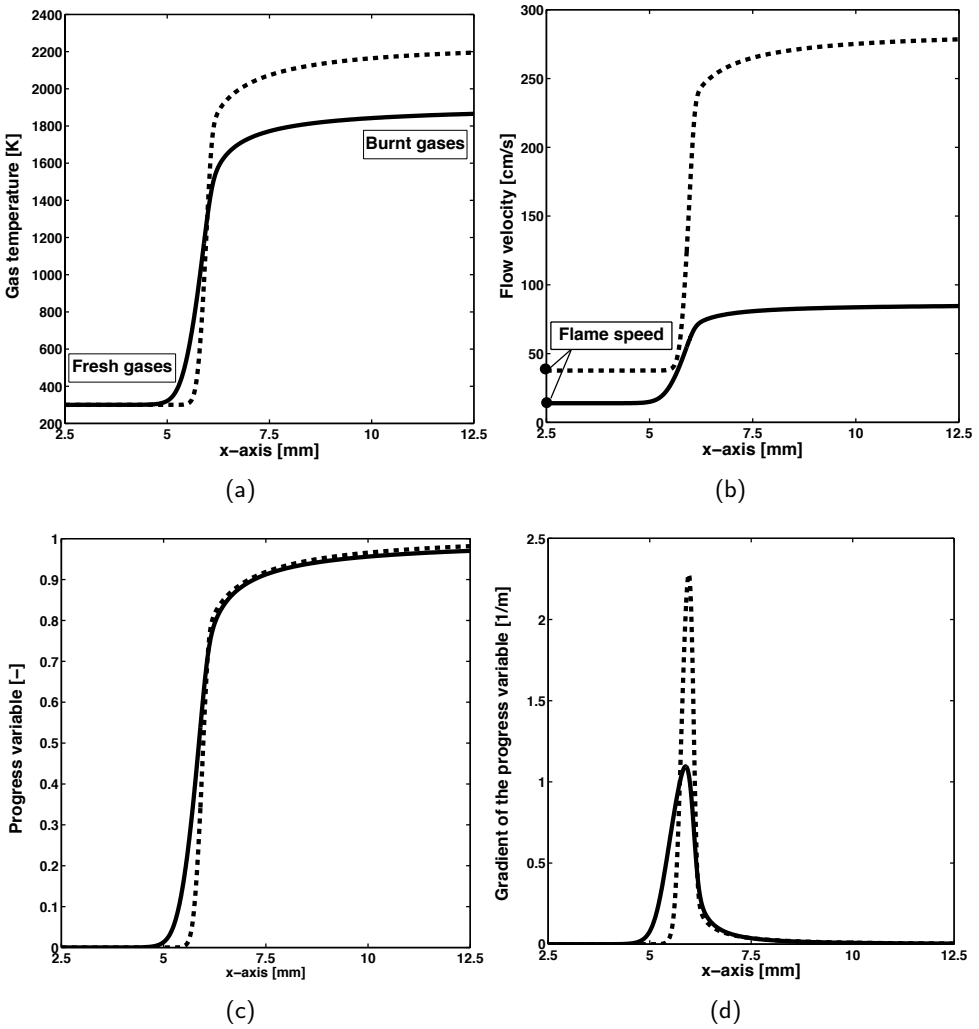


Figure 3.14: *Comparison of 1-D premixed flame calculations using GRI-Mech3.0 between a pure-methane mixture (dashed line) and a pyrolysis-gas mixture (solid line) for reference conditions ($T_f = 300\ K$, $\Phi = 1.0$). (a) Gas temperature [K]. (b) Flow velocity [cm/s]. (c) Progress variable c [-] (see Eq. 2.5). (d) Gradient of the progress variable ∇c [1/m].*

Thermochemistry controls the maximum value of the temperature reached in the burnt gases. Accordingly, the gap in the burnt gas temperature (2231 K for the pure-methane mixture versus 1914 K for the pyrolysis-gas mixture) in Fig. 3.14(a) is due to change in gas composition and thereby, to the lower heat of combustion of CO compared to $\mathrm{CH_4}$. This change in burnt gas temperature has a direct impact on the flow velocity profile through thermal expansion, explaining the discrepancies in the flow velocity shown in Fig. 3.14(b) on the burnt gas side. In these calculations the flame is stationary (the flame is displayed in its frame of reference). The flow velocity on the fresh gas side then corresponds to the laminar flame speed s_L (13.9 cm/s for the pyrolysis-gas mixture versus 37.7 cm/s for the pure-methane

mixture). 1-D premixed flame calculations provide additional information on the flame structure. In particular, the temperature gradient is steeper for the pure-methane case. Thus, the thickness of the reaction zone that is controlled by the inverse of the maximum temperature gradient is wider for the pyrolysis-gas mixture as shown in Figs. 3.14(c)-(d). The flame thickness is indeed multiplied by a factor 2 for the pyrolysis-gas mixture (0.91 mm compared to 0.44 mm for the pure-methane case).

▷ **Sensitivity analysis.** The properties of 1-D laminar premixed flames in terms of flame speed s_L and thickness δ_L obtained with GRI-Mech3.0 are presented in Fig. 3.15 for a wide range of pyrolysis conditions: the equivalence ratio Φ varies between 0.8 and 1.6, and the initial gas temperature varies between 300 and 800 K. The lower temperature boundary corresponds to reference cases in literature, while the temperature range 500-800 K is representative of the temperature associated to pyrolysis during vegetation thermal degradation. From a numerical viewpoint, the corresponding 1-D premixed flame solutions are obtained by specifying the initial guess as a perturbation of the reference solution, this perturbation being imposed as a variation of the physical parameters (e.g., equivalence ratio, fresh gas temperature). This incremental technique ensures the convergence of the solution when increasing the equivalence ratio Φ towards richer mixtures and/or the fresh gas temperature towards the temperature range that is typical of pyrolysis.

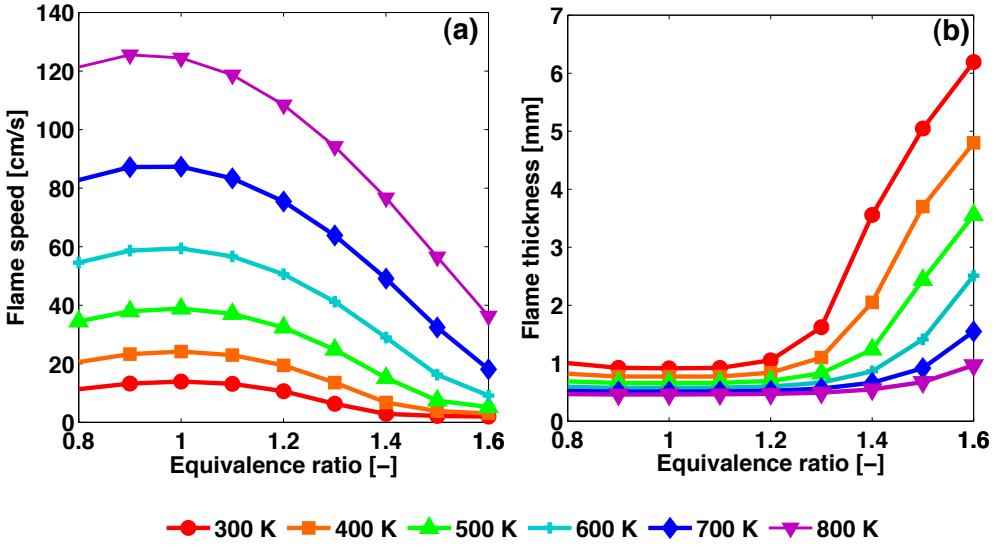


Figure 3.15: *Sensitivity analysis of 1-D premixed flames using GRI-Mech3.0 with respect to the equivalence ratio Φ, for a wide range of fresh gas temperatures T_f (from 300 to 800 K); study for the pyrolysis-gas mixture. (a) Flame speed s_L [cm/s]. (b) Flame thickness δ_L [mm].*

The laminar flame speed s_L increases with T_f for the considered range of equivalence ratios. In particular, for a given T_f, the laminar flame speed s_L reaches its

maximum value at stoichiometry (note that this maximum value is reduced by a factor 2 compared to a pure-methane mixture - not shown here) and significantly decreases for rich mixtures beyond $\Phi = 1.2$. The decreasing slope in rich mixtures increases with the fresh gas temperature such that at $\Phi = 1.6$ the flame speed is below 0.40 m/s for any considered condition. While the flame thickness δ_L remains relatively constant near 1 mm up to $\Phi = 1.2$, its behavior drastically changes above this threshold value since δ_L increases rapidly between $\Phi = 1.2$ and 1.6. For instance, between $\Phi = 1.2$ and 1.4 for $T_f = 300$ K, δ_L is multiplied by a factor 4. Thus, $\Phi = 1.2$ corresponds to a threshold in the behavior of the laminar flame structure. These results highlight the importance of studying properties of rich mixtures to improve the knowledge on the chemical kinetics underlying wildfire spread.

▷ **Behavior of the 2S-CH4-BFER global scheme.** Predictions of the pyrolysis-gas mixture behavior provided by the global scheme 2S-CH4-BFER are compared to GRI-Mech3.0 predictions for the fresh gas temperature $T_f = 300$ K in Fig. 3.16. While 2S-CH4-BFER provides realistic values of the flame speed s_L for a pure-methane mixture (i.e., the conditions for which the global scheme was calibrated, see Chapter 2), Fig. 3.16 indicates a level of accuracy depending on the equivalence ratio Φ for a pyrolysis-gas mixture. Near stoichiometry and in lean mixtures, the laminar flame speed is overestimated in comparison to GRI-Mech3.0 predictions. Still, 2S-CH4-BFER provides accurate flame speeds for rich mixtures above $\Phi = 1.2$. Since pyrolysis gases are expected to burn in rich-mixture conditions, 2S-CH4-BFER appears as a promising reduced kinetic scheme to model reaction rates in the multi-physics LES capability developed in this thesis to simulate laboratory-scale fire spread.

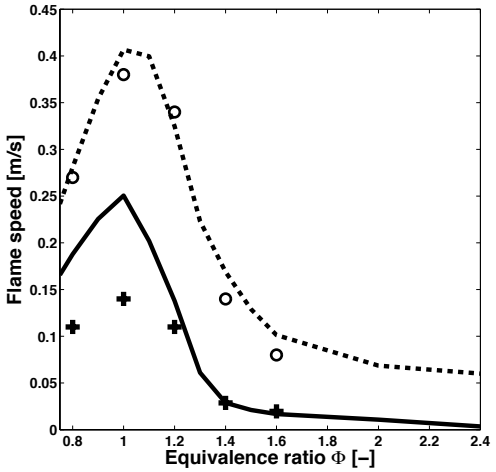


Figure 3.16: *Comparison of laminar flame speed predictions between the reduced kinetic scheme 2S-CH4-BFER (lines) and the detailed kinetic scheme GRI-Mech3.0 (symbols) for a pure-methane mixture (dashed line/circles) and a pyrolysis-gas mixture (solid line/crosses) with $T_f = 300\ K$.*

↪ Counter-flow diffusion flames and strain effects

While 1-D laminar premixed flames are a common configuration to evaluate the ability of a chemical kinetic scheme to reproduce the macroscopic features of the flame (since they are the most studied flame configuration from numerical and experimental viewpoints), the fine-scale features of wildfires are clearly of the diffusion flame type (see discussion in Section 2.1, Chapter 2). It is therefore of primary importance to study the effect of pyrolysis gases on the diffusion flame properties. Wildfires are buoyant-induced processes, meaning that density gradients in the flame ($\Delta\rho/\rho$) induce the rise of hot burnt gas pockets as well as air motion on both sides of the flame. As a consequence, the gas mixture is stretched with a dominant contribution of strain over curvature (Cuenot and Poinsot, 1995). The impact of the strain rate on laboratory-scale flame properties and in particular on the extinction limits may also be assessed with 1-D flame configurations (Lecoustre et al., 2011).

The common diffusion flame prototype used in this study is the counter-flow flame schematized in Fig. 3.17 and featuring the two following boundaries: air at atmospheric pressure on the left side (characterized by the temperature $T_{air,f}$), and pyrolysis gases on the right side (characterized by the temperature $T_{pyr,f}$). Note that the equivalence ratio changes from 0 (air boundary) to ∞ (fuel boundary) in this configuration. Pyrolysis gases and air diffuse towards the reaction zone, where they meet and burn, leading to a maximum temperature at stoichiometry. In contrast, far away from this reaction zone, the gas mixture is out of its flammability limits. At the interface between pyrolysis gases and air, the burning rate is controlled by the rate of diffusion and is therefore significantly affected by flow perturbations. In this configuration, the strain rate is imposed as a user-defined input. Figure 3.18 illustrates the corresponding counter-flow flame for a strain rate equal to 50 s^{-1}, air at $T_{air,f} = 298$ K and pyrolysis gases at $T_{pyr,f} = 600$ K, obtained with the detailed kinetic scheme GRI-Mech3.0 and simplified transport properties (i.e., unity Lewis numbers).

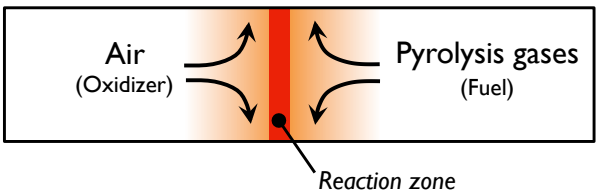


Figure 3.17: *Schematic of the counter-flow diffusion flame prototype used in the present study to evaluate the diffusion flame response to the strain rate.*

For a given counter-flow configuration, increasing the strain rate enhances heat losses and thereby, leads to lower burnt gas temperatures but faster combustion. For instance, multiplying the strain rate by a factor 4 in the configuration presented in Fig. 3.18 reduces the burnt gas temperature by 200 K.

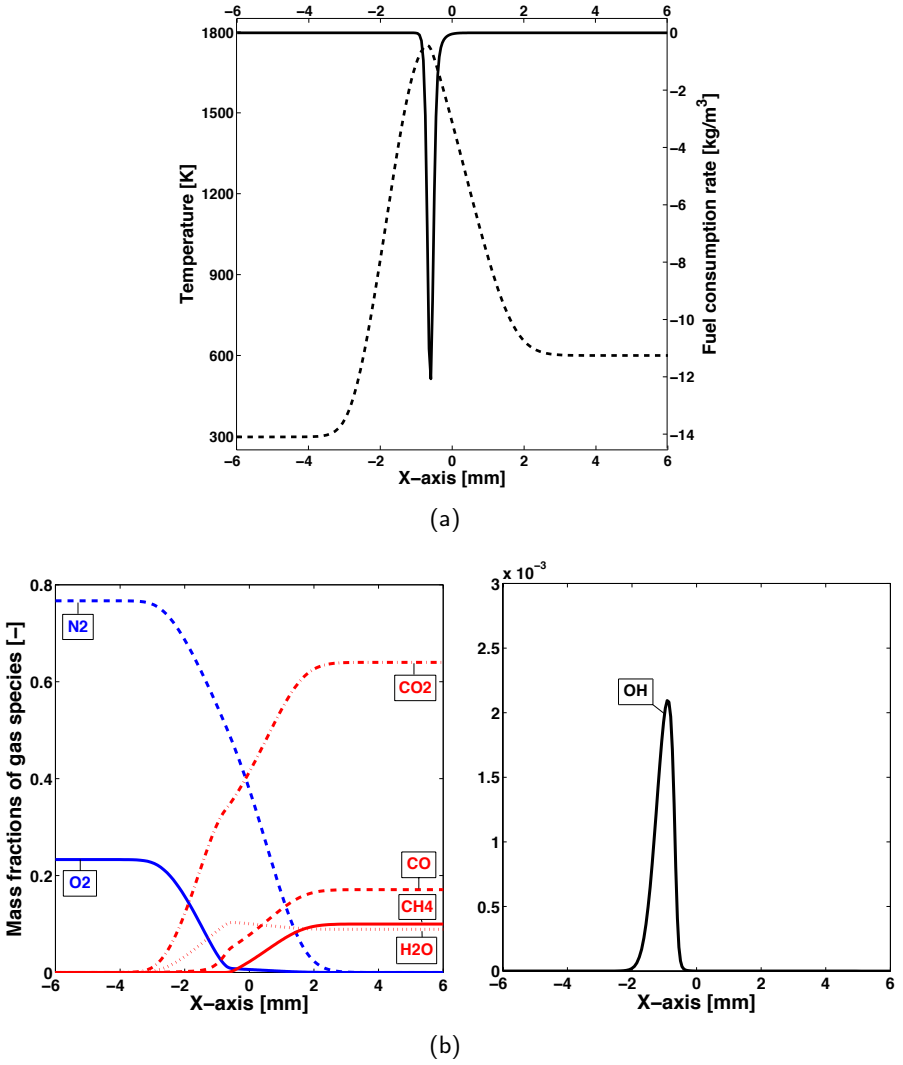
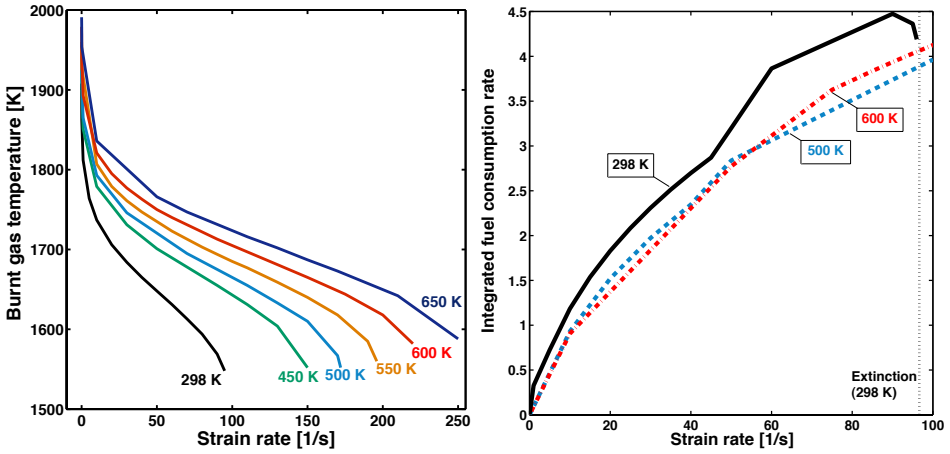


Figure 3.18: *Example of 1-D laminar counter-flow flame based on the detailed kinetic scheme GRI-Mech3.0 for $T_{air,f} = 298$ K, $T_{pyr,f} = 600$ K and a flow strain rate equal to 50 s^{-1}. (a) 1-D profile of gas temperature (dashed line) and fuel consumption term (solid line). (b) 1-D profile of mass fractions of gas species Y_k for combustion reactants (left) and with a zoom on OH radical indicating the flame location (right).*

To map all the possible states of the burnt gas temperature in a diffusion flame typical of the present laboratory-scale experiment, a sensitivity analysis is performed and shown in Fig. 3.19(a), for the initial temperature of the pyrolysis gases $T_{pyr,f}$ varying between 298 and 650 K and the strain rate increasing from 1 to 250 s^{-1}. For all considered fresh gas temperatures, the increasing strain rate promotes combustion (up to a certain threshold) since more fuel and oxidizer are transported towards the reaction zone. However, for large values of the strain rate (this threshold value depends on the fresh gas temperature T_f), chemistry becomes too slow to make

combustion possible and extinction occurs; there is indeed a competition between the supply in fuel reactants and the capacity for burning them, which induces heat losses (Cuenot and Poinsot, 1995). The maximum strain rate is 100 s^{-1} before flame extinction for pyrolysis gases at ambient temperature (i.e., $T_{pyr,f} = 298$ K) as shown in Fig. 3.19(b); the decrease of the integrated fuel consumption rate nearby 100 s^{-1} is a typical feature of extinction. In contrast, this decrease of the integrated fuel consumption rate is not observed before a strain rate equal to 250 s^{-1} for a temperature representative of the pyrolysis step during vegetation thermal degradation ($T_{pyr,f}$ between 500 and 600 K). The higher the fresh temperature of pyrolysis gases $T_{pyr,f}$, the larger strain rate the flame can undergo without extinction.



(a) (b)

Figure 3.19: *Counter-flow diffusion flame features with respect to the flow-induced strain rate with GRI-Mech3.0. (a) Burnt gas temperature for varying initial gas temperatures (between 298 and 650 K). (b) Integrated fuel consumption rate for varying initial gas temperatures (298 K in black solid line, 500 K in blue dashed line, 600 K in red dashed-dotted line). The extinction limit for 298 K is represented with the vertical line.*

In summary, macroscopic flame features are subject to significant variations when heat transfer and flow conditions are modified. In particular, they drastically vary with change in the temperature at which pyrolysis products (the fuel reactants of combustion) are released in the gas phase and change in their mixing with entrained air (through change in the fresh gas temperature and in the fuel/air equivalence ratio Φ, respectively). The resulting flame exhibits slightly slower burning velocity, lower burnt gas temperature, wider flame region and higher resistance to flow-induced strain than in industrial combustion. The next step consists in studying how these parameters are affected by the buoyancy-induced flow in CFD simulations. The PEA-based reduced kinetic scheme 2S-CH4-BFER is retained in this thesis, since it has demonstrated its ability to describe the flame thermo-chemical structure at high equivalence ratios Φ and for fuel reactants typical of pyrolysis gases (CO, $\mathrm{CH_4}$).

3.3.4 Strategies for turbulent combustion modeling

While previously-described parameter studies offer insight into the main flame features in simplified laminar flame configurations (corresponding to pyrolysis conditions simulated by PYROWO), this section is dedicated to the description of the turbulent combustion models that are effectively used in the LES solver AVBP to account for the interactions between turbulence, combustion and buoyancy in this work.

↪ Buoyant-driven diffusion flame

AVBP was originally developed for industrial engines associated with momentum-driven combustion. Gravitational effects (referred to as *buoyancy*) are negligible in these applications. However, it is not the case for wildfires, in which the characteristic speeds are much slower and in which the structure of the diffusion flame is affected by buoyancy effects as illustrated in Fig. 3.20.

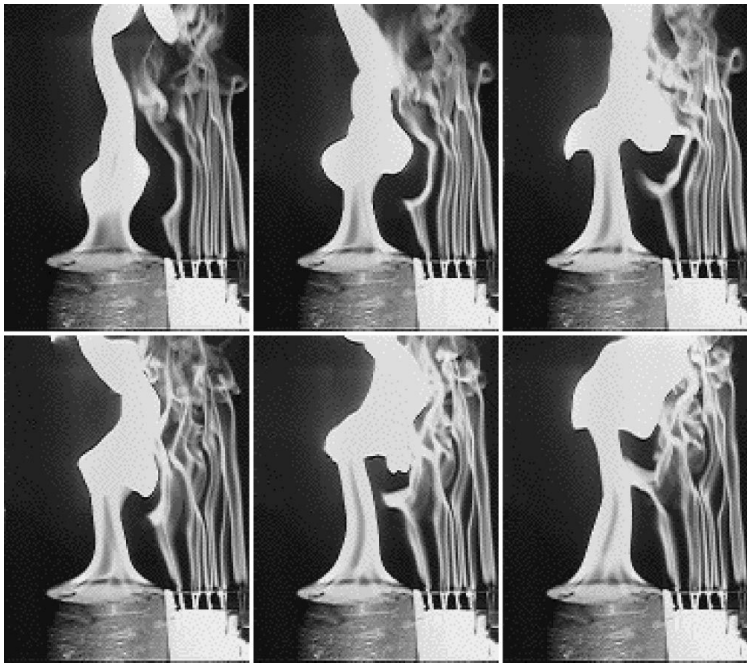


Figure 3.20: *Temporal series of experimental snapshots representing a propane/air buoyant diffusion flame from a 10 cm porous bed burner. Ambient air are visualized through smoke trails. Credit: Cetegen (1998).*

As described by Joulain (1996) and Tieszen (2001), density variations are responsible for the production of large-scale vortices in the flame region beyond a certain height, the bottom of the flame being commonly considered as a laminar flow. More precisely, these vortices are due to thermal instabilities typical of Rayleigh-Bénard instabilities, i.e., by an unstable equilibrium due to the presence of cold air above a

layer of hot burnt gas products. Figure 3.20 shows the deformations of the ambient air streamlines due to the formation of vortices in the fire plume. Large-scale eddies induce ambient air entrainment towards the flame region, thereby promoting combustion and enhancing thermal expansion through increased heat release. Cetegen (1998) proposed a macroscopic model to represent air entrainment in the near-field of turbulent diffusion flames (see Fig. 3.21). In this model, the eddy circulation acts as a periodic pumping process of ambient air and thereby, governs the rate of air entrainment that sustains oxidation reactions of pyrolysis gases. The frequency of these large-scale vortices rising by buoyancy and inducing time-fluctuations of the flame height is referred to as *puffing frequency* (Weckman and Sobiesiak, 1988; Hekestad, 1998; Mandin and Most, 2000).

From this description, buoyancy appears as a key unsteady mechanism in wildfire spread, which enhances turbulence, affects the flame structure, the heat release rate as well as the heat transfer towards the vegetation ahead of the flame front. In this work, the acceleration due to gravity was added in the AVBP momentum balance equation to account for buoyancy-driven flows and for induced dynamic thermal instabilities (the buoyancy term is included in the Navier-Stokes balance equations presented in Chapter 2).

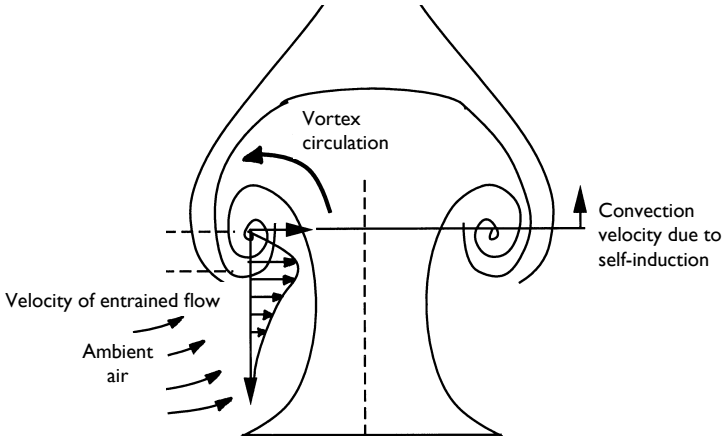


Figure 3.21: *Schematic of the air entrainment effect induced by buoyancy in fires. Credit: Cetegen (1998).*

↪ Subgrid-scale turbulent combustion modeling

Flame-scale LES capabilities are commonly based on a subgrid-scale turbulent combustion model to account for unresolved turbulent combustion features. The TFLES model introduced in Section 2.4.4, Chapter 2, has been designed for premixed flames to preserve laminar flame speed. A subgrid-scale turbulent combustion model is necessary in a LES framework since the thickness of these premixed flames is usually far below the computational grid size. In contrast, diffusion flames do not feature an intrinsic thickness (there is no velocity scale such as the burning velocity by which a characteristic length-scale such as the premixed flame thickness could be defined).

Their thickness is controlled by the convection/diffusion transport through the local flow stretch factor (including flow-induced strain and curvature effects), which leads to a scalar dissipation and thereby to the diffusive flame length-scale (also referred to as the *diffusive flame thickness*). Subsequently, diffusion flames typically have a thickness that is proportional to the resolved gradient of the flow. This behavior implies that the flame is numerically stable (i.e., the diffusive flame thickness is resolved on the LES computational grid).

Note that the only case in which the numerical stability is not secured for diffusion flames is when combustion occurs too rapidly (in this case, the flame is infinitely thin). In that case, the temporal resolution of the LES simulation might be insufficient. One efficient way to overcome this difficulty is to take advantage of the behavior of fast diffusion flames: if chemistry is fast enough, the flame burning is not controlled by chemistry but instead by transport. Subsequently, it is possible to decrease the chemical rates in the combustion kinetic scheme without changing the global flame burning.

In the present case, no turbulent combustion model is applied at the subgrid-scale since these effects are evaluated negligible at first approximation. Furthermore, since large-scale convection due to buoyancy is captured by the computational grid (on the order of 1 mm), no specific subgrid-scale buoyant combustion model is required. This aspect (Chomiak and Nisbet, 1985; Ince and Launder, 1989; Tieszen et al., 2004; Nicolette et al., 2005) needs to be confirmed in future developments towards a more physically-consistent LES solver for simulating laboratory-scale fires.

3.4 Flame/vegetation coupling interface

3.4.1 Elementary components of the coupling interface

While Avbp only solves equations for the gaseous phase, Pyrowo, developed in this thesis, describes the thermal degradation within the biomass fuel bed. Thus, the final step towards flame-scale LES of laboratory-scale fires consists in combining all the elementary components required for performing multi-physics CFD simulations:

- ▷ the combustion LES solver Avbp that describes the state of the gaseous phase and the flame behavior (see Section 2.4, Chapter 2);
- ▷ the DOM-based radiation solver Prissma (see Section 2.5, Chapter 2) that solves for the radiation source term $\dot{q}$ within the 3-D computational domain and for the radiation heat flux $\dot{q}_w$ at the upper layer of the pine needle fuel bed;
- ▷ the vegetation thermal degradation solver Pyrowo with calibrated model parameters (see Section 3.2) that solves for the mass loss rate $\dot{m}_v$ in the porous vegetation.

The coupling between these solvers is managed by the open-source dynamic coupler OPENPALM.[10] The development of OPENPALM has been committed to provide a generic environment that ensures the correct exchange and synchronization of the multiple software components on massively parallel computers. Such technique is very recent and requires the development of coupling methodologies based on HPC[11] (Duchaine et al., 2009; Amaya et al., 2010; Jauré et al., 2011; Maheu et al., 2012).

In the proposed coupling strategy, the transport of pyrolysis gases within the vegetation is not explicitly solved. Thus, an interface between the gaseous phase (i.e., AVBP combined to PRISSMA) and the porous vegetation (i.e., PYROWO) must be defined. This corresponds to a surface coupling, where data are exchanged at the boundaries of the flow/vegetation domain, indifferently referred to as wall, gas/vegetation interface, top of the vegetation layer or WALL-PYROWO in the following. This interface, schematized in Fig. 3.22(a) and corresponding to a new boundary condition in AVBP, is discretized using the AVBP computational grid in a finite number of grid nodes that define boundary cells along the 2-D interface as illustrated in Fig. 3.22(b). To each AVBP-related boundary cell corresponds the top of a control volume in PYROWO. In this control volume, the properties of pine needles are assumed uniform and the temporal evolution of the variables of the ODE system in PYROWO are solved for. This implies that Eqs. (3.1) to (3.6) are solved for each boundary cell of the AVBP computational grid; the resulting variables of PYROWO for the control volumes are then mapped at the flow/vegetation interface and interpolated at each AVBP grid node.

3.4.2 Description of the gas/vegetation interface

↪ Physical quantities of interest at the interface

The WALL-PYROWO interface is based on the boundary condition proposed by Cabrit (2009) and Cabrit and Nicoud (2010) to describe wall surface ablation in solid rocket motors, assuming there is no mechanical erosion and no geometry deformation. The development of this interface between AVBP and PYROWO consists in extending the existing isothermal condition to spatially-varying and temporally-varying profile along the gas/vegetation interface of the three following variables:

(i) the wall-normal Stephan injection velocity $v_{inj,w}$ [m/s], assumed orthogonal to the gas/vegetation interface (i.e., no tangential component) and provided by PYROWO (Dirichlet boundary condition);

(ii) the wall-normal species mass diffusive flux $\partial Y_{w,k}/\partial \mathbf{n}$ [1/m], where $\mathbf{n}$ indicates the normal direction to the gas/vegetation interface and where $Y_{w,k}$ [-] represents the mass fraction of gas species k at this interface (Neumann boundary condition);

[10] See Appendix B, www.cerfacs.fr/globc/PALM_WEB/.
[11] http://cerfacs.fr/coupling/.

(iii) the gas temperature at the wall T_w [K] (Dirichlet boundary condition).

While the gas temperature at the gas/vegetation interface is directly provided by PYROWO since a thermal equilibrium is assumed between the solid and gas phases of the vegetation (i.e., $T_w = T_v$), the calculation of the injection velocity $v_{inj,w}$ and of the wall-normal species diffusive flux $\partial Y_{w,k}/\partial \mathbf{n}$ is not straightforward and is explained in the following. By convention, all the variables at the gas/vegetation interface are projected on the wall normal $\mathbf{n}$ directed towards the gaseous phase (the index w refers to this interface, w standing for *wall*). Note also that the height at which the interface is located above the vegetation (i.e., the top of the vegetation layer) is assumed to remain constant over time. This assumption remains acceptable for simulating laboratory-scale fire of pine needles (of a few cm high).

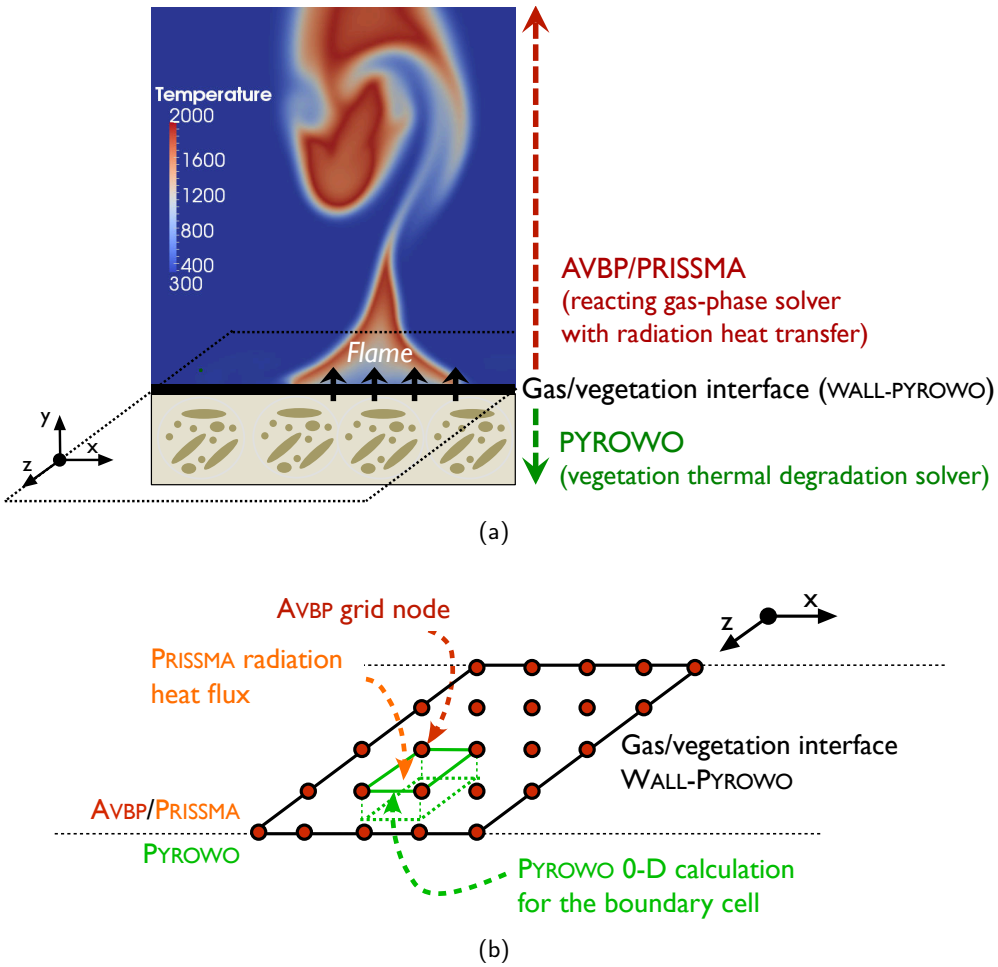


Figure 3.22: *Schematic of the* AVBP/PRISSMA/PYROWO *interface (*WALL-PYROWO*). (a) Two-dimensional (vertical) cross-section view of the simulation in the direction of flame propagation; and (b) Gas/vegetation (horizontal) interface representation.*

↪ Wall-normal injection velocity

The variable $\dot{m}_v$ represents the vegetation mass loss rate per unit volume provided by the 0-D PYROWO solver. The resulting mass flux at the gas/vegetation interface, noted $\dot{s}_w$ [kg/m^2/s], is expressed as follows:

$$\dot{s}_w = \frac{\dot{m}_v}{\delta_v}, \tag{3.21}$$

with δ_v [m] the vegetation layer depth. Thus, the surface mass production rate of each gas species k denoted by $\dot{s}_w^{(k)}$ $(k = 1, \cdots, N_g)$ satisfies:

$$\dot{s}_w = \sum_{k=1}^{N_g} \dot{s}_w^{(k)} = \sum_{k=1}^{N_g} \frac{\dot{m}_v^{(k)}}{\delta_v}, \tag{3.22}$$

with $\dot{m}_v^{(k)} = Y_{v,k}\,\dot{m}_v$ the mass rate per unit volume and $Y_{v,k}$ the mass fraction associated with each gas species k ($Y_{v,CH_4} = 0.10$, $Y_{v,CO} = 0.171$, $Y_{v,CO_2} = 0.64$ and $Y_{v,H_2O} = 0.089$, see Section 3.3.1). Then, by applying conservation of the mass flux $\dot{s}_w$ at the gas/vegetation interface, the following relation holds:

$$\dot{s}_w = \rho_w\, v_{inj,w} = \sum_{k=1}^{N_g} \dot{s}_w^{(k)}, \tag{3.23}$$

with ρ_w [kg/m^3] the gas density at the gas/vegetation interface. Thus, the wall-normal Stephan injection velocity $v_{inj,w}$ reads:

$$v_{inj,w} = \frac{1}{\rho_w} \left(\sum_{k=1}^{N_g} \dot{s}_w^{(k)} \right). \tag{3.24}$$

↪ Wall-normal species mass diffusion fluxes

Similarly to mass conservation, species conservation at the gas/vegetation interface leads to the following relation:

$$\rho_w\, V_{w,inj}\, Y_{w,k} + \rho_w\, V_{k,n,w}\, Y_{w,k} = \dot{s}_w^{(k)}, \tag{3.25}$$

where $V_{k,n,w}$ [m/s] corresponds to the wall-normal diffusion velocity of species k satisfying:

$$\sum_{k=1}^{N_g} Y_{w,k}\, V_{k,n,w} = 0$$

to ensure mass conservation. As for Navier-Stokes balance equations (presented in Chapter 2), the corrected Hirschfelder-Curtiss approximation is used to evaluate the wall-normal diffusion flux of each species k such that:

$$Y_{w,k}\, V_{k,n,w} = -D_k\, \frac{W_k}{W_w} \left(\frac{\partial X_{w,k}}{\partial \mathbf{n}} \right) + Y_{w,k}\, V_{c,n}, \tag{3.26}$$

with W_w the mean molecular weight of the gaseous mixture at the gas/vegetation interface and $V_{c,n}$ the wall-normal correction velocity satisfying:

$$V_{c,n} = -\sum_{k=1}^{N_g} D_k \frac{W_k}{W_w} \left(\frac{\partial X_{w,k}}{\partial \mathbf{n}} \right). \tag{3.27}$$

The molar fraction gradients $\partial X_{w,k}/\partial \mathbf{n}$ are related to their mass fraction counterparts in the wall-normal direction as follows:

$$\frac{\partial X_{w,k}}{\partial \mathbf{n}} = \frac{W_w}{W_k} \frac{\partial Y_{w,k}}{\partial \mathbf{n}} - \frac{W_w^2}{W_k} Y_{w,k} \sum_{q=1}^{N_g} \frac{1}{W_q} \frac{\partial Y_{w,q}}{\partial \mathbf{n}}. \tag{3.28}$$

By reformulating Eq. (3.25), the wall-normal mass species diffusion flux $\partial Y_{w,k}/\partial \mathbf{n}$ reads:

$$\frac{\partial Y_{w,k}}{\partial \mathbf{n}} = \frac{Y_{w,k}}{D_k} \left(\frac{\sum_{q=1}^{N_g} \dot{s}_w^{(q)}}{\rho_w} + V_{c,n} + W_w D_k \sum_{q=1}^{N_g} \frac{1}{W_q} \frac{\partial Y_{w,q}}{\partial \mathbf{n}} \right) - \frac{\dot{s}_w^{(k)}}{\rho_w D_k}. \tag{3.29}$$

Equation (3.29) is a system of equations of dimension $N_g \times N_g$ (N_g being the number of gas species). In practice, this equation is solved using an iterative algorithm providing at each time step the wall-normal mass fraction gradient $\partial Y_{w,k}/\partial \mathbf{n}$ for each species k of the gas mixture.

Based on the temperature, mass loss rate and composition of the gas compounds provided at the upper layer of the vegetation, the conservative variables of the AVBP solver (e.g., temperature, velocity field, mass density, composition) can be retrieved at the wall. Note that the wall temperature is corrected by accounting for wall thermal losses through the radiation solver PRISSMA.

3.4.3 Flowchart of multi-physics large-eddy simulations

In summary, the heat transfer from the flame to the pre-heated vegetation ahead of the flame zone is represented at the WALL-PYROWO interface using the radiation solver PRISSMA. PRISSMA requires the mass fraction of CO_2 and H_2O, the soot volume fraction Y_{soot} (if soot is considered), the pressure p as well as the temperature T of the gas phase from AVBP calculations. It also requires the gas temperature T_w at the wall from PYROWO. In return, it provides the radiation source term $\dot{q}$ [W/m^3] to AVBP and the radiation heat flux $\dot{q}_w$ [W/m^2] at the wall to PYROWO. In-turn, PYROWO provides the wall-normal injection velocity $v_{inj,w}$, the wall-normal species mass diffusive flux $\partial Y_{w,k}/\partial \mathbf{n}$ and the gas temperature T_w at the wall to AVBP. These variables act then as boundary condition for AVBP.

The complexity of radiation heat transfer and their non-linear interaction with reacting turbulent flows require a parallel coupling strategy between PRISSMA and AVBP, presented in Fig. 3.23. Since AVBP is a compressible solver, its time step is limited by the acoustic time step. There is no need to solve for radiation at this acoustic time step since the radiation source term $\dot{q}$ is only modified by the slower convection motions of fresh and burnt gas pockets. This is governed by a convection time step that is much larger than the acoustic time step in the present configuration. Thus, one iteration of the radiation solver PRISSMA is typically performed every 1,000 iterations of the combined AVBP/PYROWO solver (i.e., $N_{it} = 1,000$).

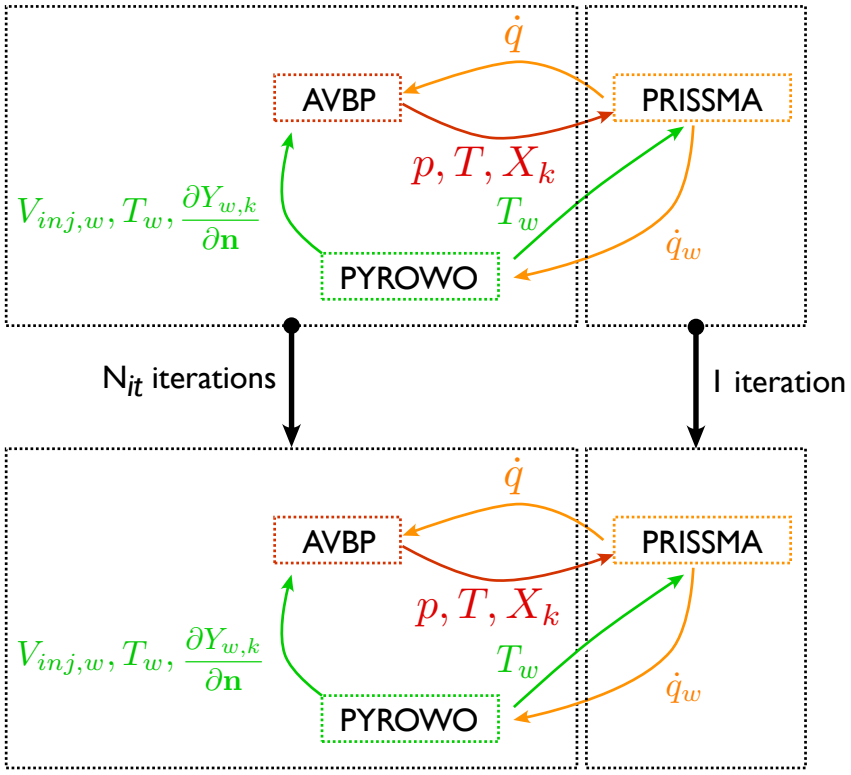


Figure 3.23: *Parallel-coupling strategy with a coupling frequency $1/N_{it}$ (N_{it} being the number of LES iterations between two radiation calculations).*

The flowchart of the coupling capability proposed in this thesis between AVBP, PRISSMA and PYROWO is provided in Fig. 3.24. Note that the coupling between AVBP and PRISSMA via OPENPALM is based on prior work due to Amaya et al. (2010) and Poitou et al. (2012).

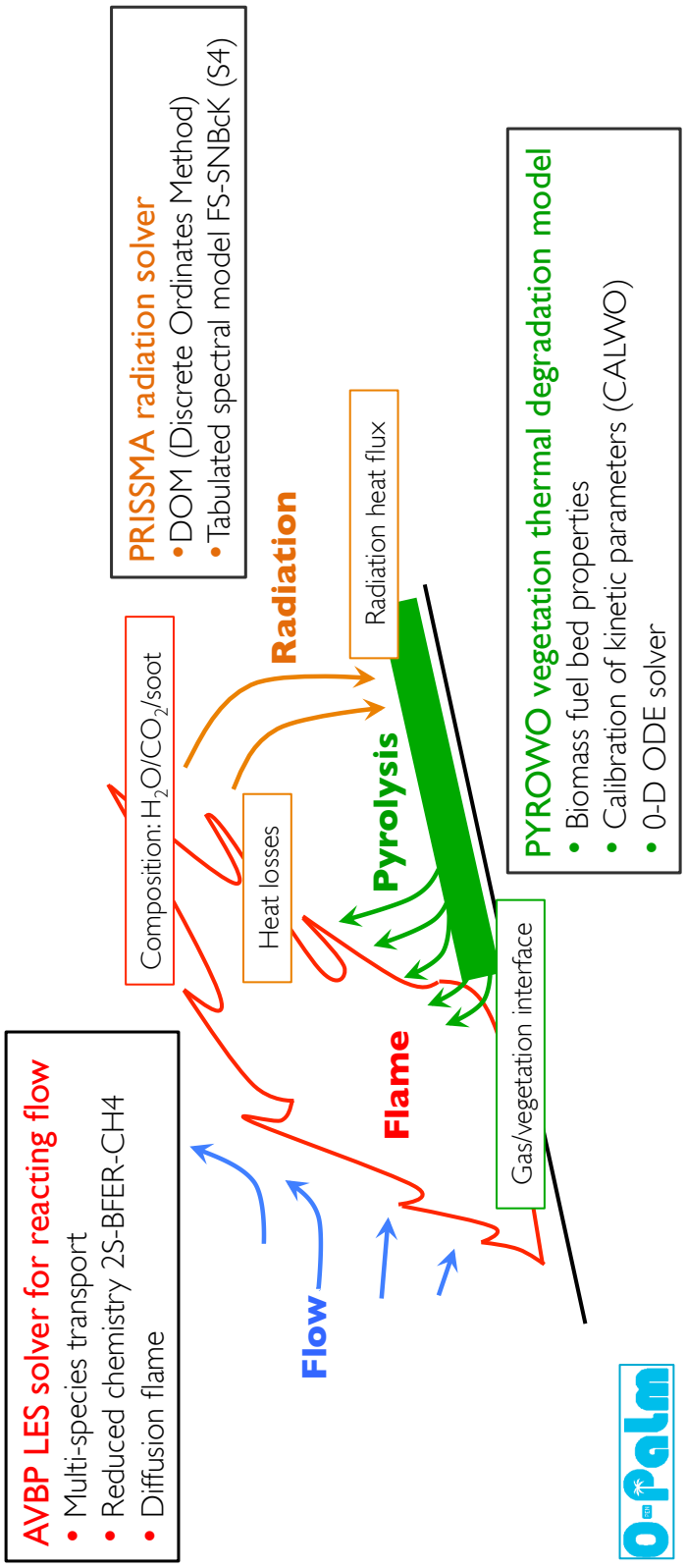


Figure 3.24: *Schematic of the* OpenPalm *coupling strategy for multi-physics flame-scale LES of laboratory-scale fire propagation. Data exchanges between the three different models occur simultaneously to avoid loss of information.*

Conclusion

This chapter presents the strategy developed in Part I of this thesis to perform multi-physics flame-scale LES that are representative of the interactions between turbulence, combustion, radiation and pyrolysis in a laboratory-scale fire propagation.

Two different chemical mechanisms underlying fire spread are represented through this coupling simulation capability: (1) the thermal degradation of the porous vegetation that produces pyrolysis gas reactants for flaming combustion, and (2) the oxidation of these gas reactants that sustains the flame and its propagation towards unburnt biomass fuel. They are both incomplete in reality and are therefore responsible for the large amount of emissions associated with wildfire. The proposed strategy includes a LES approach solving for the flame structure (AVBP), a radiation DOM approach that calculates the radiation heat transfer at the flame/vegetation interface (PRISSMA), and a thermal degradation model providing a macroscopic description of the heated porous vegetation (PYROWO). The following aspects are important to mention.

- The newly-developed biomass thermal degradation model PYROWO was calibrated against TGA experimental data in order to reproduce the observed mass loss of the vegetation and to represent the release of pyrolysis gases (mainly CH_4, CO, CO_2 and H_2O).
- The flame structure due to oxidation of pyrolysis gases was studied in laminar one-dimensional premixed and counter-flow flame configurations. In particular, the impact of the fuel/oxidizer equivalence ratio on the flame features (such as burnt gas temperature, flame speed and thickness) was highlighted. However, this equivalence ratio is subject to significant uncertainties since it depends on the flame-induced air entrainment through the porous vegetation, itself affected by the vegetation packing ratio and by time-dependent buoyancy effects.
- The most critical point underlying these multi-physics large-eddy simulations remains the gas/vegetation interface WALL-PYROWO between the gaseous phase and the porous vegetation. In practice, this boundary condition is described through the OPENPALM-based parallel coupling between PYROWO and AVBP/PRISSMA solvers.

As a preliminary step towards a research tool for exploring wildfire behavior, the objective is to demonstrate the potential of such modular capability to improve our knowledge on the physical processes governing wildfire and on their interactions. For this purpose, results are compared to laboratory-scale fire of pine needle litters (see Fig. 3.1).

Chapter 4

Analysis of laboratory-scale fire simulations

Multi-physics coupled simulations of fire propagation are performed at laboratory flame scale and are compared to measurements to provide a comprehensive understanding of the mechanisms underlying fire spread. In particular, the assumptions used to estimate the semi-empirical rate of spread (ROS) in regional-scale wildfire spread simulators are examined. To our knowledge, it is one of the first attempts, with studies due to Zhou et al. (2005, 2007), to apply a large-eddy simulation (LES) approach solving for the flame structure and including radiation heat transfer as well as biomass fuel pyrolysis to natural fire propagation. Insight into the flame-induced flow entrainment is also provided through the analysis of particle image velocimetry (PIV) measurements as highlighted in Fig. 4.1.

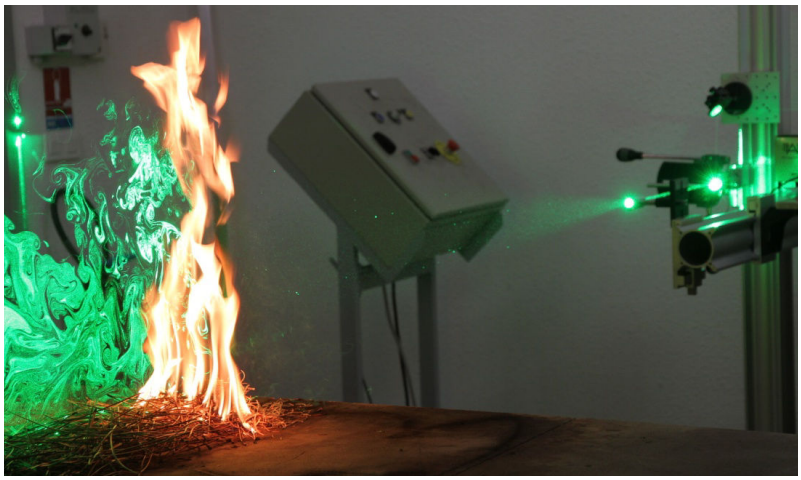


Figure 4.1: *PIV application to laboratory-scale fire spread. Credit: EM2C laboratory.*

Outline

4.1 Description of the laboratory-scale fire experiment

4.1.1 Overview of experimental measurements

The target configuration of LES computations is a laboratory-scale experiment that consists in a biomass fuel bed made of maritime pine needles (*Pinus Pinaster*) performed at the CNRS-SPE laboratory (France). As illustrated in Fig. 4.2, this fuel bed lies on a 2 m $\times$ 1 m surface, which can be tilted to study slope-aided fires and over which there is a 3 m $\times$ 3 m hood extractor. The fire is ignited as a 1-m line; the side of the fuel bed that is 2 m long corresponds to the direction of the fire propagation. The biomass fuel typically exhibits the properties listed below (Santoni et al., 2010).

- ▷ Pine needles are characterized by a surface-area-to-volume ratio Σ_v varying between 3,000 and 4,000 1/m and a mass density ρ_p varying between 500 and 520 kg/m^3.
- ▷ Their moisture content M_v typically varies between 4 and 5 $\%$ after a 24-hour drying in oven at $60°\mathrm{C}$ (due to equilibrium with ambient air).
- ▷ The layer of pine needles is 5-10 cm high, with a surface loading m_v'' varying between 0.6 and 1.2 kg/m^2 and a packing ratio β_v varying between 0.03 and 0.05.

Over the fire duration (that lasts between 300 and 400 s depending on the slope angle), the heat release rate was measured by a 1 MW large-scale calorimeter based on the oxygen consumption calorimetry principle, valid for both steady and unsteady fire propagation (Huggett, 1980; Santoni et al., 2010). This principle relies on a proportional relationship between the heat release and the oxygen consumed for

complete biomass combustion. Thus, the heat release rate was retrieved based on the measurement of O_2 flow in the exhaust duct (Santoni et al., 2010). The time-evolving mass loss of the biomass fuel was also measured through the weight balance (see Fig. 4.2).

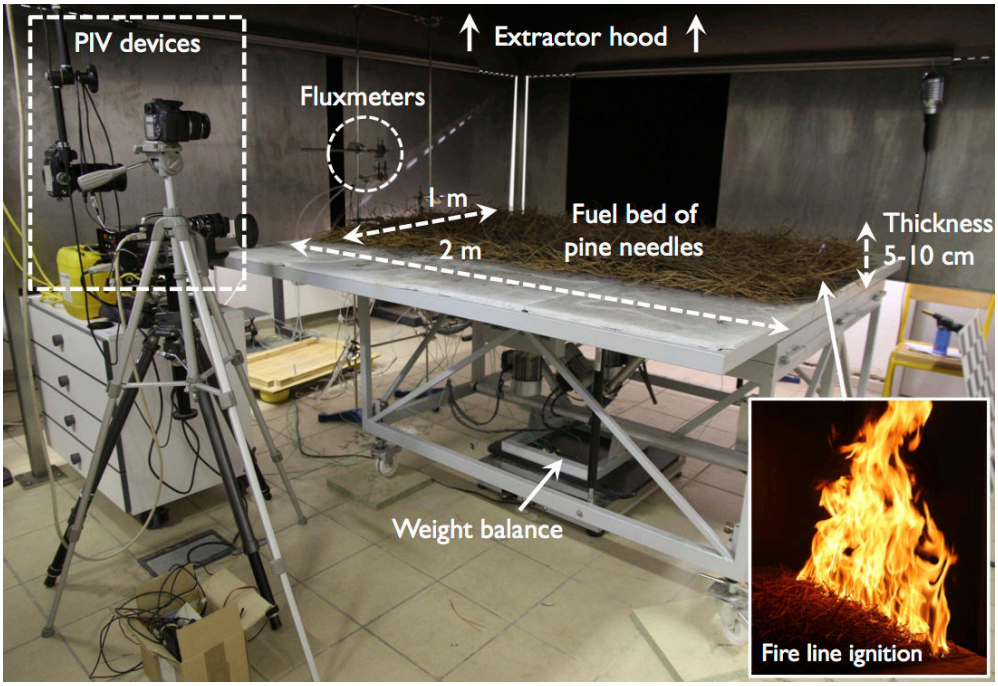


Figure 4.2: *Snapshot of the laboratory-scale experiment, with a hood extractor above the biomass fuel bed. Credit: CNRS-SPE laboratory.*

Figure 4.3 shows the evolution of the heat release and mass loss rates as a function of time, measured for flat and $20°$-upslope configurations of the laboratory-scale fire propagation. These results indicate that combustion rapidly reaches a quasi-steady state for the flat configuration, with a heat release rate nearly 100 kW and a mass loss rate maintained at 6 g/s over the fire propagation. In contrast, combustion exhibits highly unsteady features for the $20°$-tilted configuration since the heat and mass release rates do not reach stationary values. These release rates exhibit a peak value corresponding to 300 kW and 15 g/s, respectively, between 100 and 150 s after fire ignition. This behavior is partly due to increasing flame front surface over time. Due to increased buoyant effects induced by the upslope configuration, larger heat exchanges at the head of the fire make the propagation much faster than in the flat configuration (i.e., 13.1 mm/s versus 5 mm/s) but still at a quasi-constant value. The edges of the fire front propagate at a much reduced ROS due to important heat losses and thereby, the flame front surface increases over time. This is confirmed by Fig. 3.1, Chapter 3, which highlights the curved shape of the fire front for the $20°$-upslope configuration. Since the flame front surface increases, biomass fuel burns at a faster rate and more heat is released. Thus, heat and mass

release rates are enhanced and no quasi-steady state can be reached.

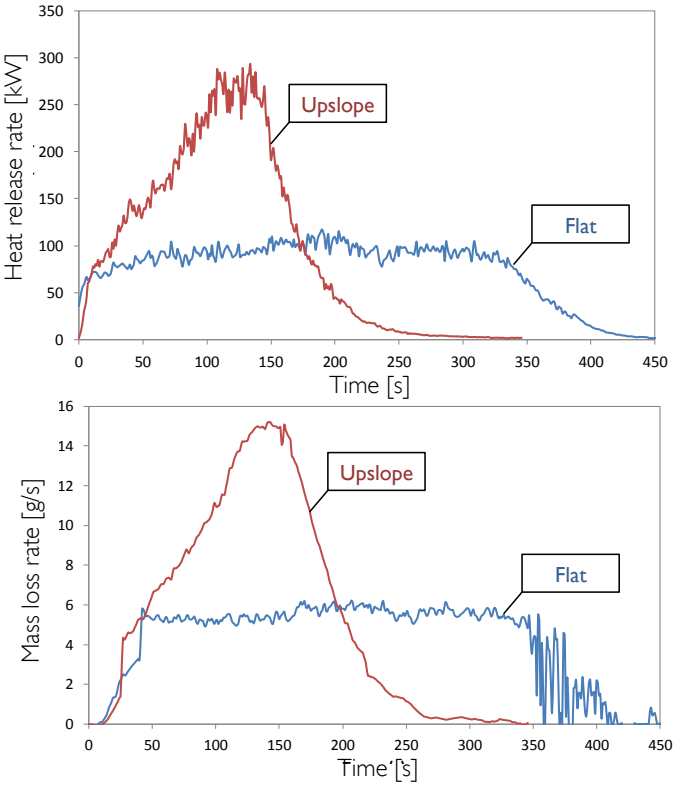


Figure 4.3: *Measurement of heat release rate (top) and mass loss rate (bottom) over time for the laboratory-scale fire: comparison between flat and upslope (20°-tilted) configurations. Credit: CNRS-SPE laboratory (private communication).*

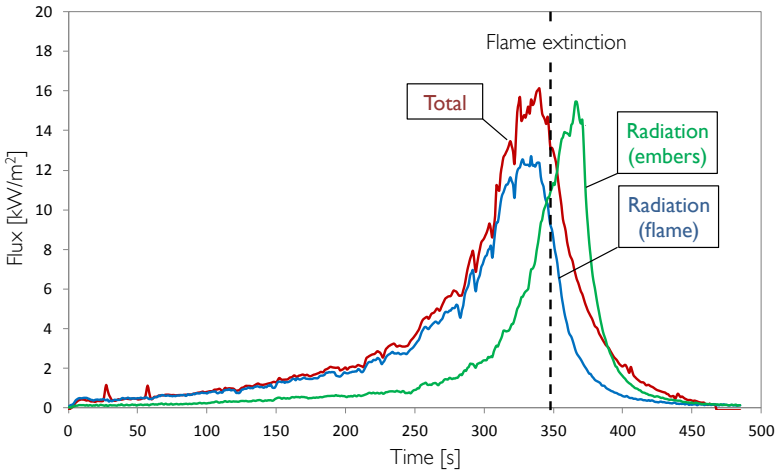


Figure 4.4: *Heat flux and radiation contributions in the flame and from embers as a function of time in flat configuration. Credit: CNRS-SPE laboratory (private communication).*

As shown in Fig. 4.3, the heat release rate and mass loss rate are highly correlated for both flat and $20°$-upslope configurations; they are indeed related by the effective heat of combustion $\Delta h_{c,ef}$ that is between 17 and 18 kW/kg. This value is consistent with the heat of combustion $\Delta h_c = 18.6$ kW/kg retained in the Rothermel's ROS model. Thus, the heat release rate (HRR) is related to the mass loss rate (MLR) as follows:

$$\text{HRR} = \underbrace{\Delta h_{c,ef}}_{\chi_{ef}\,\Delta h_c} \times \text{MLR}, \tag{4.1}$$

with χ_{ef} the effective burning coefficient, equal to 0.88 for the flat configuration and 0.92 for the $20°$-upslope configuration. The laboratory-scale experiment was also equipped with three fluxmeters (see Fig. 4.2) located at different heights with respect to the pine fuel bed: while one fluxmeter measured the total heat flux, the remaining two others measured only the radiation contribution of the heat flux, within the litter and in the flame. This configuration allows to estimate the radiation contribution in the heat release rate, noted $\chi_{rad,fl}$ for the flame contribution and $\chi_{rad,emb}$ for the ember contribution. Results are shown in Fig. 4.4; discrepancies between total and flame-induced radiation heat fluxes evidence the presence of convection heat fluxes. The convection contribution χ_{conv} to the heat release rate is estimated based on the analysis of the smoke flow and composition in the exhaust duct (Santoni et al., 2010).

Measurements performed during the laboratory-scale fire experiments are summarized in Table 4.1 for both flat and $20°$-upslope configurations, in terms of flame geometry, combustion and heat transfer characteristics.

Table 4.1: *Experimental measurements corresponding to laboratory-scale fires. Credit: CNRS-SPE laboratory (private communication).*

	Flat	**Upslope**
Flame height [m]	0.47	0.67
Flame tilt angle [°]	-8.2	27.8
Mean ROS [mm/s]	5	13
Effective heat of combustion $\Delta h_{c,ef}$ [kJ/kg]	17100	17900
Effective burning coefficient χ_{ef} [-]	0.88	0.92
Mean MLR [g/s]	5.3	14.0
Total radiation heat flux [kW/m^2]	15.1	25.8
Flame-induced radiation heat flux [kW/m^2]	12.3	22.2
Radiation contributions [-]		
• $\chi_{rad,fl}$ (flame)	0.10	0.19
• $\chi_{rad,emb}$ (embers)	0.13	0.11
Convection contribution χ_{conv} [-]	0.72	0.66

The upslope configuration induces larger flame height and ROS than in flat configuration, multiplied by a factor 1.4 and 2.6, respectively. These properties are due to the tilt of the flame towards unburnt pine needles when propagating in the upslope direction, increasing the radiation heat flux up to 25.8 kW/m^2, compared to 15.1 kW/m^2 in flat configuration (in the latter, the flame is slightly tilted towards the back side of the flame). Note that these radiation values correspond to the horizontal component of the radiation heat flux since fluxmeters are horizontally-positioned.

4.1.2 Flame-induced air entrainment

Displayed in Fig. 4.1, PIV is an optical laser diagnostic to characterize flow velocities (Melling, 1997; Raffel et al., 1998; Adrian, 2005). While being common in laboratory-scale and industrially-oriented combustion applications (Reuss et al., 1989; Wolfrum, 1998), and while being used in fire research to characterize the thermal plume (Zhou et al., 2003), this PIV technique has rarely been applied to characterize flow velocity in the near-flame field of a pine needle fire (Zhou and Gore, 1996; Sun et al., 2005; Said et al., 2008). The objective of this work was to demonstrate the feasibility of the PIV optical diagnostic in a delicate environment corresponding to the laboratory-scale experiment presented in Fig. 4.1, with slow flow velocities, buoyant-induced flow, strong flame radiation and flow seeding difficulties, but not to conduct a campaign of systematic measurements. An overview of the PIV technique is presented in Appendix E.

↪ Experimental setup

▷ **Particle image velocimetry system.** The PIV system classically used a Dantec Dynamics camera FlowSense[1] and a planar cross-correlation light sheet plane with a double-pulsed Continuum Precision Nd:YAG laser.[2] This Powerlite SL3-PIV laser operates at a monochromatic wavelength of 532 nm per pulse, a pulse rate of maximum 10 Hz and an intensity of 400 mJ per pulse (350 mJ per pulse effectively in the experiment). The light sheet plane is 50 cm high and 0.5 mm wide.

- At about 1.50 m ahead from fire ignition (see Fig. 4.5), the flow is seeded with oil liquid droplets (whose nominal diameter is lower than 1 μm) by injecting them at the top of the vegetation layer, ahead of the flame front. Even though they are subject to strong evaporation, this type of seed particles is highly recommended for health considerations (solid particles being toxic and more difficult to seed the flow).
- The seeded flow is illuminated based on two successive laser pulses within a short time interval fixed at $\Delta t = 3$ ms; this value allows to retrieve instanta-

[1] www.dantecdynamics.com/ccd-and-scmos-cameras.
[2] www.continuumlasers.com/products/pulsed_default.asp.

neous features of the gas flow since the flame propagates at a rate of 5 mm/s in flat configuration (the displacement of the flame during this time interval is 0.015 mm and is therefore negligible compared to the length-scales of the problem, e.g., the height of the flame is nearby 50 cm).

- The PIV frame in the gas flow is of size 115 mm $\times$ 59 mm corresponding to 13.6 pixel/mm. This visualization window is located at 1.50 m ahead from fire ignition. The gas flow velocity field is reconstructed based on statistical post-processing using the PIV software Dynamic Studio from Dantec Dynamics;[3] an additional filtering procedure is performed to remove non-physical features (e.g., noise, interrogation areas with a low amount of tracer particles) and focus on the zones in which valuable signals were obtained. In particular, velocity magnitudes lower than 0.01 m/s and characterized by less than 1 % occurrence are not considered in post-processing.

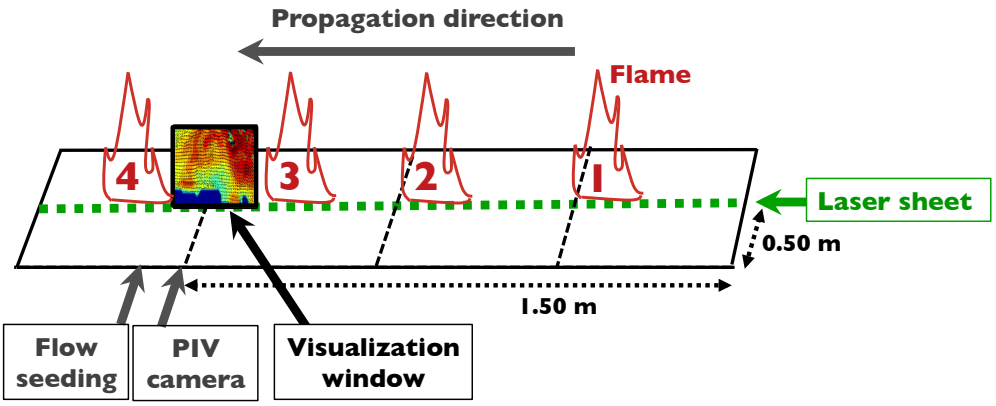


Figure 4.5: *Schematic of data acquisition for the PIV fire spread application.*

▷ **Measurement of ambient air flow conditions.** Since the experimental enclosure is connected to an extractor hood to evacuate combustion products, the impact of this hood on the air flow must be quantified before applying PIV measurements to flame propagation. For this purpose, PIV measurements were performed on the ambient gas flow (in the absence of flame). Figure 4.6 shows a typical example of PIV-based flow velocity field at a given time. There is no directional preference of air flow, implying that the extractor hood has a limited effect on this flow. The mean flow velocity magnitude was computed for the whole time-series of PIV frames (one value of velocity magnitude per PIV frame as shown in Fig. 4.7), leading to an ambient air flow velocity equal to 2.7 cm/s and characterized by a standard deviation of 0.8 cm/s. Globally, air flow velocity below 5 cm/s is typical of ambient conditions.

[3] www.dantecdynamics.com/particle-image-velocimetry.

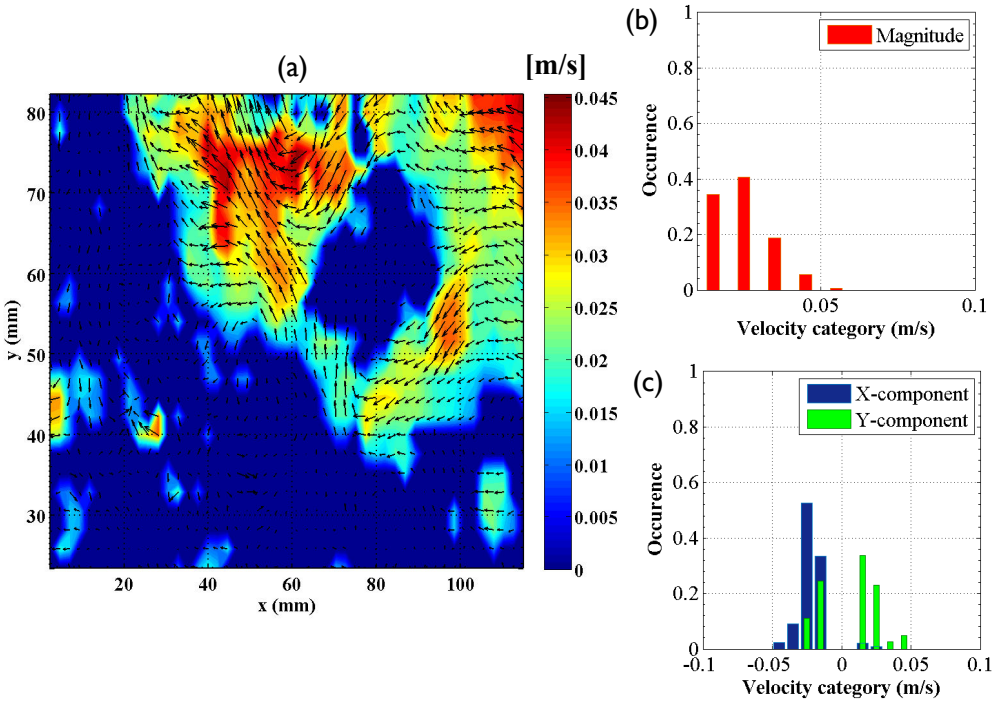


Figure 4.6: *Example of gas flow velocity field reconstructed from PIV analysis in ambient flow conditions (no flame). The x-axis corresponds to the horizontal direction, and the y-axis corresponds to the vertical direction. (a) 2-D velocity field. (b) Distribution of velocity magnitude. (c) Distribution of velocity magnitude along x- and y-directions (in blue and green colors, respectively).*

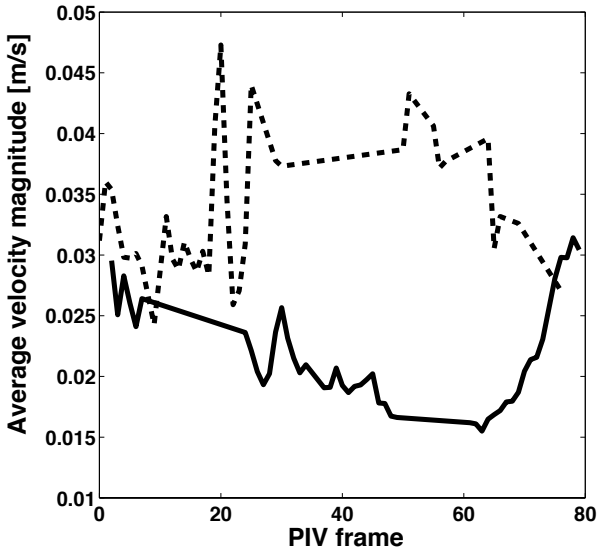


Figure 4.7: *Series of spatially-averaged flow velocity magnitude per PIV frame for two repetitive PIV experiments (solid and dashed lines).*

↪ Measurements of flame-induced air entrainment

As illustrated in Fig. 4.5, PIV measurements were performed for different positions of the flame with respect to the PIV camera (the visualization window), ahead from the flame front in positions indexed by 1, 2 and 3 as well as on the back side of the flame front in the position indexed by 4. For each position of the flame, 80 set of images (A/B) were recorded during 40 s, meaning that the flame propagates over 20 cm during each PIV acquisition. Figure 4.8 shows a typical example of PIV-based flow velocity field on the head side of the flame front (position 3); the flow is entrained towards the basis of the flame at a velocity magnitude nearby 0.10 m/s. This magnitude is increased by a factor more than 3 compared to ambient flow conditions. For comparison, Fig. 4.9 shows a typical example of PIV-based flow velocity field on the back side of the flame front (position 4); the flow is also entrained towards the flame, but at a much higher velocity magnitude (nearby 0.2 m/s) compared to the head side of the flame. Due to the lower density of hot gases, the velocity magnitude is higher on the back side of the fire by nearly a factor 2 than on the head side of the fire (at an equal distance to the flame). These discrepancies between the head and back sides of the fire approximately correspond to the ratio of the ambient air mass density to the pre-heated air mass density (at 575 K as shown in Fig. 4.10), meaning that this increased velocity magnitude is mainly due to thermal expansion. This explanation relies on the assumption that the gas entrained towards the flame is composed of air. While the head side of the flame most likely captures fresh air at ambient temperature, the back side of the flame could also be made of a gas mixture among whom CO_2. In this case, the temperature of the gas mixture would be higher than 575 K (see Fig. 4.10 for a pure CO_2 mixture). Further analysis is required to correlate these data to the gas composition and temperature on both sides of the flame.

PIV results for the different flame positions are summarized in Fig. 4.11, where each symbol corresponds to the mean velocity magnitude per PIV frame. Linear fits were performed between each dataset, i.e., between each flame position, in order to obtain the global change of the velocity magnitude with respect to the distance to the flame. These results confirm that the closer to the flame, the stronger the air entrainment. They also show that the flame induces significant gas flow towards the flame on both head and back sides of the fire. While the spatially-averaged velocity magnitude (one value of velocity magnitude per PIV frame as in Fig. 4.7) is varying between 0.08 and 0.14 m/s for the PIV acquisition at location 3, it is varying between 0.14 and 0.22 m/s at location 4, consistently with the instantaneous results presented in Figs. 4.8 and 4.9.

These measurements of gas flow velocity are limited here due to the difficulties in seeding the flow (e.g., evaporation of oil liquid droplets, difficult injection of particles in the zones of low velocities without resorting to an artificial flow), which provide partial information on the PIV frame (certain zones of the PIV frame remain blank due to the lack of tracer particles in corresponding interrogation areas).

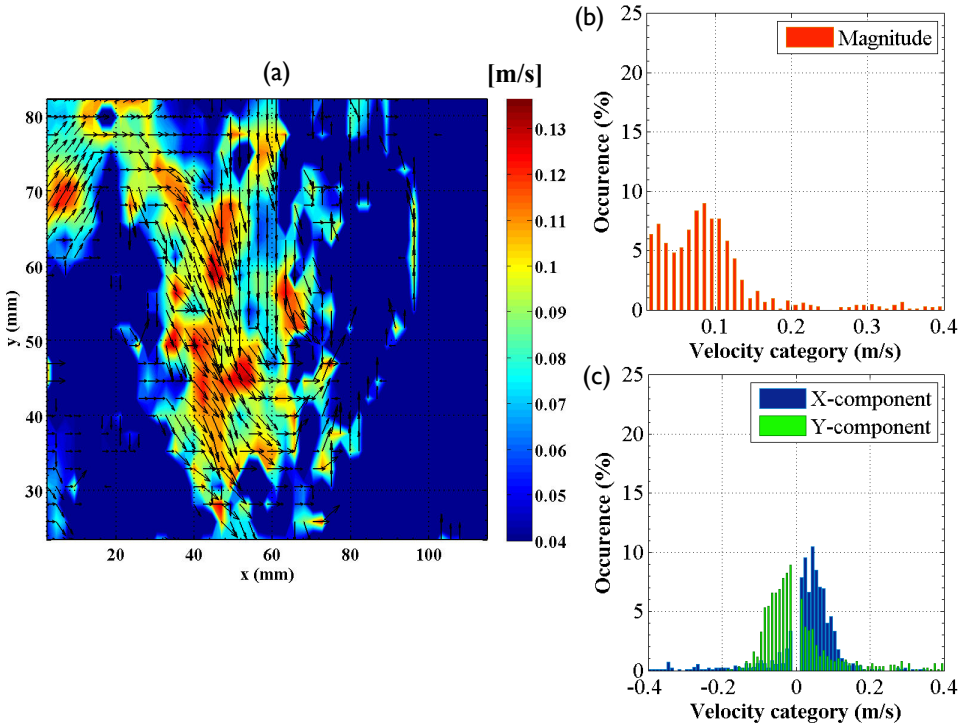


Figure 4.8: *Flow velocities for the fire head side (position 3). See caption of Fig. 4.6.*

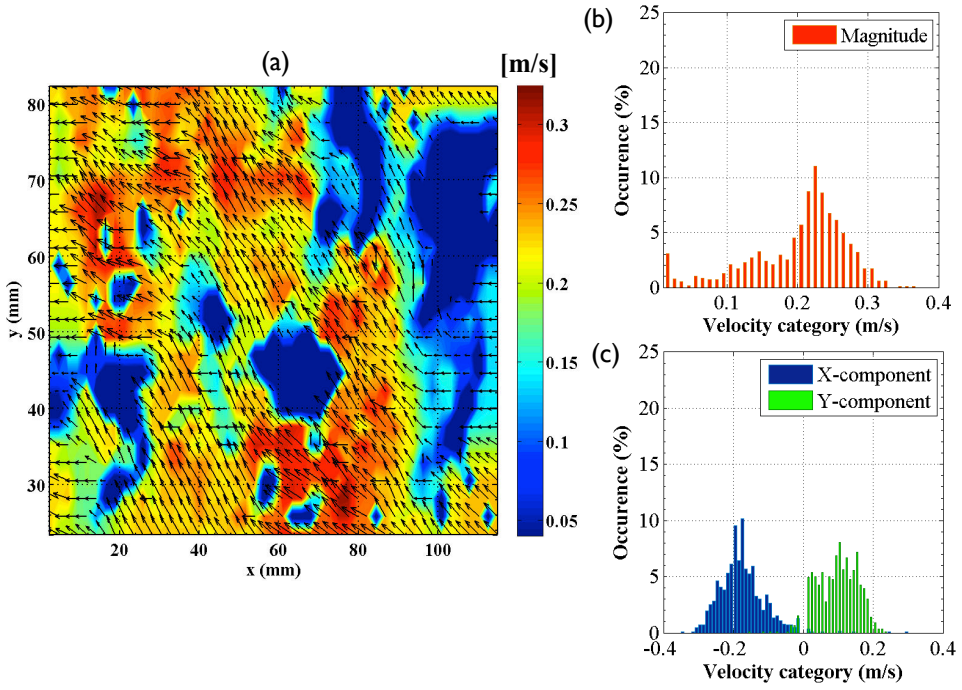


Figure 4.9: *Flow velocities for the fire back side (position 4). See caption of Fig. 4.6.*

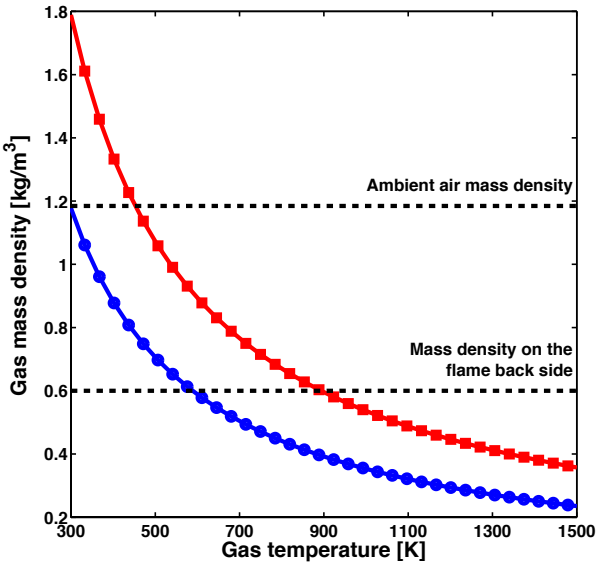


Figure 4.10: *Evolution of mass density with respect to temperature, based on the ideal gas state equation. The blue circled-solid line corresponds to air; the red squared-solid line corresponds to* CO_2*; horizontal dashed lines correspond to reference and target values.*

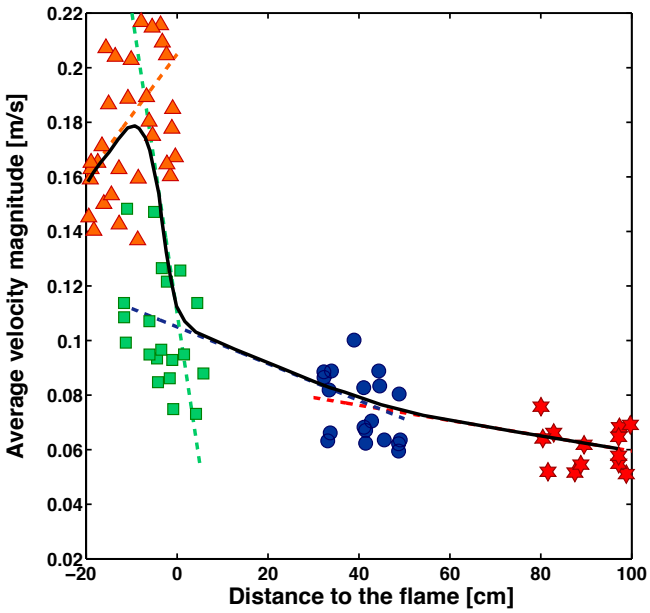


Figure 4.11: *Fit of the flow entrainment with respect to the distance to the flame against PIV measurements for 4 PIV datasets (corresponding to 4 different positions of the flame with respect to the PIV frame located at 0 cm). Symbols correspond to a subset of PIV measurements (one color per flame position at the time of PIV acquisition): red stars correspond to position 1, blue circles to position 2, green squares to position 3 and orange triangles to position 4; dashed lines correspond to analytical fits and the black solid line shows the global trend of the spatially-averaged velocity magnitude (per PIV frame).*

These measurements are also limited due to the restricted size of the visualization window in comparison to the length-scales of the problem. This limitation is due to the resolution of the camera and the laser power. In particular, the vertical profile of gas flow velocity at a given distance from the flame needs to be further analyzed to improve our knowledge on the mechanisms inducing air entrainment and vortex formation that are typical of buoyant flames. Still, this preliminary study of PIV applied to laboratory-scale fires has already showed that (1) air entrainment occurs on both head and back sides of the fire; (2) it is possible to quantify the related gas flow velocity due to thermal expansion. These information are essential for a quantitative comparison with detailed simulations of natural fire propagation.

It is worth mentioning that flame spectroscopy is not suitable for identifying the composition of radicals or burnt gases on the head/back sides of the flame. Indeed, the discontinuous emissions of minerals such as Na^* at 588 nm and K^* at 766 nm are much more important in biomass fuels than for instance radicals CH^* or C_2 as illustrated in Fig. 4.12. This implies that gas emissions are difficult to detect and to capture above the continuous baseline attributed to soot formation in the flame.

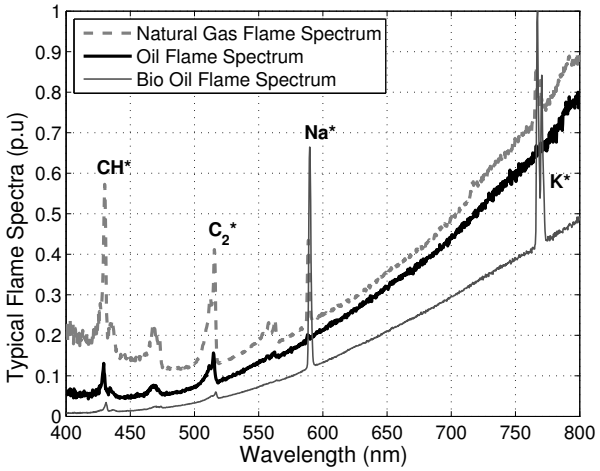


Figure 4.12: *Typical (normalized) flame spectrum of natural gas, oil and bio-oil hydrocarbon fuels, in which discontinuous emissions superimpose to the continuous baseline (soot). Credit: Arias and Pezoa (2012).*

In summary, PIV results evidenced the presence of gas flow entrainment on the back side of the flame region. The closer the flame, the stronger the air entrainment and in particular on the back side of the flame region due to thermal expansion. Thus, entrained gases are made of pre-heated air and possibly of hot burnt products. Further experimental analysis is required to infer the detailed composition of these gases. Still, these PIV measurements provide valuable information in the near-flame region that are useful for the development and validation of CFD simulations.

4.2 Radiation-forced two-dimensional case

As a preliminary step, two-dimensional (vertical) simulations combining the AVBP LES fluid solver and the PYROWO biomass thermal degradation solver, following the strategy presented in Chapter 3, are performed to investigate the flame response to buoyancy and to external radiation forcing, i.e., without coupling with the PRISSMA radiation solver.

The combustion of pyrolysis gas products (CH_4, CO, CO_2 and H_2O) with ambient air is described using the 2S-CH4-BFER reduced kinetic scheme presented in Section 2.4.3, Chapter 2. This 2-step scheme involves one reaction for CH_4 oxidation and one reaction for the equibribrium between CO and CO_2. The combustion of biomass fuel is presumed to occur for high equivalence ratios ($\Phi > 1$) at the scale of biomass fuel solid particles, where pyrolysis gases are emitted. For this purpose and as the actual equivalence ratio Φ is not known, the operating point studied here corresponds to a limit case, in which no air is premixed with pyrolysis gases before flame ignition. A pure diffusion flame is then considered, while in reality fires are more of the partially-premixed type (see discussion in Section 2.1.2, Chapter 2).

4.2.1 Numerical configuration

↪ Computational domain

The two-dimensional computational domain corresponds to a vertical cross-section of the gas flow in the frame of reference (x, y) as shown in Fig. 4.13(a), the x-axis representing the direction of fire propagation and the y-axis the (vertical) normal direction to the ground surface. Thus, $y = 0$ m corresponds to the WALL-PYROWO interface, i.e., the interface between the gas-phase solved by AVBP and the vegetation phase solved by PYROWO. For increasing values of y, the gas phase corresponds to the atmosphere, in which the flame is located in the near-ground region, referred to as *flame region* in Fig. 4.13(a), and in which the thermal plume can grow vertically over the fire duration.

↪ Computational grid

The 10 m $\times$ 10 m computational grid is meshed with about 235,000 triangular cells using the CENTAUR software[4] and with a refined zone in the flame region of size 4 m $\times$ 2.25 m as shown in Fig. 4.13(b). Therefore, the minimal cell size (1.2 mm) is obtained at the WALL-PYROWO interface, where the pyrolysis of the vegetation is initially located. In the adjacent mesh volume (of size 0.40 m $\times$ 0.60 m), the cell size is 2.5 mm on average. Then, the cell size is slowly increased up to 2 cm in the region of interest for LES, 10 cm at the inflow/outflow conditions and 30 cm in ambient atmosphere. The characteristics of the present computational domain are summarized in Table 4.2. Note that in the following simulations, the flame mainly

[4] www.centaursoft.com/.

remains in the zone meshed with a resolution of 2.5 mm and that the mesh would need to be modified for propagating flames over a large distance (for instance, by moving the zone of highest spatial resolution consistently with the flame propagation to save computational time). The second-order accurate cell-vertex Lax-Wendroff scheme is used in AVBP for numerical integration (to save computational time and because the grid resolution is fine enough to justify a second-order approximation).

Table 4.2: *Computational domain characteristics.*

x-axis		y-axis		Cell size		Cell numbers
Min.	Max.	Min.	Max.	Min.	Max.	
0 m	10 m	0 m	10 m	1.2 mm	30 cm	235000

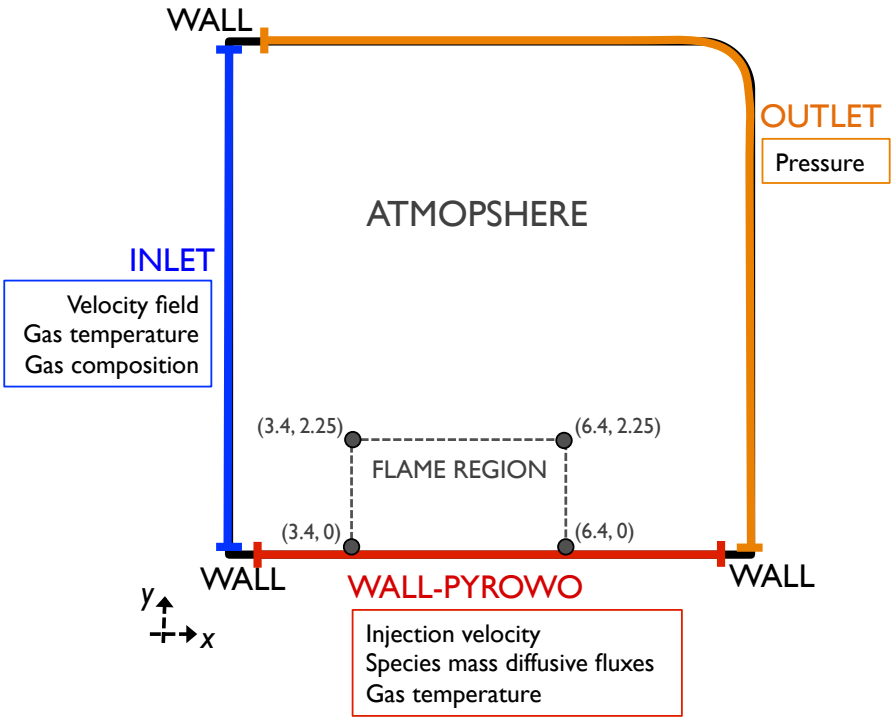
↪ Initial and boundary conditions

Navier-Stokes Characteristic Boundary Conditions (NSCBC) are used to impose the inflow/outflow conditions (Poinsot and Lele, 1992). On the left side corresponding to the inlet, fresh air at 300 K is injected with a low horizontal velocity equal to 0.01 m, see Fig. 4.13(a). On the right side and at the top, an imposed pressure outlet boundary condition is applied (the pressure is relaxed towards the ambient pressure $p_\infty = 1$ bar). At the bottom, the boundary condition corresponds to the gas/vegetation coupling interface (WALL-PYROWO), explained in detail in Section 3.4.2, Chapter 3. Note that these different boundary conditions are separated by no-slip adiabatic walls (i.e., zero-value velocity of the gas phase relative to the boundary) to avoid conflicts in the definition of the variables of interest (e.g., velocity vector) at corners.

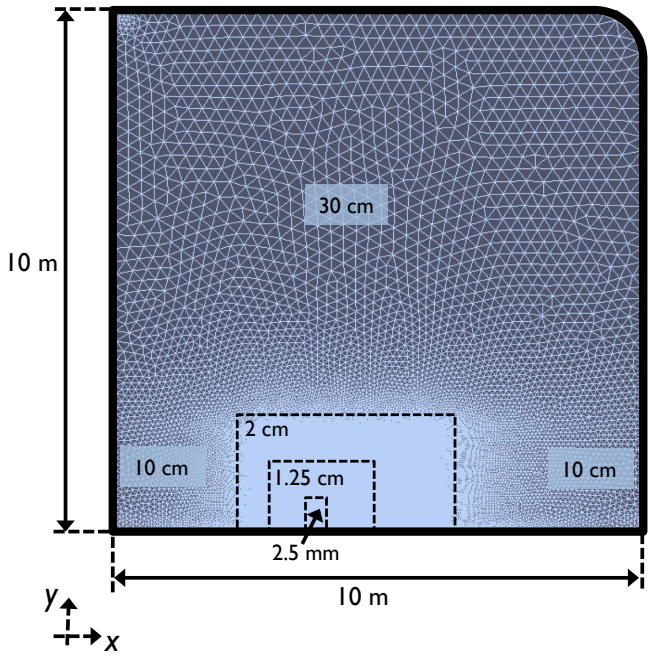
Note also that the computational domain is large enough (10 m $\times$ 10 m) to avoid interactions between the hot thermal plume and boundary conditions and thereby, to suppose that simulation results are not affected by boundaries. Initially, the gas phase is made of air ($Y_{N_2} = 0.767$, $Y_{O_2} = 0.233$) at 300 K and atmospheric pressure 1 bar, with a 0.01-m horizontal velocity from left to right (consistently with the inflow condition).

↪ Gas/vegetation interface

▷ **Radiation external forcing.** The AVBP/PYROWO coupling is tested by submitting the vegetation to a constant-imposed radiation profile over the simulation (see Fig. 4.14). This external forcing noted $\dot{q}_{rad,fl}$ enters in Eq. (3.14) of the PYROWO model (see Section 3.2, Chapter 3). The maximum value for the radiation source term $\dot{q}_{rad,fl}$ is 3.0×10^6 W/m^3 between $x = 3.8$ and 4.2 m, and is consistent with measurements. This region of the x-axis corresponds to the zone where pyrolysis gases are released during the simulation.



(a) Boundary conditions.



(b) Computational grid.

Figure 4.13: *Numerical configuration of two-dimensional simulations of* Avbp *coupled to* Pyrowo *through the gas/vegetation interface* Wall-Pyrowo. *The coordinates* (x, y) *of the boundaries of the flame region are indicated [m].*

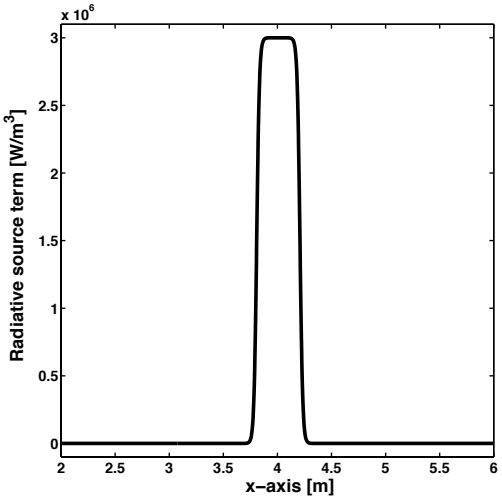


Figure 4.14: *Spatial variation of the radiation source term $\dot{q}_{rad,fl} = \dot{q}_{rad,fl}(x)$ along the x-axis in the* PYROWO *biomass fuel thermal degradation model. This view corresponds to a zoom since the actual x-axis is 10 m long.*

▷ **Boundary condition.** The pyrolysis gas composition at the flame/vegetation interface is set to $Y_{w,CH_4} = 0.10$, $Y_{w,CO} = 0.171$, $Y_{w,CO_2} = 0.64$ and $Y_{w,H_2O} = 0.089$ (the subscript w refers to the flame/vegetation interface). The related injection velocity $v_{inj,w}$ and temperature T_w at the gas/vegetation interface result from the time-integration of PYROWO at each AVBP boundary cell along the one-dimensional gas/vegetation coupling interface. This 10-m-long interface is discretized with 771 nodes, corresponding to 770 boundary faces (1-D segments). This implies that one PYROWO calculation is performed for any of the 770 boundary faces[5] using the calibrated parameters presented in Table 3.4, Chapter 3, with $c_{p,v} = 1,500$ J/K/kg, $\Sigma_v = 4,000$ 1/m, $\delta_v = 0.10$ m and $k_{conv} = 10$ W/K/m^2 (assumed constant over the simulation).

▷ **Initial condition.** Initially, the vegetation is at ambient temperature $T_v^0 = 300$ K; pine needle properties are specified as $\rho_p^0 = 500$ kg/m^3 (fuel bulk mass density), $\beta_v^0 = 0.05$ (fuel packing ratio), $Y_{v,wat}^0 = 0.10$ (fuel moisture content), $Y_{v,dry}^0 = 0.90$ (dry material content) and $Y_{v,char}^0 = 0$ (char material content), the superscript 0 referring to the initial state.

↪ Flame ignition

Since our objective is to study flame propagation, the flame is artificially ignited at $x = 4$ m in a three-step process:

(1) **Injection of non-reacting gas flow.** Radiation external forcing specified as $\dot{q}_{rad,fl} = 3.0\times10^6$ W/m^3 activates pyrolysis and gases are emitted through the

[5] The calibration performed in Section 3.2.3, Chapter 3, was performed for a unique boundary face, also referred to as *control volume*.

WALL-PYROWO interface between $x = 3.8$ and 4.2 m during the vegetation pre-heating phase. These gases emitted at an increasing temperature T_w are non-reacting in this preliminary step and induce the growth of a buoyant thermal plume as shown in Fig. 4.15.

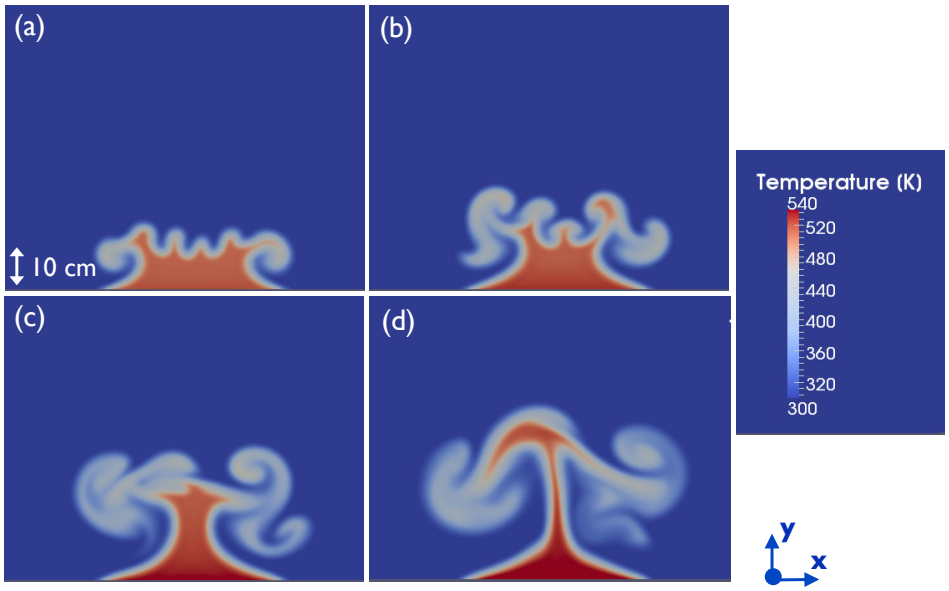


Figure 4.15: *Time-series of the two-dimensional (vertical) temperature field associated to pyrolysis gases (derived by* PYROWO*) at 25 s intervals: (a) 1.32 s. (b) 1.47 s. (c) 1.72 s. (d) 1.97 s. Images are of size* 95 *cm* $\times$ 75 *cm (width* $\times$ *height).*

Due to natural convection, injected pyrolysis gases are subject to perturbations that lead to the formation of convective cells, see Fig. 4.15(d). Corresponding Navier-Stokes fields are shown in Fig. 4.16. As they are injected separately, oxidizer (i.e., O_2 in Fig. 4.16(b)) and fuel reactants (e.g., CH_4 in Fig. 4.16(c)) have disjoint distributions; there is no O_2 where CH_4 reaches its maximum value $Y_{w,CH_4} = 0.10$. The velocity magnitude field in Fig. 4.16(d) highlights the vertical acceleration undergone by pyrolysis gases, which can reach values between 1.25 and 1.7 m/s as confirmed by the velocity (vertical) y-component in Fig. 4.16(f). The velocity x-component in Fig. 4.16(e) shows the formation of convective cells, clockwise on the right side of the thermal plume and counterclockwise on its left side. They are due to mass density gradients $\Delta\rho$ through the flow as shown in Fig. 4.16(a), resulting from changes in chemical composition (oxidizer/fuel reactants) and temperature (ambient/pyrolysis).

From an analytical viewpoint, the characteristic velocity induced by natural convection can be estimated by the following expression (see Section 2.1, Chapter 2):

$$U = \sqrt{g \left(\frac{\Delta\rho}{\rho_\infty}\right) H} = \sqrt{g \left(\frac{\rho_\infty - \rho}{\rho_\infty}\right) H}, \tag{4.2}$$

with H [m] the vertical length-scale corresponding here to the size of convection cells, ρ_∞ [kg/m^3] the air density at ambient temperature and U [m/s] the bulk flow velocity corresponding to the vertical velocity. At time shown in Fig. 4.15(d), $H \cong 45$ cm and $\Delta\rho \cong 0.50$ kg/m^3. Using Eq. (4.2), the characteristic buoyant velocity is estimated at 1.40 m/s; this value is consistent with the simulated velocity magnitude obtained in the plume with AVBP in Fig. 4.16(d). It also corresponds to a Richardson number equal to Ri $= 1$, with the Reynolds number Re $= 40,000$ (considering the air kinematic viscosity $\nu = 15.68 \times 10^{-6}$ [m^2/s]) and the Grashof number Gr $= 1.6 \times 10^9$, a value typical of the transition to turbulent flows.

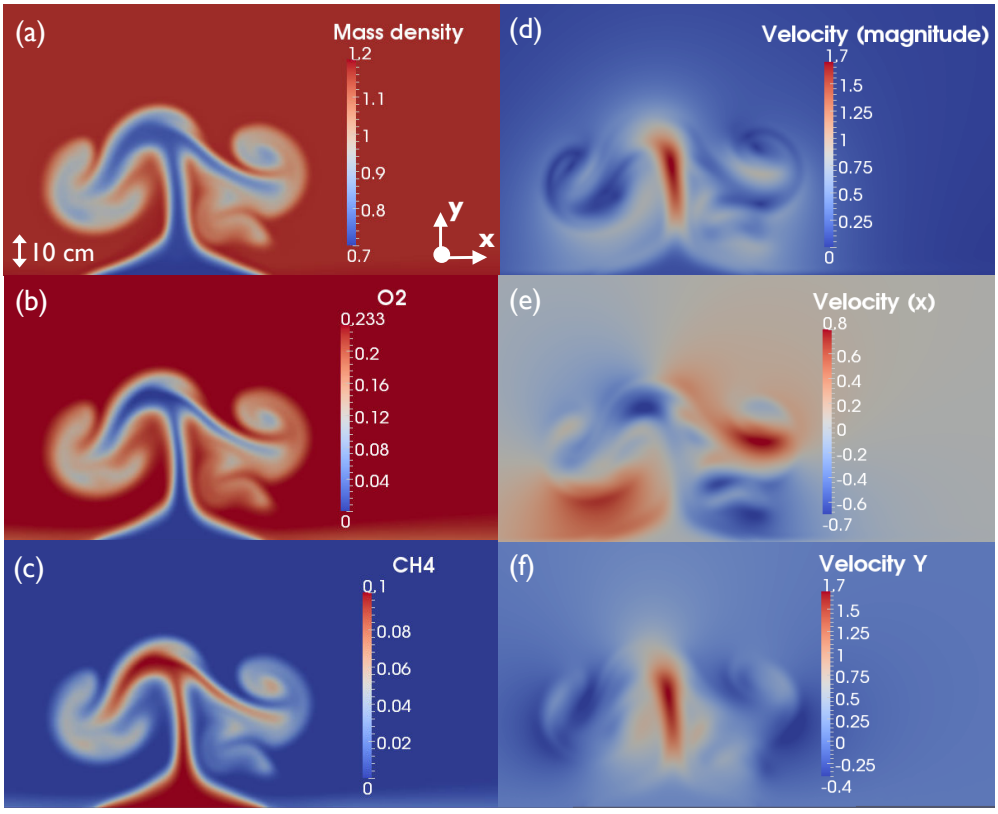


Figure 4.16: *Two-dimensional fields associated to the injection of pyrolysis gases into the non-reacting gas phase corresponding to the temperature field shown in Fig. 4.15(d). (a) Mass density [kg/m^3]. (b) Mass fraction of O_2 [-]. (c) Mass fraction of CH_4 [-]. (d) Velocity magnitude [m/s]. (e) Velocity component along x-axis [m/s]. (f) Velocity component along y-axis [m/s]. Images are of size 150 cm $\times$ 75 cm (width $\times$ height).*

(2) **Ignition using a burnt gas pocket.** To start combustion and ensure numerical stability, a pocket of hot burnt gases (burnt gas temperature and mass fractions based on 0-D calculations in adiabatic conditions) is deposited at mid-plume height. Then, AVBP solves for the reacting gas flow with a time step $\Delta t^A = 5 \times 10^{-7}$ s using the standard isotropic Smagorinsky model for

subgrid-scale turbulent combustion modeling and no flame thickening. Thus, the chemistry of the gas phase is described through Arrhenius reaction rates with pre-exponential adjustment (PEA) via the 2S-CH4-BFER reduced kinetic scheme (Franzelli, 2011).

(3) **Flame attachment/stabilization.** Through diffusion processes, the flammable gas mixture surrounding the burnt gas pocket heats and starts burning. The flame front then propagates upstream towards the injection of pyrolysis gases. At the end of the propagation phase, the flame stabilizes above the pyrolysis zone and anchors to the Wall-Pyrowo interface at locations corresponding to stoichiometry between the incoming air flow and released pyrolysis gases as shown in Fig. 4.17(a). Note that this flame attachment is difficult to handle from a numerical viewpoint, due to the ongoing competition between the vegetation supply in pyrolysis gases (low injection velocities) and buoyancy that convects burnt gases upwards and thereby, stretches the flame.

This ignition procedure is obviously not realistic and must be regarded as a numerical procedure to establish a flame. The activation of combustion reaction rates constitutes indeed a drastic shock for the simulation (since the thermodynamical properties of the flow in terms of temperature and mass density abruptly change), which needs to be carefully handled numerically.

4.2.2 Flame structure

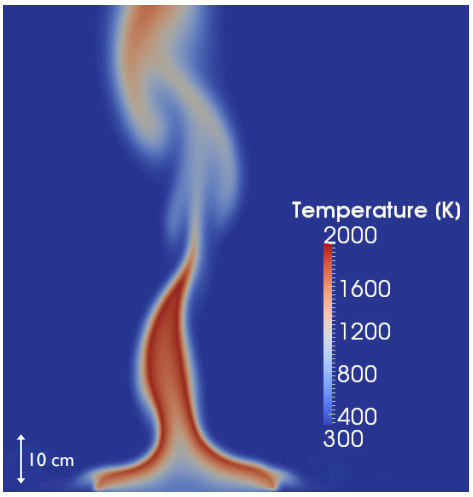
↪ Macroscopic flame properties

An instantaneous snapshot of the temperature field after ignition is presented in Fig. 4.17(a). Figure 4.18 shows the corresponding Navier-Stokes variables: since the gas mass density is drastically reduced in burnt gases (see Fig. 4.18(a)), they are subject to a much stronger buoyant acceleration (see Fig. 4.18(b)). Figures 4.18(c) and (d) illustrate the two reaction rates associated with the 2S-CH4-BFER reduced chemical kinetic scheme: since CO and CO_2 are present in pyrolysis gases before flame ignition, they strongly enhance the second reaction corresponding to the equilibrium reaction between CO and CO_2. This reaction rate is more important than that of the first reaction corresponding to the dissociation of CH_4 (the mass fraction of CH_4 is significantly lower than that of CO_2 and CO). This chemical behavior is consistent with the results provided by the one-dimensional laminar flame analysis (see Section 3.3.1, Chapter 3).

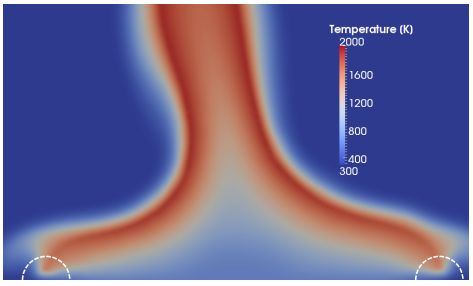
The mass density of pyrolysis gases is significantly reduced at the flame/vegetation interface compared to that of the surrounding ambient air ($\rho_{w,pyr} = 0.58$ kg/m^3 versus $\rho_{w,air} = 1.17$ kg/m^3) and the temperature $T_w = 654$ K is representative of the actual pyrolysis temperature. The wall-normal injection velocity is $v_{inj,w} = 0.114$ m/s, corresponding to a maximum wall-normal mass rate of $\dot{s}_w = 0.066$ kg/m^2/s (see Eq. 3.23, Chapter 3).

This configuration corresponds to a global equivalence ratio $\Phi = 2.14$ (see Eq. 3.19,

Chapter 3) based on the composition of pyrolysis gases and air. However, due to unconfined environment, this value does not describe for which fuel/air proportion combustion actually occurs in LES. The thin flame region (a few mm) produces hot gases at a temperature slightly below 2000 K (1920 K at time 2.275 s), which is consistent with the burnt gas temperature obtained in the laminar counter-flow diffusion flame studies (see Fig. 3.19, Chapter 3) for strain rate values below $10\ s^{-1}$. The flame height is about 55 cm, which is consistent with experimental data (see Table 4.1). Besides, the fuel reactants are completely burned; there is no CH_4 and CO left in the thermal plume above the flame zone, while a large amount of O_2 is still present due to buoyancy and air entrainment, thereby promoting mixing with surrounding ambient air. The zoom on the temperature field presented in Fig. 4.17(b) also indicates that the flame is detached from the top of the vegetation layer. This flame detachment may be due to the non-infinitely fast chemical kinetics, implying that there is a time delay before combustion starts.



(a) Temperature field.



(b) Zoom on pyrolysis zone.

Figure 4.17: *Instantaneous snapshot of the* AVBP *temperature field at time 2.275 s (time is reset to 0 at ignition) corresponding to the combustion of pyrolysis gases in a diffusion flame configuration, with the flame front (the red zone between 1900 and 2000 K) attached to the top of the vegetation layer. The top-view is of size* $95\ cm \times 100\ cm$.

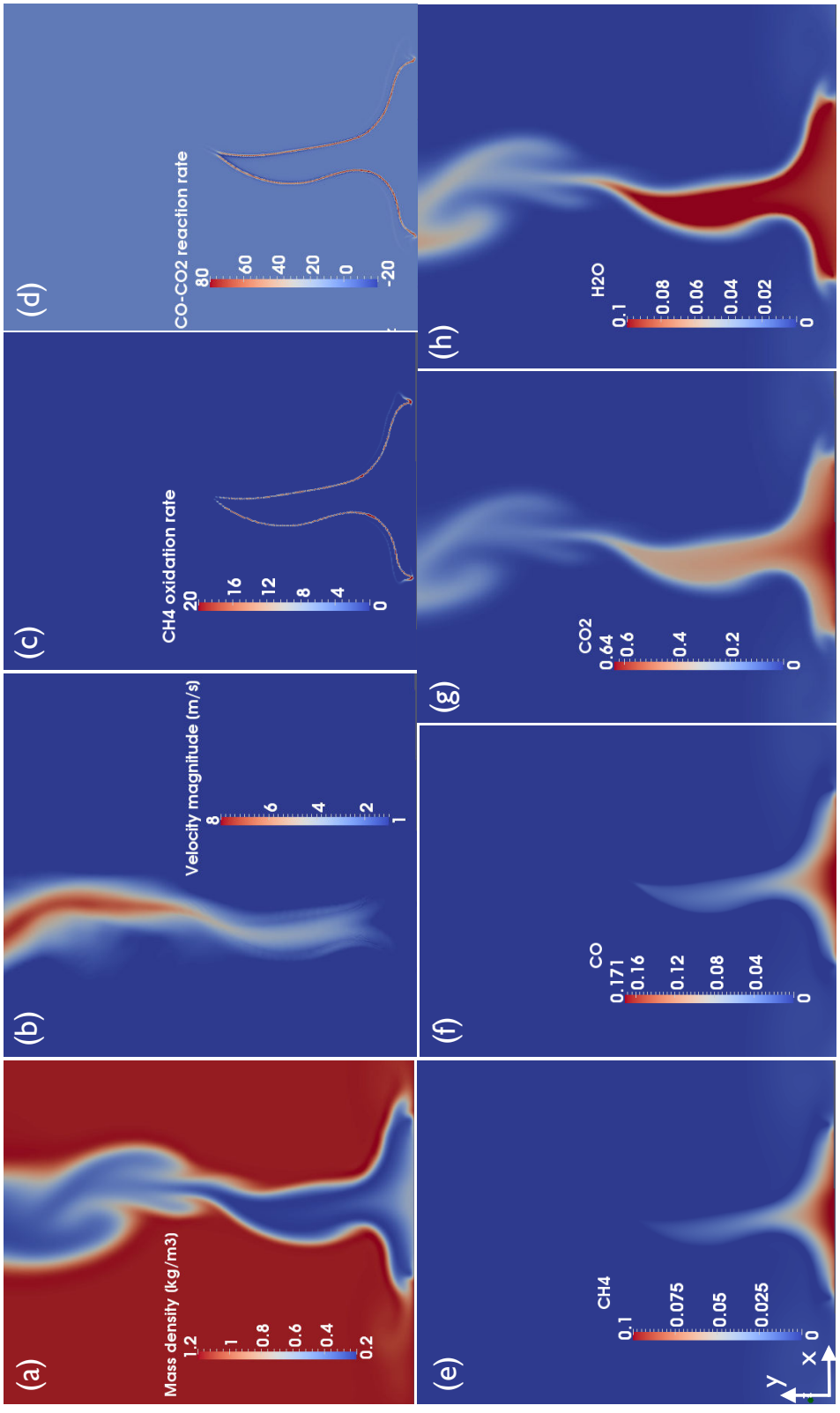


Figure 4.18: *Two-dimensional fields for the combustion of pyrolysis gases corresponding to the temperature field shown in Fig. 4.17. (a) Mass density [kg/m^3]. (b) Velocity magnitude [m/s]. (c) Reaction rate $\dot{\omega}_1$ [kg/m^3]. (d) Net reaction rate $\dot{\omega}_2$ [kg/m^3]. (e)-(f)-(g)-(h) Mass fractions Y_k [-]. Images represent an actual height of* 100 *cm.*

↪ **Analysis of the flame structure**

▷ **Analysis tools for diffusion flame.** A classical approach taken from the diffusion flame literature is adopted, in which a mixture fraction z is introduced to locally measure the fuel/oxidizer ratio: $z = 0$ in pure oxidizer, $z = 1$ in pure fuel reactants. Variables $T = T(z,t)$ and $Y_k = Y_k(z,t)$ define the structure of a diffusion flame. Since in the present configuration, fuel reactants are made of several gas compounds; the mixture fraction z is defined based on the atomic mass fraction (Bilger, 1989):

$$z = \frac{2\,Z_C + \frac{1}{2}\,Z_H + (Z_O^O - Z_O)}{2\,Z_C^F + \frac{1}{2}\,Z_H^F + Z_O^O}, \tag{4.3}$$

where the superscripts F and O indicate pure fuel reactants and air, respectively, and where the atomic mass fraction Z_i of atom i satisfies:

$$Z_i = \sum_{k=1}^{N_g} \frac{n_i^k\,Y_k}{W_k}, \tag{4.4}$$

with n_i^k the number of atoms i in the k-th gas compound. In addition, the scalar dissipation rate χ $[1/s]$ proportional to $|\nabla z|^2$ controls mixing. The Takeno flame index is also defined to identify local combustion regimes (i.e., indicator of how much a flame tends towards a premixed or diffusion flame) and thereby, understand the flame structure (Takeno and Nishioka, 1993). This dimensionless index is normalized to be 1 in premixed flames and -1 in diffusion flames; it is conditioned by the reaction zones as follows:

$$\text{Takeno} = \frac{\nabla Y_F \cdot \nabla Y_O}{\max\,(\nabla Y_F \cdot \nabla Y_O)}. \tag{4.5}$$

▷ **Simulation results.** The field of mixture fraction z is shown in Fig. 4.19(a), where the isoline $z_{st} = 0.32$, corresponding to the mixture fraction at stoichiometry $z = z_{st} = 1/(1+\Phi)$ ($\Phi = 2.14$ taking each reactant in its respective boundary state, i.e., $Y_{w,CH_4} = 0.10$, $Y_{w,CO} = 0.171$ and $Y_{air,O_2} = 0.233$), is superimposed as an indicator of the flame region. Figure 4.19(b) shows the map of the Takeno index. The flame region corresponds to a negative Takeno index, showing that the combustion regime is of the diffusion type. The Takeno index also reveals the presence of one zone near the flame/vegetation interface where the Takeno index is positive, indicating locally the occurrence of premixed combustion (due to the flame detachement from the top of the vegetation, allowing for premixing between ambient air and pyrolysis gases). This mixing is confirmed in Fig. 4.19(c) by the maximum values of the scalar dissipation rate reached in this region. This figure also highlights that the flame section (between $y = 0.10$ and 0.20 m) subject to the strongest buoyancy effects is also where the scalar dissipation rate is maximum. This confirms that buoyancy enhances mixing.

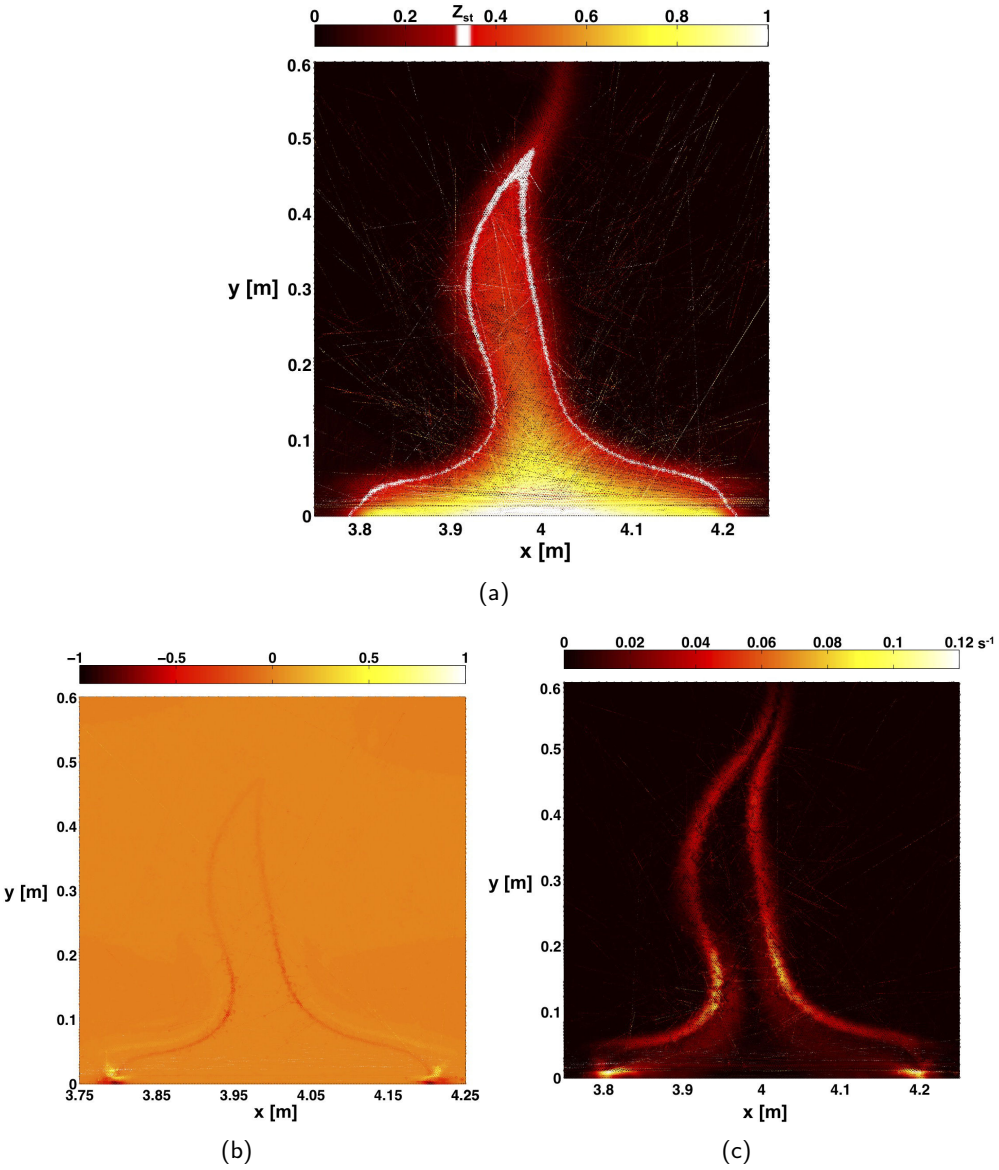


Figure 4.19: *Instantaneous snapshot of the* Avbp *fields at time 2.275 s corresponding to Fig. 4.17(a). (a) Mixture fraction z $[-]$ (the white zone corresponds to the mixture fraction at stoichiometry $z = z_{st} = 0.32$) with $z = 0$ in air and $z = 1$ in fuel reactants. (b) Takeno index. (c) Scalar dissipation rate χ [1/s].*

The particular flame structure is characterized with a scatter plot of temperature versus mixture fraction z, colored by the Takeno index, in Fig. 4.20. This plot shows that the 1921-K maximum temperature is reached at the stoichiometric mixture fraction $z = z_{st}$ and is typical of a diffusion flame: the bottom line between the non-reacting air at 300 K and the pyrolysis gases injected at 654 K corresponds to the mixing line; the two curve portions joining at the stoichiometric point z_{st}

correspond to fully burnt gases. Figure 4.21 shows the corresponding scatter plots of CH_4 (as an example of fuel reactants) and O_2 (oxidizer) versus mixture fraction z, colored by the gas temperature. They are also typical of a diffusion flame structure.

These results highlight that the present buoyant flame exhibits a structure typical of a diffusion flame. Its thermal thickness can be estimated from the gradient of the mixture fraction z to $\delta_L = 1.3$ cm, which is consistent with the length-scales obtained with the one-dimensional laminar flame calculations (see Section 3.3.3, Chapter 3). Note that the resolution of the computational grid in the flame zone is on average 2.5 mm, which is suitable for $\delta_L = 1.3$ cm.

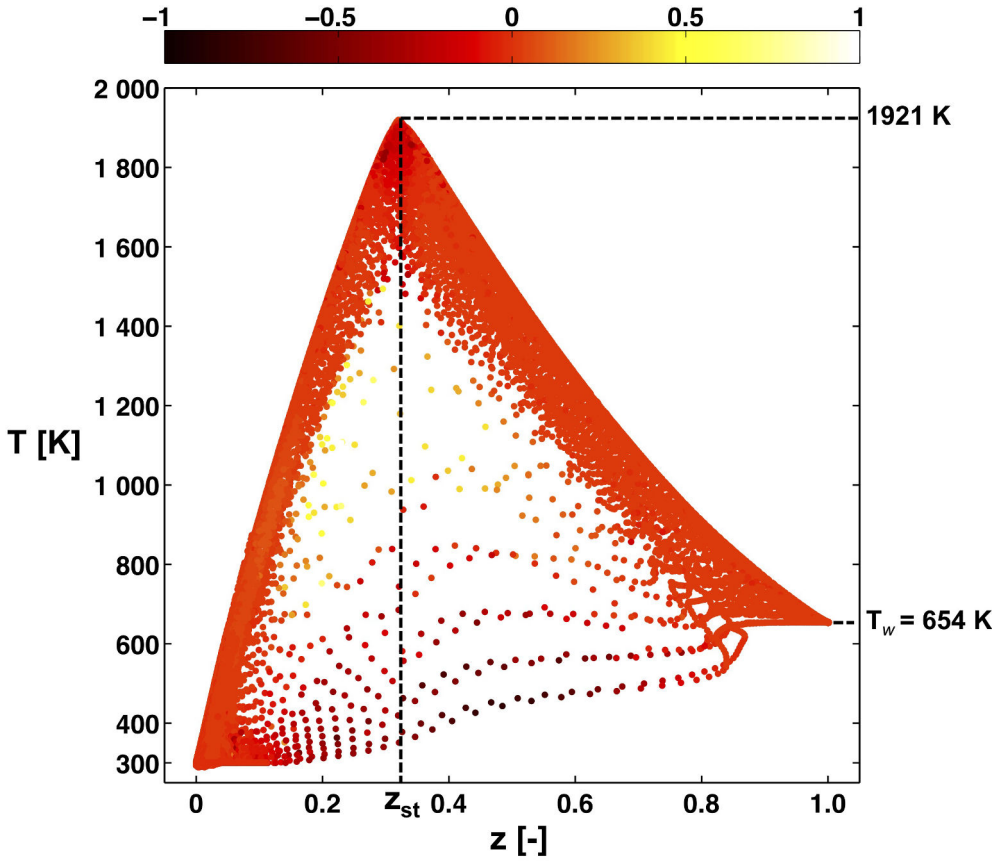
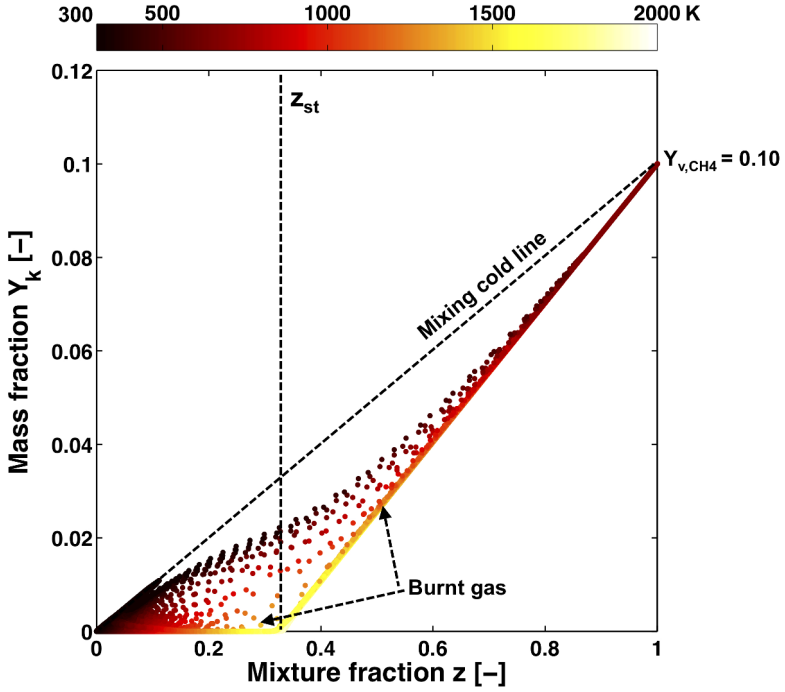
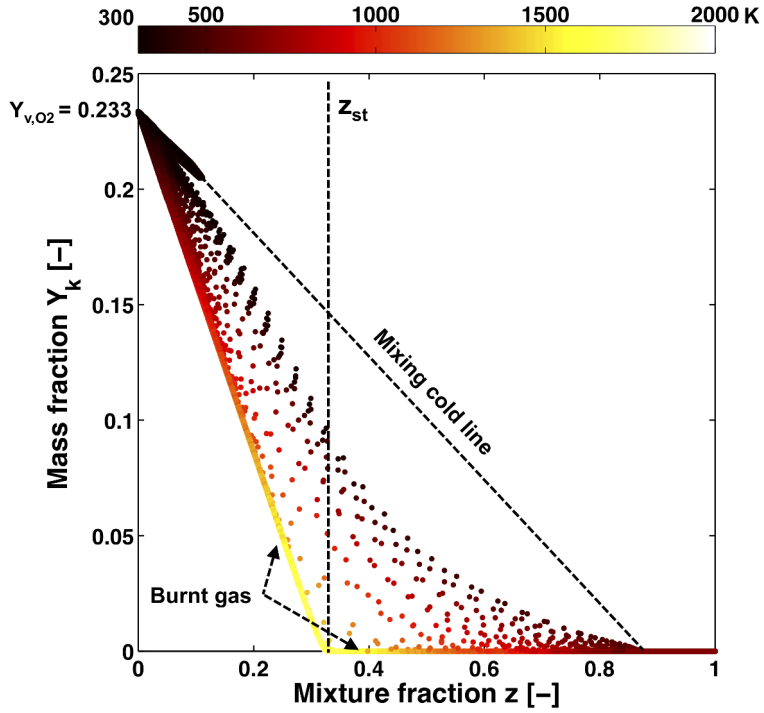


Figure 4.20: *Scatter plot of temperature versus mixture fraction z at time 2.275 s corresponding to Fig. 4.17(a), colored by the Takeno index (ranging from -1 to 1); $z = 0$ corresponds to ambient air, and $z = 1$ corresponds to pyrolysis gases released at $T_w = 654$ K.*



(a) Fuel reactant.



(b) Oxidizer.

Figure 4.21: *Scatter plot of pyrolysis reactants and oxidizer versus mixture fraction z at time 2.275 s corresponding to Fig. 4.17(a), colored by the gas temperature [K]. (a) Mass fraction of CH_4. (b) Mass fraction of O_2.*

4.2.3 Flame puffing

The flame height significantly fluctuates over time as shown in Fig. 4.22. Consistently, the temporal evolution of the reacting gas phase obtained in the simulation is presented for the two-dimensional configuration in Fig. 4.23, from 1.40 to 2.525 s. Figure 4.23 highlights the transient behavior of the flame enhanced by buoyancy effects (due to large density gradients between ambient air and burnt gases). The flame height significantly fluctuates over time (the flame height is multiplied by up to a factor 2): due to high flow strain and large-scale vortices in the fire plume, the flame is stretched out and then, pockets of hot gases are detached from the flame (attached to the top of vegetation). Vortices induce ambient air entrainment towards the flame region, increasing the surface between oxidizer and fuel reactants and thereby, promoting combustion. The detachment of hot gas pockets, also known as *puffing*, occurs periodically at 0.4 s intervals, as observed experimentally in Fig. 3.20, Chapter 3. The mean flame height is 68 cm, which is higher than experimental measurements (one reason for these discrepancies may be the lack of heat losses in this simulation).

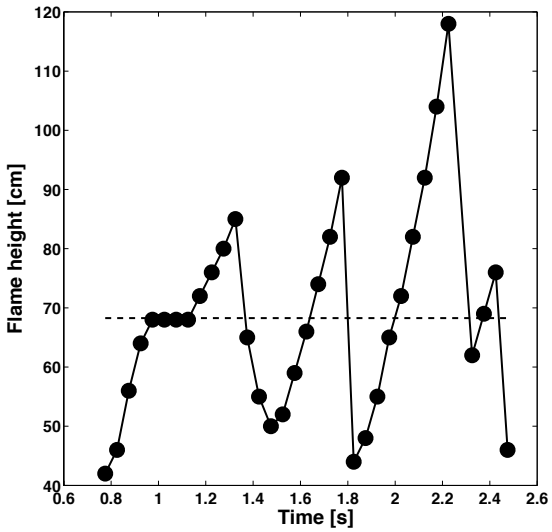


Figure 4.22: *Time-evolution of the flame height [m], measured as the height of the 1200 K iso-temperature of the main flame component anchored at the gas/vegetation interface. Solid line: instantaneous value. Dashed line: mean value over the simulation (i.e., 1.125 s). Images are of size* 50 *cm* $\times$ 100 *cm (width* $\times$ *height).*

The physical time period simulated in Fig. 4.23 (i.e., 1.125 s) corresponds to nearly 4 million iterations of the coupled AVBP/PYROWO solver (the coupling frequency between these two solvers is $N_{it} = 1$); these iterations require 24 h on 192 processors on the CORAIL supercomputer at CERFACS (190 processors are devoted to AVBP, 1 processor to PYROWO and 1 processor to the OPENPALM driver, see Appendix B).

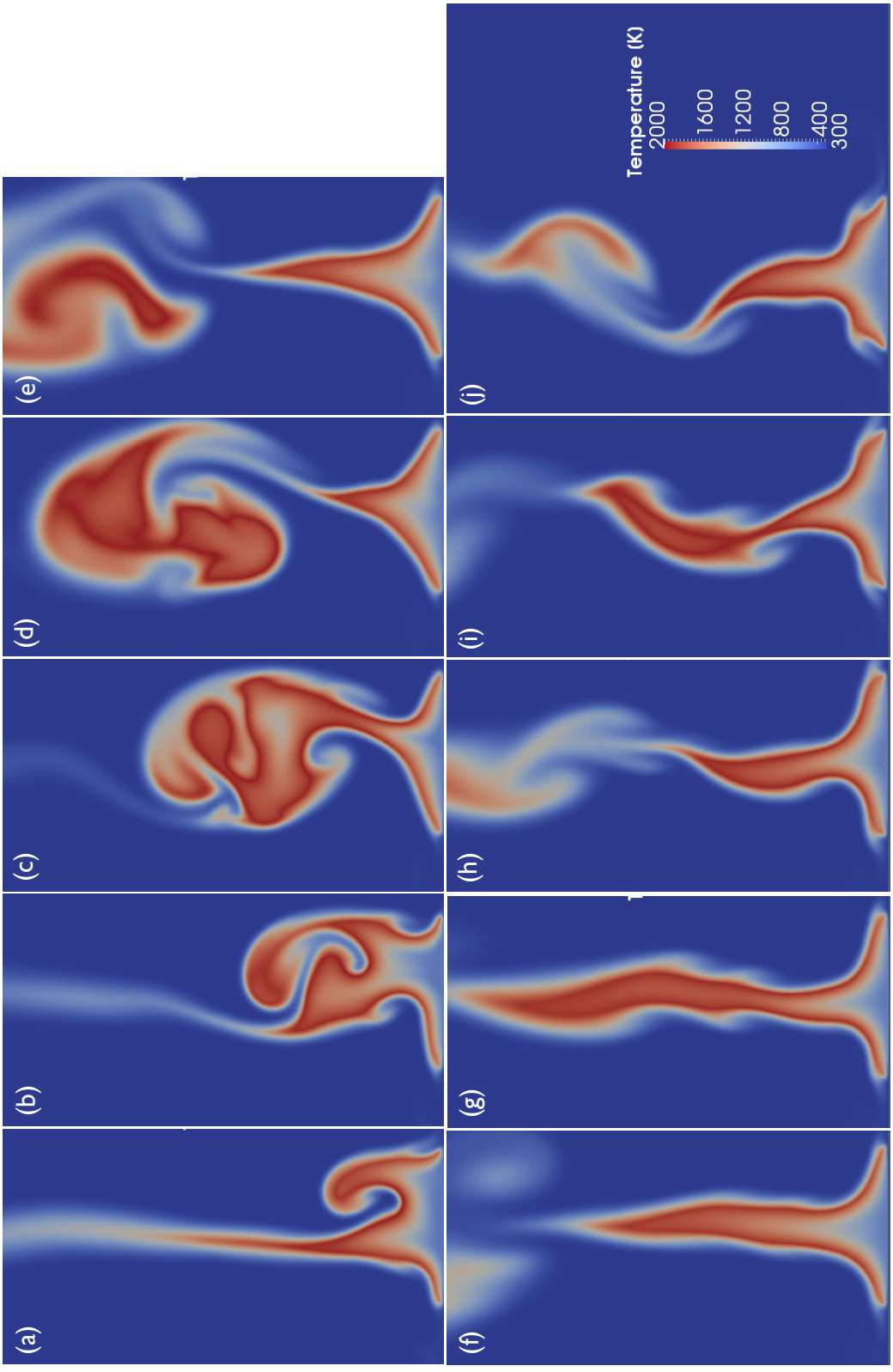


Figure 4.23: *Time-series of two-dimensional temperature field of the reacting gas phase solved by* AVBP *at 0.125 s time intervals (from 1.40 to 2.525 s).*

4.3 Multi-physics three-dimensional case

Three-dimensional flame-scale simulations are performed at laboratory scale. First, a radiation-forced three-dimensional case with a coupling between AVBP and PYROWO (i.e., without coupling with the PRISSMA radiation solver) is performed to determine the most suitable numerical strategy to obtain a robust and stable flame solution. Second, a fully multi-physics three-dimensional case with a coupling between AVBP, PYROWO and PRISSMA is run to investigate the spatial distribution of radiation that determines the potential rate of fire spread. A comparison with semi-empirical modeling approaches for macroscopic quantities of interest is provided.

As for the radiation-forced two-dimensional case, the combustion is described using the 2S-CH4-BFER reduced chemical kinetic scheme presented in Section 2.4.3, Chapter 2. The operating point studied here is also a pure diffusion flame.

4.3.1 Numerical configuration

↪ Computational domain

The three-dimensional case is set up based on the previous two-dimensional case shown in Fig. 4.13(a) by adding the z-dimension, representing the normal direction to the fire spread direction in the frame of reference (x, y, z). Thus, $y = 0$ m still corresponds to the WALL-PYROWO interface that is now two-dimensional and described in terms of x- and z-coordinates. z varies between 0 and 1 m since the fuel bed litter is 1-m-wide in the experiments (see Fig. 4.2). A schematic of the computational domain is presented in Fig. 4.24.

↪ Computational grid

As a preliminary step, the three-dimensional computational domain is meshed based on the two-dimensional unstructured mesh shown in Fig. 4.13(b). While each plane (x, y) for a given z-coordinate is still meshed with the two-dimensional unstructured mesh using 235,000 triangular cells, reaching a resolution of 2.5 mm in the flame region; the z-direction is discretized with 21 nodes corresponding to a cell size of 5 cm. Globally, the computational grid is composed of 4.7 millions of cells. In particular, the WALL-PYROWO at the gas/vegetation interface is discretized with 15,400 cells. As for the two-dimensional case, the second-order accurate cell-vertex Lax-Wendroff scheme is used in AVBP for the numerical integration with a time step $\Delta t^A = 3 \times 10^{-7}$ s. The PYROWO time step Δt^{PY} is set to Δt^A, since the coupling frequency is $N_{it} = 1$ and both solvers simulate the same physical time. Note that the PYROWO time step Δt^{PY} could be much larger (on the order of 10^{-3} s), but for simplicity purposes and since its computational cost is negligible in comparison to AVBP, $\Delta t^{PY} = \Delta t^A$.

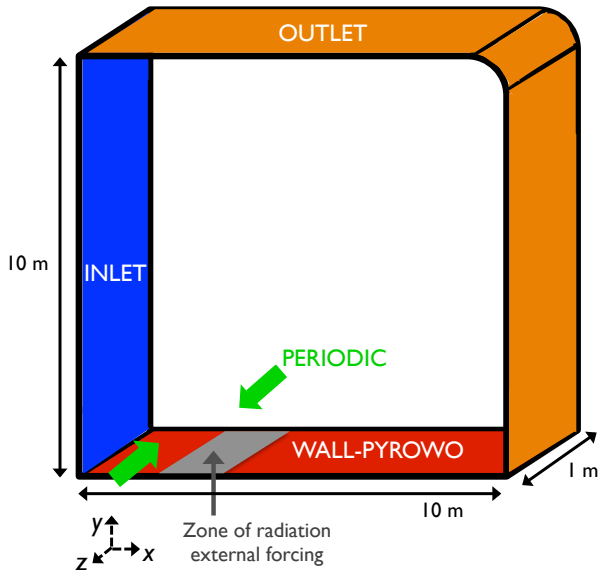


Figure 4.24: *Numerical configuration (geometry and boundary conditions) of three-dimensional simulations of* Avbp *coupled to* Pyrowo *through the two-dimensional gas/vegetation interface* Wall-Pyrowo.

In this three-dimensional configuration, the flame is artificially-thickened using the TFLES model following Eqs. (2.47)-(2.48), see Chapter 2. Motivations for this choice for a diffusion flame are explained in Section 3.3.4, see Chapter 3. The adaptive thickening along the y-direction presented in Fig. 4.25 avoids numerical issues associated with the flow-induced strain that makes the flame thinner and that leads to puffing.

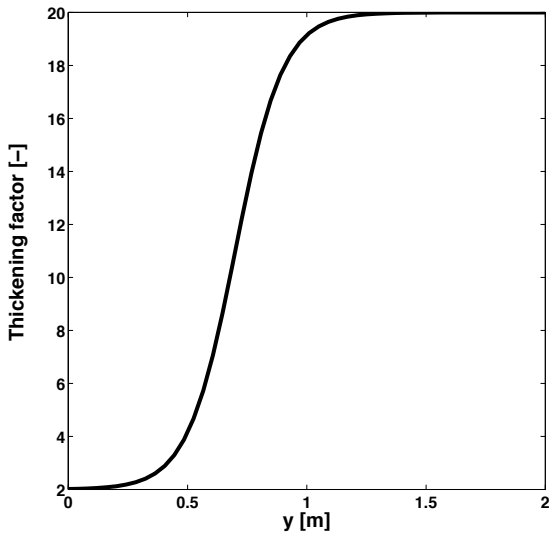


Figure 4.25: *Adaptive thickening factor $\mathcal{F}$: variation of $\mathcal{F}$ along the y-direction (vertical direction normal to the gas/vegetation interface).*

↪ Initial and boundary conditions

As shown in Fig. 4.24, boundary conditions are based on the two-dimensional case. The plane (x, z) for $y = 0$ m corresponds to the gas/vegetation coupling interface (WALL-PYROWO) presented in Section 3.4.2, Chapter 3. In addition, the planes (x, y) for $z = 0$ m and $z = 1$ m are set up as periodic boundary conditions in a preliminary step. Initially, the spatial profiles of the gas phase and vegetation are imposed as the two-dimensional solution extended along the z-direction. The radiation profile shown in Fig. 4.14 for the two-dimensional case is also extended along the z-direction.

- ▷ **Gas phase.** Figure 4.26(a) represents the temperature cross-section of the gas mixture, in which the flame height is nearby 2.75 m at the given time (the flame height is particularly high at this time due to an incoming detachment of burnt gas pockets) and in which the temperature reaches a maximum value of 2087 K. Three-dimensional 2000-K iso-temperatures of the flame colored by the mass fraction of CH_4 and by the velocity magnitude are also shown to represent the flame structure in Figs. 4.26(b)-(c). The first iso-contour in Fig. 4.26(b) is in the burning region since the mass fraction of CH_4 is far below 0.10 but still non-zero (near 0.008). The second 2000 K iso-contour corresponds to a zero-value of the mass fraction of CH_4, meaning that the first reaction in 2S-CH4-BFER (methane oxidation) is completed and that this iso-contour is in the burnt gas region. Consistently, the velocity magnitude is increased at the top of the second iso-contour in Fig. 4.26(c) due to buoyancy.

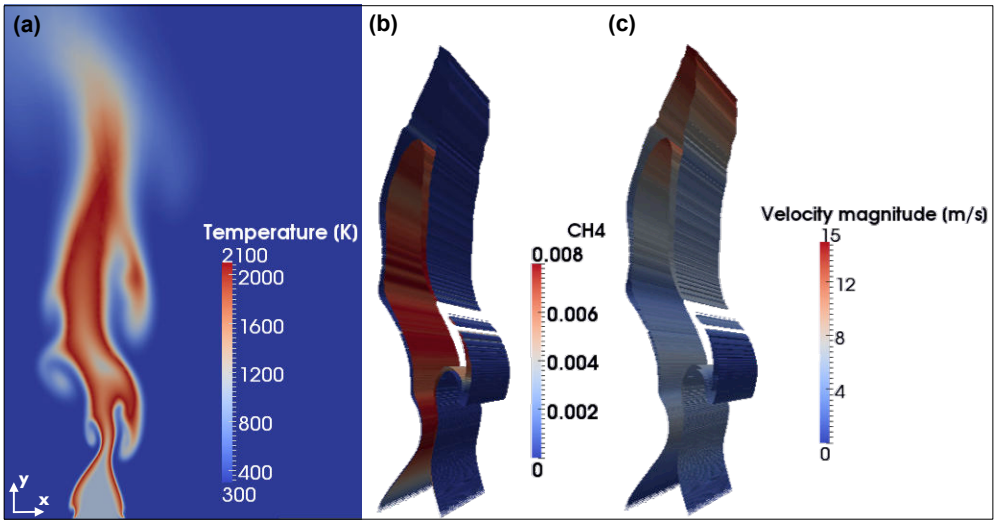


Figure 4.26: *Initial condition of the gas phase for three-dimensional simulations. (a) Cross section of the temperature field at $z = 0.5$ m. (b) 3-D iso-contour of the temperature field (2000 K) colored by the mass fraction of CH_4. (c) 3-D iso-contour of the temperature field (2000 K) colored by the velocity magnitude [m/s].*

▷ **Gas/vegetation interface.** Also, the two-dimensional gas/vegetation interface is represented in Fig. 4.27. The pyrolysis zone illustrated in Fig. 4.24 and centered at $x = 4$ m exhibits a temperature near 680 K in Fig. 4.27(a), corresponding to the zone where the radiation forcing is maximum as shown in Fig. 4.27(b). Consistently, Fig. 4.27(c) highlights that the bulk mass density of the vegetation layer $\rho_v = \beta_v \rho_p$ is significantly decreased in this zone below 10 kg/m^3, while far from the flame its value remains near 25 kg/m^3.

Figure 4.28 presents the different mass contents in the vegetation, in terms of moisture, dry material and char material contents. Figure 4.29 shows the corresponding source terms in the vegetation solid phase (i.e., drying, pyrolysis, char oxidation). The mass fraction of char materials is nearby 1 in the pyrolysis zone, meaning that vegetation is in this zone, at the beginning of the char oxidation stage of its thermal degradation, with a temperature between 600 and 700 K consistently with Fig. 4.29(c). In contrast, Figs. 4.29(a)-(b) indicate that biomass drying and pyrolysis are confined at the boundaries of the pyrolysis zone since moisture is already evaporated and dry materials are already transformed into pyrolysis gases and char materials in this zone. The cross-section of the source term fields at $z = 0.5$ m, representing their spatial distribution along the x-direction, is presented in Fig. 4.30. From a global viewpoint, the mass rate at the gas/vegetation interface is about 0.2 kg/m^2/s in the pyrolysis zone (since $\delta_v = 0.10$ m), leading to a wall-normal injection velocity approximately equal to 0.30 m/s.

Note that this configuration is a preliminary step towards fully three-dimensional simulations, whose aim is to demonstrate the feasibility of multi-physics CFD for studying the flame structure of a propagating fire.

4.3.2 Flame structure

The previously-described three-dimensional case solving for the coupling between AVBP and PYROWO is simulated during 0.135 s. A stable solution is obtained from a numerical viewpoint. In this preliminary step towards a fully multi-physics three-dimensional simulation, Fig. 4.31 already shows that the flame structure obtained for the two-dimensional simulation (see Fig. 4.20) is consistent with the present three-dimensional simulation. The scatter plot colored by the Takeno index shows that the maximum value for the gas temperature is obtained at the stoichiometric point $z_{st} = 0.32$ and that the flame is still of the diffusion type; the burnt gas temperature is more than 100 K higher than in the two-dimensional case and thereby closer to the equilibrium temperature.

Simulating such a physical time (0.135 s) corresponds to nearly 440,000 iterations of the three-dimensional coupled AVBP/PYROWO solver ($N_{it} = 1$); these iterations require 48 h on 256 processors on the GENCI-CURIE supercomputer (254 processors are devoted to AVBP, 1 processor to PYROWO and 1 processor to the OPENPALM driver, see Appendix B). This simulation is time-consuming and requires further

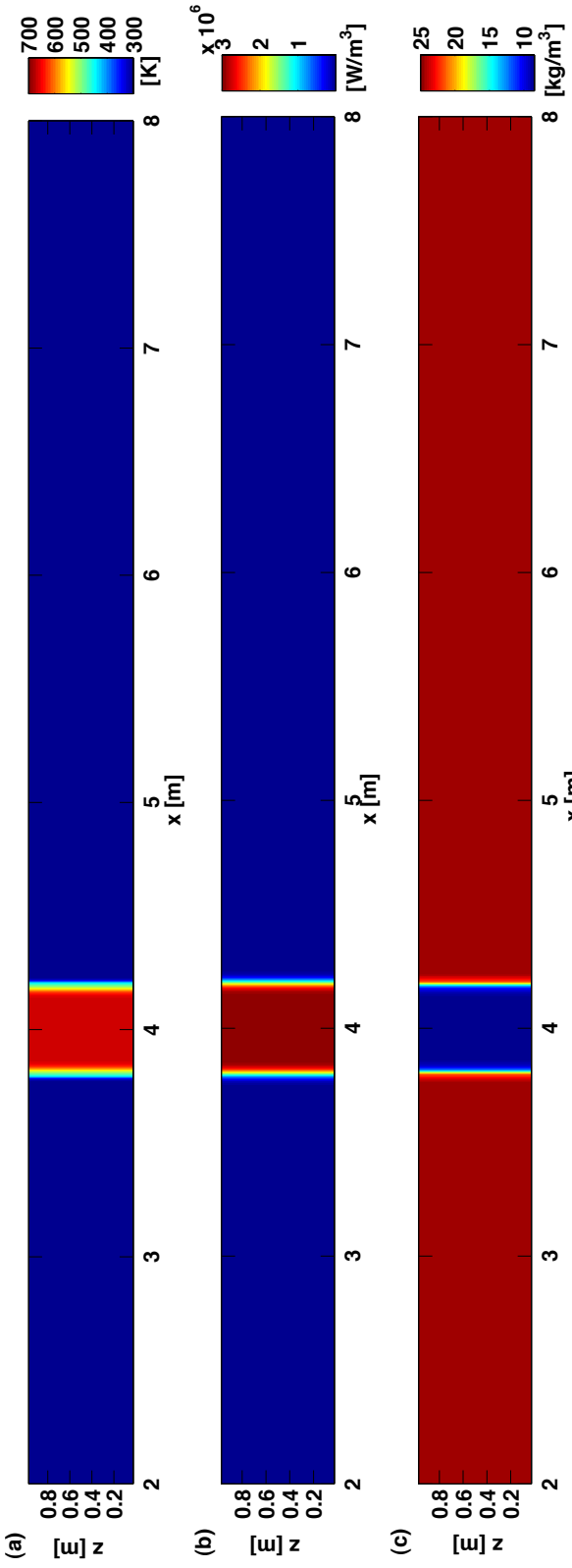


Figure 4.27: *Initial condition of the two-dimensional gas/vegetation interface* Wall-Pyrowo. *(a) Vegetation temperature* T_v $[K]$. *(b) Radiation forcing* $\dot{q}_{rad,fl}$ $[W/m^3]$. *(c) Bulk mass density of the vegetation* ρ_v $[kg/m^3]$.

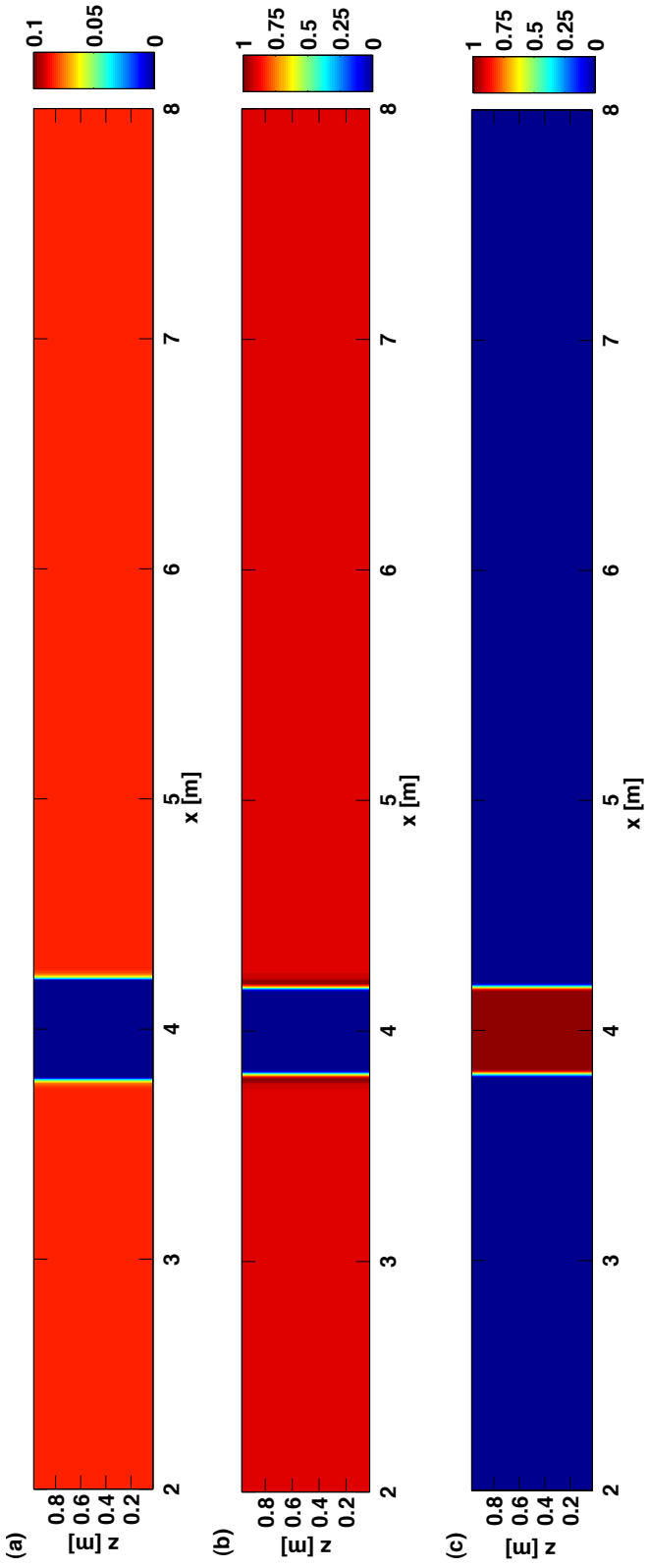


Figure 4.28: *Initial condition of the two-dimensional gas/vegetation interface* Wall-Pyrowo. *(a) Moisture content* $Y_{v,wat}$ $[-]$. *(b) Dry material content* $Y_{v,dry}$ $[-]$. *(c) Char material content* $Y_{v,char}$ $[-]$.

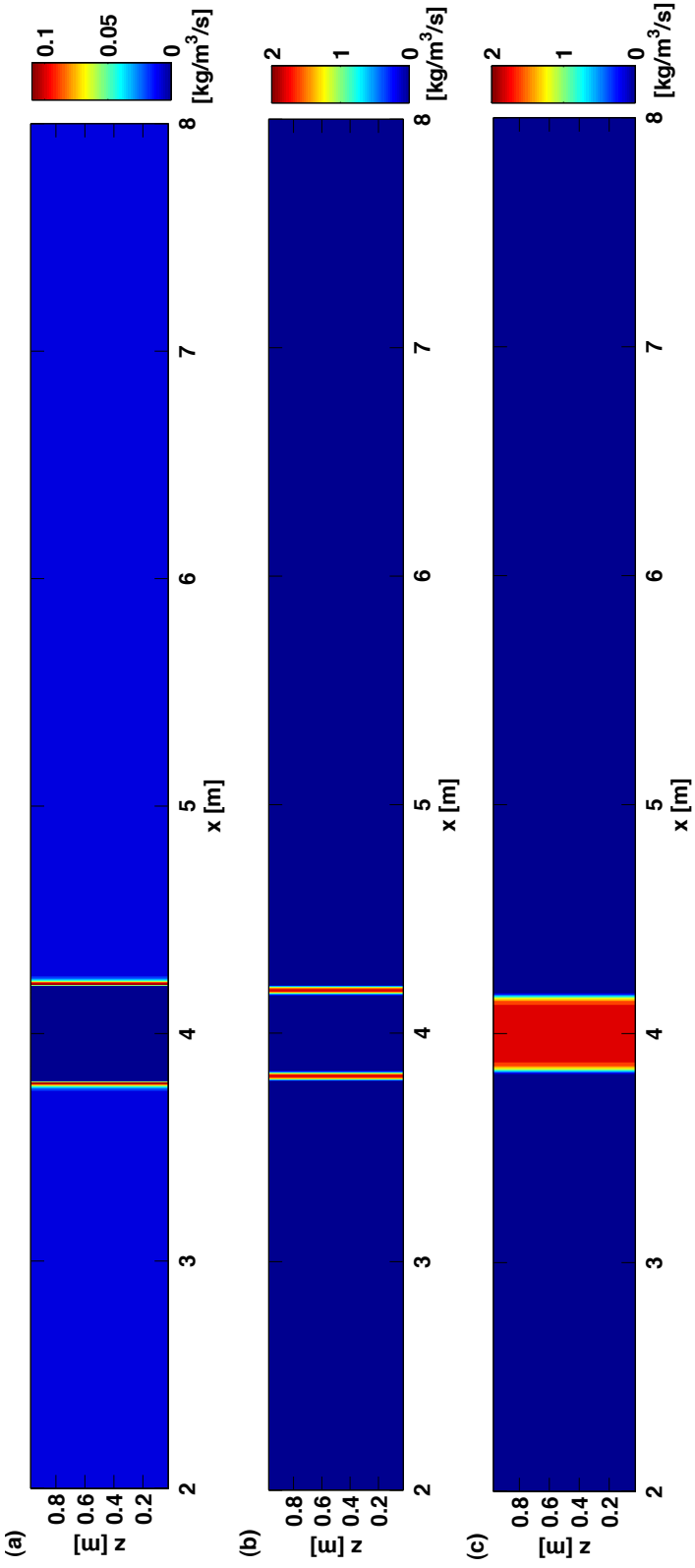


Figure 4.29: *Initial condition of the two-dimensional gas/vegetation interface* Wall-Pyrowo. *(a) Drying mass rate* $\dot{m}_{v,vap}$ $[kg/m^3/s]$. *(b) Pyrolysis mass rate* $\dot{m}_{v,pyr}$ $[kg/m^3/s]$. *(c) Char mass loss rate* $\dot{m}_{v,char}$ $[kg/m^3/s]$.

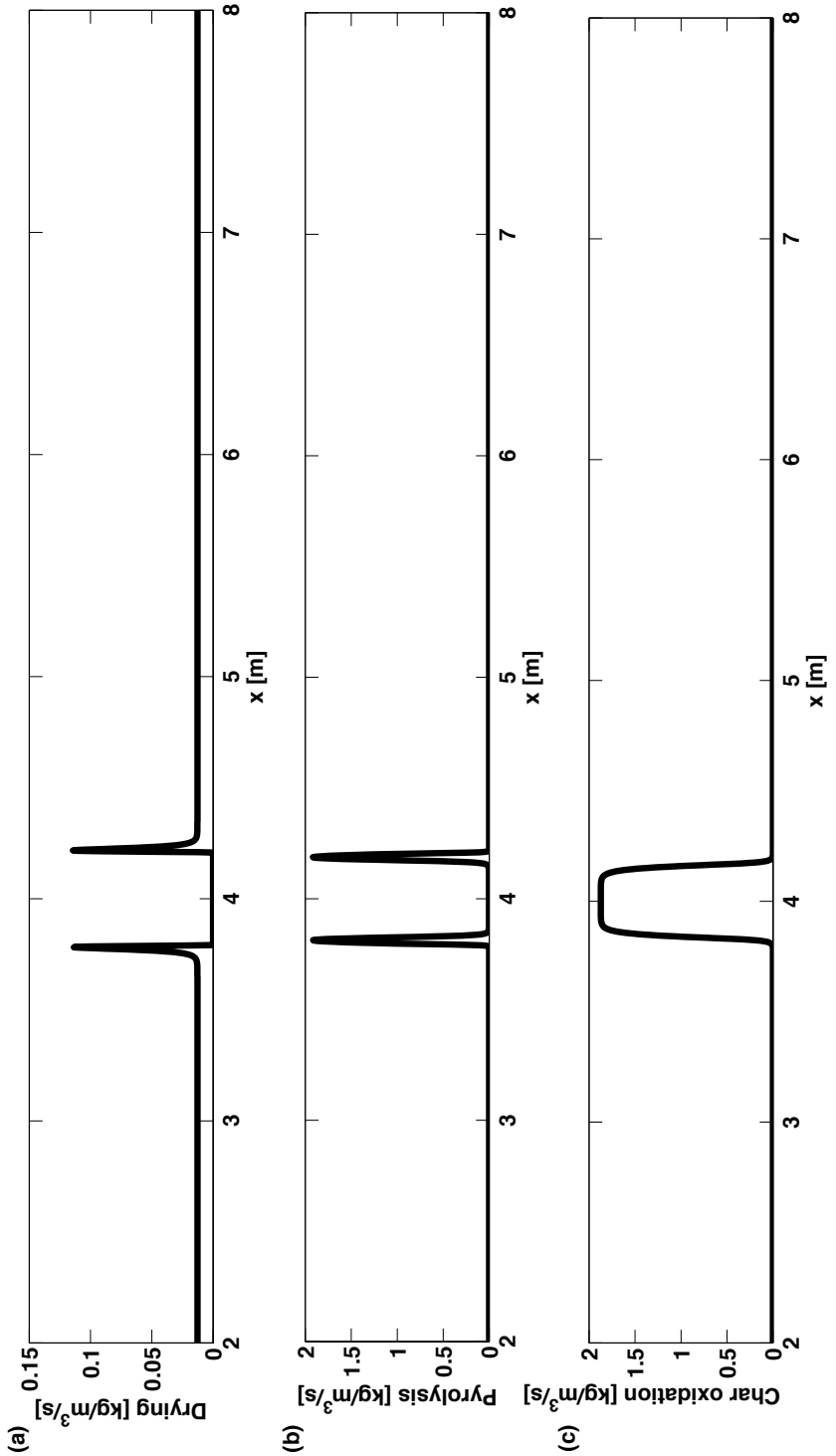


Figure 4.30: *Initial condition of the two-dimensional gas/vegetation interface* WALL-PYROWO *- Cross-section at* $z = 0.5$ *m of the two-dimensional fields presented in Fig. 4.29. (a) Drying mass rate* $\dot{m}_{v,vap}$ *[kg/m*^{3}*/s]. (b) Pyrolysis mass rate* $\dot{m}_{v,pyr}$ *[kg/m*^{3}*/s]. (c) Char mass loss rate* $\dot{m}_{v,char}$ *[kg/m*^{3}*/s].*

time-integration to obtain significant changes in flame behavior such as puffing. Figure 4.32 provides insight into the flow field (through the representation of the velocity-based field colored by the gas temperature); air entrainment is induced at the flame basis and on both sides (consistently with PIV measurements) as well as near $y = 0.5$ m (which enhances the detachment of burnt gas pockets that are subject to buoyant vertical acceleration).

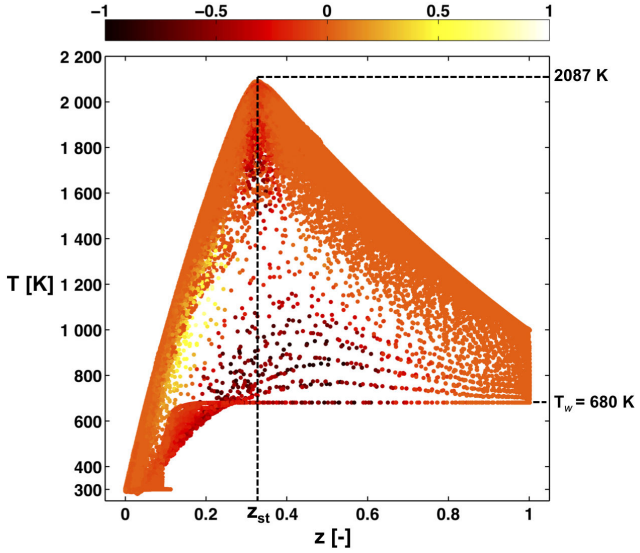


Figure 4.31: *Scatter plot of temperature versus mixture fraction z after simulating 0.135 s, colored by the Takeno index (ranging from -1 to 1); $z = 0$ corresponds to ambient air, and $z = 1$ corresponds to pyrolysis gases released at $T_w = 680\ K$.*

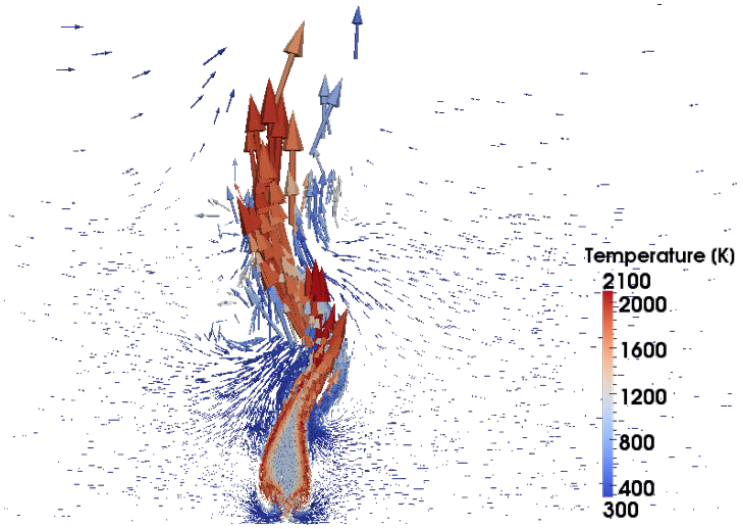


Figure 4.32: *Velocity magnitude in the flow cross-section (at $z = 0.5\ m$), with arrows colored by the gas temperature [K], showing air entrainment at the basis of the flame.*

4.3.3 Quantification of radiation heat transfer

A three-dimensional fully coupled simulation including AVBP, PYROWO and the DOM-based radiation solver PRISSMA was performed following the strategy presented in Section 3.4, Chapter 3. The simulation corresponds to the previously-described three-dimensional case, starting from the same initial condition extended from two-dimensional calculations and with the same boundary conditions. In addition, the emissivity of the gas/vegetation interface is assumed to be equal to 1 (black-body type) in a preliminary step.

Figure 4.33 shows the main variables of interest at the gas/vegetation interface WALL-PYROWO after simulating a physical time of 0.018 s. Even though this computational time is not yet sufficient to study the temporal behavior of the physical system, radiation properties of the three-dimensional flame and the spatial distribution of the heat transfer to the vegetation are investigated.

- ▷ Vegetation reaches the temperature $T_w = 675\ K$ and is subject to a mass loss rate ranging from 0.12 kg/m^2/s in the drying vegetation zone up to 0.19 kg/m^2/s due to pyrolysis and char oxidation, see Figs. 4.33(a)-(b).
- ▷ The flame is anchored at the boundaries of the pyrolysis zone, where pyrolysis gases are released towards the gas phase, and produces a radiation heat flux $\dot{q}_w$ [W/m^2] that can be decomposed into three components presented in Figs. 4.33(c)-(d)-(e), along x-, y- and z-directions, respectively. Additionally, Fig. 4.34 provides the one-dimensional profile of the radiation heat flux for its x- (horizontal) and y- (vertical) components, along the x-direction that corresponds to the direction of fire propagation. The latter gives a more quantitative viewpoint of the spatial distribution of radiation (with respect to the temperature of vegetation) and thereby, of the state of the physical system.
  - It is shown that the x-component of the radiation heat flux features a magnitude that is quasi-symmetric with respect to the flame location (itself marked by the highest values of biomass temperature). This x-component reaches a magnitude of 18 kW/m^2. This value is consistent with experimental measurements obtained from fluxmeters directed in the x-direction and evaluated to 12.3 kW/m^2 in flat configuration (see Table 4.1).
  - Along the y-direction, the radiation heat flux towards the vegetation reaches 50 kW/m^2 ahead of the flame front. The zone over which the radiation heat flux is significant is about 60 cm long on the side of the flame corresponding to increasing x-coordinates. The maximum value of the radiation heat flux is reached within 20 cm of the flame; this zone corresponds to the pre-heated zone that enhances the thermal degradation of unburnt vegetation and leads to the propagation of the fire. Note that the spatial distribution of the y-component of the radiation

heat flux on both sides of the flame is not totally symmetric; this is due to the shape of the buoyant flame similar to that shown in Fig. 4.26.

- In contrast, the radiation heat flux along the z-direction is symmetric due to periodic boundary conditions imposed both for $z = 0$ m and $z = 1$ m in AVBP. However, boundary conditions in PRISSMA do not satisfy this periodicity (there is no external heat flux) and further developments are required to ensure the compatibility of boundary conditions.

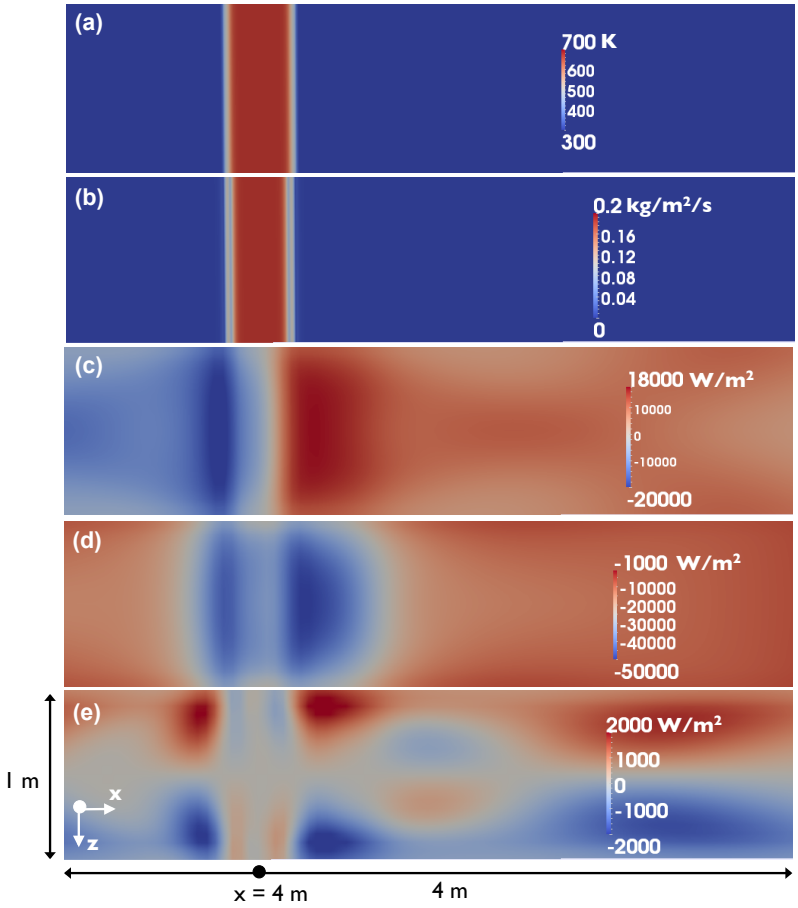


Figure 4.33: *Instantaneous snapshots of the two-dimensional gas/vegetation interface* WALL-PYROWO *colored by: (a) vegetation temperature* T_v *[K]; (b) vegetation mass loss rate* $\dot{s}_w$ *[kg/m*^{2}*/s]; (c)* x*-component of the wall radiation heat flux* $\dot{q}_w$ *[W/m*^{2}*]; (d)* y*-component of the wall radiation heat flux* $\dot{q}_w$ *[W/m*^{2}*]; and (e)* z*-component of the wall radiation heat flux* $\dot{q}_w$ *[W/m*^{2}*].*

Simulating such a physical time (0.018 s) corresponds to nearly 60,000 iterations of the three-dimensional coupled AVBP/PRISSMA/PYROWO solver. The coupling frequency between AVBP/PYROWO and PRISSMA is $N_{it} = 10,000$, corresponding to a physical time of 0.003 s, which is sufficient to capture the motion of the flow field and its impact on radiation (AVBP is limited by acoustic time scales, while the

radiation heat transfer evolves with convection time scales). These iterations require 12 h on 192 processors on the CORAIL supercomputer at CERFACS (190 processors are devoted to AVBP, 1 processor to PYROWO and 1 processor to the OPENPALM driver, see Appendix B). This fully-coupled simulation is computationally expensive and requires further time-integration to study the time-varying behavior of the flame and its consequence on the radiation heat flux at the flame/vegetation interface (in terms of intensity and spatial distribution).

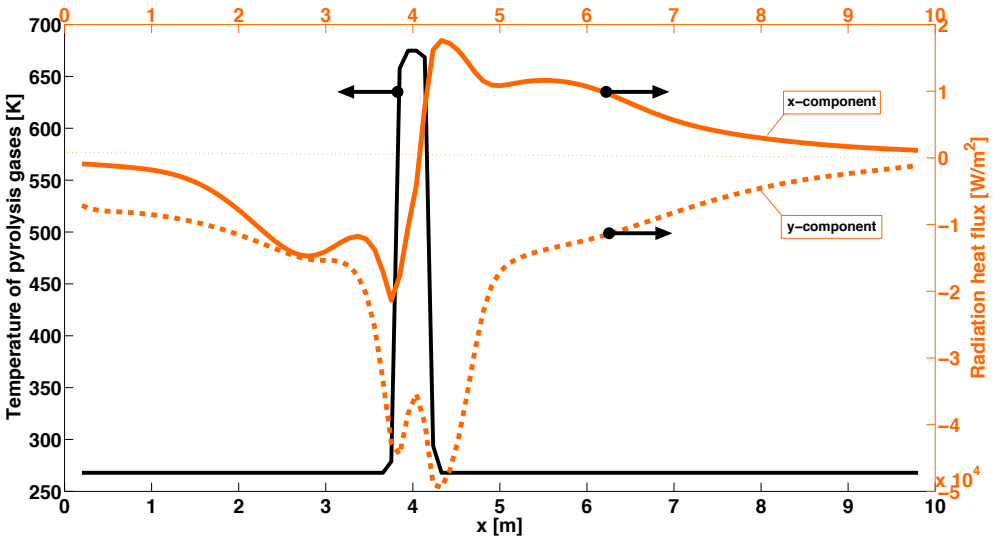


Figure 4.34: *Cross-section of the fields at the two-dimensional gas/vegetation interface ($z = 0.5$ m) presented in Fig. 4.33, corresponding to their spatial distribution along the x-direction. The black solid line represents the vegetation temperature T_v [K]; the orange solid line represents the x-component of the wall radiation heat flux $\dot{q}_w$ [W/m^2], while the orange dashed line represents its y-component.*

4.3.4 Towards comparison to empirical spread-rate modeling

The submodels underlying the Rothermel's semi-empirical model (see Appendix A) are evaluated for the conditions of the laboratory-scale experiment, i.e., $\beta_v = 0.05$, $M_v = 0.10$, $\Sigma_v = 4,000$ 1/m, $\delta_v = 0.10$ m and $\rho_p = 500$ kg/m^3 (corresponding to input parameters of the Rothermel's ROS model). The no-wind ROS given by the Rothermel's model is $\Gamma_0 = 4$ mm/s, which is coherent with the experimental ROS value, 5 mm/s. Note that Γ_0 is formulated as follows in the Rothermel's model:

$$\Gamma_0 = \frac{I_{p,0}}{\rho_b \, \epsilon \, Q_{ig}} = \frac{\chi \, I_r}{\rho_b \, \epsilon \, Q_{ig}}, \tag{4.6}$$

where $I_{p,0}$ [W/m^2] is the propagating heat flux that is expressed as a fraction χ of the reaction intensity I_r [W/m^2], and where $\rho_b \, \epsilon \, Q_{ig}$ [J/m^3] represents the heat required to ignite biomass fuel. In the present case, $\Gamma_0 = 4$ mm/s corresponds to a propagating heat flux $I_{p,0} = 4.4 \times 10^6$ W/m^2 with the release rate of combustion

$I_r = 5.5 \times 10^7$ W/m^2 and the propagating flux ratio $\chi = 0.075$ (a function of the fuel packing ratio β_v and the fuel particle surface-area-to-volume ratio Σ_v). This propagating heat flux $I_{p,0}$ encompasses all heat transfer contributions to the vegetation ahead of the flame front (due to the calibration of this term against wind-tunnel experiments in Rothermel's approach) and is therefore much higher than the radiation heat flux simulated by PRISSMA, equal to 50 kW/m^2 along the vertical direction ahead of the flame front.

In summary, the term $I_{p,0}$ is a global characterization of the heat transfer towards the unburnt vegetation in the Rothermel's semi-empirical ROS model. Multi-physics LES appear as a powerful tool to analyze and dissociate the different contributions to the heat transfer, which lead to biomass fuel pyrolysis and promote combustion. It is worth mentioning that the calculation with PRISSMA only accounts for the radiation in the gas phase and does not yet include the soot volume fraction in the gas phase. The soot contribution is expected to increase the total propagating heat flux towards unburnt vegetation. More physically-consistent simulations of the proposed multi-physics coupled strategy and extensive comparisons with existing simulation capabilities (e.g. WFDS) require to include soot formation and soot oxidation in vegetation and gas phase.

Conclusion

This chapter presents an original development of multi-physics coupled simulations of fire propagation at laboratory flame scale, solving for the flame structure (at the millimeter scale) and including radiation heat transfer as well as biomass pyrolysis. The proposed coupling relies on a multi-model strategy based on high-performance computing that involves:

- ▷ the AVBP large-eddy simulation solver, accounting for reacting multi-species gas mixture, transport and diffusion processes at flame scale: the combustion of pyrolysis gases is modeled using the 2S-CH4-BFER reduced chemical kinetic scheme with pre-adjusted pre-exponential factors to obtain consistent flame temperature and burning velocities;
- ▷ the PRISSMA radiation solver, which relies on a discrete ordinate method and a tabulated spectral model (FS-SNBcK) to account for the radiation of the gas phase (e.g., CO_2, H_2O);
- ▷ the PYROWO biomass thermal degradation solver, developed in this work, to accurately and effectively describe the release of pyrolysis gases from vegetation (CH_4, CO, CO_2, H_2O) when subject to thermal heating.

This multi-physics coupling is handled by the dynamic code coupler OPENPALM, which allows for developing the multi-physics simulation in a message passing interface environment, managing and optimizing computing resources between the different solvers, offering more flexibility in the future development of each model component. Such physics-based strategy is useful to provide a comprehensive understanding of the mechanisms as well as of the characteristic time-scales and length-scales underlying fire propagation. It is also useful to estimate wildfire emissions by quantifying the amount of pollutant emissions such as carbon monoxide, which are produced by the flame zone and released into the atmospheric thermal plume. In summary, coupled multi-physics large-eddy simulations have been found promising to simulate propagating buoyant flames typical of wildfires. Preliminary results were presented in the following conferences:

- ▷ Rochoux, M.C., Cuenot, B., Riber, E., Veynante, D. and Darabiha, N., *Turbulent combustion simulations of a laboratory-scale fire propagation*, in Conference of Numerical Simulation of Forest Fires, from Combustion to Emissions, May 13-18, Cargèse, France.
- ▷ Rochoux, M.C., Cuenot, B., Riber, E., Veynante, D. and Darabiha, N., *LES of natural fire propagation in a lab-scale configuration*, in 14th International Conference of Numerical Combustion, April 8-10, San Antonio.

However, only the feasibility of the proposed approach has been evidenced here and extensive tests are now mandatory to assess its performance. To improve the fine-scale representation of the processes involved in wildfires, future works could include the following steps:

- ▷ **Comparative study of experiments and physics-based simulators.** The proposed multi-physics coupled strategy needs to be further evaluated against experiments and current physics-based simulators such as Wfds and FIRETEC, in order to evaluate the relevance of each model component for given fire spread conditions. The assumptions used to estimate the rate of fire spread based on empirical/semi-empirical modeling could be examined in detail.
- ▷ **Improvement of the pyrolysis model.** Pyrowo is currently limited to a 0-D formulation; it could be extended to a 1-D (vertical) formulation accounting for the radiation vertical profile in vegetation, from top to bottom, due to flame and vegetation contributions. The soot formation during biomass thermal degradation could also be included.
- ▷ **Refinement of the radiation model.** One of the next steps is to include soot oxidation in the gas phase; one challenge lies in the characterization of the soot volume fraction to provide to Prissma. For this purpose, current physics-based models such as Wfds or FireStar could be helpful to provide preliminary estimates of this soot volume fraction. Also, boundary conditions in Prissma need to be carefully studied to properly represent heat losses at the boundaries as well as biomass emissivity, based on comparison to measurements. The influence of the spectral model also needs to be analyzed in detail, when accounting for the radiation of the gas phase and soot.
- ▷ **Refinement of the combustion model.** From the air quality viewpoint, quantifying CO and other pollutants lifetimes remains a challenge for fire modeling and for better predicting wildfire emissions. To better track the formation and oxidation of CO at flame scale, a specific chemical kinetic scheme needs to be integrated in Avbp; 2S-CH4-BFER was only calibrated to properly represent the flame temperature and speed in rich mixtures, but improvements in the representation of the burnt gas composition would be required to properly estimate CO emissions. The need for a subgrid-scale turbulent combustion model adapted for buoyant flows could also be discussed. For these developments, it is imperative to use simultaneously diagnostic techniques and simulations, in order to obtain physically-consistent results.

Note that these developments would not reconsider the coupling strategy based on OpenPalm, which is of great interest to refine the different submodels

underlying the multi-physics simulations, while maintaining the same coupled simulator.

In the long-term, tabulating the quantities of interest derived from flame-scale simulations (e.g., fireline intensity, rate of spread, radiation heat transfer ahead of the flame front) in a look-up table could be an alternative to estimating the rate of spread based on empirical modeling in regional-scale front-tracking simulators.

Part II

Data assimilation for regional-scale wildfire spread forecast

Introduction

The challenges found on the route to developing quantitative wildfire spread models are two-fold. First, there is the classical modeling challenge associated with providing accurate mathematical representations of the multi-physical processes that govern wildfire dynamics (involving biomass pyrolysis, combustion, flow dynamics as well as atmospheric dynamics and chemistry). Second, there is the less common challenge associated with providing accurate estimates of the input data and parameters required by the models. Current fire models are limited in scope because of the large uncertainties associated with the accuracy of physics-based models, because also of the large uncertainties associated with many of the environmental conditions required as input parameters to the fire problem.

A possible approach to overcome the limitations in numerical simulations of wildfires compatible with operational forecast is data assimilation. This approach takes advantage of recent progress made in airborne remote sensing that allows real-time monitoring of the fire front location. Data assimilation consists in combining computer simulation tools with sensor observations, or more precisely in using observations to correct and optimize computer model predictions assuming both sources of information (model and observations) are subject to uncertainties. While still original in fire and combustion research, data assimilation is an established approach in several scientific areas, for instance in the field of numerical weather predictions (Rabier, 2005) or operational oceanography (GODAE, 2009).

Whereas the source of uncertainties in observing data mostly relate to instrumental and representativeness errors, the source of uncertainties in numerical models relate to an incomplete and imperfect knowledge of boundary and initial conditions, an imperfect knowledge of physico-chemical mechanisms as well as an approximate evaluation of a large number of model parameters. These uncertainties in the model inputs translate into uncertainties in the model outputs and thus, do not allow for the development of fully deterministic models. It is therefore necessary to adjust model parameters, initial and boundary conditions in order to provide satisfying simulations and forecasts, as observations of the system become available. Similar issues can be raised in the field of wildfires. For this purpose, we aim at demonstrating, in this thesis, the merits and potential benefits of data assimilation to achieve data-driven predictive simulations of wildfire spread. The prototype data-driven wildfire simulator adopts a regional-scale viewpoint: it treats wildfires as propagating fronts at a rate of spread due to Rothermel's model and assimilates a time-series of observations of the fire front location. The data assimilation algorithm features a choice between a parameter estimation approach in which the control variables (the estimation targets) are the input parameters of the ROS model, and a state estimation approach in which the control variables are the

spatial coordinates of the discretized fire front. The performance of the prototype data-driven wildfire spread simulator is demonstrated in synthetic representative cases including spatially-varying vegetation properties and temporally-varying wind conditions as well as in a real-case study corresponding to a controlled grassland burning experiment.[6]

This PhD-level project is a collaboration between CERFACS (Drs. Sophie Ricci and Bénédicte Cuenot), France, and the Department of Fire Protection Engineering, University of Maryland (Dr. Arnaud Trouvé), USA, with the help of Blaise Delmotte (Delmotte et al., 2011) and Charlotte Emery (Emery et al., 2013) as Master students. Additional aspects were studied through a collaboration with Wellington da Silva (Mines Albi, France, and the Federal University of Rio de Janeiro, Brazil) for the application of particle filters to wildfire spread forecast and through a collaboration with Didier Lucor (Institut d'Alembert, France) for the development of a polynomial chaos strategy to reduce the computational cost of wildfire spread forecast.

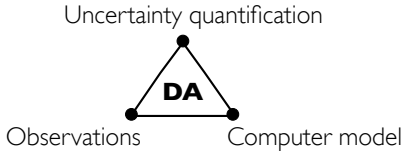
[6]The proposed data assimilation strategy is not applied here to a real-world regional-scale fire spread as Aullène for example (see Fig. 1.24, Chapter 1) due to the required treatment of complex terrain topography (that was only added recently in the FIREFLY simulator, see Appendix G) and to the required post-processing that is heavy for data assimilation, but this is definitely the ultimate goal of this research.

Chapter 5

General features of data assimilation

Data assimilation (DA) is an efficient strategy inherited from estimation theory that incorporates measurements into a computer model of a real system, while accounting for their uncertainties as weight for the resolution of an inverse problem. The objective is to formulate some feedback information to the computer model and to improve the physical understanding and/or the numerical prediction of the behavior of this physical system. Estimation theory provides a broad conceptual framework and a number of algorithms for addressing the scientific challenges and open questions of data assimilation. Still, data assimilation is far from being a simple statistical tool since physical understanding and modern computational mathematics play a large role in actually solving the practical issues of data assimilation, in particular in the domain of geosciences that involves complex, non-linear, multi-scale and multi-physical dynamic models.

In this chapter, the objective is to provide a reasonable starting point for newcomers to the field of data assimilation and its formalism as well as to provide some insight on the similitudes and differences between the classical data assimilation algorithms (e.g., variational approaches, Kalman filter and extensions, particle filters).



Outline

5.1 Data assimilation: an inverse modeling problem

5.1.1 Inverse modeling problems

↪ Definition: The dragon analogy

Consider a forward problem that is cast in the (discrete) form $\mathbf{y} = \mathcal{G}(\mathbf{x})$, with $\mathcal{G}$ the explicit operator that represents the governing equations of the physical problem and that translates the knowledge on the problem noted $\mathbf{x} \in \mathbb{R}^n$ (e.g., input

parameters, initial condition, boundary conditions, model variables) into its measurable outcomes $\mathbf{y} \in \mathbb{R}^p$. Using the dragon analogy taken from Bohren and Huffman (1983), while the *forward problem* consists in, given a dragon $\mathbf{x}$ (the control space), inferring its tracks $\mathbf{y}$ through a mapping function $\mathcal{G}$, the *inverse problem* consists in, given its tracks $\mathbf{y}$ (the observation space), retrieving the dragon characteristics $\mathbf{x}$ that cannot be directly measured. The inverse problem aims therefore at inferring a coherent picture of the dragon characteristics $\mathbf{x}$ given the available incomplete information on its tracks $\mathbf{y}$ and in some sense, at inverting the mapping function $\mathcal{G}$. A general schematic of an inverse modeling problem is given in Fig. 5.1. Note that $\mathcal{G}$ is called the *observation operator* within the data assimilation framework.

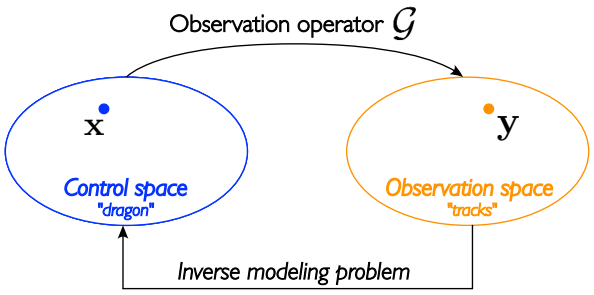


Figure 5.1: *Schematic of an inverse modeling problem.*

↪ Challenges found in inverse modeling problems

The resolution of the inverse problem is often confronted with two major difficulties: (1) the linearity or non-linearity property of the observation operator $\mathcal{G}$ and of the underlying equations; (2) the amount of information available (i.e., the dimension p of the outcomes $\mathbf{y}$) to determine the unknown inputs $\mathbf{x}$ (i.e., the control vector of dimension n), which can induce an ill-posed inverse problem. From a theoretical perspective, an inverse problem is ill-posed in the sense of Hadamard (1902) if one of the three following conditions is not satisfied: existence, unicity and stability of the solution $\mathbf{x}$. In this context, each unknown is considered as an available degree of freedom, while each equation introduced in the system $\mathcal{G}$ acts as a constraint that restricts the range of the possible solutions for the unknown $\mathbf{x}$. Different scenarios can be considered:

- ▷ **Under-determined problems.** When the number of unknowns n is larger than the number of outcomes p included in the observation operator $\mathcal{G}$, multiple solutions exist, violating the principle of unicity.
- ▷ **Over-determined problems.** When the number of outcomes p is larger than the number of unknowns n, there is an inconsistency in the solution as different series of outcomes $\mathbf{y}$ would lead to different values for $\mathbf{x}$. The existence of the solution is not guaranteed. Thus, a selection criterion has to be introduced in the resolution of the inverse problem to determine the

effective physical values for the unknown $\mathbf{x}$.

- ▷ **Threshold case.** When the number of outcomes p and the number of unknowns n are identical, it is possible to define a unique solution for the inverse problem. However, in this case, the inverse problem might suffer from numerical instability when solved with a finite precision or with errors in the measurements. For instance, if two series of measurements are embedded with significantly different errors, then the resolution of the inverse problem may lead to two different solutions $\mathbf{x}$. In this case, the inverse problem solution is not repeatable.

To overcome the issues related to inverse problems, a large body of literature presents strategies to reformulate the inverse problem for numerical treatment. One commonly-used strategy for over-determined and under-determined problems is called *regularization* (e.g., generalized Tikhonov regularization) and consists in considering the prediction of the computer model itself as a source of information on the physical system. Thus, this model prediction can be introduced as an additional term or constraint in the inverse problem formulation to give preference to the most realistic solution. This a priori information is usually referred to as the *background*. See Tikhonov and Arsenin (1977), Tikhonov and Leonov (1998), Tarantola (1987) or Aster et al. (2012) for a more detailed discussion on these aspects.

5.1.2 Key ideas of data assimilation

Data assimilation is the meeting point of physical modeling, uncertainty analysis and mathematical algorithms that aim at forecasting with accuracy the behavior of a physical system by combining observations and predictions of a computer model. Inherited from estimation theory and control theory (Gelb, 1974), it has been mainly developed in the context of geosciences, where neither observations nor computer model, when used alone, can provide a complete and certain description of the real state of a physical system. Data assimilation can therefore be considered as the probabilistic formulation of an inverse problem, where the uncertainties in the computer model and in the measurements are used as weights for the resolution of the inverse problem (Daley, 1991; Ide et al., 1997; Talagrand, 1997; Bouttier and Courtier, 1999; Kalnay, 2003; Bocquet, 2011).

- ▷ **Incompleteness of observations.** Observations (noted $\mathbf{y}^{\mathrm{o}}$ in the following) are usually incomplete and sparse in space and time. They measure a particular phenomenon, without providing information on the physical mechanism (e.g., spaceborne spectral radiance measurement). In that case, geophysical properties are only inferred through physical modeling (e.g., retrieval of surface temperature or aerosol properties). When observations are available, their sources of uncertainties mostly relate to instrumental errors (i.e., errors in raw measurements) as well as representativeness errors (i.e., inconsistency between what the instrument can measure and what the computer model can represent).

▷ **Limitations of physical modeling.** Unlike observations, the computer model (also referred to as the *forward model*) describes the temporal evolution of a physical system through a set of physical laws (e.g., empirical formulation, system of partial differential equations). It can provide a highly-resolved description of the quantities of interest (e.g., temperature, pressure, velocity fields in fluid-mechanics problems), which is useful to improve the understanding of the physics-related processes (re-analysis mode) and/or to estimate future behavior of the physical system (forecast mode). However, a computer model only provides an approximate solution of the evolution of the physical system, because of an incomplete knowledge of boundary and initial conditions as well as an approximate description of a large number of parameters (free or model parameters, commonly referred to as *input parameters*). These uncertainties in the model inputs translate into uncertainties in the model outcomes and thus, do not allow for the development of fully deterministic and generic forward models. In addition to these errors, the simplification and discretization of the physics and the use of numerical methods are also significant sources of uncertainties in a computer model; these numerical errors can even accumulate over time. The knowledge of these uncertainties and their quantification is therefore important for the development of any physical analysis based on an uncertain model (this has motivated the development of uncertainty quantification in fluid mechanics for instance, Lucor et al., 2007). In this context, the uncertain variables in the computer model are described in the form of a probability distribution (instead of a unique deterministic value), meaning that the computer model is considered as stochastic. Thus, the computer model provides some background information in the form of a prior estimate, useful to initiate the estimation process.

Given the multiple sources of uncertainties that inevitably exist in both measurements and modeling, the purpose of data assimilation is to combine, with the proper weighting, all these different sources of information on a physical system to formulate some feedback information on the system and thus, to adjust input parameters, initial and boundary conditions and/or model state variables (generally referred to as *control variables* $\mathbf{x}$) as shown in Fig. 5.2. It follows that data assimilation is a particular type of inverse modeling problem with a probabilistic dimension; the unknown quantities being represented as random variables that are associated with a probability density function (PDF). In this sense, data assimilation aims at determining the best (optimal) estimate of the unknown $\mathbf{x}$ (called analysis or posterior) using estimated statistical errors for both the prior estimate of $\mathbf{x}$ (indifferently called background, forecast or prior) and the observations $\mathbf{y}^{\mathrm{o}}$. It follows that, within the data assimilation framework, the optimality criterion is based on maximum-likelihood estimation for the PDF associated with the control vector $\mathbf{x}$. Note that this is equivalent to the variance error minimization of the analysis solution when the error statistics are Gaussian as shown further in this

manuscript. One crucial question to be addressed is which statistics of the errors are the targets of the optimization process (moments of the PDF or the full PDF).

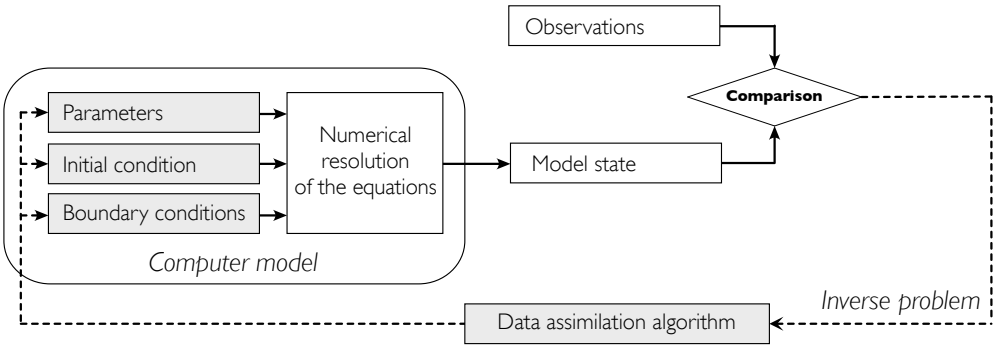


Figure 5.2: *General framework of data assimilation.*

This data assimilation approach fits into the wider domain of inverse problems and dynamic data-driven application systems, where data are used to derive more accurate and reliable simulations for improved predictions of the dynamics of a complex physical system.

5.1.3 Genesis of data assimilation

↪ A look back in history

The ideas underlying data assimilation appeared at the end of the eighteenth century thanks to the efforts of astronomers to predict the motion of our solar system's planets, moons and asteroids with a limited handful of observations. The young mathematician Karl Friedrich Gauss (1777-1855) successfully predicted in 1801 where the newly-discovered asteroid Ceres might be found after reappearing from behind the Sun, without solving the Kepler's non-linear equations of planetary motions, but using a very limited set of imperfect observations to compute the asteroid orbit. At this early stage, he already introduced the idea that the model can be used as a prior estimate and that the correction has to fit the observations within their presumed observation errors. This work at the origin of estimation theory was closely related to the principle of least-squares published in 1805 by Adrien-Marie Legendre (1752-1834) at that time.

> "If the astronomical observations and other quantities on which the computations of orbits is based, were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to actually take place exactly according to the laws of Kepler), and therefore, if other observations were used, they might be confirmed, but not corrected. But since our measurements and observations are nothing more than approximations to truth, the same must be true of all

> calculations resting upon them, and the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly be undertaken when an approximate knowledge of the orbit has been already taken into account, which is afterwards to be corrected, so as to satisfy all the observations in the most accurate manner possible."
>
> K.F. Gauss, *Theoria Motus Corporum Coelestium* (1809), English translation.

Due to the fast development of digital computers, the field of estimation theory began to mature in the 1960s and 1970s. In particular, the concept of maximum likelihood by Fisher (1890-1962) and of sequential optimal filtering by Rudolf Emil Kalman (1930-$\cdot$) were introduced. The resulting Kalman filter (KF) is particularly suited for sequential estimation in the case of a linear dynamic model with Gaussian error statistics (Kalman, 1960; Kalman and Bucy, 1961). Its predicting capability was demonstrated during the Apollo program (between 1961 and 1972) by sequentially estimating the evolving position of the Apollo module towards the Moon. A scheme of the technique is presented in Fig. 5.3: the KF algorithm generated a more realistic estimate of the module position (the analysis) using available measurements and then, used this new estimate as a prior information to forecast the trajectory of the Apollo module at future lead-times. This combination of estimation and forecast constitutes an assimilation time window and was sequentially applied as new observations became available, thus tracking the position of the Apollo module until it reached the Moon.

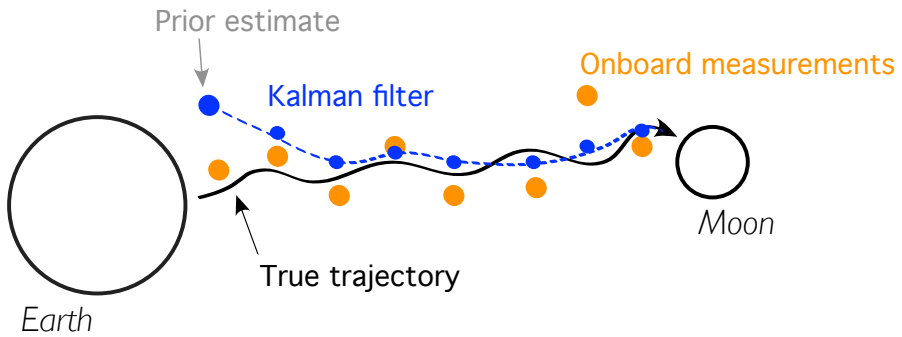


Figure 5.3: *Sequential estimation of the position of the Apollo module using the KF algorithm: at a given time, the KF compares model predictions using Newton's laws (blue dashed line) with onboard measurements (orange dots) to generate a better estimate of the module true position (black solid line).*

↪ Example of numerical weather prediction

Historically, weather forecast has been a pioneer in the operational application of data assimilation. The dynamics of the atmosphere is governed by a prognostic model based on the fundamental laws of fluid mechanics such as mass, momentum or energy balance equations; subgrid-scale processes are parameterized and are therefore embedded with uncertainties. The need for a reliable numerical weather prediction system and for forecasting natural disasters (e.g., air quality prediction, tropical cyclone forecasting) has motivated the development of advanced data assimilation techniques that incorporate the large amount of available observations (e.g., in-situ sensors, radiosondes, weather satellites, airborne sensors), which inequally cover the Earth surface. In this context, the objective of data assimilation techniques is to estimate the current weather (in terms of temperature or pressure) and subsequently, to use this estimation as the initial condition of the computer model for the next assimilation cycle.

Over the past decades, different data assimilation algorithms have been developed and their performance have been evaluated for weather forecasting. First attempts of using measurements to constrain the computer model predictions were based on simple interpolation approaches called *objective analyses* (in contrast to the subjective analyses based on the expertise of operational meteorologists), see Bergthorsson and Döös (1955). For instance, the Cressman analysis scheme (Cressman, 1959; Lorenc, 1986; Daley, 1991; Bouttier and Courtier, 1999) constitutes a basic algorithm of sequential objective analyses, in which the model state (represented as grid-point values) is set equal to the observed values in the vicinity of the observation locations. As shown in Fig. 5.4, this correction of the prior model state decreases while the distance between the simulated grid-point and the observations becomes larger. This decrease is controlled by a weight function that is parameterized with a user-defined constant called the *influence radius*, beyond which the observations have no contribution to the analysis model state. The closer the observation, the larger its weight in the analysis. In particular, this weight is equal to one if the grid-point is collocated with the observation location.

In practice, the Cressman method and its variants have severe drawbacks: (1) there is no direct method to specify the optimal weight function; (2) the precision associated with the background information and the observations is not accounted for; for instance, assimilating poor-quality measurements would degrade a good-quality estimation of the computer model; and (3) the analysis solution is not guaranteed to respect the properties of the true system when integrating further the computer model (e.g., smoothness of the fields, balance relation between the different model state variables), in which cases unobserved model variables may be significantly degraded by unphysical features.

The breakthrough of statistical estimation was achieved by Gandin (1963) with the development of the optimal interpolation (OI) technique. This technique relies on the idea that the weights given to the observations can be calculated according to

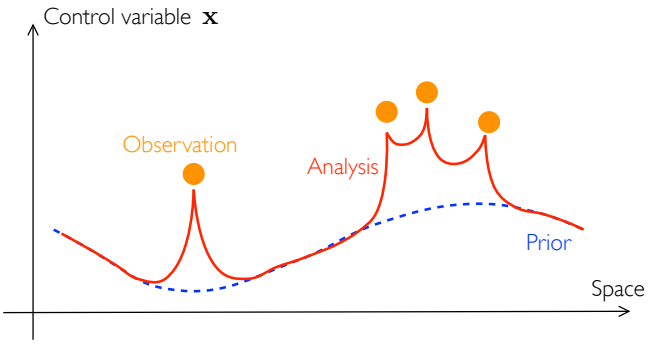


Figure 5.4: *Example of Cressman analysis for a 1-D spatially-distributed control variable* $\mathbf{x}$ *at a given time: the blue dashed curve corresponds to the prior field, orange dots correspond to observations, and the red solid curve corresponds to the analysis. Credit: Bouttier and Courtier (1999).*

the error statistics of both the prior information (i.e., a climatological information as OI considers time-invariant, homogeneous, isotropic errors) and the observations. Within this framework, error statistics are represented with the first and second moments of the PDF (mean and covariance) associated with the prior and observation errors (the complete knowledge of the PDF is not required by the OI methodology); they are numerically implemented by the means of error covariance matrices. Thus, the weights given to the observations are chosen to minimize the error variance on the analysis estimate, similarly to the best linear unbiased estimator (BLUE). In a state estimation problem, a specificity of the OI technique is that it only assimilates a selection of observations to derive the analysis for each control variable at each grid-point. The OI only considers the observations in the vicinity of this grid-point and thereby, the analysis is computed block-by-block over the computational domain and at a lower computational cost than the BLUE (Lorenc, 1981; Massart, 2003; Ricci, 2004). Within this framework, poor-quality observations have a very low weight and thus, do not have a significant impact on the OI-based analysis solution. Note that even though this technique is widely used due to its easy implementation, it is not designed for a highly non-linear observation operator $\mathcal{G}$ and in particular for tracking extreme meteorological events.

While proposed in the 1960s, the application of the KF in the numerical weather prediction framework was first investigated in the 1980s (Ghil et al., 1981). In the KF framework, the weights associated with the uncertainties in the background information and in the observations are expressed as a Kalman gain matrix. While similar to the OI technique for the computation of the analysis solution, the KF technique additionally propagates the error statistics (through the propagation of the error covariances via the dynamic numerical model) from the previous analysis time to the current forecast time. Thus, the KF is a sequential data assimilation technique that allows for dynamic error covariances (represented through error covariance matrices). However, when the computer model is non-linear, the equations need to be locally linearized for the propagation of the error statistics; thus,

the non-linear treatment implies the use of tangent-linear operators. If the non-linearities are weak (Jazwinski, 1970; Bouttier and Courtier, 1999; Evensen, 2007), the KF technique remains efficient, even though it is no longer optimal. This is called the extended Kalman filter (EKF). However, due to the high-dimensionality of meteorological fields, the application of the KF and its extension for numerical weather prediction is a complex and very time-consuming task. For this purpose, reduced-rank methodologies (Buehner and Malanotte-Rizzoli, 2003) that reduce the rank of the error covariance matrices were developed; one example is the singular evolutive extended Kalman (SEEK) filter introduced by Pham et al. (1998). One alternative to overcome the issues related to error propagation is to dynamically estimate the error covariances using a statistical sample (also called an *ensemble*). The idea is to stochastically represent the PDF of the control variables using an ensemble of realizations (based on Monte Carlo sampling techniques for instance) and thus, to replace the error covariance matrices by their sample counterparts. In this context of ensemble-based techniques, the non-linear model operator is applied to perturbations of the model state estimated at the previous analysis time (these perturbations are based on the estimated uncertainties over the forecast time period), thus leading to different realizations of the forecast (prior) errors. This leads to a prior information expressed as an ensemble of possible trajectories at the forecast time, which are directly used to model the error covariances and to stochastically formulate the Kalman gain matrix. This methodology due to Evensen (1994) is known as the ensemble Kalman filter (EnKF), which is the main data assimilation technique applied to wildfire spread forecast in this thesis.

The growing number of real-time meteorological measurements combined to the growing complexity of atmospheric computer models has required the development of time-efficient data assimilation techniques. In the 1960s, Sasaki (1958, 1970) proposed a variational approach of the estimation problem based on the minimization of a cost function, with an application in tracking hurricane trajectories. This cost function (also called objective function, penalty function or misfit) is constructed from the discrepancies between the measurements and the model predictions on the one hand, and from the control deviation from the prior atmospheric state on the other hand. The latter is a penality (regularization) term that constrains the analysis solution by the computer model dynamics. In particular, it transfers information from data-rich areas to data-sparse areas through the time-integration of the computer model (the model state summarizes in an organised way the information from earlier observations). This variational approach applied at a single observation time is called three-dimensional variational DA algorithm (3D-Var) and is still widely used in operational weather forecasting centers (Parrish and Derber, 1992; Fisher, 1998; Gauthier et al., 1999). In contrast, if the cost function contains measurements at several different times within an assimilation time interval and if the minimum of this cost function is sought for this interval (by varying the model initial condition for instance), the technique is named four-dimensional variational DA algorithm (4D-Var), where 4D refers to the three spatial dimensions

plus the time dimension. 4D-Var includes dynamic features such as the propagation of the model to the exact time of the observation as well as the evolution of the forecast error covariance along the assimilation time window. It is worth noting that the error covariance evolution is sometimes referred to as *implicit* because the assimilation is performed without explicitly computing the full error covariance matrix. The 4D-Var technique is thus more flow-dependent than 3D-Var and the quality of the estimates improves. Note that Le Dimet and Talagrand (1986) proposed iteratives techniques to solve non-linear cost function minimizations and showed in particular how the adjoint model is useful to limit their computatonal cost. A widely-used algorithm in this context is the Gauss-Newton method, known under the name of incremental 4D-Var (Courtier et al., 1994; Trémolet, 2007a). The incremental 4D-Var technique formulates the non-linear inverse problem as a sequence of quadratic minimization problems, in which case the minimum (i.e., the analysis solution) is always unique.

Currently, major weather prediction centers, such as the European Centre for Medium-Range Weather Forecasts (ECMWF), Météo-France, and the US National Center for Environmental Prediction (NCEP), produce a medium-range global weather forecast using an incremental 4D-Var algorithm, typically each six hours (Rabier, 2005; Gauthier et al., 2007; Laroche et al., 2007). This time interval that encompasses the observations available for the last six hours is the assimilation window. Thus, the operational objective is to retrieve the state of the atmosphere at the start of this six-hour time window, to reconstruct the model trajectory consistently with the observations over the whole time window and to forecast the model behavior at future lead-times. One operational issue concerns the quality of the forecast, gauged by the number of days for which the forecast is considered accurate. While the quality of the forecast has improved steadily in the recent decades, partly due to the increased model resolution, the growing number of available meteorological measurements and to the development of adapted data assimilation techniques (based notably on ensemble forecasting), one limitation in the forecast skill of numerical weather prediction systems lies in the chaotic nature of the weather behavior (i.e., aleatoric uncertainties), meaning that data assimilation has to be sequentially applied to avoid inevitable deviation from the reality.

The benefit of data assimilation has already been greatly demonstrated in meteorology over the past decades, especially for providing initial conditions for numerical forecast. Beyond weather forecast, applications of data assimilation arise in many fields, e.g., oceanography (Ricci, 2004; Weaver et al., 2005; Daget, 2008; GODAE, 2009; Mirouze, 2010; Gürol, 2013), hydrodynamics and hydrology (Moradkhani et al., 2005; Durand et al., 2008; Ricci et al., 2011; Harader et al., 2012), atmospheric chemistry (Massart, 2003), oil reservoir modeling (Oliver and Chen, 2011), biomechanics (Moireau, 2008; Beltrán, 2012). Note that engineering applications based on modelling (e.g., combustion and heat transfer) are also making a growing use of data assimilation techniques with the increase of quantitative information on the physical system (Suzuki, 2012; da Silva et al., 2011; Orlande et al., 2012).

5.1.4 Practical issues of data assimilation

In general, data assimilation appears as a powerful framework to directly improve predictive simulation results through improved initial and/or boundary conditions, or improve physical models when applied to parameters. Data assimilation can also be used to dynamically optimize an observation network (Kalnay, 2003; Bocquet, 2011). The common idea underlying all data assimilation techniques briefly described above is that they can be regarded as a generalized least-squares problem, in which each source of information (i.e., background model estimate and measurements) is weighted by its associated error statistics and in which the analysis estimate has to fit to the observed data and to the background estimate within their presumed errors. This implies that the analysis solution minimizes the sum of square errors between the observed data and the model predictions.

While general theories exist for non-linear data assimilation with non-Gaussian error statistics, most practical data assimilation techniques in geosciences rely on linear theory and assume Gaussian error distributions. Earth science systems are indeed highly-complex and exceedingly large systems with complicated error structures, which prevent the application of classical advanced optimization methodologies and which require efficient techniques for operational applications. Depending on the minimization strategy (directly or iteratively), two families of data assimilation techniques can be distinguished: filtering (e.g., OI, KF and extensions) on the one hand, variational algorithms (e.g., 3D-Var, 4D-Var and incremental counterparts) on the other hand. Both families are based on the minimization of a cost function that describes the discrepancies between the simulated and observed values as well as their associated error statistics. There exists one fundamental difference: while filtering techniques explicitly solve for the analysis using linear algebra (which is time-consuming and only affordable for an inverse problem with a limited size), variational approaches require the use of a minimizer and are therefore more adapted for high-dimensional problems. Note that they produce equivalent analysis results (under the assumption of linear models) from a theoretical perspective, but their implementation can be significantly different.

It is worth mentioning that no particular technique has been identified as the ideal choice for the resolution of a general data assimilation inverse problem (Lorenc, 1986). The choice of the technique highly depends on the features of the physical system and in this sense, data assimilation is far from being a simple statistical tool; it cannot be dissociated from a profound physical understanding of the system and of the sources of uncertainties. These uncertainties are two-fold:

(i) **epistemic uncertainty** expressing an imperfect knowledge of the key variables of the dynamic model (e.g., initial conditions, input parameters, boundary conditions), which could in theory be removed through experiments, improved numerical approximations and/or higher fidelity physics modeling;

(ii) **aleatoric uncertainty** resulting from inherent and unpredictable stochastic

variability of the physical system, which can be probabilistically characterized through a set of random variables along with their PDF.

For instance, as highlighted by Reichle (2008), atmospheric and ocean data assimilations focus on the estimation of the initial condition due to the chaotic dynamics, while land surface assimilation deals with uncertain meteorological forcing conditions and model parameterizations. Thus, the choice of estimation targets that are representative of the system uncertainties is essential. Furthermore, the ingredients for a successful analysis are listed below.

- **Dimension of the physical problem.** The resolution of physical processes at more and more detailed scales significantly increases the number of grid-points at which the model state variables are considered and thus, significantly increases the number of unknowns in a state estimation problem. The implementation of effective data assimilation algorithms to perform the analysis update is therefore essential within operational frameworks.
- **Good-quality prior estimate.** An estimation process that starts with good-quality prior information better represents the correlation between the control variables and makes the analysis update more efficient and robust, in particular for non-linear extensions of the data assimilation algorithm (e.g., EKF, incremental 4D-Var) that rely on linearization techniques in the vicinity of the prior.
- **Balance constraints.** The analysis solution has to respect the known physical features of the physical system; otherwise, the physics of the problem cannot be correctly represented. This implies that the constraint imposed by the dynamic model has to be accounted for in the inverse problem, for instance in the prior error covariance matrix if they are not implicitly taken care of in the algorithm (Weaver et al., 2005).
- **Smoothness of the solution.** The analysis solution has to be smooth, because the true state in fluid-mechanics problems is. Thus, data assimilation algorithms have to ensure that when going away from an observation, the analysis will relax smoothly to the model prediction on scales that are close to the physical problem length-scales. This issue could arise if the data assimilation correction (i.e., the analysis increment) is added to the dynamic model trajectory in a too abrupt manner. This possibly leads to strong discontinuities at analysis times. One solution to overcome this issue is to apply incremental analysis updates (IAU), meaning that the correction is spread over time, instead of being applied at a single time step. This IAU strategy therefore reduces the shocks of data assimilation on the dynamic model and on the unobserved variables, and gradually forces the model integration throughout the assimilation window.
- **Quality control of the measurements.** Within an operational framework, it is essential to develop procedures that remove particularly poor-quality

measurements to avoid degrading the prior information provided by the computer model and/or measurements that the computer model cannot represent. Along with on-line quality control routines, these procedures are crucial to avoid the failure of the data assimilation procedure.

- ▷ **Bias correction.** Unbiased errors are a key assumption in all data assimilation techniques, meaning that on average the model estimates and the observations must agree with the true control variables. In practice, however, it is extremely difficult to provide bias estimates and to identify the sources of the bias (Fertig et al., 2009). If this identification is possible, a bias in the observations can be removed prior to assimilation, while a bias in the model can be accounted for including model parameters in the control vector.
- ▷ **Non-linearities and non-gaussianity of the error statistics.** While most data assimilation algorithms rely on the assumptions of linear models and Gaussian error statistics, the computer model and the observation operator are usually non-linear, meaning that the relationship between what is estimated and what is assimilated noted $\mathcal{G}$ is non-linear. If these non-linearities are limited, they can be approximated using a linearization technique (e.g., EKF, incremental 4D-Var) or ensemble techniques (e.g., EnKF). However, if these non-linearities are severe, linearization techniques and/or assumptions of Gaussian error statistics are no longer valid. Alternative data assimilation algorithms (e.g., particle filters) need to be considered. Accounting for non-Gaussian structures and non-linearities at a computational cost that is consistent with operational framework is one of the current challenges of data assimilation.
- ▷ **Validation.** It is essential to evaluate the quality of the analysis solution provided by a data assimilation algorithm given the multiple assumptions about non-linearities and error structures. However, this remains a difficult task since the true trajectory of the system is completely unknown and therefore, the precision of the solution becomes difficult to evaluate. Still, consistency diagnostics can be developed (Talagrand, 1997; Chapnik et al., 2004; Desroziers et al., 2005; Chapnik et al., 2006), in particular in the observation space (Desroziers et al., 2005) based on a careful analysis of the innovations (defined as the discrepancies between the model estimates and the observations) and on comparisons with independent data sets (corresponding to data that have not been assimilated).

A schematic of the differences between a deterministic 3D-Var approach and the ensemble-based EnKF technique is presented in Fig. 5.5. They are representative of two different viewpoints, deterministic versus stochastic.

- ▷ **Deterministic viewpoint.** One can view variational approaches as a purely deterministic problem of error minimization, in which the error covariance matrices do not have a probabilistic interpretation and in which the minimization (under the assumption of a perfect dynamic model) can be regarded as a

least-squares curve-fitting of a deterministic model trajectory to the observed data. The analysis is then used to restart the model and to produce a new prediction until next observation time, see Fig. 5.5(a). This is a fundamental difference with ensemble-based filtering techniques such as the EnKF,

- ▷ **Stochastic viewpoint.** As shown in Fig. 5.5(b), an ensemble of model predictions is generated based on estimated modeling uncertainties in the EnKF. The prior error covariance matrix is then constructed based on error statistics defined by the scatter in the ensemble trajectories over the assimilation cycle (corresponding to the time period between two successive observation times). This is useful to stochastically characterize the Kalman gain matrix and to produce more accurate estimates of the control variables at observation times: each member of the ensemble is corrected consistently with the observations and thereby, the scatter of the analysis estimates is reduced. These more accurate estimates of the control variables are used to produce new model predictions beyond the current time.

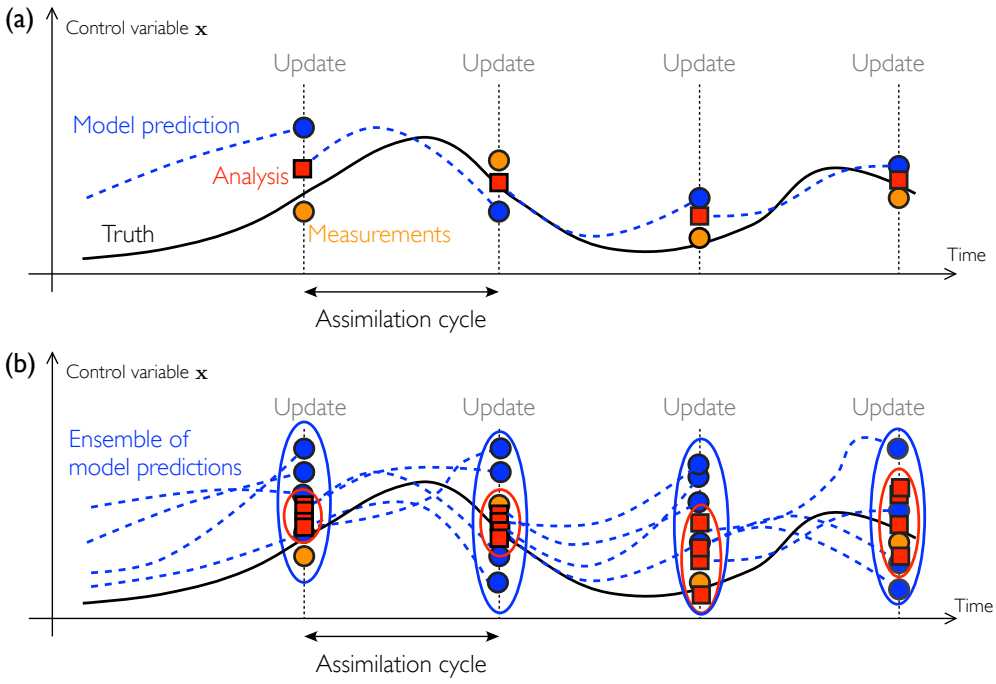


Figure 5.5: *Schematic of sequential data assimilation methodologies with successive prediction/update steps: the true trajectory is represented using a black solid curve; prior model predictions are represented using a blue dashed curve with blue dots at observation times; measurements are represented using orange dots; and analyses are represented using red-squares at observation times. (a) Deterministic viewpoint with error covariance modeling (e.g., 3D-Var). (b) Stochastic viewpoint (e.g., EnKF), for which the uncertainties in the control variables are represented using a statistical sample (ensemble) of model trajectories.*

Instead of comparing the different data assimilation techniques, Gustafsson (2007) suggested to find hybrid strategies to combine ideas of variational and EnKF approaches. The study of hybrid variational-EnKF algorithms is currently a major research topic in the weather forecast field (Liu et al., 2008; Buehner et al., 2010b); they have already shown great potential in a near-operational framework as demonstrated by Buehner et al. (2010a) for the Ensemble-4D-Var technique developed at Environment Canada.[1]

5.2 Stochastic models and data assimilation variables

Mathematical quantities required by the data assimilation framework to describe the behavior of the physical system over the assimilation window, also called *assimilation cycle* and shown in Fig. 5.6, are introduced here. The definition of the assimilation window depends on the data assimilation methodology. Since we focus primarily on sequential filtering techniques in this work, $[t-1,t]$ is referred to as the assimilation cycle (unless mentioned otherwise). While possibly including multiple time steps of the dynamic model, $[t-1,t]$ represents the time-period between two successive observation times.

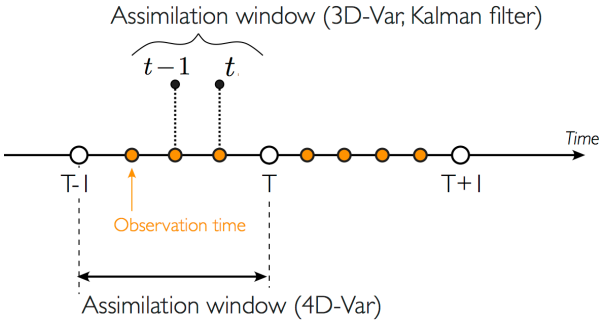


Figure 5.6: *Definition of the assimilation window:* $[T-1,T]$ *defined as the assimilation window for the 4D-Var algorithm, include several observation times; each sub-interval* $[t-1,t]$ *corresponds to the assimilation window for sequential data assimilation techniques such as the Kalman filter and the 3D-Var algorithm.*

Within this framework, the quantities of interest are considered as random variables and their inherent uncertainties are therefore described in terms of PDF (see Appendix F for further details on the notations and definitions of stochastic quantities). Note that we only present the data assimilation formalism for a discrete problem perspective and thereby, the discussion on the mapping between the continuum space and the discrete space associated with numerical models is not addressed (Bocquet, 2011).

[1] weather.gc.ca/.

5.2.1 Forward model of the physical system

We note $\mathbf{c}_t = [c_{1,t}, c_{2,t}, \cdots, c_{n_s,t}]^{\mathrm{T}} \in \mathbb{R}^{n_s}$ the prognostic state variable of a physical system at time t, with $c_{i,t}$ its value at a particular grid-point i (with $i = 1, \cdots, n_s$). The temporal and spatial evolution of this model state variable is described by a possibly non-linear computer model, denoted by $\mathcal{F}_{[t-1,t]}$ for a model integration from time $(t-1)$ to time t. This model state variable depends implicitly on a list of input parameters and external forcings, which are possibly uncertain and of different spatial and temporal resolutions. They are included in the vector

$$\boldsymbol{\theta}_t = \left[\theta_{1,t},\ \theta_{2,t}, \cdots,\ \theta_{n_{\boldsymbol{\theta}},t}\right]^{\mathrm{T}} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}.$$

These parameters can evolve over time; this evolution from time $(t-1)$ to time t is noted $\mathcal{P}_{[t-1,t]}$, also referred to as the *parameter model operator*. Formally, the dynamics of the physical system from time $(t-1)$ to time t is expressed as:

$$\begin{cases} \mathbf{c}_t = \mathcal{F}_{[t-1,t]}(\mathbf{c}_{t-1}, \boldsymbol{\theta}_{t-1}) \\ \boldsymbol{\theta}_t = \mathcal{P}_{[t-1,t]}(\boldsymbol{\theta}_{t-1}). \end{cases} \tag{5.1}$$

Note that if the input parameters and external forcing do not evolve in time or if their evolution is not known because of a lack of physical modeling or knowledge, then the operator $\mathcal{P}_{[t-1,t]}$ is reduced to an identity matrix $\mathbf{I}_{n_{\boldsymbol{\theta}}}$. In this case, the input parameters and external forcing are assumed invariant, i.e., $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$.

5.2.2 Control vector and stochastic-dynamic model

The size of the estimation problem is noted n. Thus, the control vector denoted by $\mathbf{x}_t = [\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{n,t}]^{\mathrm{T}} \in \mathbb{R}^n$ includes the n variables to be dynamically estimated by data assimilation at the discrete time t. These control variables are the estimation targets that are identified as important sources of uncertainties in the physical system and to which the model operators $\mathcal{F}_{[t-1,t]}$ and $\mathcal{P}_{[t-1,t]}$ are highly sensitive. In numerical weather prediction applications, the control vector is formed by the three-dimensional discrete fields of temperature, humidity, pressure and wind (i.e., the simulated fields defined at the grid-points of the computational domain); the dimension of the control n is then extremely high with $10^8/10^9$ control variables for $10^6/10^7$ observations to assimilate. However, the control vector $\mathbf{x}_t$ is not systematically in the same space as the model state vector; it can include input parameters and external forcing contained in the vector $\boldsymbol{\theta}_t$ and/or model state variables $\mathbf{c}_t$. Thus, the size n of the estimation problem is defined as follows:

- $n = n_v \times n_s$ for multi-variate state estimation with n_v referring to the number of model variables in the system and with n_s to the number of grid-points at which these model variables are described in the computer model (when there is a unique model variable, $n = n_s$);.

- ▷ $n = n_p \times n_s$ for multi-parameter estimation, with $n_p \leq n_{\boldsymbol{\theta}}$ only a subset of the input parameters and external forcing of the problem and with n_s the number of grid-points at which these control parameters are spatially-distributed (when these parameters are assumed uniform over the computational domain, $n = n_p$);

Note that the spatial distribution of the control parameters is not systematically provided at the same resolution as the model variables. However, for clarity purposes, we introduce a unique parameter n_s to describe the spatial distribution of both parameter and model state variables. Note also that for joint parameter/state estimation, $n = n_v \times n_s + n_p \times n_s$.

The n-dimensional space in which the control variables evolve is referred to as the *control space*. In the wildfire spread application presented in this work, we deal with both state and parameter estimations. For this purpose, we adopt a generic formalism for the presentation of data assimilation techniques.

↪ Forecast errors

All the control quantities are subject to uncertainties and their prior values at time t can result from a previous model integration or from the prior knowledge on the control vector. This prior knowledge, also called background or forecast within the framework of sequential data assimilation techniques, is denoted by $\mathbf{x}_t^{\mathrm{f}}$ (the superscript f referring to *forecast*). It incorporates the effects of a number of modeling choices (i.e. problem scales, numerical schemes, mesh resolution, subgrid-scale models, etc.). This implies that the true value of the control vector noted $\mathbf{x}_t^{\mathrm{t}}$ (the superscript t referring to *true*) is completely unknown and thereby, a probabilistic error treatment is required for $\mathbf{x}_t^{\mathrm{f}}$. It is worth noting that, while $\mathbf{x}_t^{\mathrm{f}}$ is treated as a random variable, $\mathbf{x}_t^{\mathrm{t}}$ is a deterministic variable. The associated forecast errors $\mathbf{e}_t^{\mathrm{f}}$ are defined as the difference between the forecast $\mathbf{x}_t^{\mathrm{f}}$ and the truth $\mathbf{x}_t^{\mathrm{t}}$ such that:

$$\mathbf{e}_t^{\mathrm{f}} = \mathbf{x}_t^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{t}}. \tag{5.2}$$

↪ Stochastic model and model errors

The time-integration of the model equations presented in Eq. (5.1) is associated with errors, which are related to the discretization and numerical errors in the resolution of the equations and/or to the physical assumptions and simplifications made for the original development of these equations. These errors exist, even though the true control variables are known; they are referred to as *model errors* noted $\boldsymbol{\varepsilon}_t^{\mathcal{F}}$ for the model state operator $\mathcal{F}_{[t-1,t]}$ and $\boldsymbol{\varepsilon}_t^{\mathcal{P}}$ for the model parameter operator $\mathcal{P}_{[t-1,t]}$. In this context, Eq. (5.1) can be expressed in terms of the true parameters, external forcing and model states as follows:

$$\begin{cases} \mathbf{c}_t^{\mathrm{t}} = \mathcal{F}_{[t-1,t]}(\mathbf{c}_{t-1}^{\mathrm{t}}, \boldsymbol{\theta}_{t-1}^{\mathrm{t}}) + \boldsymbol{\varepsilon}_t^{\mathcal{F}} \\ \boldsymbol{\theta}_t^{\mathrm{t}} = \mathcal{P}_{[t-1,t]}(\boldsymbol{\theta}_{t-1}^{\mathrm{t}}) + \boldsymbol{\varepsilon}_t^{\mathcal{P}}. \end{cases} \tag{5.3}$$

For the derivation of the data assimilation algorithms and the simplification of the notations, we introduce the general model operator $\mathcal{M}_{[t-1,t]}$ with the associated model error $\varepsilon_t^{\mathcal{M}}$ that encompasses the model errors related to the parameter evolution $\varepsilon_t^{\mathcal{P}}$ and to the model state evolution $\varepsilon_t^{\mathcal{F}}$. Thus, the system of equations (5.1) is regarded as equivalent to the following compact form:

$$\mathbf{x}_t = \mathcal{M}_{[t-1,t]}\left(\mathbf{x}_{t-1}\right), \tag{5.4}$$

with the following definition for the model error $\varepsilon_t^{\mathcal{M}}$:

$$\mathbf{x}_t^{\mathrm{t}} = \mathcal{M}_{[t-1,t]}\left(\mathbf{x}_{t-1}^{\mathrm{t}}\right) + \varepsilon_t^{\mathcal{M}}. \tag{5.5}$$

Note that the model variables that are not included in the control but that are required for the time-integration of the dynamic model are not mentioned in Eqs. (5.4) and (5.5) for clarity purposes.

5.2.3 Observations and observation operator

Suppose that, over the time interval $[t-1,t]$, a number p of observations are available and incorporated into an observation vector $\mathbf{y}_t^{\mathrm{o}} \in \mathbb{R}^p$, i.e.,

$$\mathbf{y}_t^{\mathrm{o}} = \left[\mathrm{y}_{1,t}^{\mathrm{o}},\ \mathrm{y}_{2,t}^{\mathrm{o}}, ...,\ \mathrm{y}_{p,t}^{\mathrm{o}}\right]^{\mathrm{T}}.$$

The p-dimensional space in which the observations evolve is named the *observation space*. Note that in most state estimation problems, the observation vector $\mathbf{y}_t^{\mathrm{o}}$ is of lower dimension than the control vector $\mathbf{x}_t$, while for parameter estimation problems, the opposite case can occur. These measurable quantities may come from various observation networks (e.g., spaceborne, airborne or in-situ measurements) at different spatial and temporal resolutions. In general, they are of different nature than the model state variables and they need post-treatments before being assimilated. They provide an indirect information on the unknown control variables $\mathbf{x}_t$ at time t. This implies that there is a possibly non-linear relationship (noted $\mathcal{G}_t$ in the following) between the control variables $\mathbf{x}_t$ and the observations $\mathbf{y}_t^{\mathrm{o}}$. This also implies that observations can provide additional information on the physical features of the true state of the physical system if this relationship between the control variables and the observations is properly characterized.

$\hookrightarrow$ Observation operator

The observation operator $\mathcal{G}_t$ maps the control space onto the observation space over the assimilation cycle $[t-1,t]$. The purpose of the observation operator $\mathcal{G}_t$ is to provide the model counterparts of the measurements $\mathbf{y}_t^{\mathrm{o}}$ at time t. These model counterparts of the observed quantities are noted $\mathbf{y}_t$ and are formally expressed as $\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t)$. $\mathbf{y}_t$ can also be regarded as the equivalent of the control vector $\mathbf{x}_t$ in the observation space. The definition of the observation operator $\mathcal{G}_t$ depends on the choice of the control variables.

- ▷ **State estimation.** $\mathcal{G}_t$ reduces to the composition of a variable substitution (i.e., from the model state variable to the observed quantity) and of an interpolation process (i.e., from simulated grid-points to the points where an actual measurement is made). This composition is denoted by the operator $\mathcal{H}_t$. Thus, for a state estimation problem, the observation operator $\mathcal{G}_t$ can be defined as follows:

$$\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t) = \mathcal{H}_t(\mathbf{x}_t). \tag{5.6}$$

- ▷ **Initial condition or parameter estimation.** $\mathcal{G}_t$ includes, on top of the variable substitution and interpolation process, the numerical integration of the model $\mathcal{M}_{[t-1,t]}$ over the assimilation cycle. In this case, the model counterparts of the observations $\mathbf{y}_t$ read:

$$\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t) = \mathcal{H}_t \circ \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}). \tag{5.7}$$

  $\mathcal{G}_t$ usually referred to as the *generalized observation operator* is non-linear if the dynamic model $\mathcal{M}_{[t-1,t]}$ exhibits non-linearities.

↪ Observation errors

The definition of the observation operator $\mathcal{G}_t$ is subject to errors. For instance, for remote sensing data from spaceborne platforms, the raw observations (e.g., radiance, radar blackscatter) are non-linear functions of the temperature field and they require the resolution of the radiation transfer equation (RTE), which is not straightforward and induces representativeness errors (Janjić and Cohn, 2006). Besides, measurements include scales of motion that are not resolved by computer models, meaning that representativeness errors in $\mathcal{G}_t$ are also due to unresolved scales. Thus, the observation errors relate to:

- ▷ **instrumental errors** defined as $\boldsymbol{\epsilon}_t^{\mu} = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t^{\mathrm{t}}$ and existing independently of the observation operator $\mathcal{G}_t$;
- ▷ **representativeness errors** defined as $\boldsymbol{\epsilon}_t^{\mathrm{r}} = \mathbf{y}_t^{\mathrm{t}} - \mathcal{G}(\mathbf{x}_t^{\mathrm{t}})$ and quantifying the imperfect knowledge in the mapping $\mathcal{G}_t$ from the control space onto the observation space, even though the true observed quantities $\mathbf{y}_t^{\mathrm{t}}$ are known.

Assuming that $\boldsymbol{\epsilon}_t^{\mu}$ and $\boldsymbol{\epsilon}_t^{\mathrm{r}}$ are uncorrelated, the total observation error $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ is estimated as the difference between the observations $\mathbf{y}_t^{\mathrm{o}}$ and the true control vector projected onto the observation space $\mathcal{G}_t(\mathbf{x}_t^{\mathrm{t}})$, see Lorenc (1986). Thus, $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ reads:

$$\boldsymbol{\epsilon}_t^{\mathrm{o}} = \boldsymbol{\epsilon}_t^{\mu} + \boldsymbol{\epsilon}_t^{\mathrm{r}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{t}}). \tag{5.8}$$

↪ Innovation vector

The discrepancies between the observation vector $\mathbf{y}_t^{\mathrm{o}}$ and the forecast model prediction in the observation space $\mathbf{y}_t^{\mathrm{f}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})$ at time t are computed through the

innovation vector $\mathbf{d}_t^{\mathrm{f}}$ that reads:

$$\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}). \tag{5.9}$$

Note that the statistical moments of $\mathbf{d}_t^{\mathrm{f}}$ (e.g., mean and standard deviation) provide a convenient measure of the deviations of model predictions from observations. From this viewpoint, Eq. (5.9) shows that the stochastic-dynamic model $\mathcal{M}_{[t-1,t]}$ (describing the behavior of the model state and parameters of the physical system over time) as well as the observation operator $\mathcal{G}_t$ (describing the relationship between the control variables and the measured quantities) are key elements at the core of all data assimilation methodologies. In particular, characterizing the nonlinearities in the observation operator $\mathcal{G}_t$ is a critical step in the formulation and efficiency of a data assimilation algorithm.

5.2.4 Stochastic treatment of errors

As the true vector $\mathbf{x}_t^{\mathrm{t}}$ is unknown, the errors on the observations $\mathbf{y}_t^{\mathrm{o}}$, on the forecast $\mathbf{x}_t^{\mathrm{f}}$ and on the dynamic model $\mathcal{M}_{[t-1,t]}$ are also unknown. As justified by Cohn (1997), these errors are represented as stochastic forcing, i.e., as random variables using PDF; these PDF describe the relative likelihood for these errors to occur at a given point in the control or observation space. These errors are defined in the framework of a white-noise analysis, having zero mean (these errors are referred to as *unbiased*) and finite variance. In most data assimilation algorithms, error statistics are described by Gaussian PDF; this assumption implies that the errors for random variables can be fully described by two characteristic variables, the expected value ($\mathbb{E}[\,\cdot\,]$ denotes the mathematical expectation operator) and an error covariance model.

↪ Forecast errors

The forecast errors $\mathbf{e}_t^{\mathrm{f}} = [\mathrm{e}_{1,t}^{\mathrm{f}}, \mathrm{e}_{2,t}^{\mathrm{f}}, \cdots, \mathrm{e}_{n,t}^{\mathrm{f}}]^{\mathrm{T}}$ defined in Eq. (5.2) are due, partly to limitations in the physical modeling and to numerical errors. Assuming the dynamic model is unbiased, these forecast errors satisfy $\mathbb{E}[\mathbf{e}_t^{\mathrm{f}}] = 0$ and thus, the statistics of $\mathbf{e}_t^{\mathrm{f}}$ are described in a square symmetric, positive definite (invertible) matrix $\mathbf{P}_t^{\mathrm{f}}$ of size $n \times n$ (with n the number of control variables). $\mathbf{P}_t^{\mathrm{f}}$ called the *forecast error covariance matrix* satisfies:

$$\mathbf{P}_t^{\mathrm{f}} = \mathbb{E}\Big[(\mathbf{x}_t^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{t}})(\mathbf{x}_t^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{t}})^{\mathrm{T}}\Big] = \mathbb{E}\Big[(\mathbf{e}_t^{\mathrm{f}})(\mathbf{e}_t^{\mathrm{f}})^{\mathrm{T}}\Big]. \tag{5.10}$$

Insight into the structure of $\mathbf{P}_t^{\mathrm{f}}$ is provided in the following box for interested readers. Since the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$ is assumed to follow the Gaussian distribution $\mathcal{N}(\mathbf{x}_t^{\mathrm{t}}, \mathbf{P}_t^{\mathrm{f}})$, the forecast error $\mathbf{e}_t^{\mathrm{f}}$ is defined as follows:

$$\mathbf{e}_t^{\mathrm{f}} \sim \mathcal{N}(0, \mathbf{P}_t^{\mathrm{f}}). \tag{5.11}$$

Detailed structure of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$

As explained in detail in Appendix F, for a multi-variate control vector $\mathbf{x}_t^{\mathrm{f}} \in \mathbb{R}^n$, either for spatially-distributed parameter estimation ($n = n_p \times n_s$) or for multi-variate state estimation ($n = n_v \times n_s$), the structure of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ that is compactly formulated in Eq. (5.10) can be decomposed as follows:

$$\mathbf{P}_t^{\mathrm{f}} = \begin{pmatrix} \mathbb{E}\Big[(e_{1,t}^{\mathrm{f}})\,(e_{1,t}^{\mathrm{f}})^{\mathrm{T}}\Big] & \cdots & \mathbb{E}\Big[(e_{1,t}^{\mathrm{f}})\,(e_{n,t}^{\mathrm{f}})^{\mathrm{T}}\Big] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\Big[(e_{n,t}^{\mathrm{f}})\,(e_{1,t}^{\mathrm{f}})^{\mathrm{T}}\Big] & \cdots & \mathbb{E}\Big[(e_{n,t}^{\mathrm{f}})\,(e_{n,t}^{\mathrm{f}})^{\mathrm{T}}\Big] \end{pmatrix}.$$

The diagonal elements represent the error variance associated with each variable of the multi-variate control vector $\mathbf{x}_t^{\mathrm{f}}$. For instance, for the i-th variable of the control vector, the error variance satisfies $(\sigma_{i,t}^{\mathrm{f}})^2 = \mathbb{E}[(e_{i,t}^{\mathrm{f}})\,(e_{i,t}^{\mathrm{f}})^{\mathrm{T}}]$. The off-diagonal terms $\mathbb{E}[(e_{i,t}^{\mathrm{f}})\,(e_{j,t}^{\mathrm{f}})^{\mathrm{T}}]$ for $i \neq j$, stand for the covariances between each pair of components of the control vector.

- ▷ The covariances between the errors in the different physical variables (called *multivariate covariances*) must be representative of the dynamic balance equations of the system (Gauthier et al., 1999; Ricci, 2004) through the constraints between the variables of the control vector $\mathbf{x}_t$. These constraints relevant to both parameter estimation ($n_p > 1$) and state estimation ($n_v > 1$) are named *balance constraints*.
- ▷ When spatially-distributed, the control variables include the same physical quantities at different spatial grid-points. The covariance terms associated with one physical quantity (called *univariate covariances*) represent its spatial correlations, which must be smoothly-defined and representative of the length-scales of the errors related to the dominant physical processes (Daley, 1991; Deckmyn and Berre, 2005; Pannekoucke et al., 2008). For instance, the univariate covariances associated with the first model variable in the control (n varying between 1 and n_s) correspond to the following block of $\mathbf{P}_t^{\mathrm{f}}$,

  $$\mathbb{E}\Big[(e_{i,t}^{\mathrm{f}})\,(e_{j,t}^{\mathrm{f}})^{\mathrm{T}}\Big], \quad i,j = 1,\cdots,n_s.$$

  As for the second model variable in the control (n varying between $(n_s{+}1)$ and $2\,n_s$), their univariate covariances correspond to $\mathbb{E}[(e_{i,t}^{\mathrm{f}})\,(e_{j,t}^{\mathrm{f}})^{\mathrm{T}}]$ for i and j varying between (n_s+1) and $2n_s$, etc.

These covariances describe the correlations between the errors in different physical variables/parameters of the system and thereby, they are responsible for transferring information between the control variables $[\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{n,t}]$ during the assimilation process. In particular, these multi-variate components are

critical for extracting information about unobserved variables from directly observed quantities (some control variables may be observed in the framework of state estimation). The forecast error covariances are therefore an effective way of incorporating multi-variate model constraints in the data assimilation system, useful to map a perturbation in the observation space onto a correction in the control space. However, it is difficult to build an accurate forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ that represents all the uncertainties present in a physical system, since the true control vector $\mathbf{x}_t^{\mathrm{t}}$ is unknown and since the number of degrees of freedom considered in a data assimilation application is reduced to overcome storage and computational cost issues. In any case, the forecast error variances $(\sigma_{i,t}^{\mathrm{f}})^2$ for $i = 1, \cdots, n$ must represent as accurately as possible the variability in the control variables.

↪ Observation errors

The observation errors $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ defined in Eq. (5.8) are supposed unbiased and thereby, satisfy $\mathbb{E}[\boldsymbol{\epsilon}_t^{\mathrm{o}}] = 0$. It is worth noting that if the observations are not unbiased, they are subject to a prior treatment to remove the bias. Thus, the statistics of $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ are described in a square symmetric, positive definite (invertible) matrix $\mathbf{R}_t$ of size $p \times p$ (with p the number of observations). $\mathbf{R}_t$ is called the *observation error covariance matrix* and satisfies:

$$\mathbf{R}_t = \mathbb{E}[(\boldsymbol{\epsilon}_t^{\mathrm{o}})(\boldsymbol{\epsilon}_t^{\mathrm{o}})^{\mathrm{T}}]. \tag{5.12}$$

Since the observation vector $\mathbf{y}_t^{\mathrm{o}}$ follows a Gaussian distribution $\mathcal{N}(\mathcal{G}(\mathbf{x}_t^{\mathrm{t}}), \mathbf{R}_t)$ (provided that $\mathcal{G}(\mathbf{x}_t^{\mathrm{t}})$ is a non-random variable), the observation error $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ satisfies:

$$\boldsymbol{\epsilon}_t^{\mathrm{o}} \sim \mathcal{N}(0, \mathbf{R}_t). \tag{5.13}$$

Classically, observational measurements are assumed to have uncorrelated errors in space. This implies that $\mathbf{R}_t$ is reduced to a diagonal matrix of p elements, each element representing the error variance of one observation quantity $(\sigma_{i,t}^{\mathrm{o}})^2$ (i varying between 1 and p). $\mathbf{R}_t$ reads:

$$\mathbf{R}_t = \begin{pmatrix} (\sigma_{1,t}^{\mathrm{o}})^2 & 0 & 0 & \cdots & 0 \\ 0 & (\sigma_{2,t}^{\mathrm{o}})^2 & 0 & \cdots & \vdots \\ 0 & 0 & (\sigma_{3,t}^{\mathrm{o}})^2 & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & (\sigma_{p,t}^{\mathrm{o}})^2 \end{pmatrix}.$$

Note that this assumption of uncorrelated errors could be questionable for a data set originating from the same measurement device (e.g., spaceborne data along the pass of polar-orbiting satellites). However, this aspect is out of the scope of this

study (Brankart et al., 2009; Gorin and Tsyrulnikov, 2011). In the following, the observation error variances are noted $(\sigma_t^{\mathrm{o}})^2$.

The data assimilation framework assumes that forecast and observation errors are uncorrelated, implying that $\mathbb{E}[(\mathbf{e}_t^{\mathrm{f}})(\boldsymbol{\epsilon}_t^{\mathrm{o}})^{\mathrm{T}}] = 0$. Note that in principle, an experimental measurement considered as a random variable has no reason for being correlated with an uncertain prior knowledge on the physical system at a given time step t, except if this forecast estimate is directly reconstructed with the assimilated measurements.

↪ Model errors

The model operator $\mathcal{M}_{[t-1,t]}$ inevitably introduces model errors, independently from the observation and forecast errors, because of their inability to cover the entire range of relevant scales, because also of knowledge gaps and/or inaccuracies in the description of the physics, see Eq. (5.3). These model errors $\boldsymbol{\varepsilon}_t^{\mathcal{M}}$ can be accounted for in the data assimilation framework through a model error covariance matrix, with $\boldsymbol{\varepsilon}_t^{\mathcal{M}} \sim \mathcal{N}(0, \mathbf{Q}_t^{\mathcal{M}})$. Note that if the model is assumed perfect, $\mathbf{Q}_t^{\mathcal{M}} = 0$ and the model dynamics is imposed as a strong constraint.

5.2.5 Formulation of the inverse modeling problem

The resolution of a data assimilation inverse modeling procedure that assumes Gaussian error statistics requires the following components:

- ▷ observations of the physical system $\mathbf{y}_t^{\mathrm{o}}$ and their associated errors $(\boldsymbol{\epsilon}_t^{\mathrm{o}}, \mathbf{R}_t)$;
- ▷ a background/forecast estimate of the control vector $\mathbf{x}_t^{\mathrm{f}}$ and its associated errors $(\mathbf{e}_t^{\mathrm{f}}, \mathbf{P}_t^{\mathrm{f}})$;
- ▷ a model operator $\mathcal{M}_{[t-1,t]}$ describing the dynamics of the physical system and its associated error $(\boldsymbol{\varepsilon}_t^{\mathcal{M}}, \mathbf{Q}_t^{\mathcal{M}})$;
- ▷ an observation operator $\mathcal{G}_t$.

Starting from prior information $\mathbf{x}_t^{\mathrm{f}}$ (i.e., the best estimate of the control vector prior to assimilation) and from available measurements $\mathbf{y}_t^{\mathrm{o}}$, data assimilation identifies the optimal estimate of the true value $\mathbf{x}_t^{\mathrm{t}}$ satisfying:

$$\left\{ \begin{array}{l} \mathbf{x}_t^{\mathrm{f}} = \mathbf{x}_t^{\mathrm{t}} + \mathbf{e}_t^{\mathrm{f}} \\ \mathbf{y}_t^{\mathrm{o}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{t}}) + \boldsymbol{\epsilon}_t^{\mathrm{o}}, \end{array} \right. \tag{5.14}$$

where the distance between the forecast estimate and the observations is represented by the innovation vector $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})$. The computation of this innovation vector is detailed in Fig. 5.7 for both parameter and state estimations. The main difference between these two estimation problems is in the formulation of the observation operator $\mathcal{G}_t$. While it relies on a selection operator $\mathcal{H}_t$ for a state estimation problem (see Eq. 5.6), $\mathcal{G}_t$ includes the model time-integration $\mathcal{M}_{[t-1,t]}$

from $(t-1)$ to t (see Eq. 5.7). In the data assimilation feedback, the differences in the observation space (i.e., the innovation vector) is mapped onto a correction in the parameter space weighted by the modeling and observation error statistics. Thus, the optimal estimate called the *analysis* and $\mathbf{x}_t^{\mathrm{a}}$ can be generally formulated as a correction to the forecast $\mathbf{x}_t^{\mathrm{f}}$:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \delta\mathbf{x}_t^{\mathrm{a}}, \tag{5.15}$$

where the term $\delta\mathbf{x}_t^{\mathrm{a}}$ is referred to as the *analysis increment*. The analysis is associated with the error $\mathbf{e}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{t}}$, characterized by the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$.

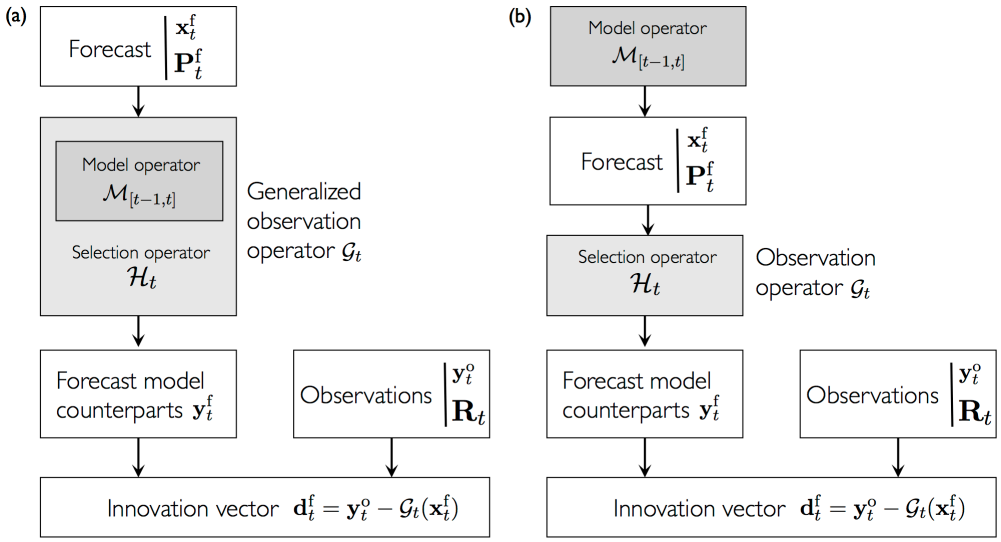


Figure 5.7: *Flowchart of the innovation vector $\mathbf{d}_t^{\mathrm{f}}$ for (a) parameter estimation; and (b) state estimation problems.*

↪ Definition of the optimal estimate

One can address the question of how to define an optimal estimate $\mathbf{x}_t^{\mathrm{a}}$ of the control vector. The complete knowledge of a physical system with a perfect confidence is inaccessible. Thus, in practice, the best estimate obtained from the data assimilation system is only a reasonable approximation of the *optimal* estimate; the analysis is sufficiently close from this knowledge to be able to provide a consistent representation of the behavior of the system without being perfect. Data assimilation cannot directly reduce the representativeness errors (these errors can for instance be addressed by increasing the spatial resolution of the model $\mathcal{M}_{[t-1,t]}$ and by improving the description of the physics). However, it can reduce forecast errors $\mathbf{e}_t^{\mathrm{f}}$. For any data assimilation algorithm, the optimality of the estimation implies that the uncertainty in the control vector is reduced through the analysis $\mathbf{x}_t^{\mathrm{a}}$, meaning that $\mathbf{e}_t^{\mathrm{a}} < \mathbf{e}_t^{\mathrm{f}}$ in some sense to be precised. Besides, the distance to the observations

is reduced. In contrast to the innovation vector $\mathbf{d}_t^{\mathrm{f}}$ defined in Eq. (5.9), we can introduce the concept of residual vector $\mathbf{d}_t^{\mathrm{a}}$ that measures the discrepancies between the observations and the analysis model counterpart $\mathbf{y}_t^{\mathrm{a}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{a}})$ as follows:

$$\mathbf{d}_t^{\mathrm{a}} = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t^{\mathrm{a}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}(\mathbf{x}_t^{\mathrm{a}}). \tag{5.16}$$

The optimality of the analysis control vector $\mathbf{x}_t^{\mathrm{a}}$ implies that the residual vector $\mathbf{d}_t^{\mathrm{a}}$ has a lower magnitude than the innovation vector $\mathbf{d}_t^{\mathrm{f}}$. Posterior diagnostics can be performed to check the consistency of the error covariance matrices (that are specified prior to assimilation) and, if necessary, to tune the error variances for the next assimilation cycles (Chapnik et al., 2004; Desroziers et al., 2005; Chapnik et al., 2006).

5.3 General Bayesian filtering formulation

Suppose that observations are available at discrete time steps over the time window $[0, T]$. The optimal solution of the inverse problem is derived from the inference of the control variable $\mathbf{x}_t$ at a time t given its past history $\{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{t-1}\}$ and the history of observations $\{\mathbf{y}_0^{\mathrm{o}}, \mathbf{y}_1^{\mathrm{o}}, \cdots, \mathbf{y}_{t-1}^{\mathrm{o}}\}$. Note that index 0 refers to the start of the time window $[0, T]$, while index T refers to its end. This is called a *Bayesian filtering problem*. The objective of this section is to provide a general conceptual estimation framework to compare different data assimilation techniques: variational approaches, KF and extensions as well as particle filters.

5.3.1 Formulation of the Bayes' theorem

↪ Assumptions

Data assimilation can be generally formulated as a Bayesian filtering problem, meaning that the formal mechanism to combine available measurements $\mathbf{y}_t^{\mathrm{o}}$ and the forecast estimate of the control vector $\mathbf{x}_t^{\mathrm{f}}$ at time t relies on the Bayes' theorem. The uncertainty in the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$ (Lorenc, 1986) can be expressed with the forecast PDF $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t)$; this represents the probability distribution that the forecast control variables represents the true value $\mathbf{x}_t^{\mathrm{t}}$. As for the observations $\mathbf{y}_t^{\mathrm{o}}$, their associated PDF is noted $\mathrm{p}(\mathbf{y}_t^{\mathrm{o}})$. Within the Bayesian framework, the sequences of the random variables are assumed to be discrete-time Markov chains with the following properties:

- ▷ The PDF of the control vector $\mathbf{x}_t$ at time t is only determined by its most recent value time $(t-1)$. This assumption is expressed as:

$$\mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{t-1}\Big) = \mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{x}_{t-1}\Big). \tag{5.17}$$

- ▷ The PDF of the observations $\mathbf{y}_t^{\mathrm{o}}$ is a Markovian process with respect to the history of the control vector $\{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_t\}$, meaning that:

$$\mathrm{p}\Big(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_t\Big) = \mathrm{p}\Big(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t\Big). \tag{5.18}$$

Note that $\mathrm{p}(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t)$ represents the data likelihood, i.e., the probability distribution of the measurements $\mathbf{y}_t^{\mathrm{o}}$ conditioned upon the control vector $\mathbf{x}_t$.

▷ The PDF of the control vector $\mathbf{x}_t$ at time t depends on the past observations only through its own history, i.e.,

$$\mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_1^{\mathrm{o}}, \mathbf{y}_2^{\mathrm{o}}, \cdots, \mathbf{y}_{t-1}^{\mathrm{o}}\Big) = \mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{x}_{t-1}\Big). \tag{5.19}$$

Note that the information provided by the observations up to time $(t-1)$ are implicitly contained in $\mathbf{x}_{t-1}$. Note also that the PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ expresses the stochastic time-evolution of the control vector from time $(t-1)$ to time t and thereby, corresponds to the forecast PDF such that:

$$\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t) = \mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}). \tag{5.20}$$

↪ **Bayes' theorem**

To find an accurate estimate of the true control vector $\mathbf{x}_t^{\mathrm{t}}$, the Bayes' theorem formulates the analysis as the PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$. This analysis is constructed based on all available information (including prior information and measurements) and satisfies:

$$\mathrm{p}^{\mathrm{a}}\Big(\mathbf{x}_t\Big) = \mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}}\Big) = \frac{\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t)\,\mathrm{p}(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t)}{\mathrm{p}(\mathbf{y}_t^{\mathrm{o}})}. \tag{5.21}$$

More precisely, the analysis is described by the PDF of the current control $\mathbf{x}_t$ conditioned upon the measurements $\mathbf{y}_t^{\mathrm{o}}$ noted $\mathrm{p}(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}})$: it represents the probability that the estimate of the control variables represents the true $\mathbf{x}_t^{\mathrm{t}}$ given that the model counterparts $\mathbf{y}_t$ are the observations $\mathbf{y}_t^{\mathrm{o}}$ at time t. This conditional PDF provides some confidence in the values of the control variables $\mathbf{x}_t$ given the information conveyed by the observations. Since $\mathrm{p}(\mathbf{y}_t^{\mathrm{o}})$ plays the role of a normalizing constant, Eq. (5.21) can be recast in the following general form:

$$\mathrm{p}^{\mathrm{a}}\Big(\mathbf{x}_t\Big) = \mathrm{p}\Big(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}}\Big) \propto \mathrm{p}^{\mathrm{f}}\Big(\mathbf{x}_t\Big)\,\mathrm{p}\Big(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t\Big), \tag{5.22}$$

where the symbol $\propto$ means *proportional to*.

5.3.2 Sequential Bayesian filtering

The Bayes' theorem can be applied sequentially to track the time-evolution of the true control vector through the time-sequence of the analysis PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ with $t = 0, \cdots, T$. This sequential Bayesian filtering is illustrated in Fig. 5.8. Each sequence, also called the assimilation cycle $[t-1, t]$, can be decomposed into two steps (Gelb, 1974; Tarantola, 1987; Todling and Cohn, 1994; Ide et al., 1997; Kalnay, 2003):

(1) **a prediction step (forecast)**, in which the PDF of the control vector $\mathbf{x}_t$ (i.e., model state and/or model parameters) is advanced in time, from time $(t-1)$ to the next analysis time t given some uncertainty ranges. The target of this step is the forecast PDF $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t) = \mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ that reflects all the uncertainties of the control variables without the information conveyed by the measurements, from time $(t-1)$ to time t.

(2) **an update step (analysis)**, in which new observations are considered at the analysis time t and the forecast PDF of the control parameters is modified consistently with the observations $\mathbf{y}_t^{\mathrm{o}}$ and their associated uncertainties (the data likelihood). Stated differently, the discrepancies between the observations $\mathbf{y}_t^{\mathrm{o}}$ and the model counterparts $\mathbf{y}_t$ (i.e., the innovation vector) weighted by the estimated uncertainties is mapped onto the control space to find a more realistic estimate of the control vector $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$. This update step is performed via the application of the Bayes' theorem presented in Eq. (5.22). Note that the analysis estimate at time t can be used as a starting point for deriving a new forecast over the next assimilation cycle $[t, t+1]$ and beyond.

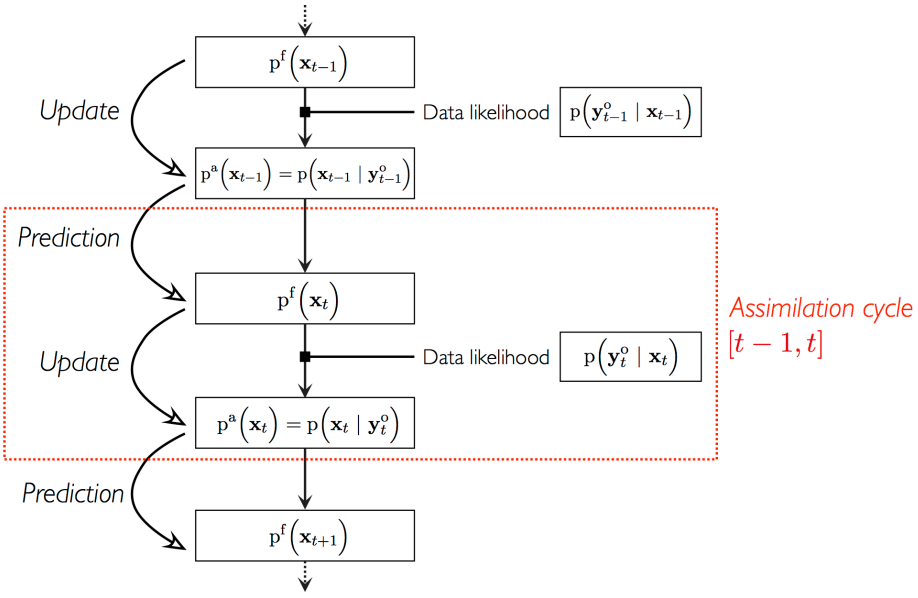


Figure 5.8: *Flowchart of sequential Bayesian filtering; each sequence $[t-1, t]$ including a prediction step and an update step.*

↪ Conditional mode estimation versus conditional mean estimation

Conceptually, the Bayesian filtering problem searches for the PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$. However, this complete PDF is difficult to track in practice, except for some restricting cases such as linearity of the models and Gaussianity of the error statistics. Thus, this PDF is generally approximated by its statistical moments. One important question

to be addressed is which statistics are essential to evolve; this question is related to the definition of an optimal estimator and thereby, of an optimality criterion for a given inverse problem. For instance, the optimal estimator could be the mean (the center) or the mode (the peak) of the PDF of interest (see Appendix F for details on PDF characteristics). Lorenc (1986) and Cohn (1997) showed that the Kalman filter (KF) yields the conditional mean estimate (which has the important property of being the minimum variance estimate), whereas the conditional mode estimate is the foundation of variational methodologies (searching for the value of highest probability of occurrence). It is worth noting that under the assumptions of linear models and Gaussian conditional PDF, these two estimates are equivalent (Jazwinski, 1970; Lorenc, 1986; Cohn, 1997; Mirouze, 2010).

5.4 Conditional mode estimation: variational approach

5.4.1 Formulation of the variational cost function

Assuming the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$ and the observations $\mathbf{y}_t^{\mathrm{o}}$ follow Gaussian PDF, their error statistics can be described using an error covariance model. The forecast PDF can be written as:

$$\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right)^{\mathrm{T}} (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right)\right\}, \tag{5.23}$$

with $\mathbf{P}_t^{\mathrm{f}}$ the forecast error covariance matrix representing the errors statistics in the forecast control variables. The data likelihood can be similarly expressed as:

$$\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t\right) \propto \exp\left\{-\frac{1}{2}\left(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\right)^{\mathrm{T}} (\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\right)\right\}, \tag{5.24}$$

with $\mathbf{R}_t$ the observation error covariance matrix representing the error statistics in the observations. Within this framework of Gaussian error statistics, Eq. (5.22) becomes:

$$\begin{aligned}\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t) \propto \exp\Bigg\{&-\frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right)^{\mathrm{T}} (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right) \\ &-\frac{1}{2}\left(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\right)^{\mathrm{T}} (\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\right)\Bigg\}.\end{aligned} \tag{5.25}$$

Conditional mode estimation searches for the mode of the PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$, i.e., the value of the control vector $\mathbf{x}_t$ that maximizes the probability to estimate the true value $\mathbf{x}_t^{\mathrm{t}}$. Thus, this technique is also known as *maximum likelihood estimation* (Maybeck, 1979). Maximizing the analysis PDF $\mathrm{p}_{\mathrm{a}}(\mathbf{x}_t)$ is equivalent to a minimization problem:

$$\max_{\mathbf{x}_t \in \mathbb{R}^n} \mathrm{p}_{\mathrm{a}}(\mathbf{x}_t) \quad \Longleftrightarrow \quad \min_{\mathbf{x}_t \in \mathbb{R}^n} \{-\ln[\mathrm{p}_{\mathrm{a}}(\mathbf{x}_t)]\} = \min_{\mathbf{x}_t \in \mathbb{R}^n} \mathcal{J}(\mathbf{x}_t). \tag{5.26}$$

with $\mathcal{J}$ the cost function of the estimation problem. Thus, under the assumptions of Gaussian error statistics, the Bayesian filtering problem can be recast in the minimization of $\mathcal{J}$ defined as follows:

$$\mathcal{J}(\mathbf{x}_t) = \frac{1}{2}\Big(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\Big)^{\mathrm{T}}(\mathbf{P}_t^{\mathrm{f}})^{-1}\Big(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\Big) + \frac{1}{2}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\Big)^{\mathrm{T}}(\mathbf{R}_t)^{-1}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\Big). \tag{5.27}$$

$\mathcal{J}$ measures the statistically-weighted square difference between the forecast $\mathbf{x}_t^{\mathrm{f}}$ and the control vector $\mathbf{x}_t$ on the one hand, between the observations $\mathbf{y}_t^{\mathrm{o}}$ and the model counterparts in the observation space $\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t)$ on the other hand. Each term is weighted by the precision taken as the inverse of the error covariance matrix, $(\mathbf{P}_t^{\mathrm{f}})^{-1}$ for the term related to the forecast and $(\mathbf{R}_t)^{-1}$ for the term related to the observations. The forecast and observation precisions define a metric, in which the distance to the observations $\mathbf{y}_t^{\mathrm{o}}$ and to the forecast $\mathbf{x}_t^{\mathrm{f}}$ can be minimized with respect to the control vector $\mathbf{x}_t$. Thus, the minimization of the cost function $\mathcal{J}$ (also referred to as the *variational approach*) can be regarded as equivalent to the following expression:

$$\min_{\mathbf{x}_t \in \mathbb{R}^n} \Big\{ || \ \mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}} \ ||^2_{(\mathbf{P}_t^{\mathrm{f}})^{-1}} + || \ \mathcal{G}(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}} \ ||^2_{(\mathbf{R}_t)^{-1}} \Big\}, \tag{5.28}$$

where the norm $|| \ . \ ||_{\mathbf{N}^{-1}}$ refers to the metric of the problem defined in the sense of the precision $\mathbf{N}^{-1}$, with $\mathbf{N}$ an error covariance matrix. Equation (5.28) can be considered as a generalized weighted least-squares problem since the variance of the different sources of information (i.e., forecast estimate of the control vector and observations) are unequal. Furthermore, a regularization is introduced with the additional term towards the forecast estimate $\mathbf{x}_t^{\mathrm{f}}$, implying that the analysis estimate must be simultaneously consistent with the forecast and the observations weighted by their respective error variances (i.e., the diagonal of the error covariance matrices). Note also that the solution of the problem is physically-constrained by the univariate and multivariate covariances described in the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$.

The direct minimization of the cost function $\mathcal{J}$ solves for the control vector $\mathbf{x}_t$ and leads to the analysis satisfying $\nabla\mathcal{J}(\mathbf{x}_t^{\mathrm{a}}) = 0$. Thus, the solution that minimizes the least-square estimate also maximizes the analysis density $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ and is called the *Bayesian estimate of the filtering problem*. It is worth mentioning that errors statistics are only described through their covariances (i.e., moments of second-order) in the variational approach. While Gaussian error statistics are exactly represented by their mean and covariances, higher-order moments of non-Gaussian error statistics (e.g., skewness, kurtosis) are disregarded for this approach. Note also that no assumption on the linearity of the dynamic model $\mathcal{M}_{[t-1,t]}$ or on the linearity of the observation operator $\mathcal{G}_t$ is made in the general formulation of the cost function (5.27). Still, the minimization process could be a difficult task if non-linearities are present, resulting in a non-quadratic cost function and in possibly multiple local

minima. Different variational techniques exist in the weather forecast literature, 3D-Var, 4D-Var and their incremental counterparts.

5.4.2 Three-dimensional variational algorithm

$\hookrightarrow$ Model and observation tangent-linear operators

We consider the time window $[t-1,t]$. Recall that the generalized observation operator $\mathcal{G}_t$ is expressed for both parameter and state estimations as follows:

$$\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t) = \mathcal{H}_t \circ \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}). \tag{5.29}$$

This non-linear operator can be linearized in the vicinity of the control vector $\mathbf{x}_t$. If we assume that non-linearities in the model operators $\mathcal{M}_{[t-1,t]}$ and $\mathcal{G}_t$ remain limited over the time period $[t-1,t]$, then it is possible to approximate the non-linear operators by their tangent-linear operators using a first-order Taylor expansion in the vicinity of a reference control vector usually taken as the forecast $\mathbf{x}_t^{\mathrm{f}}$. Then, the trajectory of the control vector $\mathbf{x}_t$ from $(t-1)$ to t can be approximated as follows:

$$\begin{aligned}
\mathbf{x}_t &= \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}) \\
&= \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}^{\mathrm{f}} + \delta\mathbf{x}_{t-1}) \\
&= \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}^{\mathrm{f}}) + \mathbf{M}_{[t-1,t]}\delta\mathbf{x}_{t-1} + \bigcirc(\| \delta\mathbf{x}_{t-1} \|^2), \\
\Rightarrow \mathbf{x}_t &= \mathbf{x}_t^{\mathrm{f}} + \delta\mathbf{x}_t + \bigcirc(\| \delta\mathbf{x}_{t-1} \|^2),
\end{aligned} \tag{5.30}$$

where the increment is integrated over time through the linearized model operator $\mathbf{M}_{[t-1,t]}$ such that $\delta\mathbf{x}_t = \mathbf{M}_{[t-1,t]}\delta\mathbf{x}_{t-1}$. $\mathbf{M}_{[t-1,t]}$ is called the *model tangent-linear operator* and is formally defined as follows:

$$\mathbf{M}_{[t-1,t]} = \frac{\partial \mathcal{M}_{[t-1,t]}}{\partial \mathbf{x}_{t-1}}\Big|_{\mathbf{x}_{t-1}=\mathbf{x}_{t-1}^{\mathrm{f}}}. \tag{5.31}$$

As for the observation operator, its linearized form reads:

$$\begin{aligned}
\mathbf{y}_t &= \mathcal{G}_t(\mathbf{x}_t) \\
&= \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}} + \delta\mathbf{x}_t) \\
&= \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) + \mathbf{G}_t\delta\mathbf{x}_t + \bigcirc(\| \delta\mathbf{x}_t \|^2), \\
\Rightarrow \mathbf{y}_t &= \mathbf{y}_t^{\mathrm{f}} + \delta\mathbf{y}_t + \bigcirc(\| \delta\mathbf{x}_t \|^2),
\end{aligned} \tag{5.32}$$

where the increment is derived at time t using the linearized observation operator $\mathbf{G}_t$ such that $\delta\mathbf{y}_t = \mathbf{G}_t\delta\mathbf{x}_t$. $\mathbf{G}_t$ is called the *observation tangent-linear operator* and reads:

$$\mathbf{G}_t = \frac{\partial \mathcal{G}_t}{\partial \mathbf{x}_t}\Big|_{\mathbf{x}_t=\mathbf{x}_t^{\mathrm{f}}}. \tag{5.33}$$

Bouttier and Courtier (1999) discussed the validity of the tangent-linear operators and the underlying assumptions on the physical system. It is worth noting that for a state estimation problem, the observation operator $\mathcal{G}_t$ reduces to the selection operator $\mathcal{H}_t$ as presented in Eq. (5.6) and thus, $\mathbf{G}_t = \mathbf{H}_t$. In contrast, for a parameter estimation problem, $\mathcal{G}_t$ involves the dynamic model integration over the time period $[t-1,t]$ as indicated by Eq. (5.7). This implies that $\mathbf{G}_t = \mathbf{H}_t\mathbf{M}_{[t-1,t]}$ and that the minimization requires the use of the adjoint operator of the dynamic model $\mathbf{M}^{\mathrm{T}}_{[t-1,t]}$ since $\mathbf{G}^{\mathrm{T}}_t = \mathbf{M}^{\mathrm{T}}_{[t-1,t]}\mathbf{H}^{\mathrm{T}}_t$.

↪ Variational analysis

Assuming model errors $\varepsilon^{\mathcal{M}}_t$ are negligible over the time period $[t-1,t]$, the 3D-Var cost function (noted $\mathcal{J}_{3D}(\mathbf{x})$ and called the strong-constraint formulation) reads:

$$\mathcal{J}_{3D}(\mathbf{x}_t) = \underbrace{\frac{1}{2}\Big(\mathbf{x}_t - \mathbf{x}^{\mathrm{f}}_t\Big)^{\mathrm{T}}(\mathbf{P}^{\mathrm{f}}_t)^{-1}\Big(\mathbf{x}_t - \mathbf{x}^{\mathrm{f}}_t\Big)}_{\mathcal{J}^{\mathrm{f}}_{3D}} + \underbrace{\frac{1}{2}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big)^{\mathrm{T}}(\mathbf{R}_t)^{-1}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big)}_{\mathcal{J}^{\mathrm{o}}_{3D}}. \tag{5.34}$$

The model physics $\mathbf{y}_t = \mathcal{G}(\mathbf{x}_t)$ is imposed as a strong constraint in the minimization. In contrast, the 3D-Var cost function accounting for the model error $\varepsilon^{\mathcal{M}}_t$, noted $\mathcal{J}_{3D^*}(\mathbf{x}_t)$, is named the weak-constraint problem and includes an additional term $\mathcal{J}^{\mathrm{q}}_{3D^*}$ such that:

$$\mathcal{J}_{3D^*}(\mathbf{x}_t) = \underbrace{\frac{1}{2}\Big(\mathbf{x}_t - \mathbf{x}^{\mathrm{f}}_t\Big)^{\mathrm{T}}(\mathbf{P}^{\mathrm{f}}_t)^{-1}\Big(\mathbf{x}_t - \mathbf{x}^{\mathrm{f}}_t\Big)}_{\mathcal{J}^{\mathrm{f}}_{3D}} + \underbrace{\frac{1}{2}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big)^{\mathrm{T}}(\mathbf{R}_t)^{-1}\Big(\mathcal{G}(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big)}_{\mathcal{J}^{\mathrm{o}}_{3D}} + \underbrace{\frac{1}{2}\eta^{\mathrm{T}}_t(\mathbf{Q}_t)^{-1}\eta_t}_{\mathcal{J}^{\mathrm{q}}_{3D^*}}, \tag{5.35}$$

with η_t the estimate of the model error at time t (Trémolet, 2007b) and with the weak constraint $\mathbf{y}_t = \mathcal{G}(\mathbf{x}_t) = \mathcal{H} \circ [\mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1} + \eta_t)]$. Note that the implementation of the weak-constraint formulation is currently an active research area in the weather forecast field and that the discussion on this technique is out of the scope of this work. In the following, the evolution models for the input parameters and the model state are considered as perfect; we only present the analysis solution for the strong-constraint formulation.

Under the assumption of uncorrelated observations errors, the observation error covariance matrix $\mathbf{R}_t$ is diagonal, meaning that the term $\mathcal{J}^{\mathrm{o}}_{3D}$ can be regarded as a linear combination of the innovations $\mathbf{d}^{\mathrm{f}}_t = \mathbf{y}^{\mathrm{o}}_t - \mathcal{G}_t(\mathbf{x}_t)$, over space, at time t, weighted by the variances at the observation points for each variable of the control

vector. Furthermore, the forecast $\mathbf{x}_t^{\mathrm{f}}$ is generally obtained through the dynamic model integration starting at the analysis of the previous time window $\mathbf{x}_{t-1}^{\mathrm{a}}$. The minimization of Eq. (5.34) is related to the gradient of the cost function $\mathcal{J}_{3D}$ as follows:

$$\nabla \mathcal{J}_{3D}(\mathbf{x}_t) = (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right) + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}_t^{\mathrm{o}}\right), \tag{5.36}$$

with $\mathbf{G}_t$ the tangent-linear operator defined in Eq. (5.33) and $\mathbf{G}_t^{\mathrm{T}}$ its adjoint. Thus, the analysis $\mathbf{x}_t^{\mathrm{a}}$ reads:

$$\begin{aligned}\nabla \mathcal{J}_{3D}(\mathbf{x}_t^{\mathrm{a}}) = 0 &\Longleftrightarrow (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}}\right) + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t^{\mathrm{a}}) - \mathbf{y}_t^{\mathrm{o}}\right) = 0,\\ &\Longleftrightarrow \mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}} = \mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\left(\mathbf{y}^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{a}})\right).\end{aligned} \tag{5.37}$$

$\mathcal{J}_{3D}$ is not quadratic with respect to the control vector $\mathbf{x}_t$ because the observation operator $\mathcal{G}_t$ is usually non-linear, in particular for a parameter estimation problem. The role of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ in the estimation of the analysis $\mathbf{x}_t^{\mathrm{a}}$ is highlighted in the following box in the context of a state estimation problem. Due to the computational expense in the tangent-linear and adjoint operators, the 3D-Var minimization can be time-consuming. For this purpose, incremental approaches based on an iterative procedure are implemented in practice.

Role of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ (Fisher, 1998)

The forecast error term of the cost function $\mathcal{J}_{3D}^{\mathrm{f}}$ is crucial to the performance of the data assimilation system and involves the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$. We illustrate its role for a simple example within a state estimation framework. We consider a single observation of the value of a model field (e.g., temperature in weather forecast) at one grid-point of the computational domain, corresponding to the k-th variable of the control vector ($p = 1$). Note that the control vector is equivalent to the model state vector in the case of state estimation. Thus, the observation operator $\mathcal{G}_t$ reduces to a linear selection operator H_t of dimension $1 \times n$ and whose tangent-linear $\mathbf{H}$ is constant over time. The k-th element of $\mathbf{H}$ is equal to one and other elements are all zero, i.e., $\mathbf{H} = (0, 0, \cdots, 0, 1, 0, \cdots, 0)$. The analysis satisfies Eq. (5.37). In this example, the matrix product $\mathbf{P}_t^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}$ is simply equal to the k-th column of $\mathbf{P}_t^{\mathrm{f}}$. Also, since we deal with a single observation, the term $\mathbf{R}_t^{-1}(\mathbf{y}^{\mathrm{o}} - H_t\,\mathbf{x}_t^{\mathrm{a}})$ is a scalar value noted $(y_t^{\mathrm{o}} - (\mathrm{x}_t^{\mathrm{a}})_k)/(\sigma_t^{\mathrm{o}})^2$, where $(\mathrm{x}_t^{\mathrm{a}})_k$ represents the analysis grid-point value corresponding to the observation and where $(\sigma_t^{\mathrm{o}})^2$ represents the variance of the observation error. Thus, Eq. (5.37) can be written as:

$$\delta\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}} = \left(\frac{y_t^{\mathrm{o}} - (\mathrm{x}_t^{\mathrm{a}})_k}{(\sigma_t^{\mathrm{o}})^2}\right)\begin{pmatrix}(\mathbf{P}_t^{\mathrm{f}})_{1k}\\(\mathbf{P}_t^{\mathrm{f}})_{2k}\\\vdots\\(\mathbf{P}_t^{\mathrm{f}})_{nk}\end{pmatrix}.$$

This relation means that the analysis increment $\delta\mathbf{x}_t^{\mathrm{a}}$ is proportional to a column of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$. This implies that $\mathbf{P}_t^{\mathrm{f}}$ controls how information is spread out from the single available observation, to provide statistically-consistent increments at the neighboring grid-points and to ensure that an observation of one model variable produce consistent increments in the other model variables.

↪ Incremental form of the cost function

The 3D-Var formulation presented in Eq. (5.34) is difficult to solve for a non-linear observation operator $\mathcal{G}_t$, in particular within the framework of parameter estimation. In this case, the cost function $\mathcal{J}_{3D}$ can exhibit multiple local minima and thus, its minimization becomes a challenging and time-consuming task (without any guarantee that the solution is optimal). This implies that the conditional mode estimate is generally not unique, in contrast to the conditional mean estimate. Considering that the extent of non-linearities remains limited, Courtier et al. (1994) showed that the Gauss-Newton method can be used to transform the non-quadratic minimization problem onto a sequence of quadratic function minimizations.

Based on the linearization of the observation operator $\mathcal{G}_t$ presented in Eq. (5.32), the 3D-Var cost function presented in Eq. (5.34) can be formulated with respect to the correction increment $\delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}$, instead of the control vector $\mathbf{x}_t$. In the incremental form, the discrepancies between the observations and the model counterparts of the observed quantities can be expressed as:

$$\mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t) = \underbrace{\mathbf{y}^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})}_{\mathbf{d}_t^{\mathrm{f}}} - \mathbf{G}_t\delta\mathbf{x}_t. \tag{5.38}$$

It follows that the incremental 3D-Var cost function noted $\mathcal{J}_{inc-3D}$ is a quadratic function in the increment $\delta\mathbf{x}_t$ such that:

$$\begin{aligned}\mathcal{J}_{inc-3D}(\delta\mathbf{x}_t) =& \frac{1}{2}\,\delta\mathbf{x}_t^{\mathrm{T}}\,(\mathbf{P}_t^{\mathrm{f}})^{-1}\,\delta\mathbf{x}_t \\ &+ \frac{1}{2}\Big(\mathbf{G}_t\delta\mathbf{x}_t - \mathbf{d}_t^{\mathrm{f}}\Big)^{\mathrm{T}}(\mathbf{R}_t)^{-1}\Big(\mathbf{G}_t\delta\mathbf{x}_t - \mathbf{d}_t^{\mathrm{f}}\Big).\end{aligned} \tag{5.39}$$

The gradient of the incremental cost function satisfies:

$$\begin{aligned}&\nabla\mathcal{J}_{inc-3D}(\mathbf{x}_t^{\mathrm{a}}) = 0 \\ \Leftrightarrow 0 &= (\mathbf{P}_t^{\mathrm{f}})^{-1}\left(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}}\right) + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) + \mathbf{G}_t(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}}) - \mathbf{y}_t^{\mathrm{o}}\right), \\ \Leftrightarrow 0 &= \left((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\mathbf{G}_t\right)\left(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{f}}\right) + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\left(\mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) - \mathbf{y}_t^{\mathrm{o}}\right), \\ \Leftrightarrow \mathbf{x}_t^{\mathrm{a}} &= \mathbf{x}_t^{\mathrm{f}} + \left((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\mathbf{G}_t\right)^{-1}\mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\left(\mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})\right),\end{aligned} \tag{5.40}$$

or equivalently,

$$\delta\mathbf{x}_t^{\mathrm{a}} = \left((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\mathbf{G}_t\right)^{-1}\mathbf{G}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\,\mathbf{d}_t^{\mathrm{f}}. \tag{5.41}$$

In practice, the minimization of the incremental cost function presented in Eq. (5.40) is obtained through an iterative algorithm (e.g., conjugate gradient, quasi-Newton) starting from the forecast $\mathbf{x}_t^{\mathrm{f}}$, each iteration requiring the evaluation of the quadratic cost function $\mathcal{J}_{inc-3D}$ and its gradient. However, there is in general no guarantee that the iterations will converge. Typically, about 50 iterations are performed to produce the solution $\mathbf{x}_t^{\mathrm{a}}$. To move to the next assimilation cycle $[t, t+1]$, the forecast $\mathbf{x}_{t+1}^{\mathrm{f}}$ results from the model integration of the analysis $\mathbf{x}_t^{\mathrm{a}}$ at time t as in sequential data assimilation approaches. If no term of the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$ is evaluated (as in most of the 3D-Var applications), the forecast error covariance matrix is not evolved and by default, $\mathbf{P}_{t+1}^{\mathrm{f}} = \mathbf{P}_t^{\mathrm{f}}$. Since the minimization is performed at a given time t, the 3D-Var approach corresponds to a *static* data assimilation technique.

5.4.3 Four-dimensional variational algorithm

↪ Generalization of the three-dimensional variational algorithm

One issue with the 3D-Var formulation is the discontinuity in the model trajectory induced by the analysis increment at the analysis time. Besides, all observations cannot be processed at once in a continually operating data assimilation system. Thus, the sequential assimilation of these data may not be the best solution to obtain a smooth model trajectory of the system behavior and to produce physically-consistent forecasts. To overcome this issue, the four-dimensional variational technique (referred to as 4D-Var) has been introduced as a generalization of 3D-Var for observations that are distributed over the time window $[0, T]$, with $\mathbf{y}^{\mathrm{o}} = \{\mathbf{y}_0^{\mathrm{o}}, \mathbf{y}_1^{\mathrm{o}}, \cdots, \mathbf{y}_t^{\mathrm{o}}, \cdots, \mathbf{y}_T^{\mathrm{o}}\}$ the $(T+1)$ observation vectors at different times (t varying between 0 and T) and with $\mathbf{y}_t^{\mathrm{o}} \in \mathbb{R}^p$ the spatially-varying observation vector at a particular time t. Within the framework of 4D-Var, $[0, T]$ is the assimilation window.

In the context of numerical weather predictions, the 4D-Var technique aims at estimating the initial condition of the atmosphere at the start of the assimilation window (referred to with *index 0*) given all the observations available over the time window $[0, T]$. We define the sequence of control vectors as $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_t, \cdots, \mathbf{x}_T\}$, with t varying between 0 and T, with $\mathbf{x}_t \in \mathbb{R}^n$ a control vector at a particular time t and with $\mathbf{y}_t \in \mathbb{R}^p$ its model counterparts in the observation space.[2] The 4D-Var objective is therefore to estimate the control vector $\mathbf{x}_0$ starting from a prior estimate $\mathbf{x}_0^{\mathrm{f}}$ and its associated error covariance matrix $\mathbf{P}_0^{\mathrm{f}}$. Note that the pair $(\mathbf{x}_0^{\mathrm{f}}, \mathbf{P}_0^{\mathrm{f}})$ is

[2] Time indices of $\mathbf{y}^{\mathrm{o}}$ and $\mathbf{x}$ are equivalent here to simplify notations. This is usually not the case in practice, implying that the mapping between the control space and the observation space requires a time interpolation.

usually denoted by $(\mathbf{x}^{\mathrm{b}}, \mathbf{B})$ in the 4D-Var literature (the superscript b referring to *background*), but the notations introduced in Section 5.2 are maintained for clarity purposes.

As a generalization of the 3D-Var approach for parameter estimation, the 4D-Var requires the use of the generalized observation operator $\mathcal{G}$ to be able to compare observations $\mathbf{y}^{\mathrm{o}}$ and their model counterparts $\mathbf{y}$ at the appropriate times over the assimilation time window.

↪ Formulation of the four-dimensional cost function

The commonly-used approach relies on the strong-constraint formulation and therefore on the assumption that the dynamic model $\mathcal{M}$ is perfect over the assimilation window $[0, T]$, meaning that model errors $\boldsymbol{\varepsilon}_t^{\mathcal{M}}$ at each time t are negligible with respect to forecast and observation errors. In this context, the variational technique is referred to as the strong-constraint 4D-Var; originally proposed by Talagrand and Courtier (1987) and later discussed in Courtier et al. (1994), it consists in minimizing the following cost function $\mathcal{J}_{4D}(\mathbf{x}_0)$:

$$\begin{aligned}\mathcal{J}_{4D}(\mathbf{x}_0) = & \underbrace{\frac{1}{2}\Big(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{f}}\Big)^{\mathrm{T}}(\mathbf{P}_0^{\mathrm{f}})^{-1}\Big(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{f}}\Big)}_{\mathcal{J}_{4D}^{\mathrm{f}}} \\ & + \underbrace{\frac{1}{2}\Big(\mathbf{y} - \mathbf{y}^{\mathrm{o}}\Big)^{\mathrm{T}}(\mathbf{R})^{-1}\Big(\mathbf{y} - \mathbf{y}^{\mathrm{o}}\Big)}_{\mathcal{J}_{4D}^{\mathrm{o}}},\end{aligned} \tag{5.42}$$

with the strong model constraint $\mathbf{y} = \mathcal{G}(\mathbf{x})$. The cost function is divided into two different terms: (1) the term $\mathcal{J}_{4D}^{\mathrm{f}}$ representing the weighted-deviation to the prior estimate $\mathbf{x}_0^{\mathrm{f}}$; and (2) the term $\mathcal{J}_{4D}^{\mathrm{o}}$ representing the weighted-deviation to the observations at the different time steps included in the time window $[0, T]$. Thus, the observation operator $\mathcal{G}_t$ at time t can be written with respect to the initial control vector $\mathbf{x}_0$ as $\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t) = \mathcal{H}_t[\mathcal{M}_{[0,t]}(\mathbf{x}_0)]$, including the non-linear model propagation. Over the assimilation window $[0, T]$, the model counterparts in the observation space $\mathbf{y}$ can therefore be recast in the following compact form:

$$\mathbf{y} = \mathcal{G}(\mathbf{x}) = \begin{bmatrix} \mathcal{G}_0(\mathbf{x}_0) \\ \mathcal{G}_1(\mathbf{x}_1) \\ \vdots \\ \mathcal{G}_t(\mathbf{x}_t) \\ \vdots \\ \mathcal{G}_T(\mathbf{x}_T) \end{bmatrix} = \begin{bmatrix} \mathcal{H}_0(\mathbf{x}_0) \\ \mathcal{H}_1\left[\mathcal{M}_{[0,1]}(\mathbf{x}_0)\right] \\ \vdots \\ \mathcal{H}_t\left[\mathcal{M}_{[0,t]}(\mathbf{x}_0)\right] \\ \vdots \\ \mathcal{H}_T\left[\mathcal{M}_{[0,T]}(\mathbf{x}_0)\right] \end{bmatrix}. \tag{5.43}$$

If we consider that the observation errors are uncorrelated in time, the observation term $\mathcal{J}^{\mathrm{o}}_{4D}$ can be expressed as:

$$\mathcal{J}^{\mathrm{o}}_{4D} = \frac{1}{2}\sum_{t=0}^{T}\Big(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big)^{\mathrm{T}}(\mathbf{R}_t)^{-1}\Big(\mathcal{G}_t(\mathbf{x}_t) - \mathbf{y}^{\mathrm{o}}_t\Big), \tag{5.44}$$

with the constraint $\mathbf{x}_t = \mathcal{M}_{[0,t]}(\mathbf{x}_0)$ for each time $t = 1, \cdots, T$ (meaning that the control vector $\mathbf{x}_t$ corresponds to an admissible evolution of the physical system, starting from the initial control vector $\mathbf{x}_0$) and with $\mathbf{R}_t$ the block of the observation error covariance matrix $\mathbf{R}$ at time t.

↪ Incremental form of the cost function

As for the 3D-Var approach, the 4D-Var formulation presented in Eq. (5.42) is difficult to solve for a non-linear generalized observation operator $\mathcal{G}$ that involves a non-linear dynamic model $\mathcal{M}$ for both parameter and state estimations. It is also possible to transform this non-quadratic minimization problem onto a sequence of quadratic function minimizations.

Based on the generalization of the use of the tangent-linear operators presented in Eqs. (5.31) and (5.33) over the time window $[0, T]$, we linearize the operators in the vicinity of a reference control vector (usually taken as the prior $\mathbf{x}^{\mathrm{f}}_0$) for the perturbation $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}^{\mathrm{f}}_0$. For the dynamic-model operator $\mathcal{M}_{[0,t]}$, we obtain at time t:

$$\begin{aligned}\mathbf{x}_t = \mathcal{M}_{[0,t]}(\mathbf{x}^{\mathrm{f}}_0 + \delta\mathbf{x}_0) &\approx \mathcal{M}_{[0,t]}(\mathbf{x}^{\mathrm{f}}_0) + \mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[0,1]}\delta\mathbf{x}_0,\\ \Rightarrow \mathbf{x}_t &\approx \mathbf{x}^{\mathrm{f}}_t + \mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[0,1]}\delta\mathbf{x}_0.\end{aligned} \tag{5.45}$$

By linearizing the operator $\mathcal{G}_t$ in the vicinity of $\mathbf{x}_t$ at time t for the perturbation $\delta\mathbf{x}_0$, the following approximation is obtained in the observation space:

$$\begin{aligned}\mathbf{y}_t = \mathcal{G}_t(\mathbf{x}_t) &\approx \mathcal{H}_t\Big[\mathcal{M}_{[0,t]}(\mathbf{x}^{\mathrm{f}}_0) + \mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[0,1]}\delta\mathbf{x}_0\Big]\\ &\approx \mathcal{H}_t\Big[\mathcal{M}_{[0,t]}(\mathbf{x}^{\mathrm{f}}_0)\Big] + \mathbf{H}_t\mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[0,1]}\delta\mathbf{x}_0\\ &\approx \mathcal{G}_t(\mathbf{x}^{\mathrm{f}}_t) + \mathbf{G}_t\delta\mathbf{x}_0,\\ \Rightarrow \mathbf{y}_t &\approx \mathbf{y}^{\mathrm{f}}_t + \mathbf{G}_t\delta\mathbf{x}_0.\end{aligned} \tag{5.46}$$

The time-sequence of the model counterparts $\mathbf{y}$ of the observed quantities over the time window $[0, T]$ can be recast in the following compact form:

$$\mathbf{G} = \begin{bmatrix}\mathbf{G}_0\\ \vdots\\ \mathbf{G}_t\\ \vdots\\ \mathbf{G}_T\end{bmatrix} = \begin{bmatrix}\mathbf{H}_0\\ \vdots\\ \mathbf{H}_t\mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[0,1]}\\ \vdots\\ \mathbf{H}_T\mathbf{M}_{[T-1,T]}\cdots\mathbf{M}_{[0,1]}\end{bmatrix}. \tag{5.47}$$

Using this formalism, the discrepancies between $\mathbf{y}^{\mathrm{o}} = \{\mathbf{y}_0^{\mathrm{o}}, \cdots, \mathbf{y}_t^{\mathrm{o}}, \cdots, \mathbf{y}_T^{\mathrm{o}}\}$ and their model counterparts $\mathcal{G}(\mathbf{x}) = \{\mathcal{G}_0(\mathbf{x}_0), \cdots, \mathcal{G}_t(\mathbf{x}_t), \cdots, \mathcal{G}_T(\mathbf{x}_T)\}$ read:

$$\mathbf{y}^{\mathrm{o}} - \mathcal{G}(\mathbf{x}) = \underbrace{\mathbf{y}^{\mathrm{o}} - \mathcal{G}(\mathbf{x}^{\mathrm{f}})}_{\mathbf{d}^{\mathrm{f}}} - \mathbf{G}\delta\mathbf{x}_0. \tag{5.48}$$

This is the analogy of Eq. (5.38) over the time window $[0, T]$. It follows that the incremental 4D-Var cost function noted $\mathcal{J}_{inc-4D}$ is not parameterized with respect to the control vector $\mathbf{x}_0$, but instead with respect to the correction increment $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^{\mathrm{f}}$ (at the start of the time window) such that:

$$\begin{aligned}\mathcal{J}_{inc-4D}(\delta\mathbf{x}_0) = &\underbrace{\frac{1}{2}\,\delta\mathbf{x}_0^{\mathrm{T}}\,(\mathbf{P}_0^{\mathrm{f}})^{-1}\,\delta\mathbf{x}_0}_{\mathcal{J}_{inc-4D}^{\mathrm{f}}} \\ &+ \underbrace{\frac{1}{2}\Big(\mathbf{G}\delta\mathbf{x}_0 - \mathbf{d}^{\mathrm{f}}\Big)^{\mathrm{T}}(\mathbf{R})^{-1}\Big(\mathbf{G}\delta\mathbf{x}_0 - \mathbf{d}^{\mathrm{f}}\Big)}_{\mathcal{J}_{inc-4D}^{\mathrm{o}}}.\end{aligned} \tag{5.49}$$

As mentioned for the 4D-Var formulation, if the observation errors are assumed to be uncorrelated in time, the term $\mathcal{J}_{inc-4D}^{\mathrm{o}}$ can be expressed as:

$$\mathcal{J}_{inc-4D}^{\mathrm{o}}(\delta\mathbf{x}_0) = \frac{1}{2}\sum_{t=1}^{T}\left\{(\mathbf{G}_t\delta\mathbf{x}_0 - \mathbf{d}_t^{\mathrm{f}})^{\mathrm{T}}(\mathbf{R}_t)^{-1}(\mathbf{G}_t\delta\mathbf{x}_0 - \mathbf{d}_t^{\mathrm{f}})\right\}, \tag{5.50}$$

with $\mathbf{G}_t$ the restriction of the tangent-linear operator $\mathbf{G}$ at time t and $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) = \mathbf{y}_t^{\mathrm{o}} - \mathcal{H}_t[\mathcal{M}_{[0,t]}(\mathbf{x}_0^{\mathrm{f}})]$ the innovation at time t. Thus,

$$\mathbf{G}_t\delta\mathbf{x}_0 - \mathbf{d}_t^{\mathrm{f}} = \mathbf{H}_t\mathbf{M}_{[t-1,t]}\mathbf{M}_{[t-2,t-1]}\cdots\mathbf{M}_{[1,0]}\delta\mathbf{x}_0 - \mathbf{d}_t^{\mathrm{f}}. \tag{5.51}$$

Note that the incremental cost function $\mathcal{J}_{inc-4D}(\delta\mathbf{x}_0)$ in Eq. (5.49) is quadratic with respect to the increment $\delta\mathbf{x}_0$: the exact minimizing solution $\delta\mathbf{x}_0^{\mathrm{a}}$ can be found by solving $\nabla\mathcal{J}_{inc-4D}(\delta\mathbf{x}_0^{\mathrm{a}}) = 0$ for the increment $\delta\mathbf{x}_0$. As for the 3D-Var incremental form, the analysis solution $\mathbf{x}_0^{\mathrm{a}}$ can be expressed as a correction of the prior $\mathbf{x}_0^{\mathrm{f}}$ such that:

$$\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}_0^{\mathrm{f}} + \delta\mathbf{x}_0^{\mathrm{a}}, \tag{5.52}$$

with the analysis increment $\delta\mathbf{x}_0^{\mathrm{a}}$ satisfying:

$$\delta\mathbf{x}_0^{\mathrm{a}} = \Big((\mathbf{P}_0^{\mathrm{f}})^{-1} + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\Big)^{-1}\,\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\,\mathbf{d}^{\mathrm{f}}. \tag{5.53}$$

This analysis increment is a generalization of the incremental 3D-Var formulation; a schematic of the incremental 4D-Var solution is shown in Fig. 5.9. The analysis trajectory is smooth over the time window $[T-1, T]$ and corresponds to the forecast trajectory integrated by the computer model, but with a corrected initial state at

the beginning of the assimilation time window. Since the observations and the forecast are subject to uncertainties, the analysis provides an intermediate solution between the forecast trajectory and the observations that is more accurate than if either of them were taken separately.

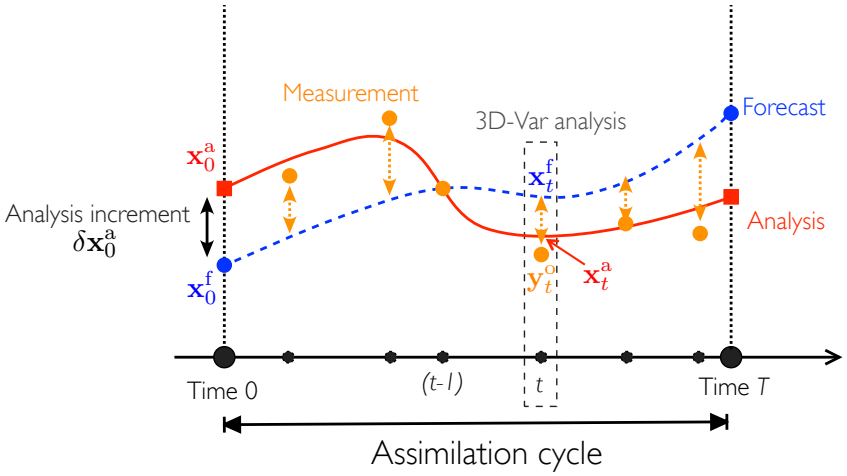


Figure 5.9: *Schematic of the incremental 4D-Var algorithm with a comparison to 3D-Var.*

Some limiting cases listed below can be identified; those highlight the role of the forecast and observation error covariance matrices in the data assimilation process.

- ▷ **Perfect forecast:** $(\mathbf{P}_0^{\mathrm{f}} \longrightarrow 0) \Longrightarrow (\delta \mathbf{x}_0^{\mathrm{a}} \longrightarrow 0)$, meaning that the assimilation of observations cannot bring more information to the physical system; the confidence in the forecast control vector $\mathbf{x}_0$ is total.
- ▷ **Worthless observations:** $(\mathbf{R}^{-1} \longrightarrow 0) \Longrightarrow (\delta \mathbf{x}_0^{\mathrm{a}} \longrightarrow 0)$, meaning that observations are subject to very high uncertainties that prevent them from capturing any physical feature and thus, from contributing to the analysis.
- ▷ **Worthless forecast:** $((\mathbf{P}_0^{\mathrm{f}})^{-1} \longrightarrow 0)$ leads to:

$$\delta \mathbf{x}_0^{\mathrm{a}} = (\mathbf{G}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{G})^{-1} \, \mathbf{G}^{\mathrm{T}} \mathbf{R}^{-1} \, \mathbf{d}^{\mathrm{f}} = \mathbf{G}^{-1} \, \mathbf{d}^{\mathrm{f}}.$$

  This is equivalent to a least-squares weighting of the observations, the forecast term does not contribute to the analysis and thus, the correction reduces to the inversion of the generalized observation operator $\mathbf{G}$. This case only makes sense if the inverse problem is over-determined (i.e., if there are more observations p than control variables n to estimate); otherwise the term $(\mathbf{G}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{G})^{-1}$ cannot be defined.

- ▷ **Perfect observations:** $(\mathbf{R} \longrightarrow 0)$ leads to:

$$\delta \mathbf{x}_0^{\mathrm{a}} = \mathbf{P}_0^{\mathrm{f}} \mathbf{G}^{\mathrm{T}} \, (\mathbf{G} \mathbf{P}_0^{\mathrm{f}} \mathbf{G}^{\mathrm{T}})^{-1} \, \mathbf{d}^{\mathrm{f}}.$$

  This case is equivalent to a data interpolation (a degenerate case of the variational cost function) and only makes sense if the inverse problem is under-

determined (i.e., if there are more control variables n than observations p); otherwise $\mathbf{G}\mathbf{P}_0^{\mathrm{f}}\mathbf{G}^{\mathrm{T}}$ cannot be defined.

↪ In practice: multi-incremental four-dimensional variational algorithm

In practice, the incremental 4D-Var technique relies on the iterative Gauss-Newton algorithm, with two different levels at which iterations (loops) are performed:

(1) **inner-loops**, whose objective is to perform the minimization of the quadratic cost function presented in Eq. (5.49) based on the model and observation tangent-linear operators. This constitutes one incremental step for an increment $\delta\mathbf{x}_0^{(k)}$; the index k referring to the incremental step.

(2) **outer-loops**, whose objective is to account for the non-linearities in the model over the full uncertain range of the control variables, meaning that the convergence towards the analysis estimate requires a sequence of minimizations of quadratic cost functions (i.e., a sequence of incremental 4D-Var steps) with k varying from 0 to $k_{\max}$. Note that $k_{\max}$ is usually less than 5 for large-dimensional systems.

The resulting multi-incremental 4D-Var algorithm is illustrated in Fig. 5.10.

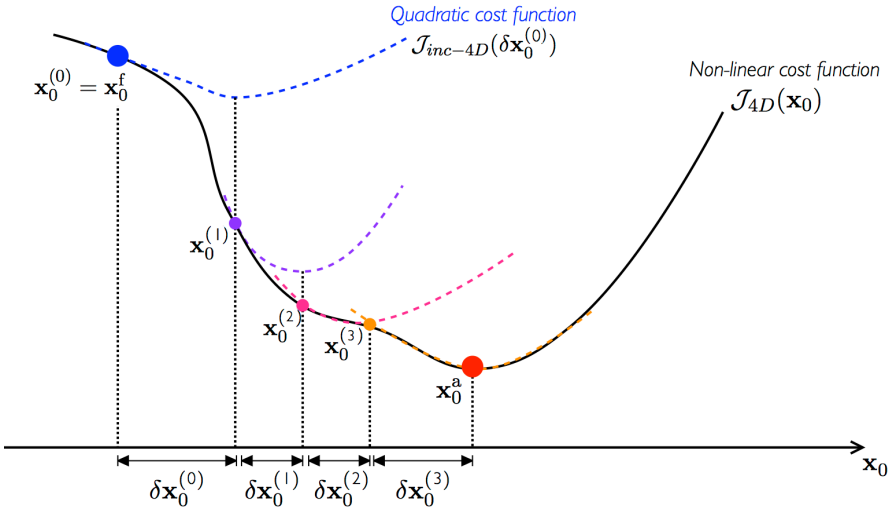


Figure 5.10: *Schematic of the outer loops in the multi-incremental 4D-Var algorithm. The black solid line represents the non-quadratic cost function to minimize due to model non-linearities and non-Gaussian error statistics. Dashed lines represents successive linearizations (outer-loops). A first linearization (blue dashed line) is performed in the vicinity of the forecast $\mathbf{x}_0^{\mathrm{f}}$, the inner loop allows to determine the increment $\delta\mathbf{x}_0^{(0)}$ of the quadratic cost function $\mathcal{J}_{inc-4D}(\delta\mathbf{x}_0^{(0)})$. This increment added to the forecast gives the point $\mathbf{x}_0^{(1)}$. A new linearization can be performed in the vicinity of this new reference point (purple dashed line), providing a new increment and a new reference point $\mathbf{x}_0^{(2)}$. This process is iterated until the minimum of $\mathcal{J}_{4D}(\mathbf{x}_0)$ is found (red dot).*

The formulation of the inverse problem as a sequence of quadratic cost functions (detailed in the following box) ensures that there is a unique minimum. Still, Courtier et al. (1994) showed that the convergence of the multi-incremental 4D-Var algorithm is not guaranteed. Note that the minimization implies that the model tangent-linear $\mathbf{M}_{[t-1,t]}$ and its adjoint $\mathbf{M}^{\mathrm{T}}_{[t-1,t]}$ for any time t are required at each inner loop of the algorithm. Note also that it is possible to perform successive assimilations with the 4D-Var algorithm beyond the time window $[0,T]$. However, in practice the analysis error covariance matrix $\mathbf{P}^{\mathrm{a}}_{T}$ is not explicitly computed and the forecast error covariance matrix for the next assimilation window $[T,2T]$ is taken by default as $\mathbf{P}^{\mathrm{f}}_{T} = \mathbf{P}^{\mathrm{f}}_{0}$. The need for a dynamic estimation of the forecast error covariance matrix motivates the use of hybrid ensemble/variational techniques.

Multi-incremental 4D-Var algorithm over the time window $[0,T]$

▷ Start with the forecast at initial time: $\mathbf{x}^{(0)}_0 = \mathbf{x}^{\mathrm{f}}_0$.

▷ Outer loops: for $k = 1, \cdots, k_{\max}$

(1) Computation of the innovation at initial time 0:

$$\mathbf{d}^{(k-1)}_0 = \mathbf{y}^{\mathrm{o}}_0 - \mathcal{G}_0\Big(\mathbf{x}^{(k-1)}_0\Big).$$

(2) Model trajectory for any time t:

$$\mathbf{x}^{(k-1)}_t = \mathcal{M}_{[0,t]}\Big(\mathbf{x}^{(k-1)}_0\Big), \quad t = 0, \cdots, T.$$

(3) Computation of the innovation at any time t:

$$\mathbf{d}^{(k-1)}_t = \mathbf{y}^{\mathrm{o}}_t - \mathcal{G}_t\Big(\mathbf{x}^{(k-1)}_t\Big), \quad t = 0, \cdots, T.$$

(4) Linearizations of the model and observation operators in the vicinity of $\mathbf{x}^{(k-1)}_t, t = 0, \cdots, T$ based on Eqs. (5.45)-(5.46).

(5) Inner loops (with convergence criterion):

(a) Minimization of the quadratic cost function $\mathcal{J}_{inc-4D}(\delta\mathbf{x}^{(k)}_0)$.

(b) Update of the increment $\delta\mathbf{x}^{(k)}_0$.

(6) Update of the control vector:

$$\mathbf{x}^{(k)}_0 = \mathbf{x}^{(k-1)}_0 + \delta\mathbf{x}^{(k)}_0.$$

▷ Analysis estimate at the initial time: $\mathbf{x}^{\mathrm{a}}_0 = \mathbf{x}^{(k_{\max})}_0$.

▷ Analysis model trajectory along the time window $[0,T]$:

$$\mathbf{x}^{\mathrm{a}}_t = \mathcal{M}_{[0,t]}(\mathbf{x}^{\mathrm{a}}_0), \quad t = 0, \cdots, T.$$

▷ Analysis at the end of the time window (at time T) used as the forecast for the next assimilation cycle $[T,2T]$.

5.5 Conditional mean estimation: Kalman filter

5.5.1 Kalman filter: analytical solution for Bayesian filtering

The Bayesian filtering problem presented in Eq. (5.22) has an analytical solution for $\mathbf{x}_t^{\mathrm{a}}$ under the assumptions of linear model operators $\mathcal{M}_{[t-1,t]}$ (denoted by $M_{[t-1,t]}$) and $\mathcal{G}_t$ (denoted by G_t) as well as of Gaussian error statistics. Under these assumptions, the gradient of the 3D-Var cost function defined in Eq. (5.37) can be formulated as follows:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \underbrace{\mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}}\left(\mathbf{G}_t\mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1}}_{\mathbf{K}_t}\left(\mathbf{y}^{\mathrm{o}} - G_t\,\mathbf{x}_t^{\mathrm{f}}\right), \tag{5.54}$$

with $\delta\mathbf{x}_t^{\mathrm{a}} = \mathbf{K}_t\left(\mathbf{y}^{\mathrm{o}} - G_t\,\mathbf{x}_t^{\mathrm{f}}\right)$ the analysis increment at time t and $\mathbf{K}_t$ the Kalman gain matrix. This analysis formula is the KF basis (its interpretation and demonstration follow in Section 5.5.2); the latter shares this static update with the 3D-Var approach. However, the KF also explicitly computes the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$ through an additional matrix equation. Indeed, under the linearity and Gaussianity assumptions, it can be shown that all forecast and analysis PDF remain Gaussian and thus, the analysis PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ can be exactly represented by its expected value and error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$ such that:

$$\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{a}}\right)^{\mathrm{T}}(\mathbf{P}_t^{\mathrm{a}})^{-1}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{a}}\right)\right\}. \tag{5.55}$$

This implies that the KF directly operates on the error covariances of the control vector $\mathbf{x}_t$ to produce an exact representation of the posterior PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$; $\mathbf{P}_t^{\mathrm{a}}$ is the estimation of the uncertainty in the analysis produced by the KF. Furthermore, the KF propagates the information from one update time to the next; this propagation is subject to possibly uncertain model dynamics, but leads to a sequential update of the control vector as observations becomes available. Thus, the KF exhibits a prediction step and an update step over one assimilation cycle $[t-1, t]$ as in the Bayesian filtering general approach.

↪ Prediction step (forecast)

The prediction step consists in integrating the dynamic model over time, starting from the analysis of the previous assimilation cycle $\mathbf{x}_{t-1}^{\mathrm{a}}$ at time $(t-1)$ and thereby producing the forecast $\mathbf{x}_t^{\mathrm{f}}$ at time t. The forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ is also obtained at the update time t via the propagation of the analysis error covariance matrix $\mathbf{P}_{t-1}^{\mathrm{a}}$ from the previous assimilation cycle. Thus, the prediction step can be summarized as the following set of equations:

$$\mathbf{x}_t^{\mathrm{f}} = M_{[t-1,t]}\mathbf{x}_{t-1}^{\mathrm{a}}, \tag{5.56}$$

$$\mathbf{P}_t^{\mathrm{f}} = \mathbf{M}_{[t-1,t]}\,\mathbf{P}_{t-1}^{\mathrm{a}}\,\mathbf{M}_{[t-1,t]}^{\mathrm{T}}. \tag{5.57}$$

If the model errors $\varepsilon_t^{\mathcal{M}}$ are accounted for, the prediction step of the error covariance matrix includes the model error covariance matrix $\mathbf{Q}_t$ such that Eq. (5.57) becomes $\mathbf{P}_t^{\mathrm{f}} = \mathbf{M}_{[t-1,t]}\,\mathbf{P}_{t-1}^{\mathrm{a}}\,\mathbf{M}_{[t-1,t]}^{\mathrm{T}} + \mathbf{Q}_t$.

↪ Update step (analysis)

Starting from prior information given by the forecast $(\mathbf{x}_t^{\mathrm{f}}, \mathbf{P}_t^{\mathrm{f}})$ and the observations $\mathbf{y}_t^{\mathrm{o}}$ available at time t, the Kalman update equations read:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t\Big(\mathbf{y}_t^{\mathrm{o}} - G_t\mathbf{x}_t^{\mathrm{f}}\Big), \tag{5.58}$$

$$\mathbf{K}_t = \mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}}\Big(\mathbf{G}_t\mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}} + \mathbf{R}\Big)^{-1}, \tag{5.59}$$

$$\mathbf{P}_t^{\mathrm{a}} = \Big(\mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t\Big)\,\mathbf{P}_t^{\mathrm{f}}. \tag{5.60}$$

Equation (5.58) shows that the correction of the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$ (the analysis increment) is proportional to the innovation vector $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - G_t\mathbf{x}_t^{\mathrm{f}}$; the magnitude of this correction being controlled by the gain matrix $\mathbf{K}_t$. This gain matrix defined in Eq. (5.59) accounts for the sensitivity of the observed quantities to changes in the control variables through the linear observation operator G_t, and for the influence of forecast and observation errors through the error covariance matrices $\mathbf{P}_t^{\mathrm{f}}$ and $\mathbf{R}_t$. Besides, Eq. (5.60) provides a posterior estimate of the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$.

5.5.2 Properties and interpretation of the Kalman filter

↪ Scalar example

We consider here a simple example (Maybeck, 1979; Talagrand, 1997; Massart, 2003; Reichle, 2008) to provide insight into the formulation of the KF. The control vector $\mathbf{x}_t$ is reduced to a scalar variable ($n = 1$). The forecast estimate $\mathbf{x}_t^{\mathrm{f}}$ of error variance $(\sigma_t^{\mathrm{f}})^2$ represents prior information; it can be produced from a previous forecast that is valid at the time of the new observation $\mathbf{y}_t^{\mathrm{o}}$. This observation is a scalar quantity ($p = 1$) associated with the error variance $(\sigma_t^{\mathrm{o}})^2$; it represents the same physical quantity as the control variable (meaning that the observation operator G_t is reduced to the identity operator and that its tangent-linear is the identity matrix $\mathbf{I}$). The objective of the KF is to determine the least-square estimate $\mathbf{x}_t^{\mathrm{a}}$ of the true control vector $\mathbf{x}_t^{\mathrm{t}}$ based on available information.

In the update step, the KF searches for the analysis $\mathbf{x}_t^{\mathrm{a}}$ as a linear combination of the available information $\mathbf{x}_t^{\mathrm{f}}$ and $\mathbf{y}_t^{\mathrm{o}}$, i.e.,

$$\mathbf{x}_t^{\mathrm{a}} = k^{\mathrm{f}}\,\mathbf{x}_t^{\mathrm{f}} + k^{\mathrm{o}}\,\mathbf{y}_t^{\mathrm{o}}, \tag{5.61}$$

where k^{f} and k^{o} become the unknowns of the inverse problem. Generally speaking, this equation has a wide range of possible estimators $\mathbf{x}_t^{\mathrm{a}}$. Thus, we constrain the

solution to be an unbiased estimate of the truth $\mathbf{x}_t^{\mathrm{t}}$, leading to:

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{e}_t^{\mathrm{a}}\Big] = \mathbb{E}\Big[\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{t}}\Big] &= \mathbb{E}\Big[k^{\mathrm{f}}\mathbf{x}_t^{\mathrm{f}} + k^{\mathrm{o}}\mathbf{y}_t^{\mathrm{o}} - \mathbf{x}_t^{\mathrm{t}}\Big] \\
&= \mathbb{E}\Big[k^{\mathrm{f}}(\mathbf{x}_t^{\mathrm{t}} + \mathbf{e}_t^{\mathrm{f}}) + k^{\mathrm{o}}(\mathbf{x}_t^{\mathrm{t}} + \boldsymbol{\epsilon}_t^{\mathrm{o}}) - \mathbf{x}_t^{\mathrm{t}}\Big] \\
&= k^{\mathrm{f}}\,\mathbb{E}[\mathbf{e}_t^{\mathrm{f}}] + k^{\mathrm{o}}\,\mathbb{E}[\boldsymbol{\epsilon}_t^{\mathrm{o}}] + (k^{\mathrm{f}} + k^{\mathrm{o}} - 1)\,\mathbb{E}[\mathbf{x}_t^{\mathrm{t}}] \\
&= (k^{\mathrm{f}} + k^{\mathrm{o}} - 1)\,\mathbb{E}[\mathbf{x}_t^{\mathrm{t}}] \\
&= 0, \\
\Rightarrow k^{\mathrm{f}} + k^{\mathrm{o}} &= 1,
\end{aligned} \tag{5.62}$$

using the white-noise assumptions $\mathbb{E}[\mathbf{e}_t^{\mathrm{f}}] = 0$ and $\mathbb{E}[\boldsymbol{\epsilon}_t^{\mathrm{o}}] = 0$ for the forecast and observation errors, respectively. An additional constraint is required to determine the unknowns k^{f} and k^{o}; this constraint is based on the idea that the lower the error variance of the estimate, the more accurate the estimate. The analysis $\mathbf{x}_t^{\mathrm{a}}$ is therefore obtained by minimizing the variance of its distance to the truth. Hence,

$$\begin{aligned}
\mathbb{E}\Big[(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}^{\mathrm{t}})^2\Big] &= \mathbb{E}\Big[\Big(k^{\mathrm{f}}\mathbf{x}_t^{\mathrm{f}} + k^{\mathrm{o}}\mathbf{y}_t^{\mathrm{o}} - \mathbf{x}_t^{\mathrm{t}}\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(k^{\mathrm{f}}(\mathbf{x}_t^{\mathrm{t}} + \mathbf{e}_t^{\mathrm{f}}) + (1 - k^{\mathrm{f}})(\mathbf{x}_t^{\mathrm{t}} + \boldsymbol{\epsilon}_t^{\mathrm{o}}) - \mathbf{x}_t^{\mathrm{t}}\Big)^2\Big] \\
&= (k^{\mathrm{f}})^2\,\mathbb{E}[(\mathbf{e}_t^{\mathrm{f}})^2] + (1 - k^{\mathrm{f}})^2\,\mathbb{E}[(\boldsymbol{\epsilon}_t^{\mathrm{o}})^2] + 2k^{\mathrm{f}}(1 - k^{\mathrm{f}})\,\mathbb{E}[\mathbf{e}_t^{\mathrm{f}} \cdot \boldsymbol{\epsilon}_t^{\mathrm{o}}], \\
\Rightarrow \mathbb{E}\Big[(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}^{\mathrm{t}})^2\Big] &= (k^{\mathrm{f}})^2\,(\sigma_t^{\mathrm{f}})^2 + (1 - k^{\mathrm{f}})^2\,(\sigma_t^{\mathrm{o}})^2.
\end{aligned}$$

The minimization of the variance with respect to the coefficient k^{f} leads to:

$$\frac{\mathrm{d}\mathbb{E}\Big[(\mathbf{x}_t^{\mathrm{a}} - \mathbf{x}^{\mathrm{t}})^2\Big]}{\mathrm{d}k^{\mathrm{f}}} = k^{\mathrm{f}}\,(\sigma_t^{\mathrm{f}})^2 - (1 - k^{\mathrm{f}})\,(\sigma_t^{\mathrm{o}})^2 = 0,$$

$$\Rightarrow k^{\mathrm{f}} = \frac{(\sigma_t^{\mathrm{o}})^2}{(\sigma_t^{\mathrm{f}})^2 + (\sigma_t^{\mathrm{o}})^2}, \quad k^{\mathrm{o}} = \frac{(\sigma_t^{\mathrm{f}})^2}{(\sigma_t^{\mathrm{f}})^2 + (\sigma_t^{\mathrm{o}})^2}, \tag{5.63}$$

using the assumption of uncorrelated forecast and observation errors $\mathbb{E}[\mathbf{e}_t^{\mathrm{f}} \cdot \boldsymbol{\epsilon}_t^{\mathrm{o}}] = 0$ and using Eq. (5.62). The optimal estimate $\mathbf{x}_t^{\mathrm{a}}$ in this scalar example is obtained by searching for an unbiased estimator of minimal variance $(\sigma_t^{\mathrm{a}})^2$, it is called the best linear unbiased estimator (BLUE) and it is also the solution of the 3D-Var cost function expressed here as:

$$\mathcal{J}_{3D}(\mathbf{x}_t) = \frac{(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}})^2}{(\sigma_t^{\mathrm{f}})^2} + \frac{(\mathbf{x}_t - \mathbf{y}_t^{\mathrm{o}})^2}{(\sigma_t^{\mathrm{o}})^2}. \tag{5.64}$$

Thus, the analysis $\mathbf{x}_t^{\mathrm{a}}$ reads:

$$\begin{aligned}
\mathbf{x}_t^{\mathrm{a}} &= \left(\frac{(\sigma_t^{\mathrm{o}})^2}{(\sigma_t^{\mathrm{f}})^2+(\sigma_t^{\mathrm{o}})^2}\right)\mathbf{x}_t^{\mathrm{f}} + \left(\frac{(\sigma_t^{\mathrm{f}})^2}{(\sigma_t^{\mathrm{f}})^2+(\sigma_t^{\mathrm{o}})^2}\right)\mathbf{y}_t^{\mathrm{o}},\\
\Rightarrow \mathbf{x}_t^{\mathrm{a}} &= \frac{(\sigma_t^{\mathrm{o}})^2\,\mathbf{x}_t^{\mathrm{f}} + (\sigma_t^{\mathrm{f}})^2\,\mathbf{y}_t^{\mathrm{o}}}{(\sigma_t^{\mathrm{f}})^2+(\sigma_t^{\mathrm{o}})^2},\\
\Rightarrow \mathbf{x}_t^{\mathrm{a}} &= \mathbf{x}_t^{\mathrm{f}} + \underbrace{\left(\frac{(\sigma_t^{\mathrm{f}})^2}{(\sigma_t^{\mathrm{f}})^2+(\sigma_t^{\mathrm{o}})^2}\right)}_{\mathbf{K}_t}(\mathbf{y}_t^{\mathrm{o}}-\mathbf{x}_t^{\mathrm{f}}).
\end{aligned} \tag{5.65}$$

This implies that the BLUE solution $\mathbf{x}_t^{\mathrm{a}}$ is a weighted-sum of the forecast and the observation. The weights are determined by the relative uncertainties in the model and the observation; they are recast in the Kalman gain matrix $\mathbf{K}_t$, which is here a scalar such that $0 \leqslant \mathbf{K}_t \leqslant 1$. Subsequently, if the measurement error variance $(\sigma_t^{\mathrm{o}})^2$ is small compared to the forecast error variance $(\sigma_t^{\mathrm{f}})^2$, the gain is large and the resulting analysis is close to the observation. To the contrary, if the forecast error variance $(\sigma_t^{\mathrm{f}})^2$ is small compared to the observation error variance $(\sigma_t^{\mathrm{o}})^2$, the gain is close to a zero-value and the analysis remains close to the prior control variable. Equal forecast and measurement error variances lead to the mean value $\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}}/2 + \mathbf{y}_t^{\mathrm{o}}/2$ with equal weights for each source of information (i.e., $\mathbf{K}_t = 0.5$). By reformulating Eq. (5.65), the analysis $\mathbf{x}_t^{\mathrm{a}}$ satisfies:

$$\frac{\mathbf{x}_t^{\mathrm{a}}}{(\sigma_t^{\mathrm{a}})^2} = \frac{\mathbf{x}_t^{\mathrm{f}}}{(\sigma_t^{\mathrm{f}})^2} + \frac{\mathbf{y}_t^{\mathrm{o}}}{(\sigma_t^{\mathrm{o}})^2}, \tag{5.66}$$

using the following expression for the analysis variance $(\sigma_t^{\mathrm{a}})^2$:

$$\frac{1}{(\sigma_t^{\mathrm{a}})^2} = \frac{1}{(\sigma_t^{\mathrm{f}})^2} + \frac{1}{(\sigma_t^{\mathrm{o}})^2}, \tag{5.67}$$

or alternatively,

$$(\sigma_t^{\mathrm{a}})^2 = (1-\mathbf{K}_t)\,(\sigma_t^{\mathrm{f}})^2 = \mathbf{K}_t\,(\sigma_t^{\mathrm{o}})^2. \tag{5.68}$$

If the precision is defined as the inverse of the error variance, Eq. (5.67) demonstrates that the precision of the analysis is the addition of the precisions of the available sources of information (i.e., forecast and measurement). The uncertainty in the analysis estimate is therefore systematically smaller than the error variance of either the forecast estimate or the stand-alone observation. This reflects the increased knowledge with data assimilation about the true control variables $\mathbf{x}_t^{\mathrm{t}}$ by combining available information on the model and observations along with their estimated error statistics.

↪ Generalization to multi-variate random variables

The scalar analysis can be generalized to a framework with n control variables and p measurements, which do not evolve in the same space. Still, the observation operator $\mathcal{G}_t$ is assumed to be linear and is noted G_t. Note that G_t is different from the Jacobian matrix of the observation operator $\mathbf{G}_t$ that describes the dynamics of the perturbations $\delta\mathbf{x}_t$. As in the scalar example, the analysis $\mathbf{x}_t^{\mathrm{a}} \in \mathbb{R}^n$ can be written as a linear combination of the prior information and the observations:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{K}_t^{\mathrm{f}}\,\mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t^{\mathrm{o}}\,\mathbf{y}_t^{\mathrm{o}}, \tag{5.69}$$

where $\mathbf{K}_t^{\mathrm{f}}$ is a $n \times n$ matrix and $\mathbf{K}_t^{\mathrm{o}}$ is a $n \times p$ matrix ($\mathbf{K}_t^{\mathrm{f}}$ and $\mathbf{K}_t^{\mathrm{o}}$ are linear operators). The BLUE solution minimizes the error $\mathbf{e}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{a}} - \mathbf{x}_t^{\mathrm{t}}$ in the sense that it minimizes the variance of the distance of each control variable to the true value. Thus, it minimizes the elements of the trace of the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$. Using the same constraints of zero-mean and minimum-variance estimator as in the scalar example, it can be shown that $\mathbf{K}_t^{\mathrm{f}} = \mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t$ and $\mathbf{K}_t = \mathbf{K}_t^{\mathrm{o}}$. This is the analogy of the scalar relation (5.62) for a non-identity operator $\mathbf{G}_t$. The analysis $\mathbf{x}_t^{\mathrm{a}}$ is expressed as a correction of the forecast $\mathbf{x}_t^{\mathrm{f}}$:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t \underbrace{\left(\mathbf{y}_t^{\mathrm{o}} - G_t\mathbf{x}_t^{\mathrm{f}}\right)}_{\mathbf{d}_t^{\mathrm{f}}}, \tag{5.70}$$

with G_t the linear observation operator and $\mathbf{K}_t$ the gain matrix defined as follows:

$$\begin{aligned} \underbrace{\mathbf{K}_t}_{n\times p} &= \underbrace{\mathbf{P}_t^{\mathrm{f}}}_{n\times n}\underbrace{\mathbf{G}_t^{\mathrm{T}}}_{n\times p}\left(\underbrace{\mathbf{G}_t\mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}} + \mathbf{R}_t}_{p\times p}\right)^{-1} \\ \Longrightarrow \underbrace{\mathbf{K}_t}_{n\times p} &= \left(\underbrace{(\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{G}_t}_{n\times n}\right)^{-1}\underbrace{\mathbf{G}_t^{\mathrm{T}}}_{n\times p}\underbrace{\mathbf{R}_t^{-1}}_{p\times p}. \end{aligned} \tag{5.71}$$

From a dimensional analysis, we can show that the gain matrix $\mathbf{K}_t$ is a linear operator from $\mathbb{R}^p$ to $\mathbb{R}^n$ (a $n \times p$ matrix), meaning that it maps a perturbation in the observation space onto a perturbation in the control space to determine the analysis increment $\delta\mathbf{x}_t^{\mathrm{a}}$ as illustrated in Fig. 5.11.

As in the scalar example, we can deduce from Eq. (5.70) and (5.71) that a perfect confidence in the forecast $\mathbf{x}_t^{\mathrm{f}}$ leads to a zero gain $\mathbf{K}_t$, meaning that the analysis $\mathbf{x}_t^{\mathrm{a}}$ remains equal to the forecast $\mathbf{x}_t^{\mathrm{f}}$. In contrast, if the observations are perfect, then they are the image of the true state by the observation operator G_t. The analysis $\mathbf{x}_t^{\mathrm{a}}$ is in this case the direct solution of the inverse problem $\mathbf{y}_t^{\mathrm{o}} = G_t\,\mathbf{x}_t^{\mathrm{a}}$ independently of the forecast $\mathbf{x}_t^{\mathrm{f}}$. In-between these two limiting cases, Eq. (5.70) shows that the correction of the forecast $\mathbf{x}_t^{\mathrm{f}}$ is proportional to the innovation vector $\mathbf{d}_t^{\mathrm{f}}$; the less accurate the forecast prediction, the larger the innovation vector and the larger the correction to the forecast.

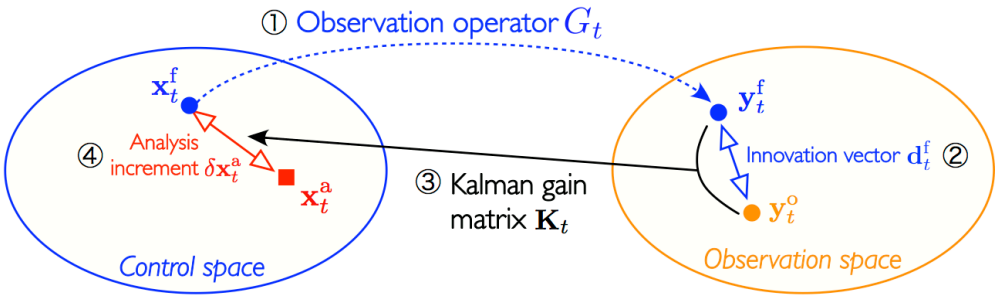


Figure 5.11: *Schematic of the 4 main steps underlying the KF algorithm to compute the gain through the Kalman gain matrix $\mathbf{K}_t$ and to derive the analysis $\mathbf{x}_t^{\mathrm{a}}$.*

In the update step of the KF, the error covariance matrices $\mathbf{P}_t^{\mathrm{f}}$ and $\mathbf{R}_t$ are fundamental ingredients as they provide an estimation of the precisions (confidence) of the forecast and observations. Thus, they directly impact the magnitude of the gain matrix $\mathbf{K}_t$. Equation (5.70) can be alternatively expressed as:

$$\mathbf{x}_t^{\mathrm{a}} = \Big((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{G}_t\Big)^{-1}\Big((\mathbf{P}_t^{\mathrm{f}})^{-1}\mathbf{x}_t^{\mathrm{f}} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{y}_t^{\mathrm{o}}\Big), \tag{5.72}$$

which is the analogy of Eq. (5.65) in a multi-dimensional case: the optimal estimator is found to be the addition of the information weighted by their precisions (taken as the inverse of the error covariance matrices), the estimate being itself weighted by the sum of all the precisions. One difficulty is that the forecast precision $(\mathbf{P}_t^{\mathrm{f}})^{-1}$ and the observation precision $(\mathbf{R}_t)^{-1}$ do not act upon the same space: the forecast is weighted by $(\mathbf{P}_t^{\mathrm{f}})^{-1}$, the observation by $\mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}$ ($\mathbf{G}_t^{\mathrm{T}}$ being applied to map the observation precision onto the control space) and the analysis by $((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{G}_t)$. Thus, the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$ reads:

$$\mathbf{P}_t^{\mathrm{a}} = \Big((\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{G}_t\Big)^{-1} = \Big(\mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t\Big)\mathbf{P}_t^{\mathrm{f}}, \tag{5.73}$$

indicating that the inverse of the error covariance matrices are additive, i.e.,

$$(\mathbf{P}_t^{\mathrm{a}})^{-1} = (\mathbf{P}_t^{\mathrm{f}})^{-1} + \mathbf{G}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{G}_t.$$

Note that the operator $(\mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t)$ measures the reduction in the innovation vector $\mathbf{d}_t^{\mathrm{f}}$ due to data assimilation. The analysis residual $\mathbf{d}_t^{\mathrm{a}}$ is indeed expressed as:

$$\mathbf{d}_t^{\mathrm{a}} = \mathbf{y}_t^{\mathrm{o}} - G_t\,\mathbf{x}_t^{\mathrm{a}} = \Big(\mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t\Big)\Big(\underbrace{\mathbf{y}_t^{\mathrm{o}} - G_t\,\mathbf{x}_t^{\mathrm{f}}}_{\mathbf{d}_t^{\mathrm{f}}}\Big). \tag{5.74}$$

↪ Observation influence in the analysis solution

To diagnose the observation influence on the data assimilation results, the degree of freedom for signal (DFS) has been introduced to quantify the contribution of the observations $\mathbf{y}_t^{\mathrm{o}}$ in the reduction of the error variance in the analysis $\mathbf{x}_t^{\mathrm{a}}$ (Rodgers,

2000; Fisher, 2003; Cardinali et al., 2004; Rabier, 2005). Desroziers and Ivanov (2001) and Chapnik et al. (2006) showed that DFS may be estimated through the evaluation of the trace of the $\mathbf{G}_t\mathbf{K}_t$ matrix at time t, i.e.,

$$\mathrm{DFS}_t = \mathrm{trace}(\mathbf{G}_t\,\mathbf{K}_t), \tag{5.75}$$

with $\mathbf{G}_t$ the tangent-linear of the observation operator $\mathcal{G}_t$ and $\mathbf{K}_t$ the Kalman gain matrix. A partial DFS, associated with a particular subset of observations, can also be determined if the associated error statistics of these observations are not correlated to the rest of the observation errors in the specified observation error covariance matrix $\mathbf{R}_t$.

5.5.3 Similitudes and differences with variational approaches

Within the framework of the KF, the analysis estimate $\mathbf{x}_t^{\mathrm{a}}$ is a feedback information for the dynamic model. $\mathbf{x}_t^{\mathrm{a}}$ is optimal when the variance of its distance to the true $\mathbf{x}_t^{\mathrm{t}}$ gets to a minimum, meaning, for Gaussian cases, that its PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ is dense around its mean. This implies that the error variances of the control variables (i.e., the trace of the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$) are minimized through the update step. This is the BLUE property of the KF that is common to the 3D-Var formulation presented in Eq. (5.37) and to the OI technique (Jazwinski, 1970; Lorenc, 1986; Cohn, 1997). For linear problems, the mean and the mode of the PDF are equivalent, meaning that there is a unique best estimate of the control vector $\mathbf{x}_t$. This also implies that the KF and the weak-constraint variational approaches produce identical estimates at the end of the assimilation time window.

Even though there are many similarities between the 4D-Var approach and the KF (Talagrand and Courtier, 1987; Bouttier and Courtier, 1999), we highlight here some fundamental differences between these two data assimilation techniques:

- ▷ **Filter versus smoother.** In the KF, observations are processed separately and information from a previous update time is explicitly propagated using the dynamic model, meaning that the analysis estimate at a given time is based on all observations up to that time. This a sequential filtering algorithm. In contrast, the 4D-Var approach is an example of smoothing algorithms (Bocquet, 2011), in which measurements at different times within an assimilation time window are processed simultaneously. The analysis estimate at the start of the time window is therefore based on past and future measurements over the time window, which is typical of a smoother. These differences between filtering and smoothing are highlighted in Fig. 5.12: while assimilation is performed at each observation time in a filtering approach, smoothing includes several observation times in one assimilation, leading to smoother model trajectory and uncertainty evolution over the considered time period.

- ▷ **Propagation of error statistics.** The KF also differs from variational approaches by the prediction step that allows for the explicit propagation of error

covariance matrices through the different assimilation cycles. In 4D-Var, the covariance propagation is implicit and only applies within the assimilation cycle; there is no propagation of the error covariance matrices from one assimilation cycle to the next.

▷ **Computational cost.** The evolution of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ is often the most computationally-demanding step of the KF, since it involves large matrix multiplications (Todling and Cohn, 1994). The 4D-Var approach is therefore an attractive operational solution, which is computationally more efficient than the KF for high-dimensional problems such as in numerical weather prediction.

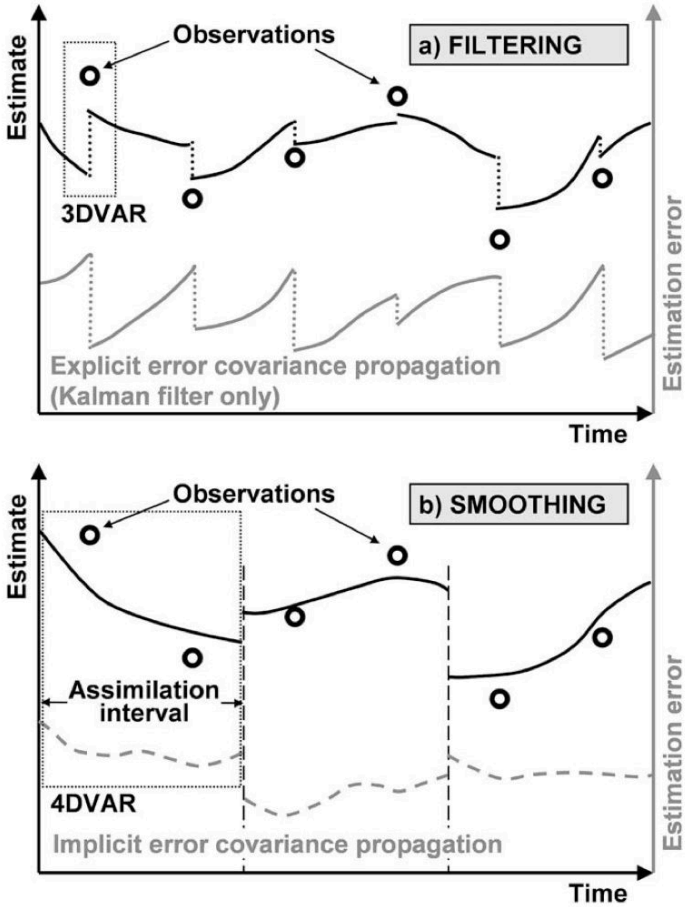


Figure 5.12: *Comparison of continually-operating data assimilation systems based on (a) filtering (e.g., 3D-Var, KF); and (b) smoothing (e.g., 4D-Var). Credit: Reichle (2008).*

5.6 Non-linear extensions of the Kalman filter

While theoretically limited to linear model dynamics and Gaussian error statistics, the KF is the most commonly-used sequential data assimilation technique. However, in many situations of interest, these assumptions do not hold. One question to be addressed is therefore the optimality of its solution when these assumptions are relaxed. Extensions of the KF that partly overcome these limitations have been proposed, for instance the extended Kalman filter (EKF) that uses local linearization techniques in a similar way as the incremental 4D-Var approach (Gelb, 1974), and the ensemble Kalman filter (EnKF) that stochastically represents the time-evolution of the model and observation error statistics (Evensen, 1994; Houtekamer and Mitchell, 1998; Evensen, 2007).

5.6.1 Extended Kalman filter

↪ Prediction and update steps of the extended Kalman filter

The EKF analysis is similar to the incremental 4D-Var analysis since it also relies on the local linearization of the observation operator $\mathcal{G}_t$ and thereby, of the dynamic model operator $\mathcal{M}_{[t-1,t]}$ (Bouttier and Courtier, 1999). Thus, the tangent-linear operators $\mathbf{G}_t$ (Eq. 5.33) and $\mathbf{M}_{[t-1,t]}$ (Eq. 5.31) are only valid in the vicinity of a reference control vector (usually taken as the forecast $\mathbf{x}_t^{\mathrm{f}}$ of the assimilation cycle $[t-1,t]$). The EKF-based prediction step is based on the following equations:

$$\mathbf{x}_t^{\mathrm{f}} = \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}^{\mathrm{a}}), \tag{5.76}$$

$$\mathbf{P}_t^{\mathrm{f}} = \mathbf{M}_{[t-1,t]}\,\mathbf{P}_{t-1}^{\mathrm{a}}\,\mathbf{M}_{[t-1,t]}^{\mathrm{T}}, \tag{5.77}$$

where Eq. (5.76) involves the non-linear model integration over the time period $[t-1,t]$ starting from the analysis at the previous assimilation time $\mathbf{x}_{t-1}^{\mathrm{a}}$. As for the update step, it becomes:

$$\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t\Big(\mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})\Big), \tag{5.78}$$

$$\mathbf{K}_t = \mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}}\Big(\mathbf{G}_t\mathbf{P}_t^{\mathrm{f}}\mathbf{G}_t^{\mathrm{T}} + \mathbf{R}\Big)^{-1}, \tag{5.79}$$

$$\mathbf{P}_t^{\mathrm{a}} = \Big(\mathbf{I}_n - \mathbf{K}_t\mathbf{G}_t\Big)\,\mathbf{P}_t^{\mathrm{f}}. \tag{5.80}$$

Equation (5.79) shows that the tangent-linear of the observation operator $\mathbf{G}_t$ is still involved in the formulation of the gain matrix $\mathbf{K}_t$, while the innovation vector $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})$ is evaluated directly using the non-linear observation operator $\mathcal{G}_t$ in Eq. (5.78). This system of equations means that a linearized and approximate equation is used in the EKF algorithm for the prediction of the error statistics as well as for the mapping between the control space and the observation space in the calculation of the Kalman gain matrix $\mathbf{K}_t$. This implies that the EKF is an approximate non-linear Bayesian filter that provides a consistent analysis if the

local linearization is sufficient to properly describe non-linearities. Note that the limitations of the EKF can be partly overcome with an iterative update of the linearized operators, also called *outer-loops* (Thirel et al., 2010; Harader et al., 2012). Note also that a higher-order EKF that retains more terms in the Taylor expansion exists, but the additional complexity prohibits its development.

↪ Iterative extended Kalman filter (with outer-loops)

In contrast to the classical KF, the EKF is not an optimal estimator due to the use of approximate tangent-linear operators. The EKF linearizes, sometimes unrealistically, the model and observation operators; these linearizations are based on local derivatives that are often difficult to compute reliably (Ros and Borga, 1997). In addition, the EKF may quickly diverge if the forecast estimate $\mathbf{x}_t^{\mathrm{f}}$ is far away from the true control vector $\mathbf{x}_t^{\mathrm{t}}$. For instance, if the associated cost function to minimize is not quadratic and not strictly monotonic, the EKF may encounter local minima and thus, not find the optimal analysis. For this purpose, the non-linearities of the observation operator and the dynamic model are partially accounted for by using an iterative process in a similar way as in the multi-incremental 4D-Var formulation with the outer-loop process.

The outer-loops allow for successive applications of the Kalman update equations, in which the gain matrix $\mathbf{K}_t$ is updated at each iteration k via the calculation of the tangent-linear $\mathbf{G}_t^{(k-1)}$ around a new reference control vector: while the forecast is used as a the reference for the first iteration ($\mathbf{x}_t^{(0)} = \mathbf{x}_t^{\mathrm{f}}$), the reference is taken as the analysis vector $\mathbf{x}_t^{(k-1)}$ from the previous iteration $(k-1)$ for the next iterations. This procedure is repeated until convergence to the optimal analysis $\mathbf{x}_t^{\mathrm{a}}$ is obtained; the number of iterations is typically between 1 and 10. This iterative process (or outer-loops) is presented in Fig. 5.13. Similar to the incremental 3D-Var and 4D-Var algorithms, the idea underlying the outer-loops is to perform a sequence of linear estimations equivalent to a sequence of minimizations of quadratic cost function. The algorithm over the time period $[0, T]$ is detailed in the following box.

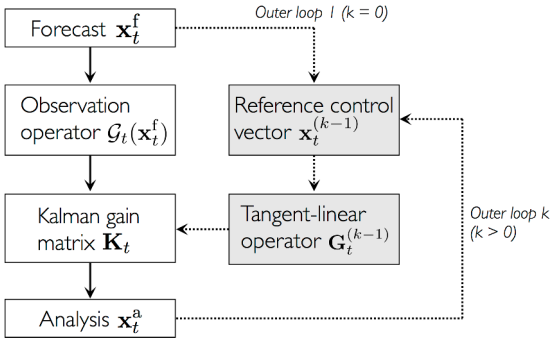


Figure 5.13: *Schematic of the iterative extended Kalman filter (EKF), with outer-loops for the iterative approximation of the tangent-linear observation operator when moving from the forecast estimate $\mathbf{x}_t^{\mathrm{f}}$ to the analysis estimate $\mathbf{x}_t^{\mathrm{a}}$.*

Extended Kalman filter algorithm over the time period $[0, T]$

▷ Start with the prior knowledge of the physical system relying on the control vector $\mathbf{x}_0$ and its error covariance matrix $\mathbf{P}_0$.

▷ Initialization: $\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}_0$, $\mathbf{P}_0^{\mathrm{a}} = \mathbf{P}_0$.

▷ Sequential estimations: for time $t = 1, \cdots, T$

(1) **Prediction step**

(a) Forecast control vector $\mathbf{x}_t^{\mathrm{f}} = \mathcal{M}_{[t-1,t]}(\mathbf{x}_{t-1}^{\mathrm{a}})$.

(b) Forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}} = \mathbf{M}_{[t-1,t]}\, \mathbf{P}_{t-1}^{\mathrm{a}}\, \mathbf{M}_{[t-1,t]}^{\mathrm{T}}$.

(2) **Analysis step**

(a) Calculation of the innovation vector $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})$.

(b) Initialization of the reference control vector, $\mathbf{x}_t^{(0)} = \mathbf{x}_t^{\mathrm{f}}$.

(c) Outer-loops: for $k = 1, \cdots, k_{\max}$

(i) Evaluation of the observation operator $\mathcal{G}(\mathbf{x}_t^{(k-1)})$.

(ii) Linearization of the observation operator in the vicinity of the reference control vector $\mathbf{x}_t^{(k-1)}$ using Eq. (5.33):

$$\mathbf{G}_t^{(k-1)} = \frac{\partial \mathcal{G}_t}{\partial \mathbf{x}_t}\Big|_{\mathbf{x}_t^{(k-1)}}.$$

(iii) Calculation of the gain matrix $\mathbf{K}_t^{(k-1)}$ using Eq. (5.79).

(iv) Update of the control vector using a modified Kalman filter update (Eq. 5.82):

$$\mathbf{x}_t^{(k)} = \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t^{(k-1)} \left(\mathbf{d}_t^{\mathrm{f}} + \mathbf{G}_t^{(k-1)} \left(\mathbf{x}_t^{(k-1)} - \mathbf{x}_t^{\mathrm{f}} \right) \right).$$

(d) Analysis for the assimilation cycle $[t-1, t]$

(i) Analysis estimate at time t: $\mathbf{x}_t^{\mathrm{a}} = \mathbf{x}_t^{(k_{\max})}$.

(ii) Calculation of the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a}}$

$$\mathbf{P}_t^{\mathrm{a}} = \left(\mathbf{I}_n - \mathbf{K}_t^{(k_{\max})} \mathbf{G}_t^{(k_{\max})} \right) \mathbf{P}_t^{\mathrm{f}}.$$

In the iterative version of the EKF, particular attention must be paid to the analysis update equation. The formulation of the analysis at iteration k noted $\mathbf{x}_t^{(k)}$ differs indeed from the classical Kalman update presented in Eq. (5.78). The multi-

incremental 3D-Var cost function at iteration k can be expressed as follows:

$$\begin{aligned}
&\mathcal{J}_{inc-3D}^{(k)}\left(\mathbf{x}_t, \mathbf{x}_t^{(k-1)}\right) \\
&= \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right)^{\mathrm{T}} \left(\mathbf{P}_t^{\mathrm{f}}\right)^{-1} \left(\mathbf{x}_t - \mathbf{x}_t^{\mathrm{f}}\right) \\
&\quad + \frac{1}{2}\left(\mathbf{G}_t^{(k-1)}\,\delta\mathbf{x}_t^{(k-1)} - \mathbf{d}_t^{(k-1)}\right)^{\mathrm{T}} \left(\mathbf{R}_t\right)^{-1} \left(\mathbf{G}_t^{(k-1)}\,\delta\mathbf{x}_t^{(k-1)} - \mathbf{d}_t^{(k-1)}\right)
\end{aligned} \tag{5.81}$$

with $\delta\mathbf{x}_t^{(k-1)} = (\mathbf{x}_t - \mathbf{x}_t^{(k-1)})$ the control deviation from the truth, $\mathbf{d}_t^{(k-1)} = \mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\mathbf{x}_t^{(k-1)})$ the deviation from the observations for the reference control vector $\mathbf{x}_t^{(k-1)}$, and $\mathbf{G}_t^{(k-1)}$ the linearized observation operator around this reference at iteration k. The gradient of this cost function satisfies:

$$\begin{aligned}
&\left(\mathbf{P}_t^{\mathrm{f}}\right)^{-1}\left(\mathbf{x}_t^{(k)} - \mathbf{x}_t^{\mathrm{f}}\right) \\
&\quad + \left(\mathbf{G}_t^{(k-1)}\right)^{\mathrm{T}}\left(\mathbf{R}_t\right)^{-1}\left(\mathbf{G}_t^{(k-1)}\left(\mathbf{x}_t^{(k)} - \mathbf{x}_t^{(k-1)}\right) - \mathbf{d}_t^{(k-1)}\right) = 0, \\
\Rightarrow \mathbf{x}_t^{(k)} &= \mathbf{x}_t^{\mathrm{f}} + \mathbf{K}_t^{(k-1)}\left(\mathbf{d}_t^{\mathrm{f}} + \mathbf{G}_t^{(k-1)}\left(\mathbf{x}_t^{(k-1)} - \mathbf{x}_t^{\mathrm{f}}\right)\right),
\end{aligned} \tag{5.82}$$

using the linear property of the tangent-linear operator $\mathbf{G}_t^{(k-1)}$ at iteration k.

↪ Successes and limitations

While irrelevant for weather forecasting due to its prohibitive computational cost (Gelb, 1974), the EKF has been applied successfully for land surface data assimilation. In particular, it has been integrated into the land surface analysis of the ECMWF global data assimilation system (de Rosnay et al., 2013). Its objective is to estimate the land surface parameters such as soil moisture, surface temperatures, snow and vegetation properties. Furthermore, this EKF approach has also demonstrated its performance for hydrodynamics applications. For instance, Harader et al. (2012) showed that an EKF strategy can correct the radar rainfall forcing of a conceptual hydrological model and be relevant for flood forecasting.

Apart from the non-linearity limits of the EKF, some difficulties might be encountered to model the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ because of storage capacities or simply because the errors do not follow a Gaussian distribution. A large body of literature addresses the modeling of $\mathbf{P}_t^{\mathrm{f}}$ at the start of the assimilation cycle, based on deformations, wavelets transformations (Deckmyn and Berre, 2005; Pannekoucke et al., 2008; Montmerle and Berre, 2010; Michel, 2013b,a) or a diffusion equation (Weaver and Courtier, 2001), in order to properly characterize the correlations between the control variables. Ensemble-based data assimilation techniques such as the EnKF (Evensen, 1994; Houtekamer and Mitchell, 1998) suggest an alternative to stochastically characterize the forecast error covariance matrix and thereby, to better account for the non-linearities in the physics, while also avoiding the storage of the matrix from one assimilation cycle to another.

5.6.2 Ensemble-based Kalman filters

Originally proposed by Evensen (1994) as an alternative to the EKF that cannot deal with severe non-linear dynamics (Evensen, 1992, 1993; Miller et al., 1999), the ensemble Kalman filter (EnKF) is a Monte Carlo approach to stochastically characterize error covariance matrices and thereby the Kalman gain matrix $\mathbf{K}_t$. The EnKF remains based on Gaussian error statistics and still works sequentially from one measurement time to the next, applying in-turn a prediction step and an update step.

A schematic comparison of EKF and EnKF is presented in Fig. 5.14. In contrast to the deterministic EKF, the EnKF does not require the explicit use of linearized operators $\mathbf{M}_{[t-1,t]}$ and $\mathbf{G}_t$ in the prediction step of the KF. The analysis error covariance matrix $\mathbf{P}_{t-1}^{\mathrm{a}}$ from the previous assimilation cycle is not directly propagated to the next update time to produce the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$. Instead, the EnKF approximates the distribution of the forecast control vector using a collection of N_e independent simulations, each simulation (called a *member*) corresponding to a realization of the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$. This collection of N_e members called the ensemble forecast reads:

$$\mathbf{x}_t^{\mathrm{f}} = \left\{ \mathbf{x}_t^{\mathrm{f},(1)}, \mathbf{x}_t^{\mathrm{f},(2)}, ..., \mathbf{x}_t^{\mathrm{f},(N_e)} \right\}.$$

In this ensemble, the k-th member is a n-dimensional multi-variate random vector such that $\mathbf{x}_t^{\mathrm{f},(k)} \in \mathbb{R}^n$. Thus, the EnKF computes the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ from the distribution of the forecast control vectors across the ensemble at time t, instead of explicitly integrating the error covariance from the previous assimilation cycle to the current time t (see Fig. 5.14). Over the assimilation cycle $[t-1, t]$, the EnKF scheme simply consists in the three following main stages:

(1) constructing an initial ensemble by adding independent random perturbations to the control variables at time $(t-1)$;

(2) advancing each ensemble member to the next observation time t to form the forecast ensemble $\mathbf{x}_t^{\mathrm{f}}$;

(3) updating each ensemble member by applying the Kalman update equation that yields the analysis of the ensemble member and contributes to the analysis ensemble $\mathbf{x}_t^{\mathrm{a}}$.

The stages 1 and 2 are part of the prediction step, while the stage 3 constitutes the update step. A detailed view of the EnKF and its relative issues follows.

↪ Prediction step

The propagation of error statistics from time $(t-1)$ to time t is simply performed by propagating each member k of the ensemble using the non-linear dynamic model as follows:

$$\mathbf{x}_t^{\mathrm{f},(k)} = \mathcal{M}_{[t,t-1]}\left(\mathbf{x}_{t-1}^{\mathrm{a},(k)}\right), \quad k = 1, \cdots, N_e. \tag{5.83}$$

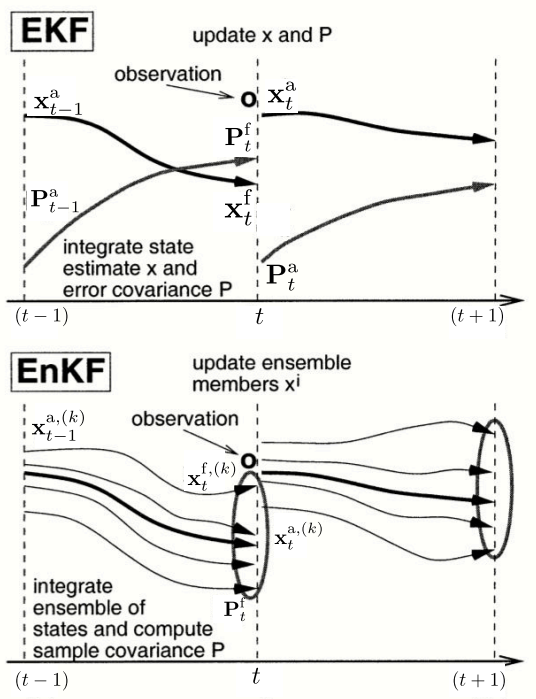


Figure 5.14: *Schematic comparison of the EKF and EnKF algorithms. Credit: Reichle et al. (2002).*

Each member in the ensemble represents a particular realization of the potential model trajectories using the (possibly) non-linear dynamic model $\mathcal{M}_{[t,t-1]}$. This means that the forecast ensemble includes some effects of the non-linear dynamics that are neglected in the EKF. It is worth mentioning that random perturbations have to be introduced in the model parameters and/or in the external forcing to account for the uncertainties present in the physical system over the time period $[t-1, t]$. From a stochastic viewpoint, this is justified by the need to generate a representative sample of the possible future behaviors of the dynamic system and thus, to avoid the ensemble to suffer from sample impoverishment. Ideally, the true control vector $\mathbf{x}_t^{\mathrm{t}}$ should fall within the predicted ensemble spread $\mathbf{x}_t^{\mathrm{f}}$ to guarantee the success of the EnKF. The generation of the ensemble is therefore a crucial step in the EnKF approach.

The forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f}}$ is computed over the ensemble, meaning that it is replaced by the sample covariance matrix of the ensemble noted $\mathbf{P}_t^{\mathrm{f,e}}$, the subscript e referring to *ensemble*. Thus, $\mathbf{P}_t^{\mathrm{f,e}}$ reads:

$$\mathbf{P}_t^{\mathrm{f,e}} = \left(\frac{1}{N_e - 1}\right) \sum_{k=1}^{N_e} \left(\mathbf{x}_t^{\mathrm{f},(k)} - \overline{\mathbf{x}_t^{\mathrm{f}}}\right) \left(\mathbf{x}_t^{\mathrm{f},(k)} - \overline{\mathbf{x}_t^{\mathrm{f}}}\right)^{\mathrm{T}}, \tag{5.84}$$

where the overline denotes the mean value over the ensemble, calculated as:

$$\overline{\mathbf{x}_t^{\mathrm{f}}} = \frac{1}{N_e} \sum_{k=1}^{N_e} \mathbf{x}_t^{\mathrm{f},(k)}. \tag{5.85}$$

More details on the structure of a sample error covariance matrix is provided in Appendix F. Equation (5.84) implies that the ensemble mean (instead of the true control vector $\mathbf{x}_t^{\mathrm{t}}$) is used as the best estimate, and that the ensemble scatter is consistently interpreted as the error standard deviation of the best estimate (Evensen, 1994; Burgers et al., 1998). The accurate estimation of $\mathbf{P}_t^{\mathrm{f}}$ depends on the size of the ensemble N_e and on the representativeness of the actual uncertainties within the ensemble. This last aspect is closely related to the generation of the forecast ensemble at the start of the assimilation cycle $[t-1, t]$.

↪ Update step

In a stochastic framework, measurements also need to be treated as random variables. Following Burgers et al. (1998), a stochastic perturbation is introduced in the original measurements $\mathbf{y}_t^{\mathrm{o}}$ for each ensemble member. This, to avoid the underestimation of the error variance in the control variables and thereby, to prevent sample impoverishment (perhaps leading to ensemble collapse) if all ensemble members are corrected with the same measurements. This process is named *data randomization* in the literature (Houtekamer and Mitchell, 1998). It leads to an ensemble of N_e realizations of the observations with a covariance equal to the measurement error covariance matrix $\mathbf{R}_t$ such that:

$$\mathbf{y}_t^{\mathrm{o},(k)} = \mathbf{y}_t^{\mathrm{o}} + \xi_t^{\mathrm{o},(k)}, \quad \xi_t^{\mathrm{o},(k)} \sim \mathcal{N}(0, \mathbf{R}_t), \quad k = 1, \cdots, N_e, \tag{5.86}$$

where the error realizations $\xi_t^{\mathrm{o},(k)}, k = 1, \cdots, N_e$, are independent of the forecast errors $\mathbf{e}_t^{\mathrm{f}}$. The ensemble of observations is then defined as:

$$\mathbf{y}_t^{\mathrm{o}} = \left\{ \mathbf{y}_t^{\mathrm{o},(1)}, \mathbf{y}_t^{\mathrm{o},(2)}, ..., \mathbf{y}_t^{\mathrm{o},(N_e)} \right\}.$$

Applying the observation operator $\mathcal{G}_t$ to each forecast ensemble member provides N_e model counterparts of the observed quantities designated as:

$$\mathbf{y}_t^{\mathrm{f}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) = \left\{ \mathbf{y}_t^{\mathrm{f},(1)}, \mathbf{y}_t^{\mathrm{f},(2)}, \cdots, \mathbf{y}_t^{\mathrm{f},(N_e)} \right\}. \tag{5.87}$$

Thus, the innovation vector $\mathbf{d}_t^{\mathrm{f},(k)}$ for each ensemble member reads:

$$\mathbf{d}_t^{\mathrm{f},(k)} = \mathbf{y}_t^{\mathrm{o}} + \xi_t^{\mathrm{o},(k)} - \mathbf{y}_t^{\mathrm{f},(k)}, \quad k = 1, ..., N_e. \tag{5.88}$$

Within this ensemble-based framework, the observation error covariance matrix $\mathbf{R}_t$ could be modeled either by its sample version or using the observation error standard deviation σ_t^{o}. Note that $\mathbf{R}_t$ has to be invertible in the KF formalism.

Using the sample version of $\mathbf{R}_t$ could lead to a non-diagonal matrix (especially if the ensemble size N_e remains small), which is more computationally expensive to invert. Thus, the observation error covariance matrix $\mathbf{R}_t$ is usually modeled assuming uncorrelated errors, independently of the ensemble size N_e. This ensures a diagonal, positive, definite matrix, where each diagonal element is equal to the observation error variance. In the following, we assume that all the observations have the same error variance denoted by $(\sigma_t^{\mathrm{o}})^2$, implying that $\mathbf{R}_t = (\sigma_t^{\mathrm{o}})^2 \mathbf{I}_p$, with p the size of the observation vector $\mathbf{y}_t^{\mathrm{o}}$.

During the analysis, each ensemble member is independently updated based on the classical Kalman update Eq. (5.58), with the difference than the generalized observation operator $\mathcal{G}_t$ is non-linear and that the gain matrix (noted $\mathbf{K}_t^{\mathrm{e}}$) is now stochastically calculated using the ensemble-based forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f,e}}$. Each analysis member satisfies:

$$\mathbf{x}_t^{\mathrm{a},(k)} = \mathbf{x}_t^{\mathrm{f},(k)} + \mathbf{K}_t^{\mathrm{e}}\Big(\underbrace{\mathbf{y}_t^{\mathrm{o}} + \xi_t^{\mathrm{o},(k)}}_{\mathbf{y}_t^{\mathrm{o},(k)}} - \mathcal{G}_t\Big(\mathbf{x}_t^{\mathrm{f},(k)}\Big)\Big), \quad k = 1, \cdots, N_e, \tag{5.89}$$

with:

$$\mathbf{K}_t^{\mathrm{e}} = \mathbf{P}_t^{\mathrm{f,e}} \mathbf{G}_t^{\mathrm{T}} \left(\mathbf{G}_t \mathbf{P}_t^{\mathrm{f,e}} \mathbf{G}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1}. \tag{5.90}$$

This step produces an ensemble of analyses $\mathbf{x}_t^{\mathrm{a}}$, from which an analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a,e}}$ can be reconstructed. The ensemble of analyses reads:

$$\mathbf{x}_t^{\mathrm{a}} = \left\{\mathbf{x}_t^{\mathrm{a},(1)}, \mathbf{x}_t^{\mathrm{a},(2)}, \cdots, \mathbf{x}_t^{\mathrm{a},(N_e)}\right\},$$

with:

$$\mathbf{P}_t^{\mathrm{a,e}} = \left(\frac{1}{N_e - 1}\right) \sum_{k=1}^{N_e} \left(\mathbf{x}_t^{\mathrm{a},(k)} - \overline{\mathbf{x}_t^{\mathrm{a}}}\right) \left(\mathbf{x}_t^{\mathrm{a},(k)} - \overline{\mathbf{x}_t^{\mathrm{a}}}\right)^{\mathrm{T}}, \tag{5.91}$$

or alternatively, it can be shown that the analysis error covariance matrix $\mathbf{P}_t^{\mathrm{a,e}}$ satisfies:

$$\mathbf{P}_t^{\mathrm{a,e}} = \left(\mathbf{I}_n - \mathbf{K}_t^{\mathrm{e}} \mathbf{G}_t\right) \mathbf{P}_t^{\mathrm{f,e}}, \tag{5.92}$$

as an analogy to Eq. (5.60) in the standard KF. The application of the EnKF with the data randomization procedure has been shown to produce correct error statistics for the analysis (Burgers et al., 1998; Evensen, 2003). Then, the updated ensemble $\mathbf{x}_t^{\mathrm{a}}$ can be propagated forward in time to the next observation time (using the non-linear dynamic model).

↪ Advantages and limitations

In contrast to EKF, the EnKF does not require the propagation and the storage of the error covariance matrices. They can be reconstructed at each observation time based on the ensemble trajectories, provided the ensemble size N_e is large enough to capture all the important physical correlations between the control variables. Furthermore, Li and Xiu (2009) showed that ensemble errors due to Monte Carlo sampling in EnKF can be dominant compared to numerical or model errors. In order to estimate converged statistics in the error covariance matrices, a sufficiently large number of members N_e is required. Since this technique is well-suited to parallel computing, the member realizations can be computed independently and therefore, simultaneously. Thus, the computational cost is limited (depending on the cost of one model integration). This explains why the development of hybrid methods combining EnKF and variational advantages are currently active research areas in the data assimilation field; they could benefit from variational approaches for the treatment of high-dimensional problems and from the ensemble for the dynamic estimation of error covariances at a limited computational cost.

Still, the EnKF is a Gaussian filter as the EKF. The error statistics are represented using an error covariance model, meaning that only the first two moments of the PDF are accounted for and that the higher-order terms (e.g., skewness, kurtosis) are disregarded. While the error covariance matrix is a perfect model for Gaussian error statistics, the higher-moments are important if the dynamic model and observation operator are highly non-linear, since a non-linear model does not translate a Gaussian PDF for the inputs into a Gaussian PDF for the outputs. Note that some variants of the EnKF have been developed to relax these Gaussian assumptions (Van Leeuwen and Evensen, 1996; Anderson and Anderson, 2003; Beezley and Mandel, 2008).

↪ Variants of the ensemble Kalman filter

▷ **Deterministic ensemble Kalman filters.** Whitaker and Hamill (2002) addressed the sampling issues of the EnKF proposed by Burgers et al. (1998) and suggested that the use of perturbed observations are partly responsible for them. Data randomization was shown to cause systematic errors in the posterior error covariance matrix for a limited size N_e of the ensemble (in particular in weather forecast applications, where N_e remains small due to the computational cost of the CFD atmospheric model). As an alternative, deterministic variants of the EnKF (also referred to as *square-root approaches*) have been proposed to eliminate the need of data randomization and improve the convergence of the analysis compared to the pure Monte Carlo sampling of the standard EnKF (Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002; Tippett et al., 2003; Evensen, 2004). The ensemble transform Kalman filter (ETKF) and the ensemble square-root filter (EnSRF) are typical examples as reviewed in Bonan (2013).

Within the framework of deterministic ensemble-based KF, the ensemble of forecasts/analyses is formulated in terms of the following perturbation matrix $\mathbf{X}_t$:

$$\mathbf{X}_t = \left[\mathbf{x}_t^{(1)} - \overline{\mathbf{x}_t},\ \mathbf{x}_t^{(2)} - \overline{\mathbf{x}_t}, \cdots, \mathbf{x}_t^{(N_e)} - \overline{\mathbf{x}_t}\right]. \tag{5.93}$$

Thus, $\mathbf{X}_t$ measures the deviation of each ensemble member from the ensemble mean $\overline{\mathbf{x}_t}$. Equation (5.84) is then rewritten in terms of the forecast perturbation matrix $\mathbf{X}_t^{\mathrm{f}}$ as follows:

$$\mathbf{P}_t^{\mathrm{f},e} = \left(\frac{1}{N_e - 1}\right) \mathbf{X}_t^{\mathrm{f}} \left(\mathbf{X}_t^{\mathrm{f}}\right)^{\mathrm{T}}. \tag{5.94}$$

The key idea underlying the deterministic ensemble KF is to express the analysis error covariances as a transformation of the forecast error covariances. Subsequently, the analysis is produced in a three-step process:

(i) The perturbation matrix $\mathbf{X}_t^{\mathrm{a}}$ is updated based on the calculation of the transform operator $\mathbf{T}_t$ such that $\mathbf{X}_t^{\mathrm{a}} = \mathbf{X}_t^{\mathrm{f}}\,\mathbf{T}_t$ with (non-unique) $\mathbf{T}_t$ satisfying:

$$\mathbf{X}_t^{\mathrm{a}} \left(\mathbf{X}_t^{\mathrm{a}}\right)^{\mathrm{T}} = \mathbf{X}_t^{\mathrm{f}}\,\mathbf{T}_t\,\mathbf{T}_t^{\mathrm{T}} \left(\mathbf{X}_t^{\mathrm{f}}\right)^{\mathrm{T}}, \tag{5.95}$$

with:

$$\mathbf{T}_t\,\mathbf{T}_t^{\mathrm{T}} = \mathbf{I}_{N_e} - \frac{1}{N_e - 1} \left(\mathbf{X}_t^{\mathrm{f}}\right)^{\mathrm{T}} \mathbf{G}_t^{\mathrm{T}} \left(\frac{1}{N_e - 1}\,\mathbf{G}_t \mathbf{X}_t^{\mathrm{f}} (\mathbf{X}_t^{\mathrm{f}})^{\mathrm{T}} \mathbf{G}_t^{\mathrm{T}} + \mathbf{R}_t\right)^{-1} \mathbf{G}_t \mathbf{X}_t^{\mathrm{f}}.$$

(ii) The mean $\overline{\mathbf{x}_t^{\mathrm{a}}}$ is computed according to the equation:

$$\overline{\mathbf{x}_t^{\mathrm{a}}} = \overline{\mathbf{x}_t^{\mathrm{f}}} + \mathbf{K}_t^{\mathrm{e}} \left(\mathbf{y}_t^{\mathrm{o}} - \mathcal{G}_t(\overline{\mathbf{x}_t^{\mathrm{f}}})\right), \tag{5.96}$$

with $(N_e - 1)\,\mathbf{K}_t^{\mathrm{e}} = \mathbf{X}_t^{\mathrm{f}} (\mathbf{X}_t^{\mathrm{f}})^{\mathrm{T}}\,\mathbf{G}_t^{\mathrm{T}} \Big(\mathbf{G}_t\,\mathbf{X}_t^{\mathrm{f}} (\mathbf{X}_t^{\mathrm{f}})^{\mathrm{T}}\,\mathbf{G}_t^{\mathrm{T}} + (N_e - 1)\,\mathbf{R}_t\Big)$.

(iii) The analysis members can be reconstructed out of deviations from the mean value $\overline{\mathbf{x}_t^{\mathrm{a}}}$ with:

$$\mathbf{x}_t^{\mathrm{a},(k)} = \overline{\mathbf{x}_t^{\mathrm{a}}} + \mathbf{X}_t^{\mathrm{a},(k)}, \quad k = 1, \cdots, N_e, \tag{5.97}$$

with $\mathbf{X}_t^{\mathrm{a},(k)}$ the k-th column of the analysis perturbation matrix $\mathbf{X}_t^{\mathrm{a}}$.

As in the standard EnKF, the ensemble of analyses is made up of linear combinations of the ensemble of forecasts. However, the EnSRF for instance is known to improve, at a fixed ensemble size N_e, the rank of the ensemble using a singular value decomposition for the determination of $\mathbf{T}_t$, since the ensemble members are deterministically chosen to exhibit less linear dependencies and to span the largest possible section of the control space with a limited number of members. The convergence of the EnSRF is therefore increased compared to the EnKF based on Monte

Carlo sampling that exhibits a slow convergence rate $\bigcirc(1/\sqrt{N_e})$ (see Appendix F). Thus, the EnSRF as any deterministic ensemble Kalman filter is an alternative to the Monte-Carlo-based EnKF to accurately estimate the error covariances in the control variables with a limited number of members.

▷ **Unscented Kalman filters.** The objective of unscented filters is to more accurately represent the propagation of the forecast error statistics than the Monte-Carlo-based EnKF, using a minimum set of carefully-chosen members (Julier, 1998; Wan and Van der Merwe, 2000). In particular, they address non-linearity issues in the propagation of error statistics from one analysis time to the next using an *unscented deterministic transformation*. We consider the analysis $\mathbf{x}_t^{\mathrm{a}}$ and its related analysis error covariance matrix $\mathbf{P}_{\mathbf{xx},t}^{\mathrm{a,u}}$ produced by the Kalman update step at time t (this analysis could be obtained by any filtering procedure). The idea underlying the unscented transformation from time t to time $(t+1)$ is to define a minimal set of members $\mathbf{x}_t^{(k)}$ for $k = 1, \cdots, N_u$ (the sigma-points) around the mean value of the analysis ensemble $\overline{\mathbf{x}_t^{\mathrm{a}}}$ at time t to represent the exact PDF of the n-dimensional multi-variate random variable $\mathbf{x}_t$ with $N_u = (2\,n+1)$ members. Then, these sigma-points undergo the non-linear transformation $\mathcal{G}_{t+1}$ up to time $(t+1)$. While capturing high-order information than the EKF and being more efficient than the EnKF, the unscented transformation is useful to reconstruct the PDF of the resulting sigma-points $\mathbf{y}_{t+1}^{(k)}, k = 1, \cdots, N_u$, in the observation space at time $(t+1)$. The related error covariance matrix is noted $\mathbf{P}_{\mathbf{yy},t+1}^{\mathrm{f,u}}$. For this purpose, the N_u sigma-vectors and their related weights are calculated as follows:

$$\mathbf{x}_t^{(1)} = \overline{\mathbf{x}_t^{\mathrm{a}}}, \quad \omega^{(1)} = \frac{\kappa}{n+\kappa}, \tag{5.98}$$

$$\mathbf{x}_t^{(k+1)} = \overline{\mathbf{x}_t^{\mathrm{a}}} + \left(\sqrt{n+\kappa\,\mathbf{P}_{\mathbf{xx},t}^{\mathrm{a,u}}}\right)^{(k)}, \quad \omega^{(k+1)} = \frac{1}{2(n+\kappa)}, \tag{5.99}$$

$$\mathbf{x}^{(k+1+n)} = \overline{\mathbf{x}_t^{\mathrm{a}}} - \left(\sqrt{n+\kappa\,\mathbf{P}_{\mathbf{xx},t}^{\mathrm{a,u}}}\right)^{(k)}, \quad \omega^{(k+1+n)} = \frac{1}{2(n+\kappa)}, \tag{5.100}$$

for k varying between 1 and n (the size of the control space). Note that κ is a scaling coefficient controlling the spread of the sigma-points (see Julier, 1998) and $\left(\sqrt{n+\kappa\,\mathbf{P}_{\mathbf{xx},t}^{\mathrm{u}}}\right)^{(k)}$ represents the k-th column of the matrix square-root. The image of the sigma-vectors in the observation space at time $(t+1)$ is computed through the non-linear observation operator $\mathcal{G}_{t+1}$ as follows:

$$\mathbf{y}_{t+1}^{(k)} = \mathcal{G}_{t+1}\left(\mathbf{x}_{t+1}^{(k)}\right) = \mathcal{H}_{t+1}\left[\mathcal{M}_{[t,t+1]}(\mathbf{x}_t^{(k)})\right], \quad k = 1, \cdots, N_u. \tag{5.101}$$

Based on this ensemble, we can derive the weighted-average of the transformed members $\overline{\mathbf{y}_{t+1}^{\mathrm{f}}}$, which can be regarded as the ensemble mean of the forecasts at

time $(t+1)$:

$$\overline{\mathbf{y}^{\mathrm{f}}_{t+1}} = \sum_{k=1}^{N_u} w^{(k)}\, \mathbf{y}^{(k)}_{t+1}, \tag{5.102}$$

as well as the sample error covariance matrix $\mathbf{P}^{\mathrm{f,u}}_{\mathbf{yy},t+1}$:

$$\mathbf{P}^{\mathrm{f,u}}_{\mathbf{yy},t+1} = \sum_{k=1}^{N_u} w^{(k)} \left(\mathbf{y}^{(k)}_{t+1} - \overline{\mathbf{y}^{\mathrm{f}}_{t+1}}\right) \left(\mathbf{y}^{(k)}_{t+1} - \overline{\mathbf{y}^{\mathrm{f}}_{t+1}}\right)^{\mathrm{T}}. \tag{5.103}$$

The projection of the error statistics from the control space onto the observation space is recast in the form of a $n \times p$ matrix noted $\mathbf{P}^{\mathrm{f,u}}_{\mathbf{xy},t+1}$ and defined as follows:

$$\mathbf{P}^{\mathrm{f,u}}_{\mathbf{xy},t+1} = \sum_{k=0}^{N_u-1} w^{(k)} \left(\mathbf{x}^{(k)}_{t} - \overline{\mathbf{x}^{\mathrm{a}}_{t}}\right) \left(\mathbf{y}^{(k)}_{t+1} - \overline{\mathbf{y}^{\mathrm{f}}_{t+1}}\right)^{\mathrm{T}}. \tag{5.104}$$

As a summary, this unscented transformation allows to accurately represent the forecast control vector $\mathbf{x}^{\mathrm{f}}_{t+1}$ at time $(t+1)$ as well as to capture the non-linear dynamics of the observation operator $\mathcal{G}_{t+1}$ in the model counterparts of the observed quantities $\mathbf{y}^{\mathrm{f}}_{t+1}$ and in its related error covariance matrix $\mathbf{P}^{\mathrm{f,u}}_{\mathbf{yy},t+1} = \mathbf{G}_{t+1}\mathbf{P}^{\mathrm{f,e}}_{t+1}\mathbf{G}^{\mathrm{T}}_{t+1}$. Besides, $\mathbf{P}^{\mathrm{f,u}}_{\mathbf{xy},t+1}$ is the unscented counterpart of the EnKF term $\mathbf{P}^{\mathrm{f,e}}_{t+1}\mathbf{G}^{\mathrm{T}}_{t+1}$. The unscented transformation avoids therefore the calculation of the tangent-linear of the observation operator $\mathbf{G}_{t+1}$ and its adjoint to estimate the Kalman gain matrix. Then, the ensemble of analyses reads:

$$\mathbf{x}^{\mathrm{a},(k)}_{t+1} = \mathbf{x}^{\mathrm{f},(k)}_{t+1} + \mathbf{K}^{\mathrm{u}}_{t+1} \left(\mathbf{y}^{\mathrm{o}}_{t+1} - \mathbf{y}^{\mathrm{f},(k)}_{t+1}\right), \tag{5.105}$$

with the unscented Kalman gain matrix $\mathbf{K}^{\mathrm{u}}_{t+1}$ defined as:

$$\mathbf{K}^{\mathrm{u}}_{t+1} = \mathbf{P}^{\mathrm{f,u}}_{\mathbf{xy},t+1} \left(\mathbf{P}^{\mathrm{f,u}}_{\mathbf{yy},t+1} + \mathbf{R}_{t+1}\right)^{-1}. \tag{5.106}$$

The resulting unscented Kalman filter (UKF) appears as a promising approach to tackle the non-linearity issues of a data assimilation framework, in particular their impact on the error statistics of the random variables projected onto the observation space through the application of non-linear operators. This approach deterministically selects members from the Gaussian approximate PDF of the control vector; these members are then propagated via the non-linear dynamic model and observation model, allowing for an update of the parameters characterizing the PDF. Julier (1998) demonstrated the performance gains of the UKF compared to the EKF in the context of state estimation. Its benefits were also shown for a parameter identification problem in a high-dimensional cardiac biomechanics system (Moireau and Chapelle, 2011). Compared to the EnKF, the convergence is obtained for fewer

ensemble members. To avoid data randomization, the unscented transformation can also been combined to the ensemble square-root approaches as shown by Wan and Van der Merwe (2000). Still, the UKF only deals with Gaussian PDF of the errors as the EKF, EnKF or EnSRF. These extensions of the KF cannot properly describe bi-modal or heavily skewed PDF. For this purpose, more generic Bayesian filters such as particle filters have been developed.

↪ Numerical treatment in ensemble-based Kalman filters

A wide range of studies showed that the straightforward implementation of the EnKF with a relatively small ensemble size can produce inaccurate estimations of covariance matrices and thereby, significantly degrade the filtering performance due to sampling errors and/or uncertainty underestimation (Whitaker and Hamill, 2002). In practice, localization and inflation techniques are commonly introduced in ensemble-based data assimilation algorithms to ensure their convergence and robustness.

▷ **Localization.** Within the framework of state estimation, the sampling error can introduce artificial correlations in the forecast/analysis error covariance matrices, for instance for a variable controlled at distant grid-points (Anderson and Anderson, 2003). Since in operational frameworks the size of the ensemble N_e is limited by computational constraints, a numerical treatment is required to remove these artificial correlations that can significantly degrade the analysis solution. This is the purpose of localized EnKF algorithms. This process of localization (or tapering) typically modifies the sample forecast ensemble covariance matrix $\mathbf{P}_t^{\mathrm{f,e}}$ by tapering-off correlations between spatially-distant grid-points. One simple technique to perform this filtering prior to the Kalman update step is to multiply $\mathbf{P}_t^{\mathrm{f,e}}$ by a correlation matrix generated by a radial shape function monotonically-decreasing when moving away from control variables. In general, the resulting tapered forecast error covariance matrix is sparse with bands of non-zeros near the diagonal elements and replaces $\mathbf{P}_t^{\mathrm{f,e}}$ in the Kalman gain matrix $\mathbf{K}_t^{\mathrm{e}}$. This localization allows to overcome the sampling issues for state estimation problems, even though the ensemble size is small. The local ensemble transform Kalman filter (LETKF) is one advanced data assimilation algorithm combining the EnSRF with localization (Hunt et al., 2007), which has been applied successfully to the WRF meso-scale atmospheric model (see Chapter 1).

▷ **Inflation.** The EnKF is known to underestimate error variances in the ensemble of analyses. To overcome this unrealistic reduction of the error variances during the Bayesian update step and to account for the presence of modeling errors that are difficult to estimate, Anderson and Anderson (2003) and Hamill et al. (2001) proposed to introduce a multiplicative inflation factor β_i ($\beta_i > 1$) in the prediction step such that the ensemble of forecasts satisfies:

$$\mathbf{x}_t^{\mathrm{f},(k)} = \beta_i \left(\mathbf{x}_t^{\mathrm{f},(k)} - \overline{\mathbf{x}_t^{\mathrm{f}}} \right) + \overline{\mathbf{x}_t^{\mathrm{f}}}, \quad k = 1, \cdots, N_e, \tag{5.107}$$

where β_i inflates the forecast ensemble variance in the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f,e}}$, while keeping the forecast mean value $\overline{\mathbf{x}_t^{\mathrm{f}}}$ identical. The optimum value for this multiplicative inflation factor is usually determined using automatic procedures based on maximum likelihood (Mitchell and Houtekamer, 2000; Anderson, 2007; Brankart et al., 2010) or error statistics diagnostics in the observation space (Desroziers et al., 2005).

To summarize, localization and inflation are efficient strategies to reduce unphysical features in the error covariance matrices required by the EnKF, due to ensemble noise and due to uncertainties that are difficult to represent in the ensemble. They highlight that the prediction step in ensemble-based Kalman filters is crucial to produce a reliable ensemble of analyses.

5.7 Non-Gaussian non-linear particle filters

Due to their computational efficiency and flexibility, the EnKF and its multiple variants have recently become a popular data assimilation technique for both parameter and state estimations. However, they still rely on Gaussian assumptions regarding the description of modeling and observation error statistics. Even though the extensions proposed by the unscented and square-root approaches can deal with strong non-linearities, they still rely on the error covariance model to describe the statistical moments of the PDF and cannot be regarded as the exact solution of the Bayesian filtering problem given in Eq. (5.22). In reality, their assumptions can introduce systematic bias if the distributions are significantly non-Gaussian and if the relationship between the control space and the observation space is strongly non-linear. A filtering approach that is fully adapted to non-linearity and non-Gaussianity (at least from a theoretical viewpoint) is the *particle filter*, also referred to as *sequential Monte Carlo filter* in the literature (Gordon et al., 1993; Doucet et al., 2001; Ristic et al., 2004). One of its advantages over the EnKF is that it provides a complete description of the PDF through a point-mass representation (i.e., particles), instead of a limited number of statistical moments as in the KF variants (the covariance is the second-order statistical moment).

5.7.1 Basic principles of particle filters

Based on the idea of combining sequential Bayesian filtering and Monte Carlo simulations, particle filters were first introduced in the 1950s (Hammersley and Hanscomb, 1964) but only became popular recently as they are computationally costly and highly benefit from parallel computing. The basis for most particle filters is called the sequential importance sampling (SIS) technique. Its key idea is to describe the forecast PDF of the control variables $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t)$ as a set of N_e random particles (i.e., equivalent to the members in the EnKF). Each particle is associated with a weight that depends on the weighted-distance to the measurements $\mathbf{y}_t^{\mathrm{o}}$ (i.e., with respect to the observation error statistics). Then, the analysis PDF

$\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ is reconstructed based on this set of weights, a higher weight being given to the particles providing more consistent model trajectories with respect to the observations. By increasing the number of particles, this Monte Carlo characterization becomes an exact representation of the analysis PDF and thereby, approaches the optimal Bayesian estimate.

↪ Importance sampling

We note $\mathbf{x}_t^{(k)}, k = 1, \cdots, N_e$, the N_e particles at time t with associated normalized weights $\omega_t^{(k)}$ satisfying:

$$\sum_{k=1}^{N_e} \omega_t^{(k)} = 1. \tag{5.108}$$

These particles characterize the analysis PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}})$ as discrete-random measures (Arulampalam et al., 2001; Moradkhani and Hsu, 2005) such that:

$$\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t) = \mathrm{p}(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}}) = \sum_{k=1}^{N_e} \omega_t^{(k)} \delta\left(\mathbf{x}_t - \mathbf{x}_t^{(k)}\right), \tag{5.109}$$

with $\delta(\cdot)$ the Dirac delta-function. Within this framework, the Bayesian filtering estimation reduces to the determination of particle weights $\omega_t^{(k)}, k = 1, \cdots, N_e$. Since direct sampling from the analysis PDF is generally difficult (in particular for non-Gaussian cases), these weights are determined in practice using the principle of *importance sampling* (Bergman, 1999; Doucet et al., 2000). Importance sampling can be regarded as a variance reduction technique, which estimates the properties of the target PDF $\mathrm{p}^{\mathrm{a}}(\mathbf{x}_t)$ using an alternative PDF $\mathrm{q}(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}})$. This alternative PDF referred to as the *proposal distribution* gives more weight to the values of the control variables that have the highest probabilities of occurrence and therefore, limits the need for sampling the regions of low probabilities associated with model extreme behaviors (that are difficult to evaluate). It follows that the (importance) weights used in Eq. (5.109) to approximate the unbiased analysis PDF are defined based on the following likelihood ratio:

$$\omega_t^{*(k)} \propto \frac{\mathrm{p}(\mathbf{x}_t^{(k)} \mid \mathbf{y}_t^{\mathrm{o}})}{\mathrm{q}(\mathbf{x}_t^{(k)} \mid \mathbf{y}_t^{\mathrm{o}})}, \tag{5.110}$$

with the following normalization:

$$\omega_t^{(k)} = \frac{\omega_t^{*(k)}}{\left(\sum_{k=1}^{N_e} \omega_t^{*(k)}\right)}. \tag{5.111}$$

The choice of the importance distribution $\mathrm{q}(\mathbf{x}_t \mid \mathbf{y}_t^{\mathrm{o}})$ is a crucial step in the design of a particle filter algorithm (the SIR and ASIR particles filters described below are

based on different importance distributions for instance); a poor-performance may result if the importance PDF is not well chosen.

Within a sequential Bayesian filtering approach, if the time-series of the random variables are considered as discrete-time Markov chains, Eqs. (5.17)-(5.19) are satisfied, meaning that the complete history of the Bayesian updates is not required to compute the importance weights; only the weights and the control variables from the previous assimilation time are necessary. The weights are indeed computed according to the following sequential formulation (Doucet et al., 2000; Arulampalam et al., 2001):

$$\omega_t^{*(k)} \propto \omega_{t-1}^{*(k)} \left\{ \frac{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t^{(k)}\right) \mathrm{p}\left(\mathbf{x}_t^{(k)} \mid \mathbf{x}_{t-1}^{(k)}\right)}{\mathrm{q}\left(\mathbf{x}_t^{(k)} \mid \mathbf{x}_{t-1}^{(k)}, \mathbf{y}_t^{\mathrm{o}}\right)} \right\}, \tag{5.112}$$

where $\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t^{(k)}\right)$ represents the data likelihood with respect to the k-th particle and where $\mathrm{p}\left(\mathbf{x}_t^{(k)} \mid \mathbf{x}_{t-1}^{(k)}\right)$ represents the transition (forecast) PDF from time $(t-1)$ to time t. Equation (5.112) implies that the SIS particle filter can be regarded as the sequential propagation of the weights and particles as each observation becomes available. At time t, two steps can be distinguished:

(1) The particles are drawn from the proposal distribution $\mathrm{q}(\mathbf{x}_t^{(k)} \mid \mathbf{x}_{t-1}^{(k)}, \mathbf{y}_t^{\mathrm{o}})$ for k varying between 1 and N_e.

(2) The associated weights $\omega_t^{(k)}, k = 1, \cdots, N_e$, are determined using Eq. (5.112) and are finally normalized using Eq. (5.111).

5.7.2 Sequential importance resampling filter

↪ Degeneracy issue of the sequential importance sampling filter

The performance of the SIS filter was shown to be significantly deteriorated by the degeneracy issue (Doucet et al., 2000). As the update step selects the particles with the highest likelihood, only a very limited number of particles are left after a few assimilation cycles with a non-zero importance weight and can participate effectively in the description of the PDF of the control variables. Stated differently, the analysis PDF becomes more and more skewed along the prediction/update cycles and thus, the PDF is not sufficiently discretized for a complete description of all its statistical moments. Besides, a large computational effort is devoted to updating particles, whose contribution to the approximation of the analysis PDF is almost zero. A measure of the degeneracy extent is provided by the effective sample size N_{ef} calculated as:

$$N_{ef} = \left[\sum_{k=1}^{N_e} \left(\omega_t^{(k)} \right)^2 \right]^{-1}, \tag{5.113}$$

where the weights are derived using Eqs. (5.111)-(5.112), and where $N_{ef} \leq N_e$ is always satisfied. A low value for N_{ef} indicates a severe degeneracy of the statistical sample.

↪ Resampling step

To overcome this degeneracy problem, Gordon et al. (1993) proposed to add to the SIS algorithm a selection step called *resampling*. The resulting sequential importance resampling (SIR) algorithm eliminates the particles that have negligible importance weights and increases the particles of high importance weights (Kitagawa, 1996). This resampling can be applied either if the number of effective particles falls below a specified threshold number N_{ef}, or systematically at each update step (Liu and Chen, 1998; Bergman, 1999).

The SIR algorithm is presented over the assimilation cycle $[t-1, t]$. $\mathbf{x}_{t-1}^{(k)}$ with k varying between 1 and N_e correspond to the N_e particles, whose normalized weights are $\omega_{t-1}^{(k)}$ at the previous assimilation time $(t-1)$; the update step consists in computing the weights $\omega_t^{*(k)}, k = 1, \cdots, N_e$, at time t. The importance distribution is conveniently taken as the prior (Doucet et al., 2000; Arulampalam et al., 2001) and reads:

$$\mathrm{q}\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)}, \mathbf{y}_t^{\mathrm{o}}\right) = \mathrm{p}\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)}\right). \tag{5.114}$$

Then, Eq. (5.112) simplifies to:

$$\omega_t^{*(k)} \propto \omega_{t-1}^{(k)}\, \mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mathbf{x}_t^{(k)}\right). \tag{5.115}$$

After normalizing the weights using Eq. (5.111), a resampling step is added to eliminate the particles with low importance weights and replicate the particles with high importance weights. This step involves a mapping of the random measure $(\mathbf{x}_t^{(k)}, \omega_t^{(k)})$ into $(\hat{\mathbf{x}}_t^{(k)}, 1/N_e)$ with uniform weights equal to $1/N_e$ as described in the algorithm in the following box. Then, the analysis PDF can be reconstructed using Eq. (5.109) for $(\hat{\mathbf{x}}_t^{(k)}, 1/N_e)$ with $k = 1, \cdots, N_e$.

↪ Successes and limitations

A large number of recent studies have highlighted the performance of the SIR filter, in particular for a state estimation in heat transfer (Orlande et al., 2012) and in hydrodynamics (Jean-Baptiste et al., 2011). For instance, Jean-Baptiste et al. (2011) showed the advantages of the SIR particle filter over the EKF algorithm to reconstruct the unknown upstream flow using real-world water-level measurements, while capturing the non-linear features of the 1-D Saint-Venant equations. Although the resampling step in the SIR filter reduces the effects of degeneracy, an issue known as *sample impoverishment* can lead to a poor-quality SIR estimation due to a lack of diversity in the particles. In this case, resampling selects many times

the same few particles with high importance weights and thus, the particle sample cannot represent the whole stochastic space. This means that for the case of small error statistics, all particles collapse to a single point within a few iterations. This also means that if the dynamic system undergoes a radical change of dynamics from one assimilation cycle to the next, the SIR filter is not able to accurately track these changes. In this case, the sample is no longer representative of all the potential trajectories of the dynamic model. In addition, in the forecast step of the SIR filter, the control space is explored without the information conveyed by the measurements. Indeed, the ensemble of particles only relies on the prior $\mathrm{p}(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})$ for $k = 1, \cdots, N_e$, which is indirectly related to the past measurements up to time $(t-1)$. Thus, the SIR filter relies on a *blind* proposal distribution that ignores the measurements $\mathbf{y}_t^o$ available at time t, in contrast to auxiliary particle filters presented in the following.

Sequential importance resampling (SIR) filter
over the assimilation cycle $[t-1, t]$

(1) **Forecast step**

Draw new particles $\mathbf{x}_t^{(k)}, k = 1, \cdots, N_e$, from the prior PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)})$ used as the importance distribution, see Eq. (5.114).

(2) **Update step**

- ▷ Calculate the corresponding weights $\omega_t^{*(k)}, k = 1, \cdots, N_e$, using Eq. (5.115) based on the likelihood $\mathrm{p}(\mathbf{y}_t^o \mid \mathbf{x}_t^{(k)})$ and using for all members $\omega_{t-1}^{*(k)} = 1/N_e$.
- ▷ Compute the normalized particle weights $\omega_t^{(k)}, k = 1, \cdots, N_e$, using Eq. (5.111).

(3) **Resampling step**

- ▷ Construct the cumulative sum of weights by computing

$$d^{(k)} = d^{(k-1)} + \omega_t^{(k)}, \quad k = 2, \cdots, N_e,$$

  with $d^{(1)} = 0$.
- ▷ Draw a starting point $u^{(1)}$ from the uniform distribution $\mathcal{U}(0, 1/N_e)$.
- ▷ For $l = 1, \cdots, N_e$:
  - (a) Move along the cumulative sum of weights by making

$$u^{(l)} = u^{(1)} + \frac{1}{N_e}(l-1).$$

  - (b) While $u^{(l)} > d^{(k)}$, $k = k + 1$.
  - (c) Assign new particle $\hat{\mathbf{x}}_t^{(l)} = \mathbf{x}_t^{(k)}$ with weight $\omega_t^{(l)} = 1/N_e$.

5.7.3 Auxiliary sequential importance resampling filter

Despite its performance and efficiency to retrieve the true control vector $\mathbf{x}_t^{\mathrm{t}}$, the SIR filter remains computationally intensive since a large number of particles N_e is required to obtain a complete and accurate statistical description of the PDF related to $\mathbf{x}_t$. To overcome these issues, Pitt and Shephard (1999) introduced the auxiliary particle filters, whose main idea is to improve the prior information using an additional (auxiliary) set of particles. For instance, da Silva et al. (2011) applied the auxiliary sequential importance resampling (ASIR) filter to solve a non-linear solification problem, in which synthetically-generated temperature measurements were assimilated to estimate a transient line heat sink as well as the solidification front. Colaço et al. (2011) compared the performance of the SIR and ASIR filters in the estimation of the heat flux applied to a square-cavity in a natural convection problem, while Orlande et al. (2012) validated the SIR/ASIR estimations against the KF and showed in particular the drastic reduction on the number of particles achieved with the ASIR filter compared to the SIR filter. Also, the sequential propagation of modeling errors was studied to improve the choice of the particles at the next assimilation cycle (i.e., at the next observation time in a sequential framework), in particular in the case of joint parameter/state estimation (West, 1993; Doucet et al., 2001).

The ASIR particle filter is an attempt to overcome the drawbacks of the SIR particle filter by performing the resampling step of time $(t-1)$ using the measurements $\mathbf{y}_t^{\mathrm{o}}$ available at time t (see the steps 1 and 2 of the following algorithm). This new resampling technique is based on the determination of some reference point estimate $\mu_t^{(k)}$ that characterizes the control vector $\mathbf{x}_t$ given the particle $\mathbf{x}_{t-1}^{(k)}$ at the previous assimilation time $(t-1)$. Stated differently, this characterization $\mu_t^{(k)}$ provides insight into the forecast PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)})$. This characterization could be taken as the expected value $\mu_t^{(k)} = \mathbb{E}[\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)}]$, or as a sample $\mu_t^{(k)} \sim \mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)})$, depending on the inverse modeling procedure under consideration. The use of such characterization $\mu_t^{(k)}$ means that the SIR and ASIR filters are not based on the same importance distribution. Here, the ASIR filter introduces an importance distribution $\mathrm{q}(\mathbf{x}_t, k \mid \mathbf{y}_t^{\mathrm{o}})$, which samples the pair $\left\{\hat{\mathbf{x}}_t^{(l)},\ k^l\right\}, l = 1, \cdots, N_e$, where k^l refers to the index of the particle called the *parent particle*. A weight is assigned to each member $l = 1, \cdots, N_e$ using the following formulation:

$$\omega_t^{*(l)} \propto \omega_{t-1}^{(k^l)} \frac{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \hat{\mathbf{x}}_t^{(l)}\right)\ \mathrm{p}\left(\hat{\mathbf{x}}_t^{(l)} \mid \mathbf{x}_{t-1}^{(k^l)}\right)}{\mathrm{q}\left(\hat{\mathbf{x}}_t^{(l)}, k^l \mid \mathbf{y}_t^{\mathrm{o}}\right)},$$

$$\Longrightarrow \omega_t^{*(l)} = \frac{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \hat{\mathbf{x}}_t^{(l)}\right)}{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mu_t^{(k^l)}\right)}. \tag{5.116}$$

The ASIR filter shares with the SIR filter, the objective of updating the importance weights of the particles to reduce the mismatch between the measurements and their model counterparts. An analogy can be performed with the calculation of the importance weights for the SIR filter given in Eq. (5.115); for the ASIR filter, the observations are conditioned upon the characterization $\mu_t^{(k^l)}$ instead of the particle $\mathbf{x}_t^{(k)}$. However, auxiliary particles are introduced to select the particles of high importance weights $\omega_t^{*(k)}$ in the steps 1 and 2 of the ASIR algorithm; the particles with very low weights at the previous assimilation time are not resampled. Thus, the effective number of particles required to perform an accurate update step is significantly reduced in comparison to the SIR filter. If the error in the dynamic model is small, then the forecast PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)})$ is generally well-characterized by the point estimate $\mu_t^{(k)}$ (the weights are not significantly spread out and the ASIR filter is less sensitive to outliers than the SIR filter). In constrast, if the error in the dynamic model is large, then the single point estimate $\mu_t^{(k)}$ is not sufficient to capture the features of the forecast PDF and thus, the ASIR filter may not be as effective as the SIR filter. One advantage of the ASIR over the SIR is that the forecast PDF $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k)})$ naturally depends on the ensemble at time $(t-1)$, which, conditioned upon the current measurement $\mathbf{y}_t^{\mathrm{o}}$, are closer to the truth $\mathbf{x}_t^{\mathrm{t}}$.

While particle filters are theoretically more general solutions of the Bayesian estimation problem, their application to operational framework is limited due to the large computational cost of the underlying Monte Carlo sampling technique, which grows with the complexity of the physical system. To address this issue, more advanced algorithms than the SIR and ASIR filters have been developed (Ristic et al., 2004). These algorithms can reduce the number of particles required for an appropriate description of the PDF of interest, thus resulting in the reduction of the computational time, especially when associated with HPC. Besides, surrogate models or response surfaces for the solution of the forward dynamic model and the observation operator appear as promising approaches to limit the computational cost of Monte-Carlo-based Bayesian filtering.

Auxiliary sequential importance resampling (ASIR) filter over the assimilation cycle $[t-1, t]$

(1) **Forecast step**

▷ Start with the ensemble $\left\{\mathbf{x}_{t-1}^{(k)},\ \omega_{t-1}^{(k)}\right\}, k = 1, \cdots, N_e$ deriving from the previous analysis time $(t-1)$.

▷ For $k = 1, \cdots, N_e$:

- Calculate some characterization $\mu_t^{(k)}$ of the control vector $\mathbf{x}_t$ given the particle $\mathbf{x}_{t-1}^{(k)}$ at the previous analysis time $(t-1)$.
- Calculate the corresponding weight $\omega_t^{*(k)}$ based on the observations conditioned upon the characterization $\mu_t^{(k)}$, i.e.,

$$\omega_t^{*(k)} = \mathrm{q}\left(k \mid \mathbf{y}_t^{\mathrm{o}}\right) \propto \mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mu_t^{(k)}\right) \omega_{t-1}^{(k)}.$$

- Compute the normalized weight $\omega_t^{(k)}$ using Eq. (5.111).

(2) **Resampling step**

▷ Construct the cumulative sum of weights by computing:

$$d^{(1)} = 0, d^{(k)} = d^{(k-1)} + \omega_t^{(k)}, \quad k = 2, \cdots, N_e,$$

▷ Draw a starting point $u^{(1)}$ from the uniform distribution $\mathcal{U}(0, 1/N_e)$.

▷ For $l = 1, \cdots, N_e$:

(a) Move along the cumulative sum of weights as follows:

$$u^{(l)} = u^{(1)} + \frac{1}{N_e}(l-1).$$

(b) While $u^{(l)} > d^{(k)}$, $k = k + 1$.

(c) Assign new particle $\hat{\mathbf{x}}_t^{(l)} = \mathbf{x}_t^{(k)}$ with weight $\omega_t^{(l)} = 1/N_e$.

(d) Assign particle parent $k^l = k$.

(3) **Update step**

▷ Draw new particles $\hat{\mathbf{x}}_t^{(l)}, l = 1, \cdots, N_e$, from the importance distribution $\mathrm{p}(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(k^l)})$ using the parent k^l.

▷ Use the likelihood density to assign the corresponding weights:

$$\omega_t^{*(l)} = \frac{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \hat{\mathbf{x}}_t^{(l)}\right)}{\mathrm{p}\left(\mathbf{y}_t^{\mathrm{o}} \mid \mu_t^{(k^l)}\right)}, \quad l = 1, \cdots, N_e.$$

▷ Compute the normalized importance weight $\omega_t^{(l)}, l = 1, \cdots, N_e$, using Eq. (5.111).

Conclusion

This chapter provides an overview of data assimilation methodologies that have been developed in a wide range of geosciences and engineering applications over the past decades. The Kalman filter (a conditional mean estimator) is the most widely used sequential data assimilation technique. However, this filter assumes model linear dynamics and a Gaussian statistical distribution for both modeling and observation errors. Extensions of the Kalman filter that overcome in part these limitations have been proposed, for instance the extended Kalman filter (EKF) based on local linearization techniques or the ensemble Kalman filter (EnKF) and its variants based on a stochastic description of the model behavior and error statistics. The principles of Kalman filtering has been extensively compared to variational approaches (3D-Var, 4D-Var and incremental versions) used for instance in numerical weather prediction and oceanography. Variational approaches correspond to a conditional mode estimation that is more suitable for high-dimensional problems and that is usually regarded as a deterministic viewpoint of data assimilation, in contrast to EnKF.

In contrast to the Kalman filter and its extensions, particle filters have been specifically developed to deal with non-linear models and non-Gaussian errors (as summarized in the following table) through the determination of particle weights that are useful to retrieve any shape of probability density functions.

Technique	**Gaussian error statistics**	**Model linearity**
Variational	x	
KF	x	x
EKF	x	(linearization)
EnKF	x	(stochastic representation)
Particle filter		

The merits of data assimilation have already been greatly demonstrated in meteorology and oceanography for providing initial conditions for numerical forecast. Since recent progress made in airborne and spaceborne remote sensing provides new ways to monitor real-time fire front positions, data assimilation appears as an efficient framework to formulate some feedback information on the wildfire dynamics and to achieve data-driven forecasts of regional-scale fire spread.

Chapter 6

Data assimilation strategy for wildfire spread

Because wildfire spread involves multiple physical processes through multiple scales, our ability to predict the behavior of wildfires at large regional scales (i.e., at scales ranging from a few tens of meters up to several kilometers) remains limited. In this work, a data assimilation methodology is considered to overcome some current limitations of wildfire spread forecast. In fire research data assimilation is particularly attractive given the large uncertainties associated with many of the input variables of the models, in particular in the representation of fuel sources. For wildfires, these uncertain input variables represent the environmental conditions in which the fire propagates (e.g., wind conditions, terrain topography, moisture content and intrinsic properties of the vegetation). Thus, data assimilation provides an attractive framework for integrating wildfire sensor observations into computer models; this framework explicitly accounts for the effects of both observation and modeling errors and thereby, aims at improving predictions of wildfire behavior. This idea of data assimilation has been explored recently by several research groups, both for wildland and building fire applications, but still remains original in the field of fire and combustion.

In this chapter, the objective is to provide an overview of the recent developments in data-driven fire spread modeling as well as to present the data assimilation strategy and contributions of this thesis.

Outline

6.1 Recent developments in data-driven fire modeling

The present work builds upon several recent studies that have considered data assimilation and inverse modeling for improved fire modeling performance, in both building fire and wildland fire applications.

6.1.1 Building fire applications

In building fire applications, data-driven models have received significant interest because of their potential benefits for detection, spotting and sizing of incipient, growing or fully-developed fires. Early studies (Richards et al., 1997; Davis and Forney, 2001; Lee and Lee, 2005) typically considered ceiling-mounted heat or smoke detectors providing information on the temperature or composition of the ceiling jet generated by small fires. Sensor data are continuously monitored and compared to results from a fire model (i.e., a ceiling jet algorithm or a zone model[1]) and this comparison yields information on the fire size and location. More recent studies (Neviackas and Trouvé, 2007; Leblanc and Trouvé, 2009; Koo et al., 2010;

[1] A zone model corresponds to a solution with an intermediate level of complexity that makes it faster to run than a CFD model.

Cowlard et al., 2010; Jahn, 2010; Beji et al., 2011, 2012; Jahn et al., 2012) considered extensions of earlier work to steady or unsteady fully-developed fire conditions, multi-compartment fire scenarios, under-ventilated fire conditions, using both zone models (Neviackas and Trouvé, 2007; Leblanc and Trouvé, 2009; Koo et al., 2010; Beji et al., 2011, 2012) and CFD models (Jahn, 2010; Jahn et al., 2012). Beji et al. (2011) motivated the use of a zone model to produce fire forecasts with a reasonable lead-time. These studies typically provide estimates of the time variations of the heat release rate, including estimates of future variations. Note that none of these studies considered the full framework provided by data assimilation theory; for instance the effects of modeling and observation errors are neglected (these errors were implicitly assumed to be small).

6.1.2 Wildland fire applications

In wildland fire applications, data-driven modeling is proposed as one of the two cornerstones of the fire spread forecasting capacity (Mandel et al., 2008). The other cornerstone corresponds to the integration of a variety of in-situ and remote sensors providing (real-time) information on fire location, vegetation (i.e., the fuel sources), terrain topography and atmospheric conditions (Mandel et al., 2011, 2012).

↪ State estimation applied to the surface temperature field

Mandel et al. (2008), Beezley and Mandel (2008) and Beezley (2009) applied data assimilation to wildfires as a state estimation problem, in which the control variable is the ground surface temperature field and in which (synthetic) measurements are taken as the surface temperature at a finite number of locations in the computational domain. Early studies (Mandel et al., 2008) implemented a standard EnKF algorithm and generated the members of the ensemble based on random perturbations of the temperature field (corresponding to one model state variable in a system of reaction-diffusion partial differential equations, in which the temperature and the mass fraction of fuel were evolved in space and time). The ensemble of temperature fields resulted not only from perturbations in magnitude, but also from spatial shifts of the burnt area as illustrated in Fig. 6.1. This means that perturbed temperature fields were moved spatially in both x- and y-directions to allow for spatial displacement of the burnt area by the ensemble Kalman filter (EnKF) update. It was found that the standard EnKF is able to correctly track synthetically-generated measurements, even though the forecast ensemble started at an erroneous ignition location that was far away from the location of the true fire, see Fig. 6.2.

While Mandel et al. (2008) showed the feasibility of applying data assimilation to wildfire spread, further studies (Beezley and Mandel, 2008; Beezley, 2009) showed that the standard EnKF may fail to track the true location of the active flame zones in a certain number of configurations and also produce non-physical features.

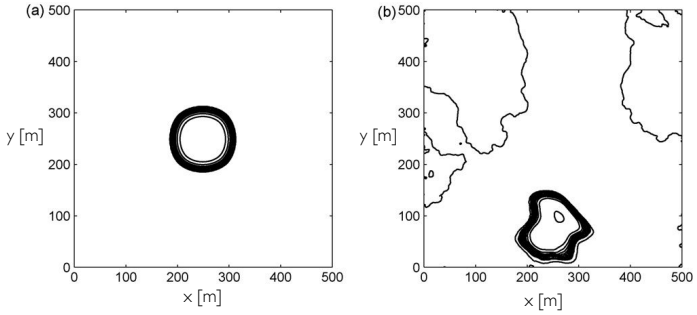


Figure 6.1: *Iso-temperature contours each 100 K. (a) True (reference) temperature profile of a fire ignited as a circular burnt area. (b) Perturbed temperature profile (in magnitude and spatial positions) corresponding to a member of the ensemble. Credit: Mandel et al. (2008).*

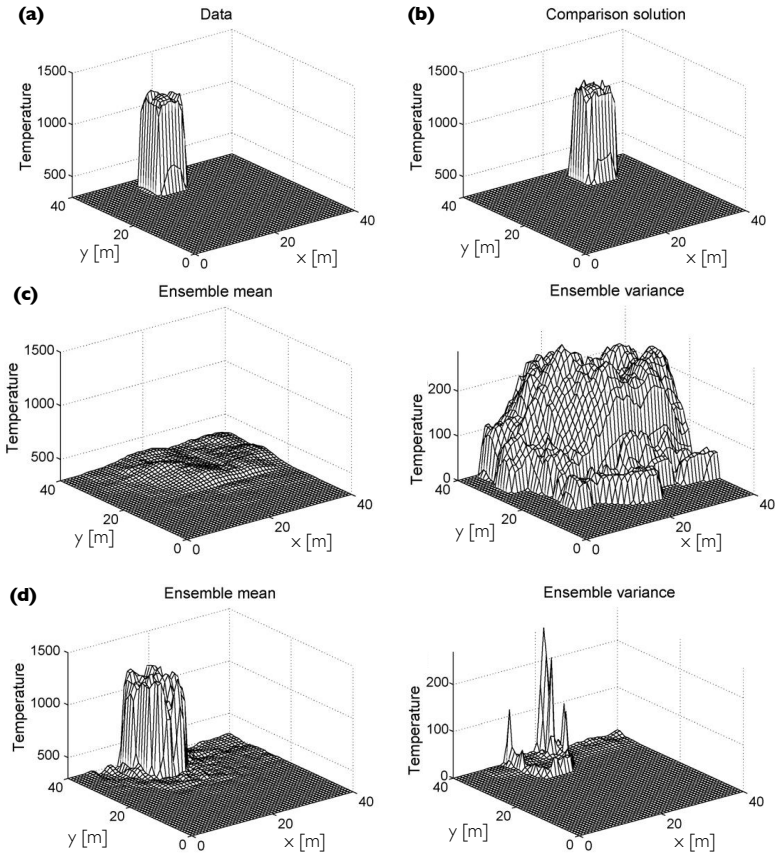


Figure 6.2: *Temperature profiles 100 s after ignition. (a) True (reference) solution. (b) Unperturbed member of the ensemble for comparison. (c) Ensemble mean and standard deviation of the forecast ensemble (prior to data assimilation). (d) Ensemble mean and standard deviation of the analysis ensemble (posterior to data assimilation). Credit: Mandel et al. (2008).*

As highlighted by several studies in hurricane position forecasting (Chen and Snyder, 2006; Wu et al., 2010), the standard EnKF works only when the increment in the location remains small and fails when there are uncertainties in the spatial location of time-evolving sharp coherent features. The main reasons for failures of the standard EnKF when applied to a state estimation problem in wildfires are that: (1) the EnKF mainly corrects the temperature magnitude and not the location of the fireline; and (2) the EnKF relies on a Gaussian assumption for the error statistics of the temperature field. However, this control variable is characterized by a bi-modal PDF in the vicinity of the flame zone (i.e., burning state or non-burning state). Thus, in order to satisfy the Gaussian assumption in the EnKF, the idea of morphing from image processing was introduced by Beezley and Mandel (2008). The resulting morphing ensemble Kalman filter (MEnKF) manipulates Gaussian random variables within the EnKF implementation, while still indirectly distorting and deforming the temperature fields in space (Hoffman et al., 1995; Davis et al., 2006; Alexander et al., 1998; Lawson and Hansen, 2005; Ravela et al., 2007). The morphing procedure, also called *registration* or *warping* in image processing, consisted in mapping the ensemble of perturbed temperature fields onto a unique reference frame. After the spatial transformation of the temperature fields, additive magnitude corrections could be efficiently performed. In the MEnKF algorithm, the estimation targets are the ensemble of morphing transformations of simulated temperature fields. A simple example of the MEnKF algorithm is given in Fig. 6.3.

The performance of the MEnKF over the standard EnKF is illustrated in Fig. 6.4 for a data assimilation prototype based on the coupled simulator WRF-Fire (see Chapter 2). In this synthetic data assimilation experiment, the ensemble was formed by $N_e = 25$ members and the sensible heat flux was used as the morphing variable. Iso-contours of the heat flux in Fig. 6.4(c) show the unphysical results produced by the standard EnKF. While assimilating field variables at a limited number of computational grid-points, the standard EnKF is not able in this experiment to correct the position of the fire front (the burning area remains close to the forecast mean) since the ensemble generation is based on perturbations of the heat flux magnitude. Such ensemble cannot provide a correction on the topology of the fire front, but a linear combination of heat flux magnitudes that leads to non-zero heat flux values over a large fraction of the computational domain. These non-zero heat flux values correspond to the previously-mentioned unphysical values. To the contrary, the morphing EnKF (MEnKF) produced an analysis ensemble that is closer to the true burning area in Fig. 6.4(d). Still, the morphing correction was imperfect and led to technical difficulties in the EnKF implementation, as it relied on an expensive non-linear optimization algorithm. In addition, the MEnKF algorithm was not tested against actual measurements of wildfire spread. Actual measurements is expected to be more noisy than synthetically-generated measurements that were used in these studies and to exhibit a more complex wildfire behavior (e.g., splitting and merging of flame fronts), making the morphing procedure even more difficult to handle.

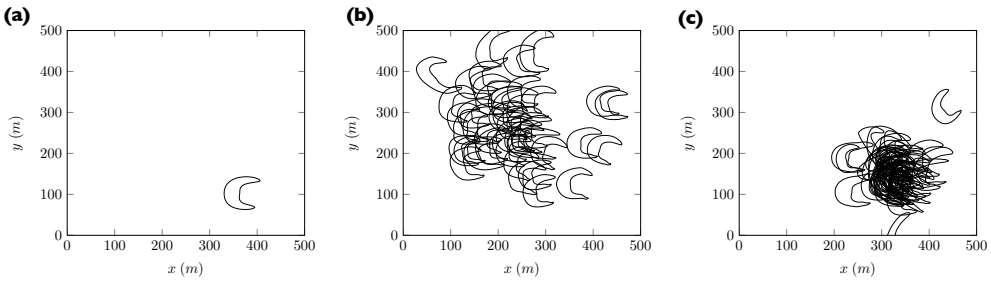


Figure 6.3: *Example of MEnKF application. The iso-temperature contour at 800 K indicates the location of the fire front. (a) True (reference) solution. (b) Ensemble of forecasts generated by smooth random morphing of the initial temperature field (prior to data assimilation). (c) Ensemble of analyses (posterior to data assimilation). Credit: Beezley and Mandel (2008).*

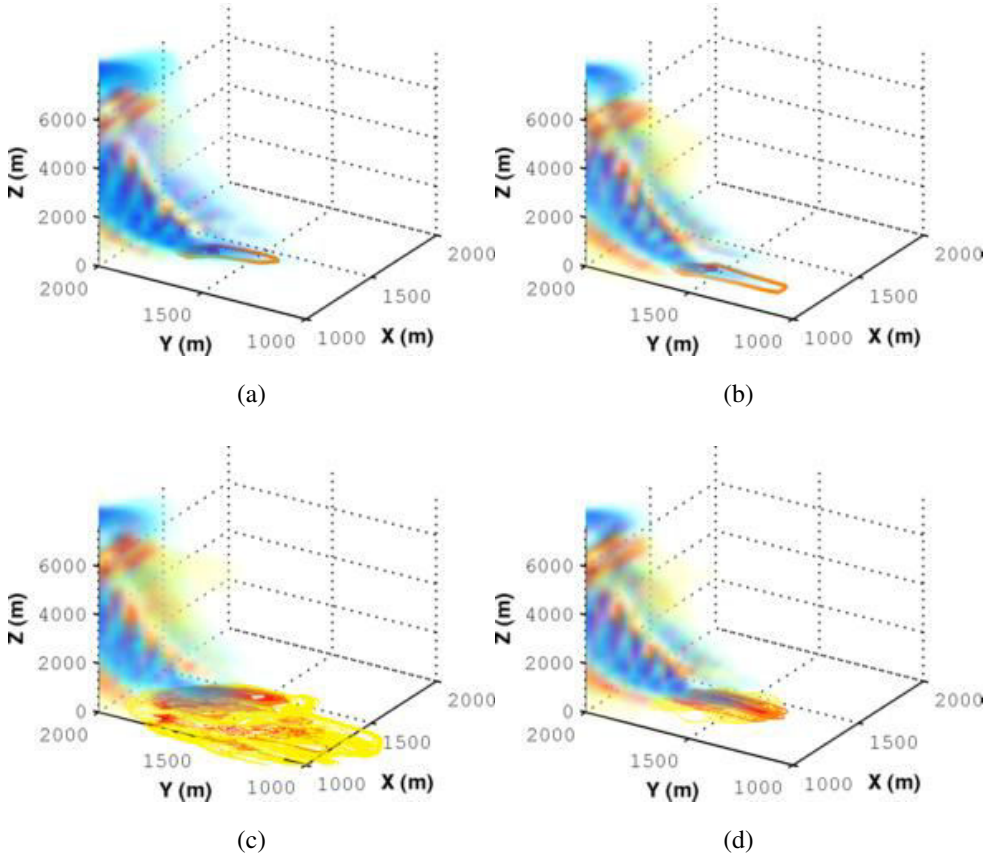


Figure 6.4: *Example of ensemble-based KF applications to the* WRF-Fire *coupled fire/atmosphere simulator. Colors on the horizontal plane corresponds to the fire output heat flux; the volume shading corresponds to atmospheric vorticity in the ensemble mean. (a) True (reference) solution. (b) Example of perturbed temperature fields. (c) Analysis given by the standard EnKF. (d) Analysis given by the morphing EnKF (MEnKF). Credit: Beezley (2009).*

↪ Parameter batch-calibration/estimation of environmental conditions

An alternative to state estimation is to directly estimate the variables that are responsible for the uncertainties in wildfire spread simulations, namely the input parameters and/or the external forcing of the rate of fire spread (ROS). Parameter batch-calibration is commonly achieved manually in Farsite (Finney, 1998) through specification of a ROS adjustment factor, leading to an improved agreement between model simulations and measurements. Beyond this manual calibration, a wide range of techniques can be used to perform parameter batch-calibration (static) or sequential parameter estimation (dynamic).

Denham et al. (2012) showed that automatic optimization algorithms have the potential to significantly improve the accuracy of wildfire spread simulations and thereby, to improve the forecast of the location of wildfire active areas. This study demonstrates, for a simplified scenario of wildfire spread, the good-quality performance of a parallel data-driven genetic algorithm to retrieve the input parameters of a surface fire spread model (fireLib[2], i.e., a cellular-automata software derived from Behave). In this study, the calibrated parameters were the wind direction and magnitude as well as the moisture content of the dead and/or living biomass fuels. This proposed sequential genetic algorithm selects the most realistic combinations of spatially-uniform control parameters among the initial population (i.e., equivalent to ensemble members in the EnKF and to sample particles in particle filters) using a two-stage prediction. A pre-search of the probable values for the input parameters was shown to improve the quality of the update and to significantly reduce the error in the analysis. This idea is similar to the use of an auxiliary set of particles in the ASIR particle filter that improves the quality of prior information (see Section 5.7, Chapter 5).

Lautenberger (2013) followed the same idea as Denham et al. (2012) and applied a genetic algorithm to the 2007 Moonlight fire. The proposed genetic algorithm was validated for the calibration of material properties in pyrolysis modeling (Lautenberger et al., 2006). Here, in order to track the real-world wildfire event, this study calibrated simultaneously 10 input parameters of the fire spread simulator ELMFire[3], using in-situ measurements from the US Forest Service combined with satellite-based fire detection data from MODIS. It is worth noting that ELMFire is a stand-alone surface fire spread simulator and the weather conditions were specified as external forcing using the outputs of the WRF numerical weather model in a non-coupled mode. ELMFire relies on a ROS parameterization due to Rothermel, including complex effects such as canopy torching, ember spotting and suppression actions. The calibrated parameters were the ROS adjustment factor, the wind reduction factor, the canopy properties, the torching time and the ember ignition probability, among others. Figure 6.5 shows that the simulation derived from parameter batch-calibration is in good agreement with the measurements (used as

[2] www.frames.gov/rcs/0/935.html.
[3] Eulerian Level-set Model for Fire spread, reaxengineering.com/trac/elmfire.

the calibration targets in the genetic algorithm).

In contrast to classical data assimilation techniques (e.g., Kalman filter and variants, particle filters), genetic algorithms do not introduce a forecast term in the optimization process; they only search for the combinations of input parameters that best fit the available measurements and do not account for the probabilistic dimension of the control variables and observations. Also, in the tangent-linear model proposed by Rios (2013) following work by Cowlard et al. (2010) and Jahn et al. (2012), the input parameters (e.g., wind magnitude and direction) are calibrated based on the minimization of a cost function that only accounts for the discrepancies between observations and model predictions. No background term that measures the deviation to the prior estimate is introduced to regularize the inverse problem. Furthermore, observation and modeling error statistics are not taken into account to find the most optimal set of control parameters. However, due to the complexity of multi-scale multi-physical wildfire spread, these measurement errors are significant within an operational framework.

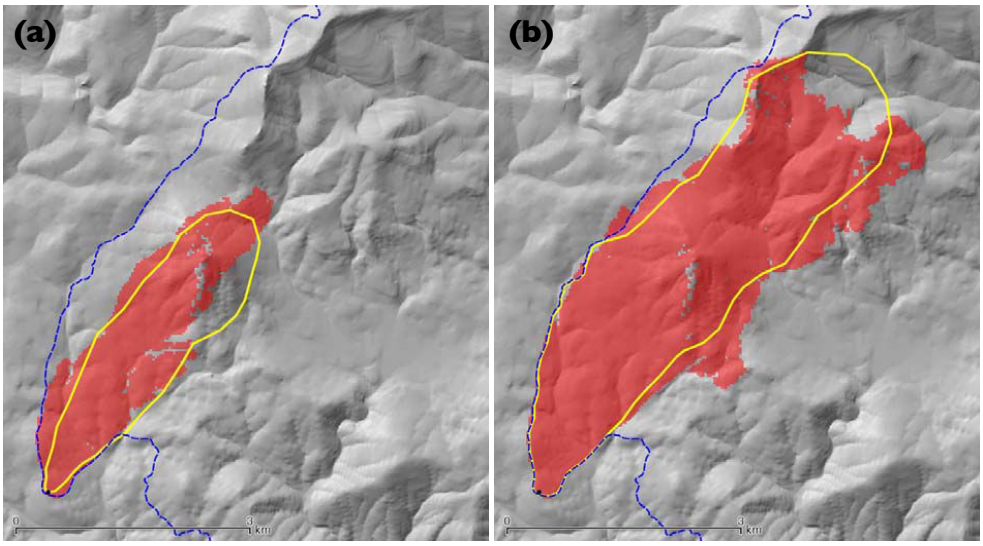


Figure 6.5: *Comparison of simulated and calibrated positions of the fire front when applying a genetic algorithm to the 2007 Moonlight fire (9000 ha, California): the yellow solid line corresponds to the real fire front position, the red-shaded area corresponds to the simulated burnt area, and the blue-dashed line is the final position of the real fire front. (a) 7 hours after simulation ignition. (b) 22 hours after simulation ignition. Credit: Lautenberger (2013).*

Considering that it is important to account for these measurement errors, Gu (2010) demonstrated the feasibility of the SIR particle filter for estimating and reducing the uncertainties in the input parameters of a semi-empirical ROS model. Based on synthetically-generated measurements of temperature at ground sensors deployed over the wildfire area, this study estimated either the wind conditions (i.e., wind magnitude and direction) or the location of the biomass fuel. It showed that the SIR

particle filter was able to produce more consistent simulations of the propagation of the fire front, but at a high computational expense since a large number of particles was required to accurately represent model uncertainties.

This review highlights that efforts are currently made in the fire research community to adapt optimization and data assimilation techniques that are widely used in geosciences and beyond, for tracking fire propagation in both building and wildfire applications. Still, wildfire spread forecast is at an early stage of development and the proposed solutions in the literature do not demonstrate the full potential of data assimilation to overcome current limitations of regional-scale wildfire modeling and to build predictive simulations that are compatible with operational applications.

6.2 Strategy guidelines for wildfire spread forecast

The objective of this study is to address challenges specific to the development of a robust inverse modeling approach for realistic wildfire spread. For this purpose, the following aspects are believed to be of primary importance.

- **Parameter estimation.** Current operational wildfire spread models rely on a front-tracking solver that includes a semi-empirical ROS model (e.g., Rothermel, 1972) and simulates the wildfire spread as a front propagation at regional scales (i.e., at scales ranging from a few tens of meters up to several kilometers). This viewpoint has some important limitations, partly because of its inability to explicitly account for the fire/atmosphere interaction and also due to knowledge gaps and/or inaccuracies in the description of the controlling input parameters (i.e., vegetation, topographical and meteorological properties). These uncertainties are a combination of epistemic uncertainty that expresses an imperfect knowledge of the input parameters of the ROS model (that could in theory be removed), and aleatoric uncertainty that results from natural and unpredictable stochastic variations of the physical system. These uncertainties translate inevitably into errors in the model outputs of interest (e.g., time-evolving position of the flame front of the fire, size of the burning area).

  While studies presented by Mandel et al. (2008) and Beezley (2009) showed the potential benefits of a state estimation approach, uncertainties in the environmental conditions required as input data to the semi-empirical ROS model (i.e., biomass fuel properties and wind external forcing) must be quantified and reduced in order to develop an efficient data assimilation prototype for wildfire spread forecast, as indicated by the batch-calibration results derived from genetic algorithms (Lautenberger, 2013). A parameter estimation approach is expected to increase the knowledge on the environmental conditions and significantly improve the quality of wildfire spread forecasts. In contrast, a state estimation approach does not provide a complete feedback on the sources of uncertainties in the physical problem under consideration,

but, instead, provides a detailed snapshot of the simulation errors made at a given time step (i.e., at the assimilation times). We consider that parameter and state estimations are complementary approaches and that it is important to discuss their benefits and drawbacks for future developments in wildfire spread forecasting.

- **Stochastic characterization of parameter and modeling uncertainties.** The uncertainties inherent in wildfire spread modeling go beyond the limits of deterministic forecast ability with the dynamical model and thus, suggest the use of ensemble forecasts to stochastically characterize the non-linear response of the wildfire spread model to variations in the input environmental parameters. For instance, Finney et al. (2011) describes an ensemble-based forecasting method, in which a large number of fire spread scenarios (i.e., the ensemble members) are generated based on a probabilistic uncertainty in the weather conditions and in the vegetation moisture content as shown in Fig. 6.6. In this study, the statistics of the probability for the burnt area computed over this ensemble of forecasts (i.e., with no data assimilation feedback) is shown to be consistent with the observation statistics among 91 real-world fire events.

  The use of stochastic approaches for uncertainty quantification is supported by the MEnKF algorithm introduced in Beezley and Mandel (2008) for solving a state estimation problem. While ensemble-based data assimilation methodologies have been widely used as a state estimation approach, they can also be applied to parameter estimation. Indeed, they avoid the computation of the tangent-linear of the physical model with respect to the parameters, while partly accounting for model non-linearities (Tarantola, 1987). Recent studies (Moradkhani et al., 2005; Andreadis et al., 2007; Durand et al., 2008) have demonstrated the successful application of EnKF methodologies in hydrology and hydrodynamics for estimating river bathymetry depth and slope parameters as well as friction coefficients. However, the transfer of such ensemble-based data assimilation techniques to wildfire applications has not been yet investigated and tested against realistic measurements of wildfire spread. Still, we are convinced that these techniques provide a new and powerful framework to combine a physical model with remote sensing measurements, in order to reliably deliver an accurate forecast of wildfire spread.

- **Consideration and treatment of non-linearities.** An additional difficulty in parameter estimation results from the fact that most input parameters of the ROS model cannot be measured directly, thus increasing their uncertainty. These parameters are indirectly related to wildfire measurements through the observation operator $\mathcal{G}$ that includes the temporal integration of the wildfire numerical model. Since the wind direction and magnitude may vary and the vegetation properties may be strongly heterogeneous in realistic cases of regional-scale wildfire spread, this numerical model is highly non-linear.

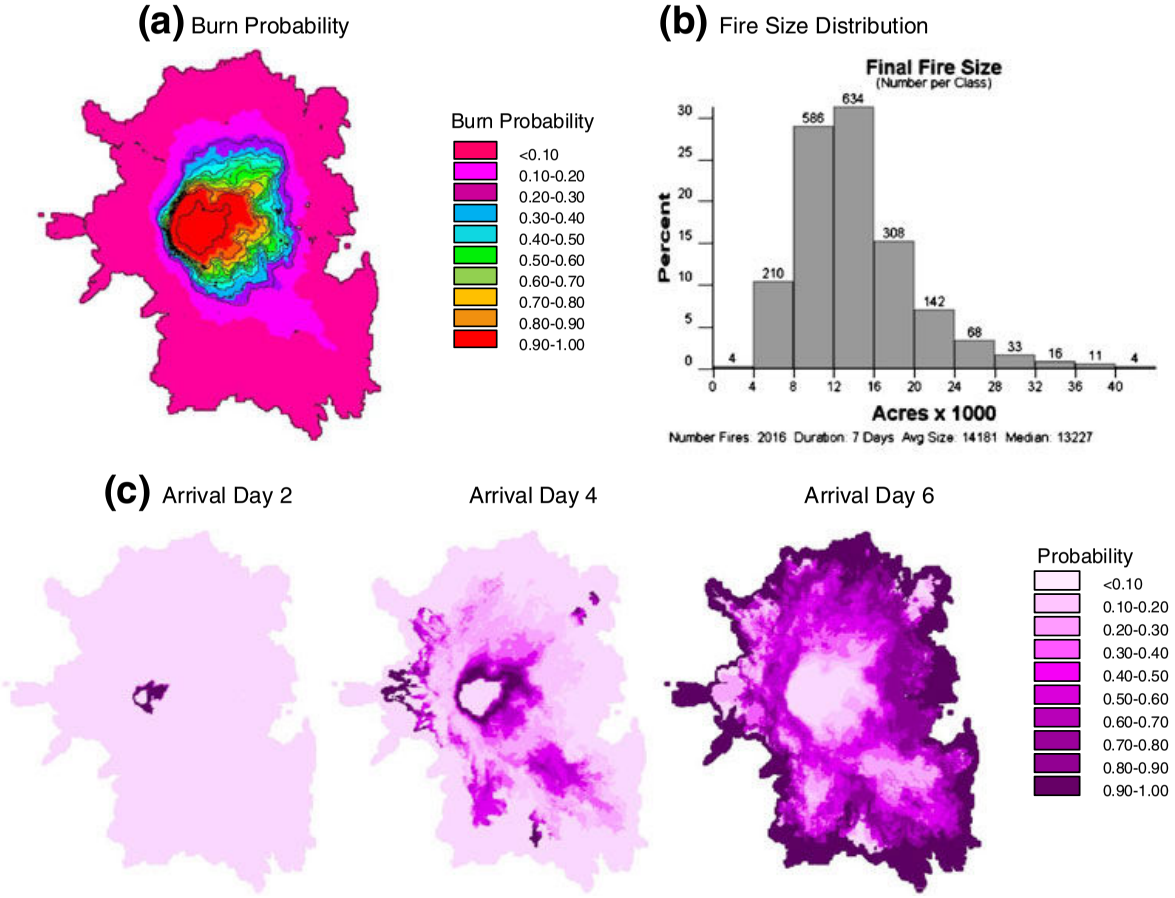


Figure 6.6: *Example of FSPro (Fire Spread Probability) predictions, a Monte Carlo fire spread simulator used to produce an ensemble of fire simulations based on perturbed weather scenarios. (a) Ensemble burn probabilities for a 7-day forecast period. (b) (Right skewed) fire size distribution reconstructed from ensemble members. (c) Probabilities of fire arrival for each day in the simulation period. Credit: Finney et al. (2011).*

We therefore need to develop a data assimilation strategy that is able to account for non-linearities in the observation operator $\mathcal{G}$, while maintaining the optimality of the analysis solution. EnKF and particle filters are appropriate data assimilation techniques in this context as discussed by Evensen (1994) and Doucet et al. (2001).

- **Consideration of measurement errors.** Due to the complexity of wildfire spread and to the practical difficulties to track wildfire using in-situ, airborne or spaceborne remote sensors, observations provide an incomplete picture of wildfire behavior with a limited spatial and temporal resolution (see Section 1.4, Chapter 1). For instance, spaceborne observations are available with a coarse temporal resolution (i.e., a few times per day for polar-orbiting satellites). In addition, the active fire front is mainly a sub-pixel phenomenon for spaceborne remote sensors, meaning that there is an indirect relationship between the raw measured quantity (i.e., fire radiation power, FRP) and the quantities of interest for the wildfire spread model (e.g., time-evolving location of the flame front). As for airborne observations, they are also based on FRP measurements and, due to the development of a thermal plume above the active fire areas and the resulting turbulence surrounding the wildfire, they are usually limited to a certain section of the wildfire active burning areas. The latter point implies that they only provide a partial picture of the wildfire as is the case with in-situ measurements deployed on the ground terrain. Wildfire measurements are then subject to significant uncertainties, which must be accounted for in a data assimilation strategy that is intended for operational applications.

- **Sequential algorithm for state and/or parameter estimations.** Even though certain input parameters of the fire ROS model can be assumed constant over the fire duration (in particular, intrinsic properties of the vegetation), other parameters exhibit a dynamic behavior due to the presence of the propagating fire front. The wind magnitude and direction at mid-flame height change over time due to the fire/atmosphere interactions and their retroactive effect on local wind conditions. In addition, the moisture content of the vegetation also varies, in particular that of the dead vegetation, which can be considered in thermal equilibrium with the atmosphere. These atmospheric conditions in terms of temperature, wind and air humidity are indeed significantly modified by the wildfire propagation and emissions. In this context, it is difficult to rely only on parameter calibration; a dynamic estimation of the time-varying parameters is necessary to produce an accurate feedback to the wildfire spread model and thereby, reliable wildfire spread forecasts. A sequential data assimilation technique that strikes a balance between accuracy and computational time and that provides forecasts at future lead-times seems appropriate for the estimation of time-varying input parameters of the ROS model.

- ▷ **Validation against real-world wildfire spreads.** Except for the study presented in Lautenberger (2013), the different approaches described in the previous state-of-the-art data assimilation applications for wildfire spread were only validated against synthetically-generated measurements. Even though these approaches show great potential for fire spread forecast in wildland fire applications, they still require a detailed evaluation against actual measurements of wildfire spread, starting from controlled burn experiments and extending to real-world wildfire events, in order to explore their benefits and drawbacks in experiments that are consistent with an operational framework (in terms of computational cost, accuracy and performance, for different wildfire scenarios subject to a wide range of biomass fuels, terrain topography and weather conditions).

In this work, we propose and explore a new paradigm for improving wildfire spread forecasts as new wildfire observations become available, using real-time data assimilation. The objective is to develop a prototype data-driven wildfire simulator capable of:

- ▷ explicitly accounting for the effects of both measurement and modeling errors and overcoming some of the current limitations of regional-scale wildfire modeling;
- ▷ accounting for the main sources of uncertainty in regional-scale wildfire modeling;
- ▷ sequentially estimating input parameters and/or model state variables of a wildfire spread model, to account for the temporal variabilty of the errors and to allow for accurate forecasts at different lead-times (i.e., at different time steps beyond the current observation time);
- ▷ forecasting reliable wildfire spread scenarios at a limited computational cost, consistently with an operational framework;
- ▷ assimilating realistic measurements of wildfire, including synthetic representative cases with spatially-varying vegetation properties and temporally-varying wind conditions as well as a controlled grassland fire experiment.

The different components required by the prototype data-driven wildfire simulator are explained in detail in the following, with a focus on the type of assimilated wildfire spread measurements and on the simulator of regional-scale wildfire spread, named Firefly and used to deliver the forecast of the wildfire behavior at different lead-times.

6.3 Assimilated measurements of wildfire spread

6.3.1 Airborne measurements

The review of the current status of operational wildfire tracking and forecasting presented in Chapter 1 shows that in-situ, airborne and spaceborne measurements are currently available. While in-situ sensors can provide some insight on the location of the active burning areas and evaluate the local values of some physical variables (e.g., temperature, wind, air humidity), they remain sparse in time and space. They highly depend on the deployment of ground sensors and on the strategy of emergency services, and thus, they cannot be systematically performed for all wildfire events. They also require a complex post-treatment process that is difficult (if not impossible) to perform in real-time. It is therefore currently difficult to rely on in-situ measurements for the development of a data assimilation strategy (sensor network may become available in the future). Airborne and spaceborne observations produce a more global and frequent picture of the wildfire event and seem therefore more suitable for data assimilation. However, wildfire front-tracking requires high-spatial resolution imagery (that is no longer out of reach for satellite missions such as Pléiades 1-A and 1-B) as well as high revisit frequencies. For this last reason, only airborne platforms provide currently spatial and temporal resolutions suitable for real-time geolocation of active fire contours. A typical example of airborne remote sensing systems is the LIVEFIRE system (Merlet, 2008; Crombette, 2010). However, in-situ and spaceborne data could be used for validation and calibration of models and data assimilation procedures in off-line mode. Also spaceborne data can be useful to monitor wildfires deployed on very large areas as highlighted by the International Charter *Space and Major Disasters*.[4]

In conjunction with the current development in airborne remote sensing technologies, we assume, in the present study, that observations of the fire front location are available and can be made at different relevant times with a low measurement error (e.g., 0-30 m for the LIVEFIRE system). In the following, the observed fire front is represented as a segmented line using a pre-defined number of equally-spaced markers (i.e., the N_{fr}^{o} observation points); the observation vector $\mathbf{y}_t^{o}$ contains the two-dimensional coordinates (x_i^{o}, y_i^{o}) of the fire front markers observed at time t (i.e., the assimilation time). Note that the subscript i is the index of a particular marker in the observation vector, with $i = 1, \cdots, N_{fr}^{o}$. A schematic of observed time-evolving location of the fire front over the time window $[t-1, t+1]$ is presented in Fig. 6.7. The fire front coordinates are assumed to have independent Gaussian-like random errors $\boldsymbol{\epsilon}_t^{o}$ with zero mean and with standard deviation (STD) noted σ_t^{o}.

Within this framework, the observation space $\mathbb{R}^p$, defined theoretically in Chapter 5, represents in this wildfire application the space spanned by the different potential observed locations of the wildfire front (in terms of x- and y-coordinates of the front

[4] www.disasterscharter.org/web/charter/home.

marker) such that $p = 2\,N_{fr}^{\mathrm{o}}$ within the proposed data assimilation framework (p being the dimension of the observation space).

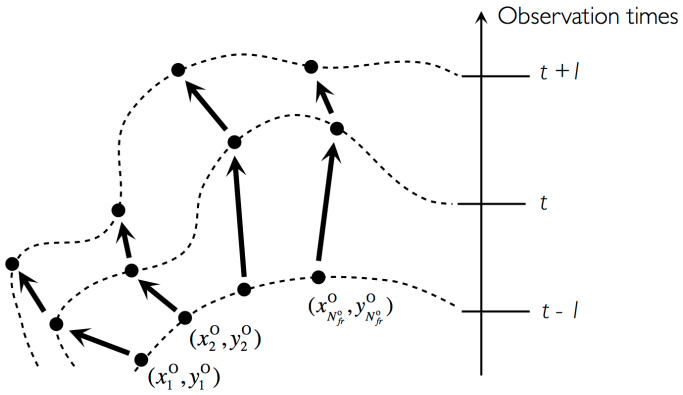


Figure 6.7: *Schematic of the observed fire front location over the time window* $[t-1, t+1]$.

6.3.2 Available data sets for data assimilation experiments

In this work, two types of data assimilation experiments are presented: (1) observation system simulation experiments (OSSE), in which observations are synthetically-generated using a reference solution of the wildfire spread model (called the *true trajectory*) modified by random observation errors $\boldsymbol{\epsilon}^{\mathrm{o}}$; and (2) a controlled grassland burn experiment, in which observations are reconstructed from measured temperature maps and using a definition of the fire front as the 600 K iso-temperature contour (corresponding to the apparent temperature provided by infrared camera).

↪ Observation system simulation experiments

Within the OSSE framework, the true control vector $\mathbf{x}_t^{\mathrm{t}}$ is supposed to be known. The observations over the assimilation time window $[t-1, t]$ are synthetically-generated using the observation operator $\mathcal{G}_t(\,\cdot\,)$ applied to the true value of the control vector $\mathbf{x}_t^{\mathrm{t}}$. The application of the observation operator $\mathcal{G}_t$ provides, at the observation time t, the location of the markers along the true simulated fire front $[(x_1^{\mathrm{t}}, y_1^{\mathrm{t}}), \cdots, (x_{N_{fr}^{\mathrm{o}}}^{\mathrm{t}}, y_{N_{fr}^{\mathrm{o}}}^{\mathrm{t}})]$. A random noise $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ of zero mean and STD σ_t^{o} is then added at each marker position (x- and y-coordinates), in order to account for observation errors and define the observation vector $\mathbf{y}_t^{\mathrm{o}}$. To define the forecast control vector $\mathbf{x}_t^{\mathrm{f}}$, a perturbation $\mathbf{e}_t^{\mathrm{f}}$ of zero mean and STD σ_t^{f} is added to the true value of the control vector $\mathbf{x}_t^{\mathrm{t}}$. In this context, the true control vector $\mathbf{x}_t^{\mathrm{t}}$ as well as the statistics of the observation and forecast errors $\boldsymbol{\epsilon}_t^{\mathrm{o}}$ and $\mathbf{e}_t^{\mathrm{f}}$ are known. Diagnostics comparing $\mathbf{x}_t^{\mathrm{a}}$ and $\mathbf{x}_t^{\mathrm{t}}$ can therefore be developed and used as verification tools of the proposed data assimilation prototype. Thus, OSSE experiments are a powerful tool to quantify the quality of the correction to the forecast and thus, ensure the optimality of the analysis $\mathbf{x}_t^{\mathrm{a}}$. A schematic of the OSSE framework is shown in Fig. 6.8.

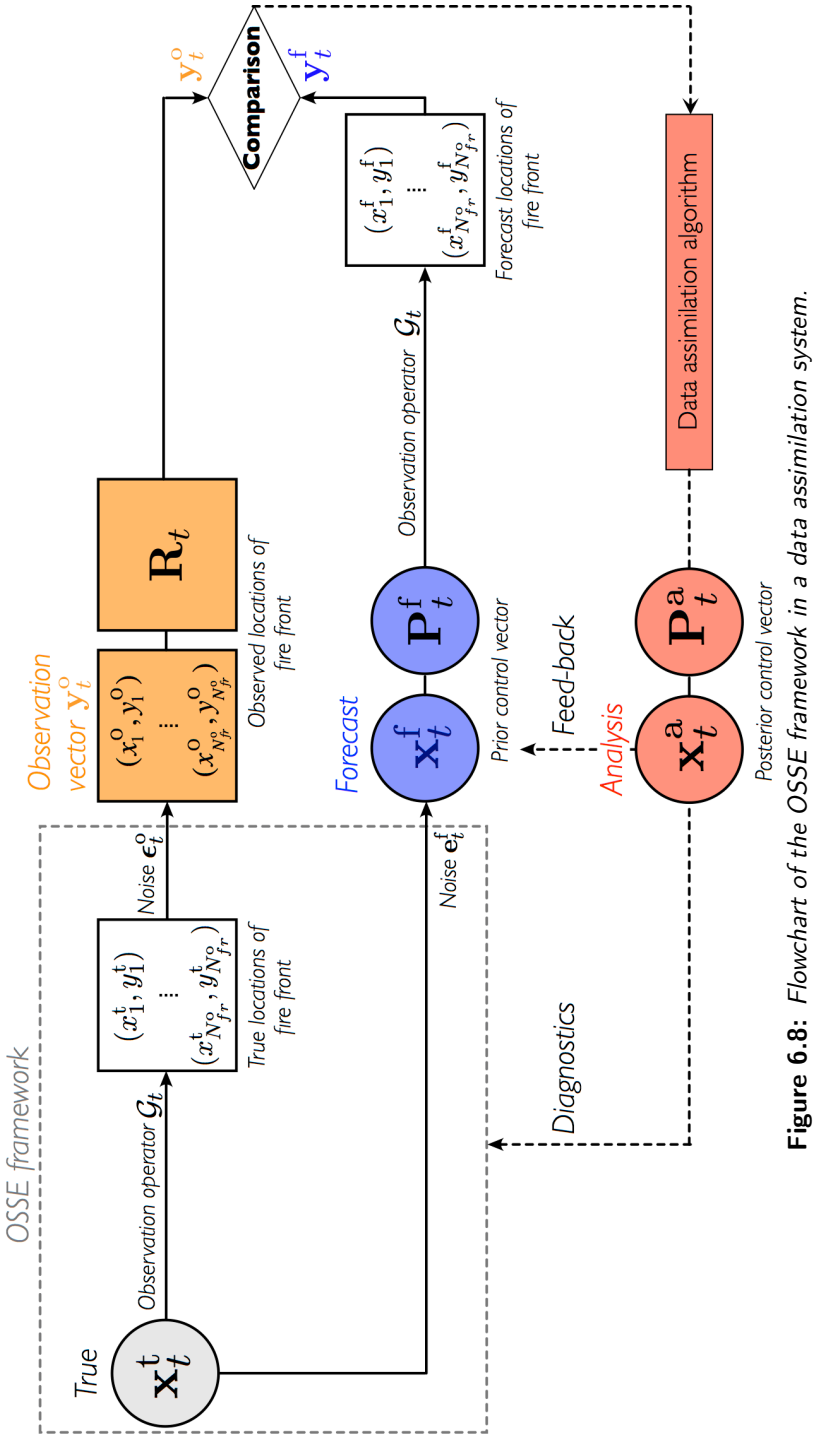


Figure 6.8: *Flowchart of the OSSE framework in a data assimilation system.*

↪ Controlled grassland burning experiment

The real case used for assessing the performance of data assimilation methodologies in this thesis corresponds to a grassland controlled burning experiment, performed by the wildfire research group at the Department of Geography of King's College London.[5] This small-scale experiment aimed at testing and validating the technique consisting in retrieving the flame temperature from fire radiation power (FRP) measurements (Wooster et al., 2005; Paugam et al., 2013).

▷ **Experimental configuration.** As illustrated in Fig. 6.9, the experimental configuration corresponds to a small-scale (4 m $\times$ 4 m), flat and horizontal, open-field grassland lot.

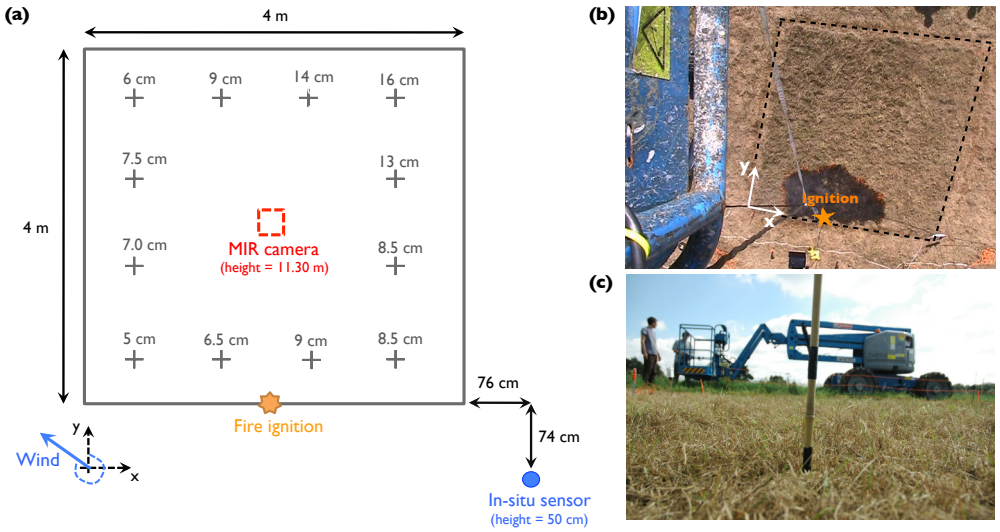


Figure 6.9: *Configuration of the grassland controlled burning experiment. (a) Top-view schematic of the burn field with the location of the flame ignition (yellow star), the MIR camera (red square) and the in-situ sensor (blue circle), at which the wind conditions and the air humidity were measured. The height of the vegetation layer was measured at different locations (gray crosses) and varies from 5 cm to 16 cm. (b) Top-view of the burning field delimited by the black-dotted line (extracted from the MIR camera aboard the cherry-picker at 11.30-m height). (c) Snapshot of the short grass vegetation.*

The short grass was characterized according to field measurements by a mean layer thickness $\delta_v = 9$ cm, a mean surface loading $m''_v = 0.4$ kg/m^2 and an approximate moisture content $M_v = 21.7$ %. Before fire ignition, the mean wind conditions are moderate, i.e., $u_w = 1.0$ m/s blowing into a western direction (i.e., $\alpha_w = 307°$, in a clockwise representation where $0°$ indicates the North direction). An in-situ sensor (see Fig. 6.9) recorded the wind conditions (in terms of magnitude and direction) as well as air humidity over the fire duration; the time-variations of these environmental conditions are shown in Fig. 6.10. The wind conditions significantly fluctuated over

[5] wildfire.geog.kcl.ac.uk/.

the fire duration, with a wind magnitude varying between 0 and 2 m/s and a wind direction decreasing from $325°$ to $175°$. Furthermore, air humidity had a tendency to increase during the first part of the fire (the first 100 s), partly due to the strong water evaporation from the grassland vegetation that is enhanced by a large flame zone, and started decreasing after reaching a peak at time 120 s, confirming a full fire development at this time.

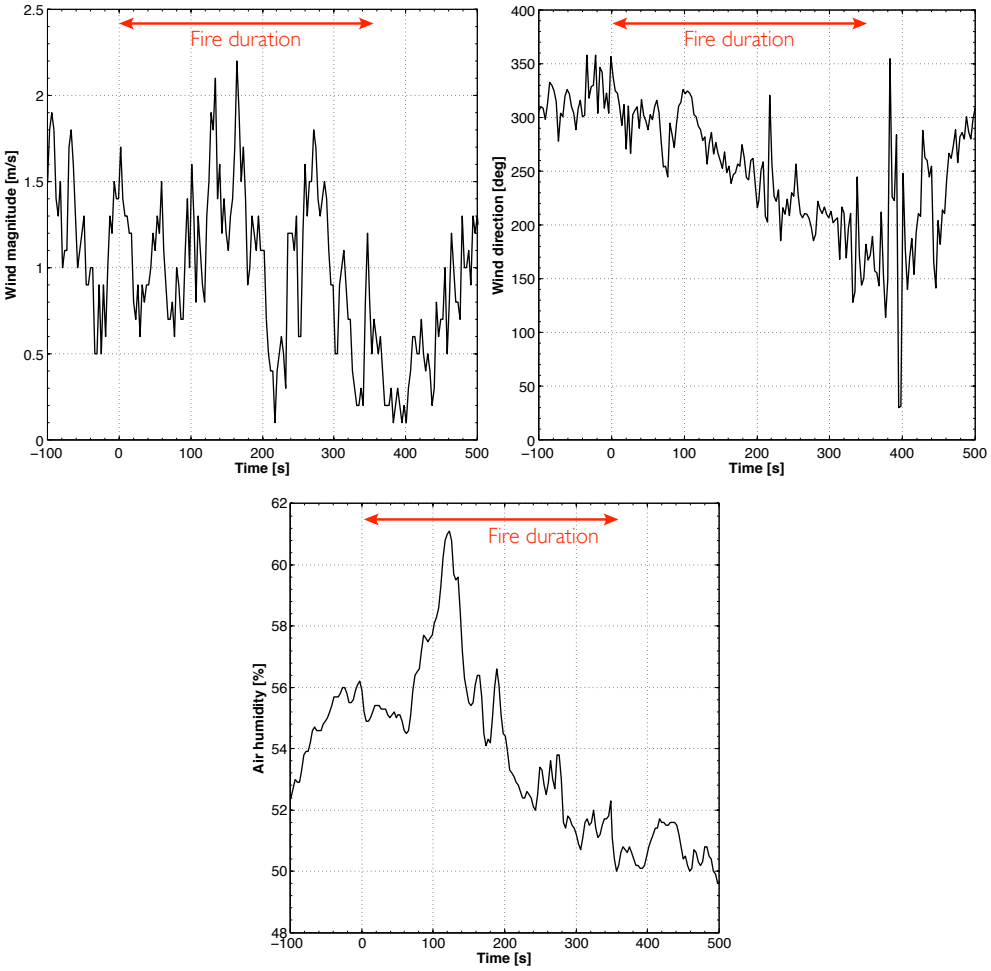


Figure 6.10: *In-situ measurements at the sensor represented by a blue circle in Fig. 6.9, before/during/after the fire (negative times indicate the time period before the fire ignition). (a) Wind magnitude [m/s]. (b) Wind direction [°]. (c) Air humidity [%]. Credit: R. Paugam (private communication).*

▷ **Experimental database.** The fire spread was recorded during 350 s using a MIR camera (see Chapter 1 for a detailed description on wildfire remote sensing). This MIR camera was aboard a cherry picker 11.30 m above the burning field and recorded the fire instantaneous field of view at a rate of four images per second. Details of the measurement technique to retrieve the temperature field from thermal

imaging are given in Wooster et al. (2005). Figure 6.11 represents the time-evolving temperature on the 4 m $\times 4$ m-domain, which is extracted from MIR imaging at 14 s intervals with a spatial resolution of 2.5 cm. This time-series of temperature maps shows that the maximum apparent temperature reached 900 K in the flame region and that the width of the active flame region was not uniform along the fireline. The latter was highly dependent on the angle between the fire spread direction and the wind direction. This fire was essentially a flank fire, meaning that only the western part of the fire propagated in the same direction as the north-western wind and this, only during the first 100 s of the fire. This implies that the apparent width of the flame region was larger when the wind blew in the same direction as the fire front propagated, while it became thinner when the fire front continued spreading in other directions than the wind direction, indicating different fire spread mechanisms for the two regions of the flame front. This change in the width of the flame region is also due to the view-angle at which the fire is observed (i.e., at nadir in the current experiment). However, this effect is currently an active research area in the fire remote sensing field and is therefore not addressed here.

▷ **Post-processing of temperature maps.** Within the data assimilation framework, the observations are the time-evolving locations of the fire front, identified as the zones where the temperature reaches the value 600 K, assumed to represent the mean temperature at which flammable compounds are released from the solid vegetation (see Section 3.2, Chapter 3). The subsequent reconstruction of the fire front location is illustrated for the observed time $t = 92.16$ s in Fig. 6.12.

In a preliminary step, the burn domain is reframed to simplify the simulation of the grassland controlled burn, resulting in a square 4 m $\times\, 4$ m-domain discretized with a Cartesian mesh, see Fig. 6.12(b). This figure shows that the burn domain was delimited by counter-fires; these counter-fires did not move over the fire duration and did not interfere with the grassland fire, implying that they can be removed from the representation of the observed fire as shown in Fig. 6.12(c). We ensure that the fireline is continuous by filling in missing burning pixels along the fire front. In addition to the geolocated fire front, the burnt area is reconstructed by binarizing the two-dimensional field, meaning that the fire front represents the interface between the burned and unburnt regions of the grassland field.

By applying this treatment at all observation times, the arrival times of the fire front over the 350 s of the fire duration can be reconstructed as shown in Fig. 6.13. The associated fire front locations are assumed to have a measurement error $\sigma^{\mathrm{o}} = 0.05$ m in both x- and y-directions; the estimation of this error is based on the spatial resolution of the camera (i.e., 2.5 cm) and on the post-processing treatment to extract the fire front location.

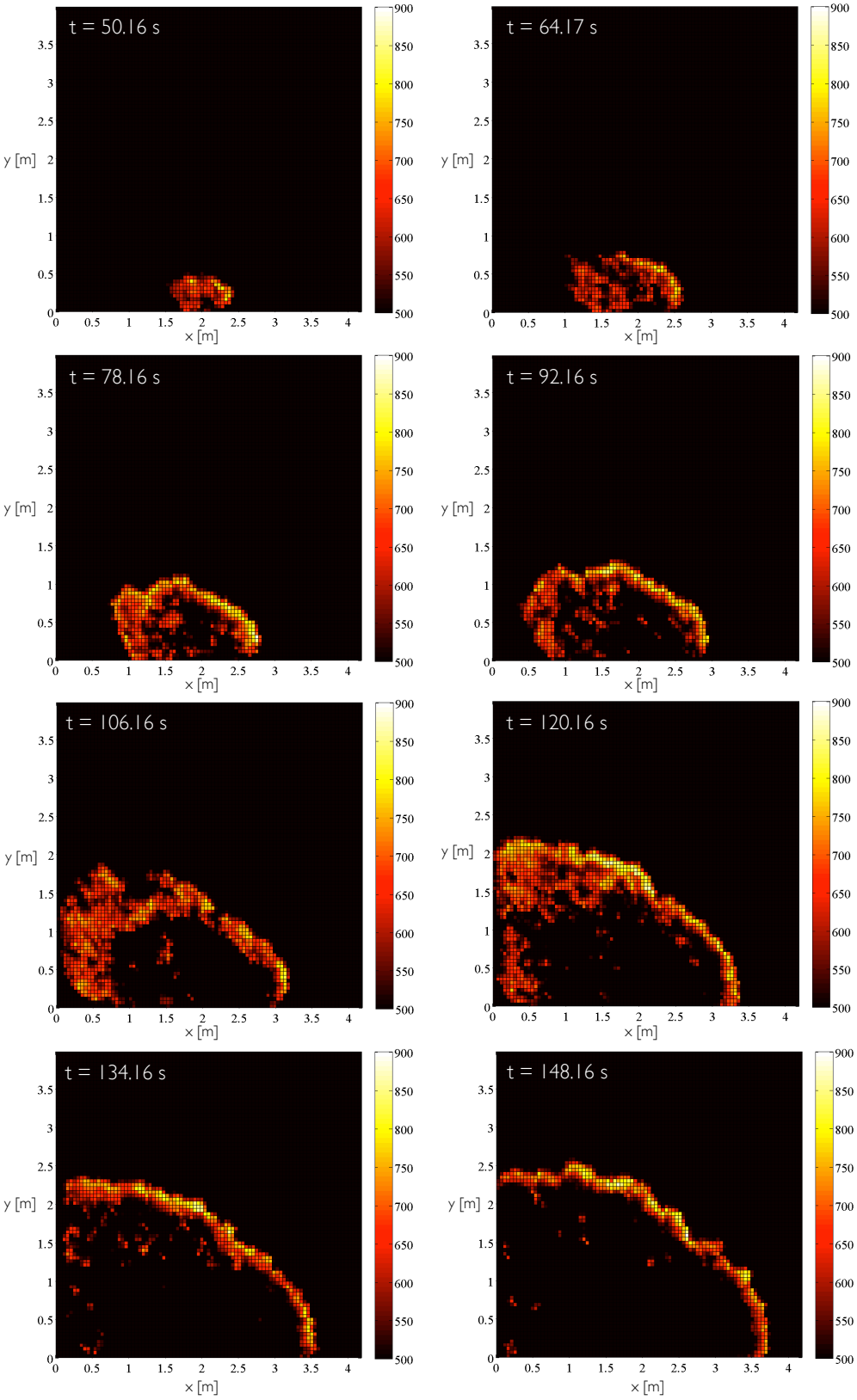


Figure 6.11: *Time-series of the fire temperature represented in the two-dimensional reference frame* (x, y) *every* $14\ s$*, from* $t = 50.16\ s$ *to* $t = 148.16\ s$.

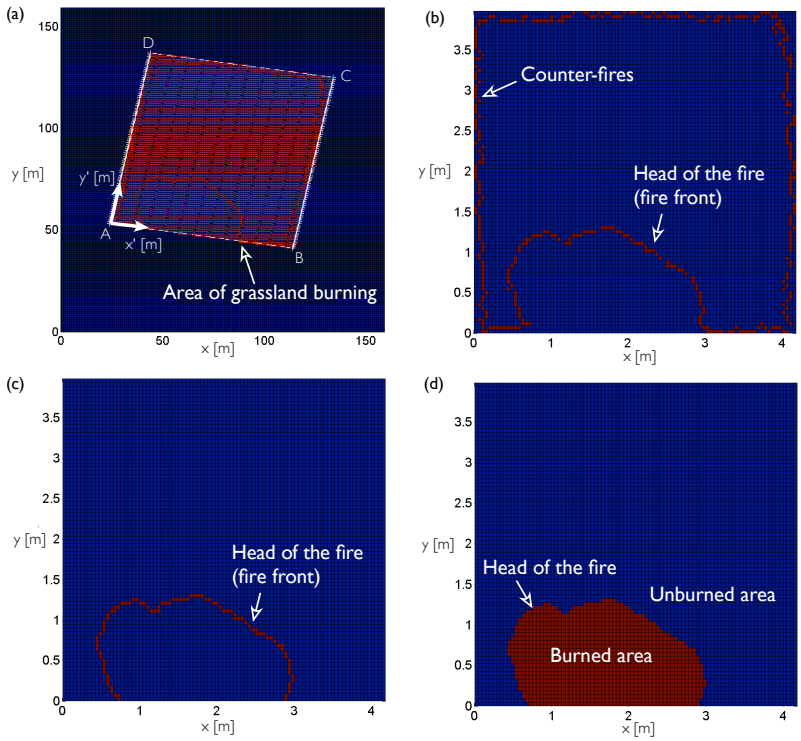


Figure 6.12: *Example of temperature map post-processing at time* $t = 92.16\ s$. *(a) Raw data. (b) Projection of the grassland burn domain (ABCD) onto the Cartesian frame* (x, y). *(c) Reframed location of the fire front. (d) Burnt area.*

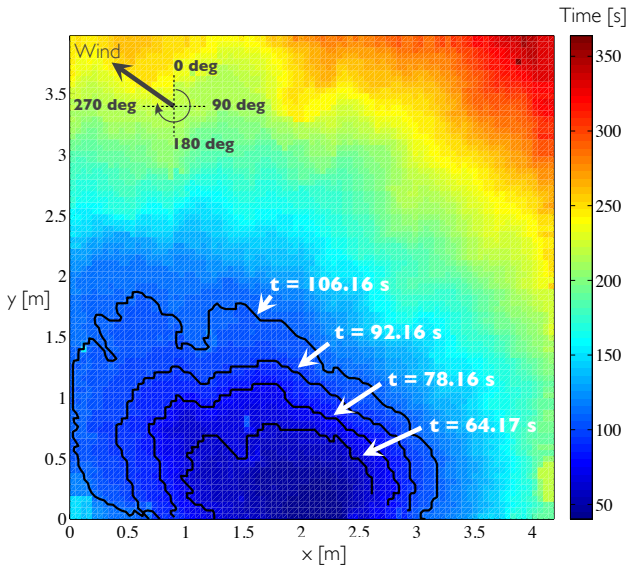


Figure 6.13: *Arrival times of the observed fire front (colormap). Observed fire fronts at* $14\ s$ *intervals are represented in black solid lines from* $t = 64.17\ s$ *to* $t = 106.16\ s$.

6.4 Regional-scale wildfire spread simulator FIREFLY

As our objective is to show the feasibility of data assimilation for wildfire spread forecast, a regional viewpoint is adopted as in current operational wildfire spread simulators (see Chapter 1 for a discussion on wildfire spread modeling approaches) and thus, wildfire spread is simulated using a front-tracking simulator based on a semi-empirical ROS model. This front-tracking strategy consists in a minimalist treatment of the fire front, idealized as an interface and consistent with the limited knowledge on the environmental conditions, implying that flame-scale processes underlying fire spread are not resolved. To allow for more flexibility and adequation with the proposed data assimilation platform, the wildfire spread simulator named FIREFLY has been developed (Rochoux et al., 2010; Delmotte et al., 2011; Emery et al., 2013) and is now presented in detail.

6.4.1 Front-tracking simulation capability

The FIREFLY front-tracking solver simulates the regional-scale propagation of surface wildfires (within the biomass fuel bed) as illustrated in Fig. 6.14. In particular, FIREFLY tracks the time-evolution of the fire front location using the following three components: (1) a submodel for the ROS noted Γ and parameterized with respect to the local environmental conditions (e.g., vegetation, terrain topography and weather properties); (2) a level-set-based solver for the fire front propagation equation that simulates the propagating fire front according to the local variations of Γ (through the integration of the progress variable c); and (3) an iso-contour algorithm for the reconstruction of the fire front.

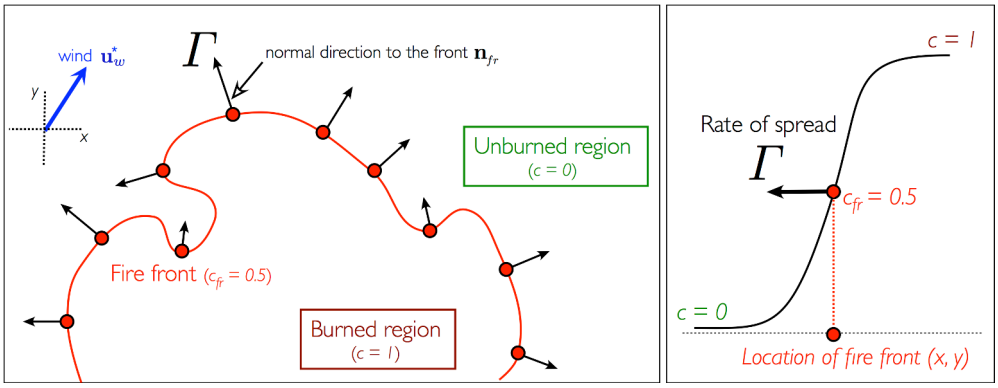


Figure 6.14: *Level-set-based fire spread simulator. Left: The fire front is the progress variable isocontour $c_{fr} = 0.5$; Γ measures the local ROS of the fire along the normal direction $\mathbf{n}_{fr}$ to the fireline. Right: Profile of the spatial variations of the progress variable c across the fire front.*

↪ The Rothermel-based rate of spread submodel

The ROS submodel is based on the reference semi-empirical model due to Rothermel (1972) since it appears as a good starting point for demonstrating the potential of data assimilation for wildfire spread forecasting. This model focuses on the propagation of the head of the fire and describes its ROS Γ as a function of vegetation properties associated with a pre-defined fuel category, topographical properties and meteorological conditions. The original Rothermel's model includes a database of 11 biomass fuel categories, among whom short grass, chaparral and timber litter.

- ▷ **Original one-dimensional formulation.** Rothermel's model is derived from the one-dimensional formulation of the energy balance equation within a unit volume of the unburnt vegetation ahead of the flame, originally proposed by Frandsen (1971) and assuming that the wildfire reaches a quasi-steady propagation rate. The initial acceleration of the wildfire (i.e., the transient phase in the propagation) is not accounted for. A schematic of this energy balance in the biomass fuel bed is presented in Fig. 6.15. The physical quantities involved in this energy balance were parameterized with respect to the measured input parameters using wind-tunnel experiments in artificial biomass fuel beds of varying properties.

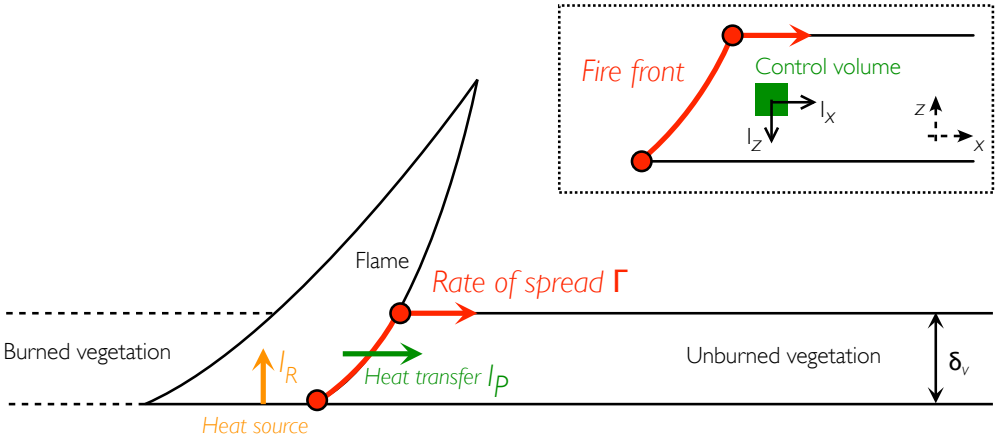


Figure 6.15: *Rothermel's ROS Γ derived from the energy balance within a unit control volume located in the unburnt vegetation ahead of the flame zone. Credit: Dupuy and Valette (1997).*

In this formulation, the ROS Γ [m/s] along the normal direction to the fire front is expressed as the ratio between the heat flux received by the unburnt vegetation I_p [W/m^2] due to the contributions of radiation, convection as well as conduction, and the energy required to ignite the fuel H_{ig} [J/m^3]. Formally, Γ reads:

$$\Gamma = \frac{I_p}{H_{ig}} = \frac{I_p}{\rho_b \, \varepsilon \, Q_{ig}}, \tag{6.1}$$

where the ignition energy H_{ig} is expressed as $H_{ig} = \rho_b \, \varepsilon \, Q_{ig}$, with Q_{ig} [J/kg] the heat of pre-ignition, ε the dimensionless effective heating number (i.e., a correction factor to only consider the amount of biomass fuel effectively involved in the combustion process) and ρ_b [kg/m^3] the bulk vegetation density (i.e., the density of the porous medium that differs from the density of the solid phase in the porous vegetation ρ_p).

In the case of a fire spread that is not induced by wind or terrain topography (also referred to as *no-wind no-slope fire spread*), Rothermel (1972) showed that the propagating heat flux I_p is proportional to the energy release rate of the combustion I_r. The proportionality coefficient is the dimensionless propagating flux ratio ξ (that describes the proportion of energy that is released by the flame and transferred to the vegetation in the non-flaming zone as illustrated in Fig. 6.15). The resulting no-wind no-slope ROS is denoted by Γ_0 [m/s] and reads:

$$\Gamma_0 = \frac{\xi \, I_r}{\rho_b \, \varepsilon \, Q_{ig}}. \tag{6.2}$$

To account for wind and slope contributions to the ROS, correction coefficients, respectively noted Φ_w^* and Φ_{sl}^*, are introduced in Rothermel's formulation as follows:

$$\Gamma = \frac{I_p}{H_{ig}} = \underbrace{\frac{\xi \, I_r}{\rho_b \, \varepsilon \, Q_{ig}}}_{\Gamma_0} \left(1 + \Phi_w^* + \Phi_{sl}^*\right). \tag{6.3}$$

This ROS formulation means that wind and/or slope positively modify the propagating heat flux I_p since the vegetation located ahead of the fire front is subject to additional convection and radiation heat fluxes (represented by the dimensionless correction coefficients Φ_w^* and Φ_{sl}^*). These coefficients were determined for the one-dimensional case of heading and upslope fire spread, meaning that their parameterization is optimal for a wind that blows in the direction of the fire spread (heading fire) and/or for a fire that spreads in the uphill direction (upslope fire). While the slope correction coefficient Φ_{sl}^* depends on the tangent of the terrain slope angle α_{sl}, the wind correction coefficient Φ_w^* non-linearly depends on the wind velocity magnitude at mid-flame height u_w such that:

$$\Phi_w^* \equiv \Phi_w^*(u_w) = C \, u_w^B \left(\frac{\beta_v}{\beta_{v,opt}}\right)^{-E}, \tag{6.4}$$

with C, B and E calibrated parameters depending on the biomass fuel surface-to-volume ratio Σ_v [1/m], with β_v the biomass fuel packing ratio and $\beta_{v,opt} \equiv \beta_{v,opt}(\Sigma_v)$ its optimum value. Note that no correlation between the wind and slope effects are accounted for in the ROS formulation since

their effects are assumed to be additive. Equation (6.3) can be written as the following compact form:

$$\Gamma \equiv \Gamma(x,y,t) = \Gamma\Big(\delta_v, M_v, M_{v,ext}, \Sigma_v, m_v^{''}, \rho_p, \Delta h_c, u_w, \alpha_{sl}\Big), \qquad (6.5)$$

or alternatively, as a linear function to the biomass fuel layer depth δ_v:

$$\Gamma \equiv \Gamma(x,y,t) = P\Big(M_v, M_{v,ext}, \Sigma_v, m_v^{''}, \rho_p, \Delta h_c, u_w, \alpha_{sl}\Big)\delta_v, \qquad (6.6)$$

where input parameters, uniformly- or spatially-distributed over the computational domain, time-invariant or time-varying, are summarized in Table 6.1. A full description of the ROS model due to Rothermel is provided in Appendix A.

Table 6.1: *Main input parameters of the Rothermel-based ROS model.*

Name	Symbol	Unit
Fuel depth (vertical thickness of the vegetation layer)	δ_v	m
Fuel moisture (mass of water divided by mass of dry vegetation)	M_v	%
Fuel moisture at extinction	$M_{v,ext}$	%
Fuel particle surface-area-to-volume ratio	Σ_v	1/m
Fuel loading	$m_v^{''}$	kg/m^2
Fuel particle mass density	ρ_p	kg/m^3
Fuel heat of combustion	Δh_c	J/kg
Wind velocity at mid-flame height projected into horizontal plane (x,y)	u_w	m/s
Terrain slope angle	α_{sl}	$^\circ$

▷ **Extension to two-dimensional surface wildfire spread.** Since it was originally calibrated for one-dimensional tunnel experiments, Rothermel's model is too restrictive to simulate two-dimensional wildfire spread in FIREFLY. We therefore adapt the original Rothermel's model to two-dimensional configurations, in order to account for the wind and slope effects on the shape of the fireline, while still maintaining a simple parameterization of the ROS with respect to local environmental conditions. It is worth mentioning that problems with complex topography are outside the scope of this thesis; the recent extension of FIREFLY to complex topography is detailed in Emery et al. (2013).

Accounting for wind-induced wildfire spread in FIREFLY is such that when the wind blows in the direction of the fire spread (i.e., a head fire configuration), the wind contribution to the ROS is maximum since the wind tilts the flame towards the unburnt vegetation and thus, enhances the pyrolysis of the vegetation. On the contrary, the wind contribution to the ROS is zero when the wind blows in the direction opposite to the direction of the fire spread (i.e., a rear fire configuration), meaning that the fire propagates at the value of no-wind ROS on this section of the fire front (i.e., $\Phi_w^* = 0$). On the flanks, the fire front advances faster than in the absence of wind (i.e., $\Phi_w^* > 0$). This implies that the ROS can drastically change along the fireline at a given time. For this purpose, characteristic angles in the horizontal plane (x, y) are defined to represent the direction angle of the wind (referred to as *wind angle* and noted α_w) and the direction angle of the fire propagation (referred to as *front angle* and noted α_{fr}). These angles are defined from the North direction, namely from the positive y-coordinates and increasing in the clockwise direction as represented in Fig. 6.16. More precisely, the front angle $\alpha_{fr}(x, y, t)$ indicates the outward-pointing normal direction to the fire front denoted by $\mathbf{n}_{fr}$. Since the shape of the fire front varies in space according to the heterogeneous environmental conditions, and since the fire front moves over time, this normal vector is not uniform along the fireline and is modified over time. Thus, $\mathbf{n}_{fr}$ is defined as follows:

$$\mathbf{n}_{fr} \equiv \mathbf{n}_{fr}(x, y, t) = \begin{pmatrix} n_{x,fr}(x, y, t) \\ n_{y,fr}(x, y, t) \end{pmatrix} = \begin{pmatrix} \sin\, \alpha_{fr}(x, y, t) \\ \cos\, \alpha_{fr}(x, y, t) \end{pmatrix}. \tag{6.7}$$

The wind angle α_w indicates the direction in which the wind is blowing. According to the wind magnitude u_w^* [m/s] and direction α_w [°] that can be spatially-distributed and time-dependent, the wind velocity vector $\mathbf{u}_w^*$ reads:

$$\mathbf{u}_w^* \equiv \mathbf{u}_w^*(x, y, t) = \begin{pmatrix} u_w^* \sin\, \alpha_w \\ u_w^* \cos\, \alpha_w \end{pmatrix}. \tag{6.8}$$

To apply Rothermel's model, the wind velocity vector $\mathbf{u}_w^*$ is projected along the (local) normal direction to the fireline $\mathbf{n}_{fr} = \mathbf{n}_{fr}(x, y, t)$. Using these notations, the wind velocity magnitude at mid-flame height u_w (see Table 6.1) corresponds to the global wind velocity vector $\mathbf{u}_w^*$ projected along the front angle α_{fr}. Thus, $u_w = u_w(x, y, t)$ is obtained using the following equation:

$$u_w = \begin{cases} \mathbf{u}_w^*(x, y, t) \cdot \mathbf{n}_{fr}(x, y, t), & \text{if } \mathbf{n}_{fr}(x, y, t) \cdot \mathbf{u}_w^* > 0 \\ 0, & \text{if } \mathbf{n}_{fr}(x, y, t) \cdot \mathbf{u}_w^* \leq 0 \end{cases}. \tag{6.9}$$

The projected wind velocity at mid-flame height $u_w = u_w(x, y, t)$ is a time-dependent and spatially-varying quantity along the propagating fireline. The wind contribution Φ_w^* is forced to a zero-value in FIREFLY when the scalar

product $\mathbf{n}_{fr}(x,y,t)\cdot\mathbf{u}_w^*$ is negative to ensure that the ROS Γ remains positive. This is consistent with the common assumption in fire research that the fire propagates at least at the no-wind no-slope ROS.

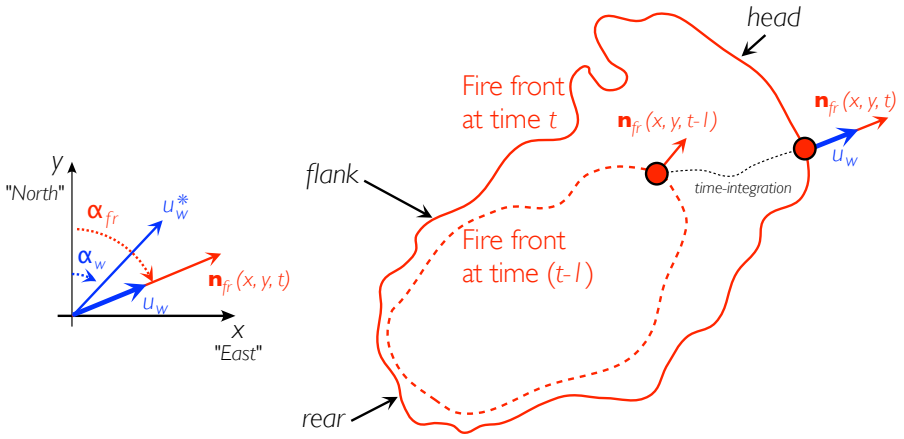


Figure 6.16: *Schematic of wind-induced fire spread for a two-dimensional configuration in the horizontal map (x,y), in which the wind direction angle α_w and the direction angle of fire propagation α_{fr} are not aligned.*

↪ The level-set-based solver

Since wildfires generally feature a front-like geometry at large regional scales, wildfire spread is described within the FIREFLY simulator as a thin flame zone that self-propagates normal to itself into unburnt vegetation. This approach relies on the assumption that a wildfire exhibits a topology similar to premixed flames at regional scales (see Chapter 2).

A large number of techniques dealing with infinitely thin interfaces are available in the literature, among whom the Eulerian front-tracking techniques and in particular the level-set method commonly used to simulate the evolution of a moving interface (Osher and Sethian, 1988; Sethian, 1999). While extensively investigated in the area of combustion (Kerstein et al., 1988) and computer vision (Chaudhury and Ramakrishnan, 2007), and while providing a robust and stable solution, the level-set method has been recently experimented on wildfire spread with promising results (Fendell and Wolff, 2001; Rehm and McDermott, 2009; Mallet et al., 2009; Mandel et al., 2011; Lautenberger, 2013). As discussed by Rehm and McDermott (2009) and Lautenberger (2013), the level-set method offers several advantages over their Lagrangian counterparts for simulating wildfire spread. In particular, spot fire formation, fire mergers and crossovers can be handled without any specific treatment unlike Lagrangian methods. For instance, Mandel et al. (2011) showed its applicability and performance for complex vegetation and terrain topography in the WRF-FIRE coupled fire/atmosphere simulator (see Chapter 1).

In FIREFLY, an Eulerian level-set approach is adopted to numerically propagate

the fire front at the Rothermel-based ROS.[6] We consider a classical approach taken from the premixed combustion literature, in which a reaction progress variable noted $c = c(x, y, t)$ is introduced as a flame marker: $c = 0$ in the unburnt vegetation, $c = 1$ in the burnt vegetation, and the flame is the region where c takes values between 0 and 1 (the flame front is identified as the progress variable iso-contour $c_{fr} = 0.5$) as illustrated in Fig. 6.14. Within the level-set framework, the reaction progress variable c is calculated as a solution of the following propagation equation over the computational domain Ω:

$$\begin{aligned} &\partial c/\partial t(x, y, t) = \mathit{\Gamma}\,|\nabla c|\,, \quad && \forall (x, y) \in \Omega,\ t \geq 0, \\ &c(x, y, 0) = c_0(x, y), && \forall (x, y) \in \Omega, \\ &\nabla c(x, y, t) \cdot \mathbf{n}_b(x, y) = 0, && \forall (x, y) \in \partial\Omega, \end{aligned} \tag{6.10}$$

where $\mathit{\Gamma}$ is specified by the Rothermel-based ROS in Eq. (6.5) using the wind projection u_w along the normal direction to the iso-contour $c_{fr} = 0.5$ of the progress variable given by Eq. (6.9), and where Ω represents the two-dimensional computational domain (with $\partial\Omega$ the boundary of this computational domain and $\mathbf{n}_b$ its normal vector). Thus, the progress variable $c = c(x, y, t)$ is a two-dimensional field solved over the whole computational domain Ω.

▷ **Numerical treatment.** In Firefly, Eq. (6.10) is solved using a second-order Runge-Kutta scheme for time-integration and an advection algorithm for spatial discretization based on a second-order total variation diminishing (TVD) scheme combined with a Superbee slope limiter. This numerical scheme originally proposed by Rehm and McDermott (2009) is detailed in Appendix G and ensures that the iso-contour $c_{fr} = 0.5$ is propagated consistently with Rothermel-based ROS $\mathit{\Gamma}$, i.e., the main physical quantity in Firefly.

↪ The iso-contour algorithm for the reconstruction of the fire front

Once the spatio-temporal variations of the progress variable c are known, the position of the fire front is extracted using a simple iso-contour algorithm such that, formally, the outputs of the Firefly model are

$$\Big[(x_i, y_i), 1 \leq i \leq N_{fr}\Big] = \mathcal{M}_{[t-1,t]}(c_{t-1}, \theta_{t-1}),$$

where (x_i, y_i) represents the two-dimensional coordinates of the N_{fr} fire front markers obtained at time t, where c_{t-1} designates the spatial distribution of the progress variable c at time $(t-1)$ (i.e., the initial condition of the time window $[t-1, t]$), and where θ_{t-1} designates the list of input parameters of the ROS model

[6] This Eulerian front-tracking approach differs from Forefire (used in the ANR-IDEA project), which is based on a Lagrangian discrete-event approach to numerically propagate the flame region, while still relying on semi-empirical ROS models, see Section 1.3.3, Chapter 1.

at time $(t-1)$. According to Table 6.1,

$$\theta = \left(\delta_v, M_v, M_{v,ext}, \Sigma_v, m''_v, \rho_p, \Delta h_c, \mathbf{u}_w, \alpha_{sl}\right).$$

The general structure of Firefly is schematized in Fig. 6.17. It is worth noting that the underlying solver requires a two-dimensional progress variable field as initial condition. If the simulated fire front is corrected at a given time t through a data assimilation technique, the forward model requires the reconstruction of the corresponding progress variable field (i.e., the burnt area) to be integrated beyond time t as explained further in the manuscript.

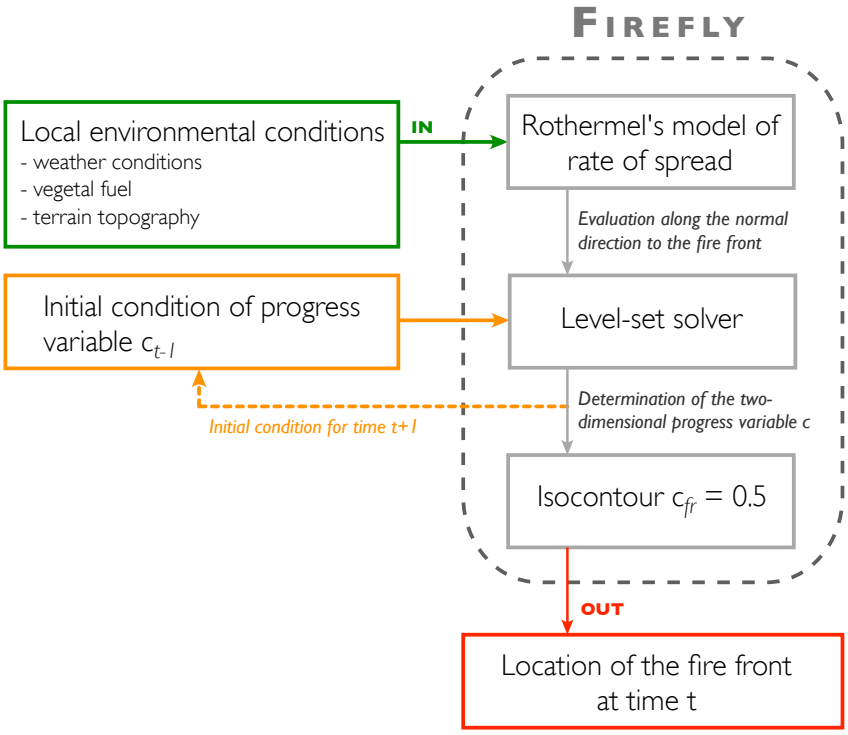


Figure 6.17: *Schematic of the* Firefly *front-tracking simulator.*

6.4.2 Validation

↪ Model performance metrics

Diagnostics on the propagating speed and thickness of the fire front, derived from Kolmogorov–Petrovsky–Piskounov (KPP) analysis (Poinsot and Veynante, 2005) and extrapolated to heterogeneous vegetations, have been developed to demonstrate the accuracy of wildfire spread simulations with Firefly[7] (Rochoux et al.,

[7]The proposed KPP-based diagnostics were originally developed for a reaction-diffusion equation in Rochoux et al. (2010) and then extrapoled for a level-set-based propagating equation in Delmotte et al. (2011).

2010; Delmotte et al., 2011). Global quantities characterizing the wildfire spread are the average speed and thickness along the fire front at a given time t.

The average thickness of the fire front is diagnosed by the inverse of the maximum gradient of the progress variable c (see Chapter 2). This diagnostic denoted by $\delta_{fr,d}$ reads:

$$\delta_{fr,d} = \left(\max \left| \frac{\partial c}{\partial \mathbf{n}_{fr}} \right|_{c=0.5} \right)^{-1}, \tag{6.11}$$

where $\mathbf{n}_{fr}$ represents the normal direction to the isocontours of the progress variable c and thereby, to the fire front.

Even though the thickness of the fire front has no physical meaning within the level-set framework, it is useful in practice to evaluate the effect of numerical diffusion. The average speed along the fire front noted $\overline{I}$ is determined by the rate of change of the progress variable $c = c(x, y, t)$ over the whole computational domain Ω. These quantities are defined as follows:

$$\overline{I} = \frac{\left(\int_{c_{fr}=0.5} I \, d\mathcal{C} \right)}{\left(\int_{c_{fr}=0.5} d\mathcal{C} \right)}, \quad \overline{I}_d = \frac{1}{\left(\int_{c_{fr}=0.5} d\mathcal{C} \right)} \frac{d}{dt} \left(\iint_{\Omega} c(x, y, t) \, dxdy \right), \tag{6.12}$$

where $d\mathcal{C}$ corresponds to a small variation in arc length along the fire contour, and where $(\int_{c_{fr}=0.5} d\mathcal{C})$ corresponds to the fireline perimeter identified as $c_{fr} = 0.5$. The formulation of $\overline{I}_d$ is based on the following arguments:

$$\begin{aligned} \frac{d}{dt} \left(\iint_{\Omega} c(x, y, t) \, dxdy \right) &= \iint_{\Omega} \frac{\partial c}{\partial t}(x, y, t) \, dxdy \\ &= \iint_{\Omega} I(x, y, t) \; |\nabla c| \; dxdy \\ &= \int_{c_{fr}=0.5} I(x, y, t) \, d\mathcal{C}, \\ \Rightarrow \frac{d}{dt} \left(\iint_{\Omega} c(x, y, t) \, dxdy \right) &= \overline{I}_d \int_{c_{fr}=0.5} d\mathcal{C}, \end{aligned}$$

using Eq. (6.10) and the following definition for the flame surface area:

$$|\nabla c| = \frac{d\mathcal{C}}{dxdy}. \tag{6.13}$$

↪ FIREFLY simulation experiments

The accuracy of the FIREFLY simulator is validated for different conditions of vegetation distribution δ_v [m] and wind magnitude u_w [m/s]. These simulations

correspond to the propagation of a semi-circular front over a two-dimensional domain of 200 m $\times$ 200 m (with a mesh resolution $\Delta x = \Delta y = 1$ m). The initial condition is described by a semi-circular front centered at ($x_0 = 100$ m, $y_0 = 0$ m) and of radius $r_0 = 5$ m. Equation (6.10) is integrated during 800 s with a time step $\Delta t = 0.5$ s. Different simulation experiments listed below are performed to show the consistency and robustness of FIREFLY.

- ▷ **Simulation 1 (uniform vegetation distribution).** This simulation represents a fire spread over a horizontal vegetation layer characterized by a uniform distribution $\delta_v = 1$ m. There is no external flow ($u_w = 0$). Thus, the proportionality coefficient P in Eq. (6.6) is uniform, i.e., $P = 0.1$ s^{-1}. This implies, theoretically, a uniform ROS $\Gamma = 0.1$ s and thereby, an isotropic propagation of the fire front. Figure 6.18 shows that the front propagates at the prescribed $\Gamma = 0.1$ m/s; the rate of change of the progress variable c over the computational domain noted $\overline{\Gamma}_d$ matches the average speed of the fire front $\overline{\Gamma}$. Also the fire front thickness $\delta_{fr,d}$ remains small and relatively constant over time, with the ratio of the thickness to the cell size satisfying $(\delta_{fr,d}/\Delta) \simeq 3$ (Δ representing the minimum mesh resolution along the x- and y-directions, i.e., $\Delta = \min(\Delta x, \Delta y) = 1$ m).

- ▷ **Simulation 2 (random vegetation distribution).** This simulation represents a fire spread over a horizontal fuel layer characterized by a randomly-distributed $\delta_v = \delta_v(x, y)$, varying between 0.4 m and 0.8 m as illustrated in Fig. 6.19(a). There is no external flow ($u_w = 0$) and $P = 0.1$ s^{-1}, leading to a slight anisotropic propagation as shown in Fig. 6.19(b).[8] The ROS Γ varies, consistently, between 0.065 m/s and 0.075 m/s, see Fig. 6.20.

- ▷ **Simulation 3 (wind-aided propagation).** This simulation represents a fire spread in presence of a moderate wind $u_w = 0.5$ m/s flowing northward ($\alpha_w = 0°$), which occurs over a horizontal fuel layer characterized by a randomly-distributed $\delta_v = \delta_v(x, y)$, varying between 0.4 m and 0.8 m similarly to Fig. 6.19(a). Figure 6.21 displays the simulated wind-aided propagation of the progress variable isocontour $c_{fr} = 0.5$ at 200 s intervals, starting from the initial condition. Consistently, the ROS $\overline{\Gamma}_d$ matching the reference $\overline{\Gamma}$ is higher than in the no-wind configuration (Simulation 2) and reaches up to 0.125 m/s as demonstrated in Fig. 6.22.

In summary, these diagnostics of ROS and direction for isotropic and anisotropic fire propagation show a consistent average fire front speed with the ROS submodel Γ and a relatively constant fire front thickness over time, which demonstrates the non-diffusive behavior of the TVD-based numerical scheme retained in FIREFLY.

[8] The key parameter to define the importance of ROS anisotropy is based here on the length-scale at which the biomass fuel layer thickness δ_v fluctuates. In the present no-wind case, this anisotropy is limited due to the relatively small size of biomass fuel pockets, see Fig. 6.19(b).

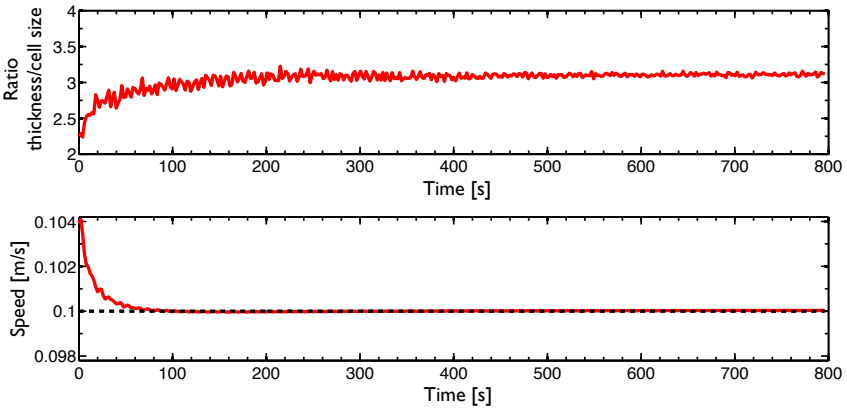
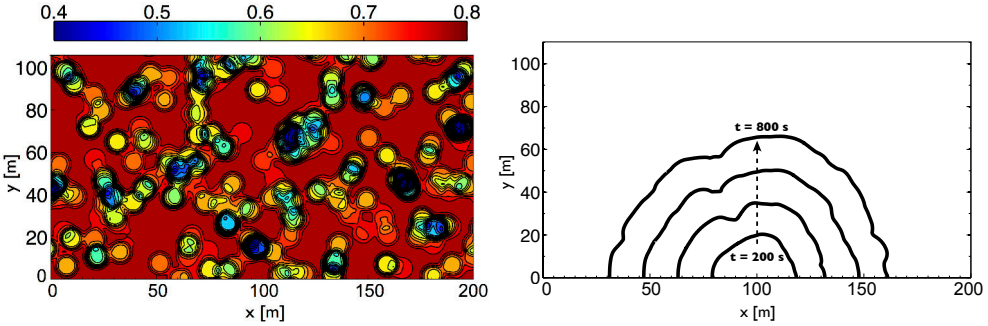


Figure 6.18: *Test of* FIREFLY *for an isotropic fire spread (Simulation 1). Top: Ratio of the front thickness $\delta_{fr,d}$ [m] to the mesh cell size Δ [m]. Bottom: Comparison of the average speed $\overline{\Gamma}$ in black dashed line with its diagnostic $\overline{\Gamma}_d$ in red solid line. Credit: Delmotte et al. (2011).*



(a) Random distribution of the vegetation layer depth $\delta_v(x, y)$ [m].

(b) Time-evolving location of the fire front from $t = 200$ s to 800 s at 200 s intervals.

Figure 6.19: FIREFLY *simulation of a wildfire spread with no wind and a randomly-distributed vegetation layer depth (Simulation 2).*

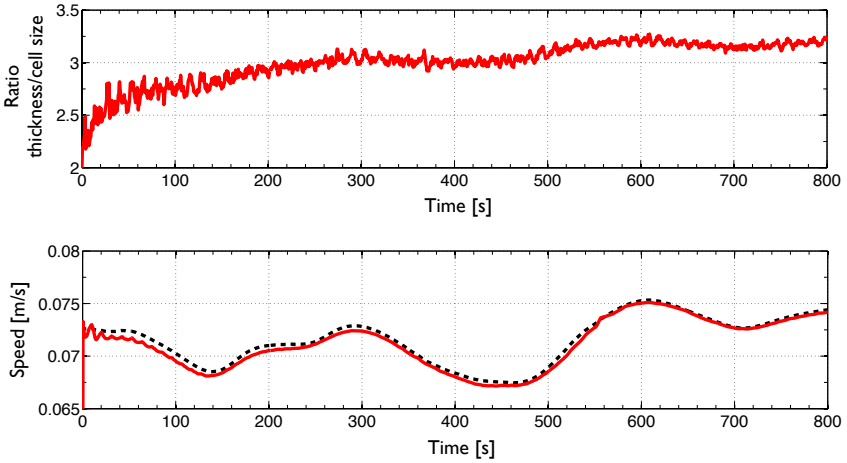


Figure 6.20: *See caption of Fig. 6.18 for a no-wind fire spread configuration with a biomass random distribution (Simulation 2).*

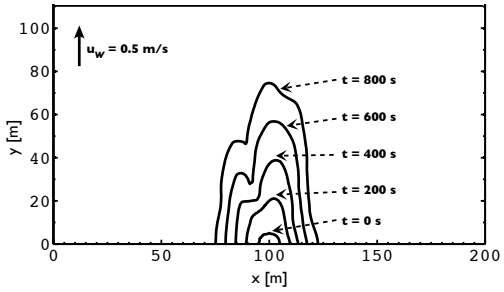


Figure 6.21: *Time-evolving location of the fire front at 200 s intervals until $t = 800$ s in presence of wind, with $u_w = 0.5\ m/s$ and $\alpha_w = 0°$ (Simulation 3).*

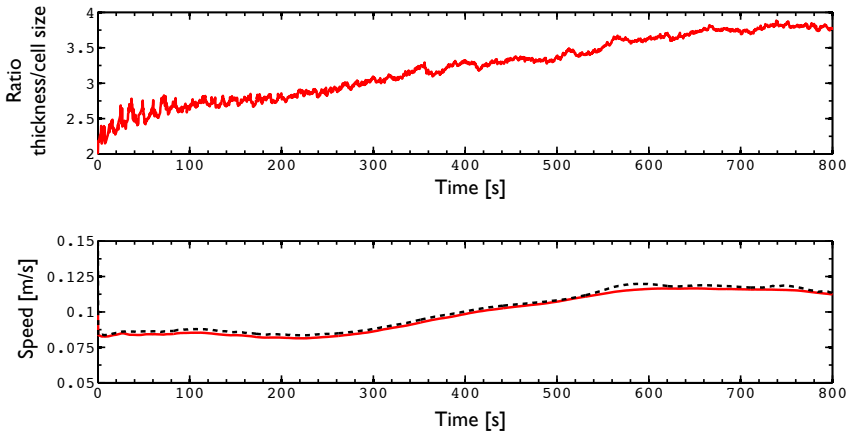


Figure 6.22: *See caption of Fig. 6.18 for a wind-aided fire spread configuration with a biomass random distribution (Simulation 3).*

6.4.3 Sensitivity study with respect to environmental conditions

To illustrate the sensitivity of the FIREFLY outputs to the ROS input parameters (in terms of fireline topology and behavior), an ensemble of simulations is performed over a two-dimensional domain of 400 m $\times$ 400 m (with a mesh cell size $\Delta x = \Delta y = 1$ m). The fire is ignited as a circular front centered at $(x_{ign}, y_{ign}) = (200\text{ m}, 200\text{ m})$ and of radius 5 m. It propagates over a flat terrain, in presence of a moderate wind $u_w = 1$ m/s flowing westward ($\alpha_w = 315°$). The vegetation is characterized by a uniform moisture content $M_v = 20\ \%$ and Rothermel's standard values for short grass (see Appendix A). Besides, the biomass fuel layer thickness is non-uniform, with $\delta_{v,1} = 0.5$ m (for $x < 200$ m) and $\delta_{v,2} = 1.0$ m (for $x \geq 200$ m). These values are referred to as the *nominal environmental conditions*.

The wildfire spread model is integrated during 200 s (with a time step $\Delta t = 0.5$ s) for different perturbations of the nominal conditions. Figure 6.23 presents the

resulting scatter of the simulated fire fronts. It was found that changes in the wind magnitude only affect the head of the fire, while changes in the moisture content and the vegetation layer depth induce modifications all along the fireline (with a maximum change in the wind direction). Changes in the wind direction significantly modify the direction of wildfire spread, but only slightly the shape of the fireline in the present case. In contrast, the spatially-distributed vegetation layer depth modifies the shape of the fireline (in particular at the interface between the values $\delta_{v,1}$ and $\delta_{v,2}$) and thereby, the direction of wildfire spread. These results imply that combining perturbations in the different parameters $[u_w, \alpha_w, M_v, \delta_{v,1}, \delta_{v,2}]$ results in a wide range of fire front shapes and behaviors, which is necessary to describe modeling uncertainties within a data assimilation system.

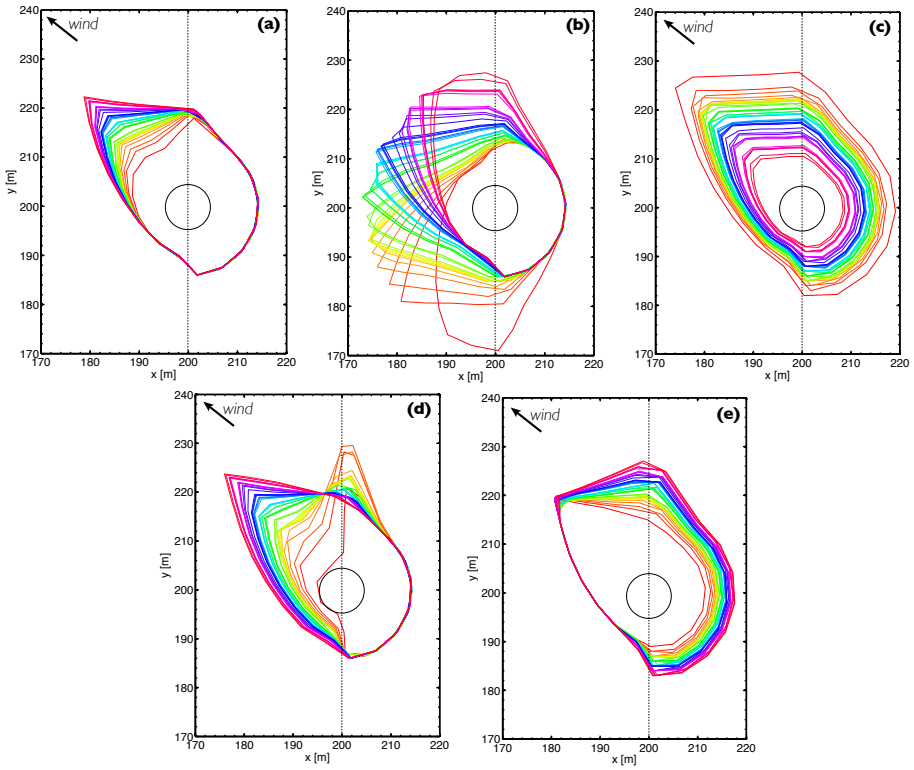


Figure 6.23: *Sensitivity of the simulated fire front locations (the colored fire fronts are discretized using $N_{fr} = 20$ markers) using* FIREFLY *with respect to perturbations in nominal environmental conditions. (a) Wind magnitude u_w (0.85 m/s $\pm$ 0.15 m/s). (b) Wind direction α_w ($280° \pm 40°$). (c) Fuel moisture content M_v ($20\,\% \pm 4\,\%$). (d) Fuel depth in zone 1 $\delta_{v,1}$ (0.4 m $\pm$ 0.15 m). (e) Fuel depth in zone 2 $\delta_{v,2}$ (1.10 m $\pm$ 0.15 m). The black circle represents the initial circular front, the vertical dotted line represents the interface between the two zones of distinct biomass layer depth, $\delta_{v,1}$ and $\delta_{v,2}$.*

In order to identify to which input parameters the ROS Γ is the most sensitive among the environmental conditions, Fig. 6.24 presents the ROS values of the head of the fire when uncertainties are assumed in six parameters: in addition to

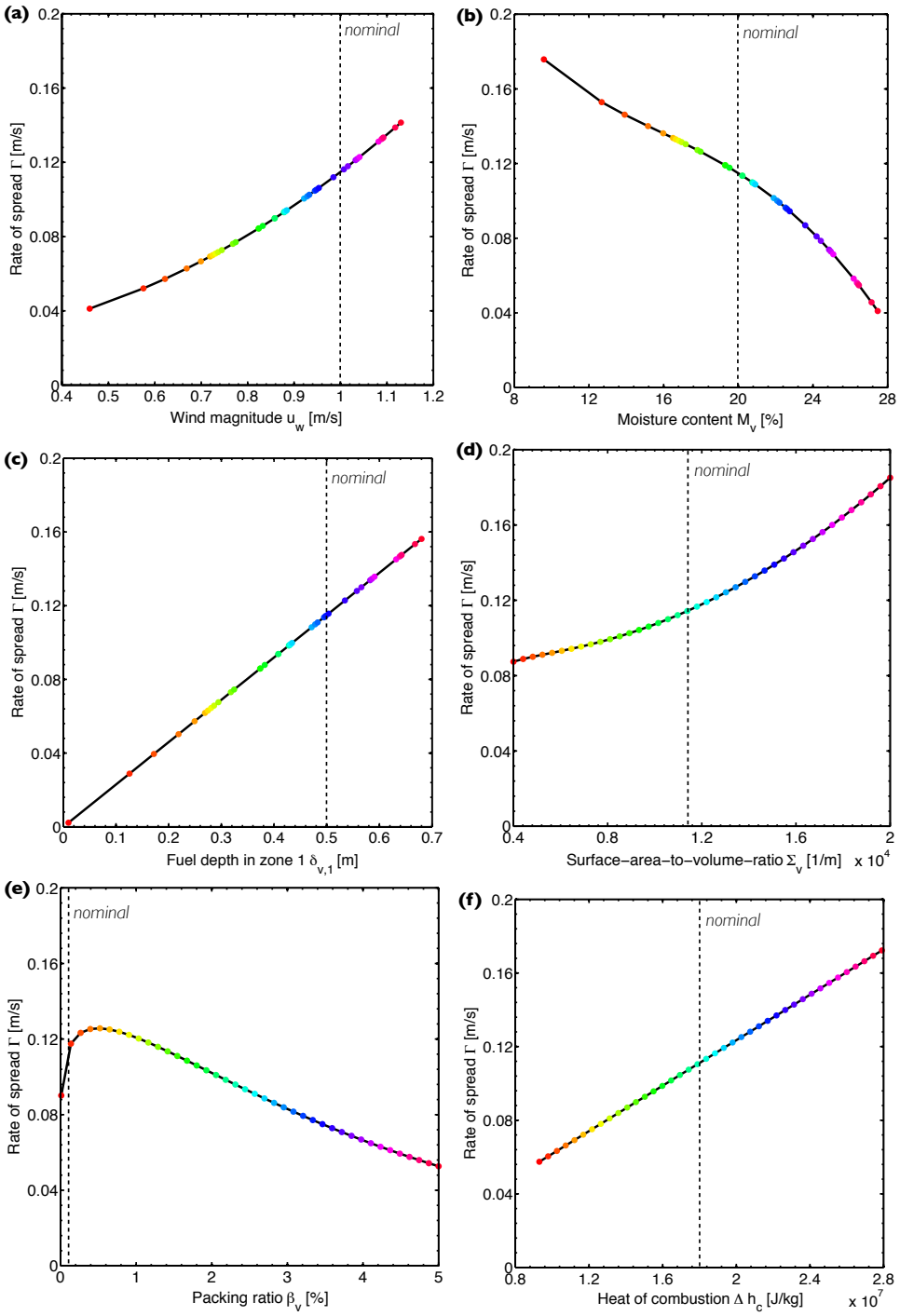


Figure 6.24: *Sensitivity of the head-fire ROS with respect to perturbations in nominal environmental conditions. (a) Wind magnitude u_w. (b) Fuel moisture content M_v. (c) Fuel depth in zone 1 $\delta_{v,1}$. (d) Fuel particle surface-area-to-volume ratio Σ_v [1/m]. (e) Fuel packing ratio β_v [%]. (f) Heat of combustion Δh_c [J/kg]. The vertical dotted line represents the nominal value for each input parameter of the ROS model (either the nominal value in the present simulation, or the standard value for short-grass in Rothermel's fuel database).*

u_w, M_v and $\delta_{v,1}$ (presented in Fig. 6.23), we also study the ROS sensitivity to the fuel particle surface-area-to-volume ratio Σ_v, the fuel packing ratio β_v and the fuel heat of combustion Δh_c. Indeed, the identification of which parameters are important to include in the control parameters $\mathbf{x}$ is an essential step towards the application of data assimilation to the FIREFLY wildfire spread simulator. In particular, the key idea when dealing with parameter estimation is to focus the correction on a reduced set of parameters that have significant uncertainties and to which FIREFLY is the most sensitive.

It was found that the ROS values feature a wide, nearly identical, scatter for the six parameters. However, it was also shown that the ROS depends non-linearly on the variations in $[u_w, M_v, \Sigma_v, \beta_v]$. Since their variations can induce significant changes in the wildfire behavior, these four parameters are critical to estimate in order to correctly forecast wildfire spread and anticipate future responses. These model non-linearities will be more important when the wind magnitude fluctuates over time or when the fire active area is covered heterogeneously by different types of vegetation. This highlights the importance of applying a data assimilation methodology able to handle multiple sources of non-linearity in the wildfire spread model.

Note that this sensitivity study did not assume any particular type for the PDF related to the errors in the ROS model parameters. If this study were performed using a Monte Carlo sampling technique (in a data assimilation framework for instance), a PDF on the ROS model parameters would be chosen and thereby, the PDF of the model outputs of interest such as the head fire spread-rate, the burnt area or the location of the fireline would be studied. Similar plots as Fig. 6.23 would be obtained in addition to quantile- or moment-based plots.

6.4.4 Comparison between simulations and observations

To apply a data assimilation algorithm for both state and parameter estimations, we need to map the outputs of the FIREFLY computer model onto the observation space that is spanned by discrete observed fire front contours. Thus, the data assimilation technique uses a discretization of both the simulated and observed fire fronts, called SFF and OFF, respectively.

- ▷ The discretization of SFF is a set of N_{fr} markers that are characterized at a given time by the following two-dimensional coordinates

$$\mathbf{y}_t = \Big[(x_1, y_1), \ldots, (x_{N_{fr}}, y_{N_{fr}})\Big].$$

- ▷ Similarly, the discretization of OFF is a set of N^{o}_{fr} markers; the observation vector $\mathbf{y}^{\mathrm{o}}_t$ is defined as:

$$\mathbf{y}^{\mathrm{o}}_t = \Big[(x^{\mathrm{o}}_1, y^{\mathrm{o}}_1), \ldots (x^{\mathrm{o}}_{N^{\mathrm{o}}_{fr}}, y^{\mathrm{o}}_{N^{\mathrm{o}}_{fr}})\Big].$$

Note that the FIREFLY solver uses a high-resolution computational grid that allows for a detailed representation of the local conditions (the spatial resolution is on

the order of 1 m). In contrast, observations of the fire front position are likely to be provided with a much coarser resolution; in addition, observations may be incomplete and cover only a fraction of the fire front perimeter. Thus, we may expect N_{fr}^{o} to be much lower than N_{fr}. In the following, we assume for simplicity that $N_{fr}^{\mathrm{o}} = (N_{fr}/r)$, where r is an integer taking values (much) larger than 1. In order to map the state variable space (SFF) onto the observation space (OFF), a selection operator $\mathcal{H}$ (see Section 5.2, Chapter 5) is introduced that selects a subset of N_{fr}^{o} markers among the fine-grained discretization of SFF and pairs each one of those markers with one of those used in the coarse-grained discretization of OFF (see Fig. 6.25).

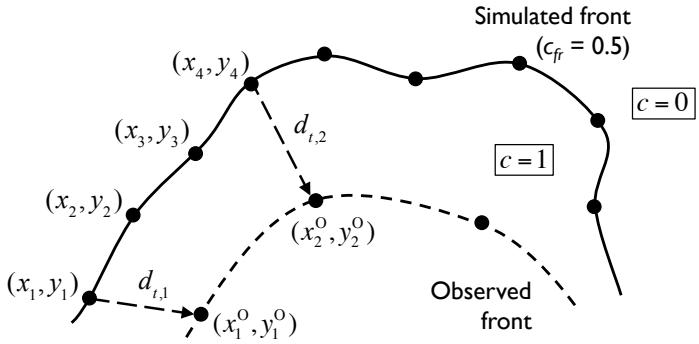


Figure 6.25: *Construction of the innovation vector $\mathbf{d}_t$ introduced to quantify the differences between simulated and observed fire fronts. In this illustration, $r = 4$.*

The selection operator $\mathcal{H}$ may be defined in several ways (for instance using a projection scheme as explained in Appendix G) but preliminary tests have shown that a simple treatment (taking 1 out of every r points) provided reasonable results. The distance between simulated and observed fire fronts is recast in the vector $\mathbf{d}_t$ of dimension $2\,N_{fr}^{\mathrm{o}}$ and is now simply defined as the vector formed by the directed distances between the paired SFF-OFF markers as illustrated in Fig. 6.25, with:

$$\mathbf{d}_t = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t = \begin{pmatrix} x_1^{\mathrm{o}} - x_1 \\ x_2^{\mathrm{o}} - x_2 \\ \vdots \\ x_{N_{fr}^{\mathrm{o}}}^{\mathrm{o}} - x_{N_{fr}} \\ y_1^{\mathrm{o}} - y_1 \\ y_2^{\mathrm{o}} - y_2 \\ \vdots \\ y_{N_{fr}^{\mathrm{o}}}^{\mathrm{o}} - y_{N_{fr}} \end{pmatrix}. \tag{6.14}$$

6.5 Flowchart for parameter and state estimations

6.5.1 Objectives of the data-driven simulation capability

The prototype data-driven wildfire simulator we propose relies on a sequential data assimilation algorithm that assimilates observations of the time-evolving location of the fire front. It features a choice between a parameter estimation approach (in which the estimation targets are the ROS model parameters θ) or a state estimation approach (in which the estimation targets are the positions of the fire front $[(x_i, y_i), 1 \leq i \leq N_{fr}]$ that can be regarded as the model state in FIREFLY). Figure 6.26 highlights the general differences between parameter estimation and state estimation approaches.

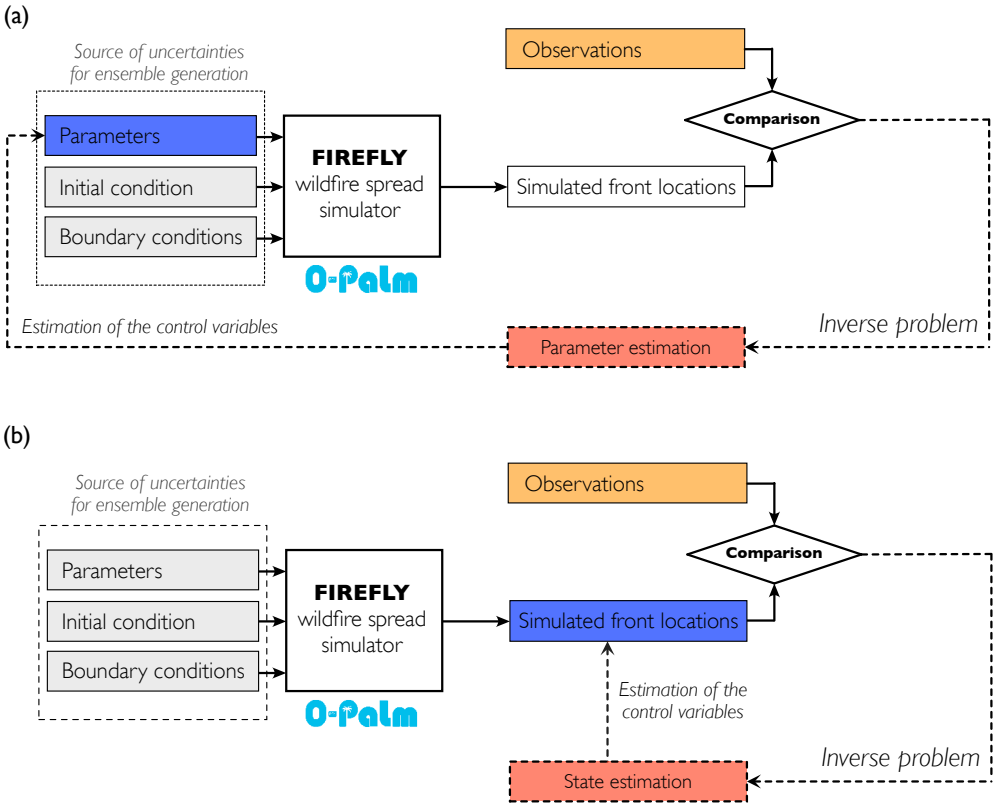


Figure 6.26: *Data assimilation flowchart for (a) parameter estimation and (b) state estimation (control variables are colored in blue).*

The cornerstone of this data-driven simulation capability is to find a data assimilation algorithm that:

(1) accounts for non-linearities in the wildfire spread behavior;

(2) is suitable for the dimensionality of the problem;

(3) is consistent with the operational framework (i.e., low computational cost)

and the available numerical tools (the tangent-linear and adjoint operators of FIREFLY are not available);

(4) reliably delivers an accurate forecast of the time-evolving location of the fire front with a positive lead-time (i.e., achieving a performance faster than real-time).

For this purpose, the performance of different data assimilation algorithms is compared in this thesis:

- ▷ **Extended Kalman filter versus ensemble Kalman filter.** The extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) are tested for wildfire spread forecasting, in order to highlight the advantages of Kalman filtering for parameter estimation and to point out the limitations of non-linear treatments, in particular in the EKF (the model and observation operators are linearized using finite differences within an iterative algorithm, see Section 5.6.1, Chapter 5). The morphing EnKF proposed by Beezley and Mandel (2008) and Beezley (2009) led to technical difficulties in the data assimilation implementation due to the registration procedure applied on two-dimensional temperature and sensible heat flux fields, these variables exhibiting a bi-modal PDF. In the present study, the time-evolving locations of the fire front are taken as the observed quantities and these data are expected to feature an approximately Gaussian PDF, which allows for a straightforward application of the classical EnKF as presented in Evensen (1994) and Houtekamer and Mitchell (1998).
- ▷ **Ensemble Kalman filter versus particle filters.** A comparison of the EnKF with particle filters (SIR and ASIR algorithms) is provided in the context of parameter estimation to ensure that the Gaussian assumption on the error statistics used in the derivation of the Kalman filter equations does not degrade over time the quality of the EnKF forecast and analysis (the Kalman filter is an analytical solution of the Bayesian filtering problem for a linear model and Gaussian error statistics, see Section 5.5, Chapter 5).
- ▷ **Reduced-cost ensemble-based data assimilation strategy.** Both EnKF and particle filter approaches require a large ensemble to properly describe uncertainties and correlations between the physical variables that are controlled, and to allow for an anisotropic correction of the fire front location. To reduce the computational cost of such ensemble-based approaches, we develop a new cost-effective data assimilation strategy for parameter estimation, inspired by uncertainty quantification techniques and relying on the generation of an approximate surface response of the wildfire spread model FIREFLY to the control variables. This surface response based on a polynomial chaos (PC) expansion, also referred to as *surrogate model*, is used in place of the forward model (i.e., the wildfire spread simulator FIREFLY). This, in order to build the ensemble of forecasts and analyses with a dramatically reduced computational cost compared to the classical EnKF and

without loss of accuracy.

- ▷ **Parameter estimation versus state estimation.** An EnKF-based state estimation approach has been developed to directly estimate the location of the fire front markers and to account for all the possible sources of uncertainty in FIREFLY, i.e., in the input parameters of the ROS model as well as in the parameterization of the ROS. In this approach, attention must be paid in the ensemble generation to properly characterize the error correlations in the front marker location along the fireline. and thereby, allowing for a spatialized correction of the fire front topology and behavior. A comparison of the EnKF-based performance for parameter and state estimations is performed on the real-case controlled grassland burn experiment.

This work aims at demonstrating the feasibility of data assimilation for surface processes of wildfire spread and therefore, does not directly consider fire/atmosphere interactions. Even though the ultimate goal of this research is to provide real-time fire forecasts using thermal-infrared imaging including a description of both wildfire dynamics and plume emissions, multi-scale fire/atmosphere interactions are beyond the scope of this work. The issues related to data assimilation for coupled physical systems are therefore not addressed here. Thus, we provide a data assimilation strategy that is suitable for surface wildfire spread, but that would need further developments for a coupled surface/atmosphere system.

6.5.2 Technical implementation

In practice, combining FIREFLY with a data assimilation algorithm is managed by the OPENPALM[9] dynamic coupling software. While OPENPALM is used in Part I to perform multi-physics flame-scale LES (data parallelism), it is used here in the context of data assimilation as a task-parallelism manager to handle communications and data exchanges between FIREFLY and the different mathematical units required to sequentially perform Bayesian prediction and update steps. In particular, ensemble-based data assimilation algorithms (e.g., EnKF, SIR/ASIR particle filters) require the generation of members (or particles) to stochastically characterize modeling uncertainties in the prediction step. This consists in the computation of the FIREFLY model trajectory for each ensemble member to be compared to the observed fire front. Since each member of the ensemble can be integrated independently, we use the PARASOL functionality of OPENPALM to efficiently launch FIREFLY model integrations, in parallel, on the available processors. A schematic of the PARASOL functionality based on master/slave principles is presented in Fig. 6.27: the Master processor of PARASOL spawns multiple copies of the same computer program (the slaves), each on one or several processors with a different set of input parameters, while each slave processor is in charge of executing one FIREFLY instance and producing the associated fire front position. Thus, starting from an ensemble of input parameters $\mathbf{x}_t^{\mathrm{f}} = [\theta^{\mathrm{f},(1)}, \cdots, \theta^{\mathrm{f},(N_e)}]$, PARASOL

[9] See Appendix B, www.cerfacs.fr/globc/PALM_WEB/.

provides an ensemble of forecast simulated marker positions designated as:

$$\mathbf{y}_t^{\mathrm{f}} = [(x_i, y_i)^{\mathrm{f},(1)}, (x_i, y_i)^{\mathrm{f},(2)}, \cdots, (x_i, y_i)^{\mathrm{f},(N_e)}]$$

for each associated observed marker i, with i varying between 1 and N_{fr}^{o}.

6.5.3 Estimation of the input parameters of the ROS model

In the context of parameter estimation, the estimation targets $\mathbf{x}$ are the input parameters of the Rothermel's ROS model, in particular the wind magnitude u_w and direction angle α_w as well as the properties of the vegetation (e.g., moisture content M_v, solid particle surface-area-to-volume ratio Σ_v) to which FIREFLY is sensitive.

▷ **Forecast step.** For ensemble-based data assimilation methodologies, an ensemble of realizations of these control parameters is generated based on random perturbations of the uncertain values (provided by field measurements or physical analysis and following Gaussian distributions). This ensemble of control parameters corresponds to the forecast ensemble $\mathbf{x}_t^{\mathrm{f}}$. Then, a series of N_e independent FIREFLY integrations up to the analysis time t (based on these N_e realizations of the control parameters) is performed using the PARASOL functionality of OPENPALM. This provides N_e fire front positions at time t designated as:

$$\mathbf{y}_t^{\mathrm{f}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}) = \left[\mathbf{y}_t^{\mathrm{f},(1)}, \mathbf{y}_t^{\mathrm{f},(2)}, \cdots, \mathbf{y}_t^{\mathrm{f},(N_e)}\right].$$

▷ **Update step.** During the analysis, each ensemble member is updated, either based on the determination of weights for particle filters, or based on the classical KF update equation presented in Eq. (5.89), Chapter 5, for the EnKF algorithm. In EnKF, there is an important difference in the stochastic calculation of the gain matrix $\mathbf{K}_t^{\mathrm{e}}$ between the parameter estimation approach and the widely-used state estimation approach. In the context of parameter estimation, following Durand et al. (2008) and Moradkhani et al. (2005), $\mathbf{K}_t^{\mathrm{e}}$ reads:

$$\mathbf{K}_t^{\mathrm{e}} = \underbrace{\mathbf{P}_t^{\mathrm{f,e}}\mathbf{G}_t^{\mathrm{T}}}_{\mathbf{C_{xy}}} \left(\underbrace{\mathbf{G}_t\mathbf{P}_t^{\mathrm{f,e}}\mathbf{G}_t^{\mathrm{T}}}_{\mathbf{C_{yy}}} + \mathbf{R}_t \right)^{-1}, \tag{6.15}$$

where $\mathbf{C_{xy}} \in \mathbb{R}^{n \times 2N_{fr}^{\mathrm{o}}}$ and $\mathbf{C_{yy}} \in \mathbb{R}^{2N_{fr}^{\mathrm{o}} \times 2N_{fr}^{\mathrm{o}}}$ are, respectively, the covariance matrix of the model parameters with the predicted measurements of fire front positions (that represents the stochastically-based relationship between the control space of size n and the observation space of size $p = 2N_{fr}^{\mathrm{o}}$) and the covariance matrix of the predicted measurements. In practice, $\mathbf{C_{xy}}$ and $\mathbf{C_{yy}}$ are directly

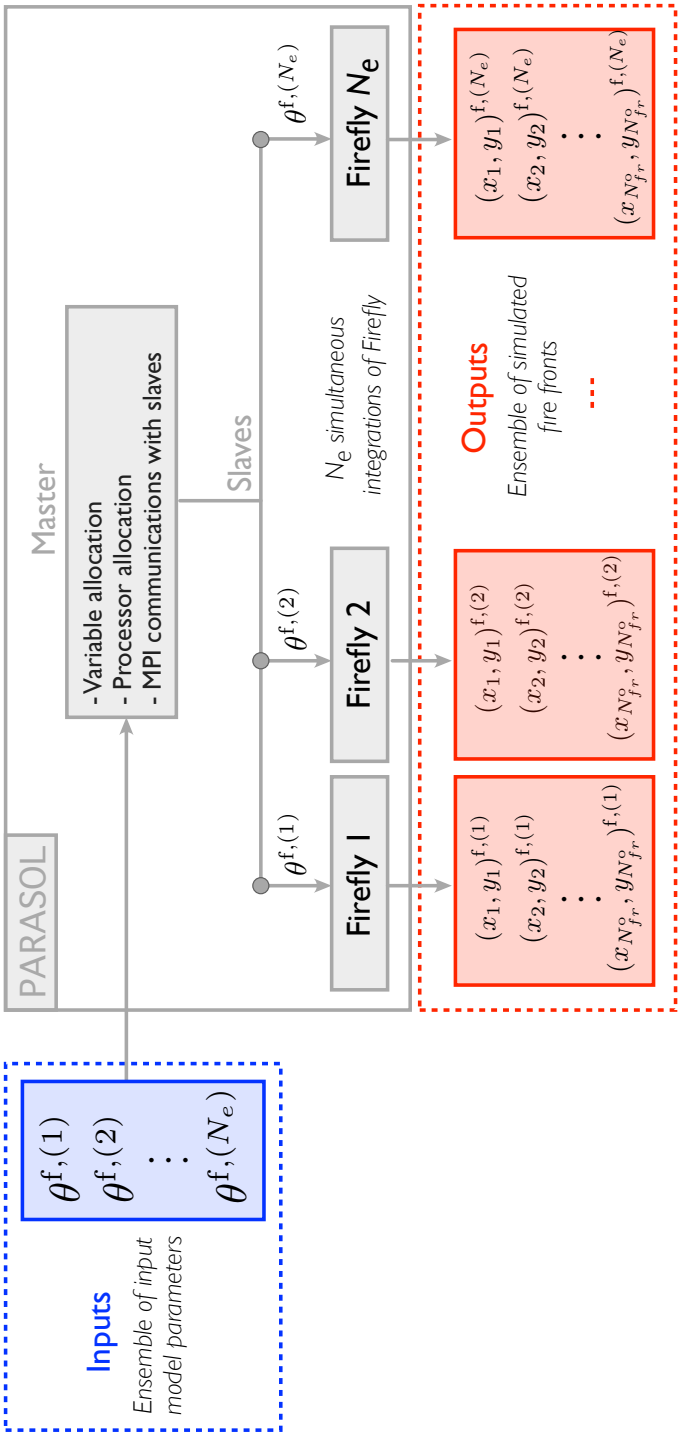


Figure 6.27: *Schematic of the* Parasol *functionality applied to* Firefly.

derived from the ensemble of forecasts:

$$\mathbf{C_{xy}} = \left(\frac{1}{N_e - 1}\right) \sum_{k=1}^{N_e} \left(\mathbf{x}_t^{\mathrm{f},(k)} - \overline{\mathbf{x}_t^{\mathrm{f}}}\right) \left(\mathcal{G}(\mathbf{x}_t^{\mathrm{f},(k)}) - \overline{\mathcal{G}(\mathbf{x}_t^{\mathrm{f}})}\right)^{\mathrm{T}}, \tag{6.16}$$

$$\mathbf{C_{yy}} = \left(\frac{1}{N_e - 1}\right) \sum_{k=1}^{N_e} \left(\mathcal{G}(\mathbf{x}_t^{\mathrm{f},(k)}) - \overline{\mathcal{G}(\mathbf{x}_t^{\mathrm{f}})}\right) \left(\mathcal{G}(\mathbf{x}_t^{\mathrm{f},(k)}) - \overline{\mathcal{G}(\mathbf{x}_t^{\mathrm{f}})}\right)^{\mathrm{T}}, \tag{6.17}$$

where the overline denotes the mean value over the ensemble. This approach avoids the explicit estimation of $\mathbf{P}_t^{\mathrm{f,e}}$ and $\mathbf{G}_t$, which is difficult to compute reliably with respect to control parameters due to model non-linearity. The update step leads to the estimation of more accurate control parameters

$$\mathbf{x}_t^{\mathrm{a}} \;=\; \left[\mathbf{x}_t^{\mathrm{a},(1)}, \mathbf{x}_t^{\mathrm{a},(2)}, \cdots, \mathbf{x}_t^{\mathrm{a},(N_e)}\right].$$

Posterior to data assimilation, FIREFLY is integrated for the analysis ensemble over the same time period as in the forecast step, in order to simulate the retrospective fire front locations $\mathbf{y}_t^{\mathrm{a}}$ associated with the newly-obtained analysis $\mathbf{x}_t^{\mathrm{a}}$ as well as an ensemble of forecasts of the fire spread beyond time t. The flowchart for parameter estimation is schematized in Fig. 6.28.

▷ **Assimilation cycling: Artificial evolution model for ROS input parameters.** In order to allow for a temporal correction of the model parameters between the assimilation cycles, ensemble-based algorithms are sequentially applied. Along the assimilation cycles, since there is no dynamic model $\mathcal{M}$ available to describe the time-evolution of the control parameters and since there is a need to cover multiple possible scenarios for wildfire spread evolution, the parameter evolution is artificially set up with a random walk model (West, 1993; Moradkhani et al., 2005). Each member/particle (indexed by k) reads at the next observation time $(t+1)$:

$$\mathbf{x}_{t+1}^{\mathrm{f},(k)} = \overline{\mathbf{x}_t^{\mathrm{a}}} + \mathbf{e}_t^{(k)}, \tag{6.18}$$

where $\overline{\mathbf{x}_t^{\mathrm{a}}}$ is the mean of the analysis estimates obtained at the previous analysis time t, and where $\mathbf{e}_t^{(k)}$ is a randomly-generated white-noise following a Gaussian distribution of zero mean and given STD; this STD is taken here as the forecast error STD σ_t^{f} over the assimilation cycle $[t-1, t]$. This approach is referred to as the *artificial evolution method* for control parameters.

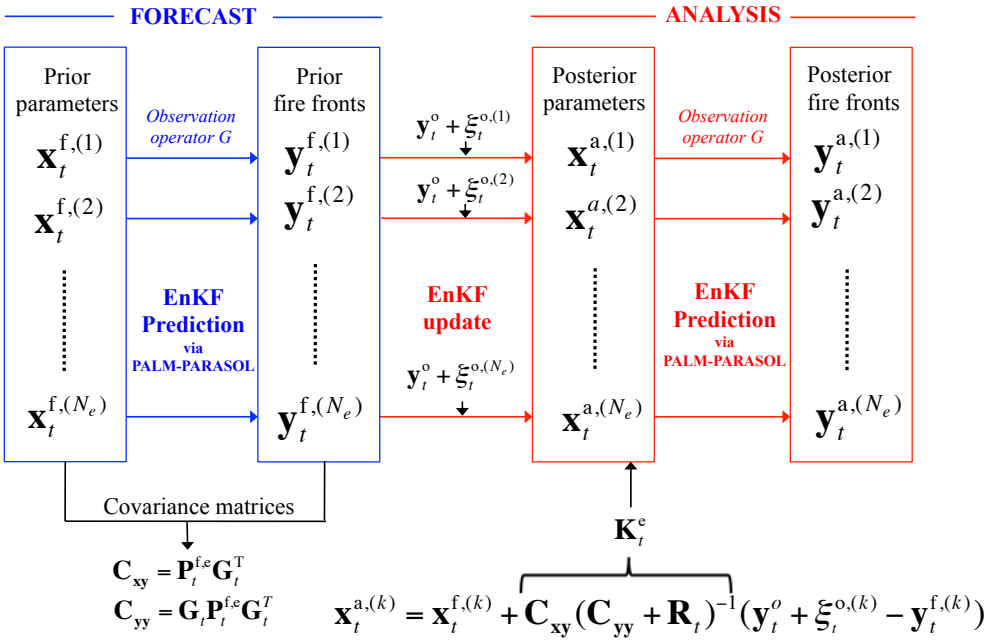


Figure 6.28: *Flowchart of the EnKF algorithm during the* $[t-1,t]$ *assimilation cycle for a parameter estimation approach. Data randomization (Burgers et al., 1998) is used in the EnKF with* $\xi_t^{o,(k)}$ *following observation error statistics for each member* $k=1,\cdots,N_e$.

6.5.4 Estimation of the state of the wildfire spread model

One alternative to parameter estimation is to directly estimate the FIREFLY model state, i.e., the time-evolving location of the fire front. This is referred to as a *state estimation approach* in the following. This approach is attractive for wildfire spread applications. Identifying the different sources of uncertainties in wildfire spread modeling is indeed particularly difficult. For instance, differentiating modeling errors from inaccuracies in input parameters of the ROS model is a complex task. On top of these inaccuracies (that express an imperfect knowledge and that could in theory be removed or at least substantially decreased), aleatoric errors (that result from natural and unpredictable stochastic variability of the physical system) can even accumulate. Since all these uncertainties translate inevitably into errors in FIREFLY outputs, and since aleatoric errors cannot be diagnosed through parameter estimation, one possible strategy is to deal with the consequences of those uncertainties by considering the marker positions along the simulated fire front as control variables. More precisely, the control vector $\mathbf{x}$ includes the x- and y-coordinates of the N_{fr} markers and is therefore of size $n=2\,N_{fr}$. This approach is also relevant for observations made with significant error and/or cases in which the observations are incomplete, e.g., when only a fraction of the fireline perimeter is observed. The flowchart for state estimation is shown in Fig. 6.29.

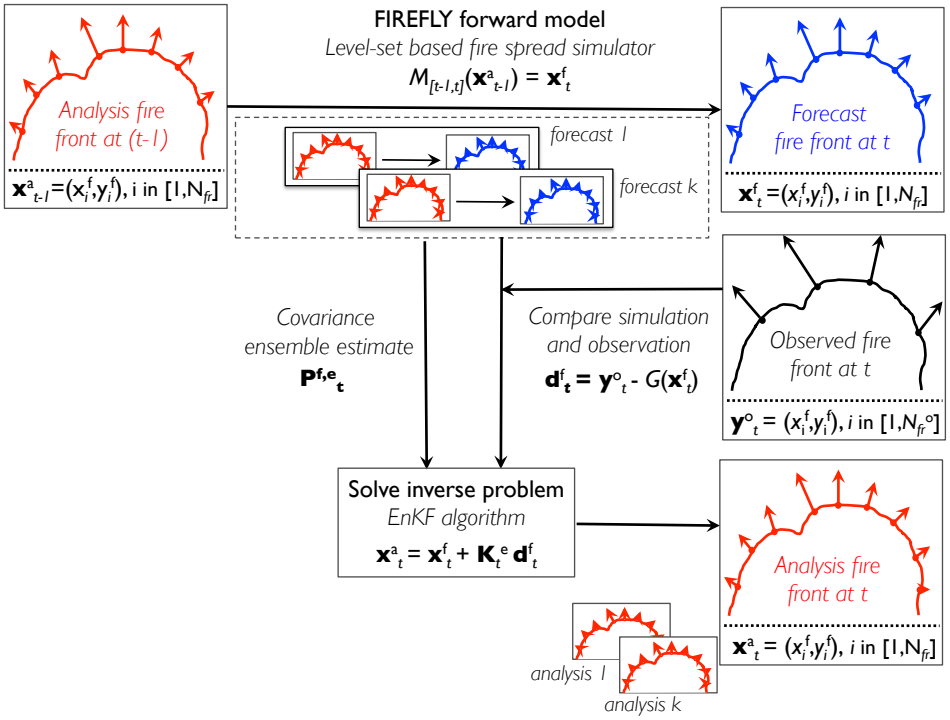


Figure 6.29: *Flowchart of the EnKF algorithm during the* $[t-1,t]$ *assimilation cycle for a state estimation approach.*

▷ **Forecast step.** In an ensemble-based approach, the first step is to generate the forecast ensemble for each control variable, while accounting for all sources of uncertainties. For front-tracking applications, a simple way to perform this forecast is to randomly perturb each front marker coordinate x and y (with a zero expectation value and a given STD corresponding to the forecast error STD σ^{f}_t over the assimilation cycle $[t-1,t]$). However, if each marker is perturbed separately (meaning that the error of one marker is uncorrelated to the errors of its neighbors), the resulting front does not exhibit coherent features. This ensemble generation is therefore conflicting with the physics of the fireline propagation. An alternative strategy is to generate an ensemble of simulated fire fronts by randomly perturbing the input parameters θ_t of the ROS model (e.g., wind magnitude u_w and direction angle α_w, fuel moisture content M_v) and by integrating FIREFLY using the PARASOL functionality of OPENPALM for each set of parameters as in the parameter estimation approach. This leads to an ensemble of N_e fire front positions at time t designated as $\mathbf{x}^{\mathrm{f}}_t$ and used to describe the error covariance matrices $\mathbf{P}^{\mathrm{f,e}}_t$ and $(\mathbf{G}_t\mathbf{P}^{\mathrm{f,e}}_t\mathbf{G}^{\mathrm{T}}_t)$. For a state estimation approach, the observation operator $\mathcal{G}_t$ is reduced to the selection operator $\mathcal{H}_t$, which is straightforward to compute (a selection of lines and columns of $\mathbf{P}^{\mathrm{f,e}}_t$ is sufficient to estimate the matrix $\mathbf{G}_t\mathbf{P}^{\mathrm{f,e}}_t\mathbf{G}^{\mathrm{T}}_t$). This selection operator $\mathcal{H}_t$ is also applied to obtain the model counterparts of the observations designated as $\mathbf{y}^{\mathrm{f}}_t = [\mathbf{y}^{\mathrm{f},(1)}_t, \mathbf{y}^{\mathrm{f},(2)}_t, \cdots, \mathbf{y}^{\mathrm{f},(N_e)}_t]$.

▷ **Update step.** Each ensemble member is updated, either based on weights for particle filters or based on the EnKF formulation in Eq. (5.89), Chapter 5. This EnKF update provides corrected positions $\mathbf{x}_t^{\mathrm{a}}$ for the N_{fr} simulated markers along the fireline at time t, but there is no feedback on the ROS model parameters used to generate variability in the ensemble of forecasts.

▷ **Insight into the forecast error covariance matrix.** The theoretical and numerical structure of the forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f,e}}$ is highlighted here for a spatially-uniform test with constant ROS but uncertain fire ignition location (x_{ign}, y_{ign}).[10] Since the ROS is isotropic, the simulated fire front keeps its initial circular shape. This implies that the location of the N_{fr} markers along the forecast simulated fronts can be parameterized as a function of this ignition location as follows:

$$\forall k = 1, \cdots, N_e, \quad \mathbf{x}^{\mathrm{f},(k)} = \begin{pmatrix} x_{ign}^{(k)} + r_t \cos(\alpha_1) \\ x_{ign}^{(k)} + r_t \cos(\alpha_2) \\ \vdots \\ x_{ign}^{(k)} + r_t \cos(\alpha_{N_{fr}}) \\ y_{ign}^{(k)} + r_t \sin(\alpha_1) \\ y_{ign}^{(k)} + r_t \sin(\alpha_2) \\ \vdots \\ y_{ign}^{(k)} + r_t \sin(\alpha_{N_{fr}}) \end{pmatrix}, \tag{6.19}$$

with r_t the radius of the circular fire front at time t (identical for all the ensemble members N_e) and $\alpha_i, i = 1, \cdots, N_{fr}$, the direction angle with respect to the center of the initial fire front (assuming that markers with the same index on the simulated fire fronts are taken at the same location on the circle for each member). Thus, the difference between each simulated fire front and the mean front sums up to a difference between the perturbed ignition locations $(x_{ign}^{(k)}, y_{ign}^{(k)})$ and the mean location over all the members $(\overline{x_{ign}},\ \overline{y_{ign}})$ for k varying between 1 and N_e:

$$\begin{aligned} \forall i = 1, \cdots, N_{fr}, \quad & x_i^{\mathrm{f},(k)} - \overline{x_i^{\mathrm{f}}} &&= x_{ign}^{(k)} - \overline{x_{ign}}, \\ \forall i = (N_{fr}+1), \cdots, 2N_{fr}, \quad & y_i^{\mathrm{f},(k)} - \overline{y_i^{\mathrm{f}}} &&= y_{ign}^{(k)} - \overline{y_{ign}}, \end{aligned} \tag{6.20}$$

with $(x_i^{\mathrm{f},(k)}, y_i^{\mathrm{f},(k)})$ the location of the marker i associated with the ensemble member k, with $(\overline{x_i^{\mathrm{f}}}, \overline{y_i^{\mathrm{f}}})$ the mean location of the marker i over the ensemble, and with

$$(\overline{x_{ign}},\ \overline{y_{ign}}) = \left(\frac{1}{N_e} \sum_{k=1}^{N_e} x_{ign}^{(k)},\ \frac{1}{N_e} \sum_{k=1}^{N_e} y_{ign}^{(k)} \right).$$

[10] The role of error covariances in the KF update step applied to wildfire spread is explained further in the manuscript, see Chapter 7.

Theoretically, the stochastically-estimated forecast error covariance matrix $\mathbf{P}_t^{\mathrm{f,e}}$ is constant in each block (of dimension $N_{fr} \times N_{fr}$) and can be written as:

$$\mathbf{P}_t^{\mathrm{f,e}} = \frac{1}{N_e - 1} \begin{pmatrix} \sum_{k=1}^{N_e} \left(x_{ign}^{(k)} - \overline{x_{ign}} \right)^2 & \sum_{k=1}^{N_e} \left(x_{ign}^{(k)} - \overline{x_{ign}} \right) \left(y_{ign}^{(k)} - \overline{y_{ign}} \right) \\ \sum_{k=1}^{N_e} \left(x_{ign}^{(k)} - \overline{x_{ign}} \right) \left(y_{ign}^{(k)} - \overline{y_{ign}} \right) & \sum_{k=1}^{N_e} \left(y_{ign}^{(k)} - \overline{y_{ign}} \right)^2 \end{pmatrix},$$

where the first diagonal block represents the error covariances of the x-coordinates (within this block, each diagonal element represent the error variance of one marker x-coordinate), where the second diagonal block represents the error covariances of the y-coordinates (within this block, each diagonal element represent the error variance of one marker y-coordinate), and where the cross-diagonal blocks represent the multi-variate error crossed-covariances between the x- and y-coordinates (see Section 5.2.4, Chapter 5, for details on the structure of error covariance matrices). We can define the associated error correlation matrix $\mathbf{C}_t^{\mathrm{f,e}}$ composed of 4 uniform blocks, i.e.,

$$\mathbf{C}_t^{\mathrm{f,e}} = \begin{pmatrix} \mathbf{I}_{N_{fr} \times N_{fr}} & \mathbf{0}_{N_{fr} \times N_{fr}} \\ \mathbf{0}_{N_{fr} \times N_{fr}} & \mathbf{I}_{N_{fr} \times N_{fr}} \end{pmatrix}. \tag{6.21}$$

From this correlation matrix $\mathbf{C}_t^{\mathrm{f,e}}$, one-dimensional correlation functions can be extracted (i.e., a column of the matrix), indicating how the error on one coordinate of a particular marker is correlated with the error on the x- or y-coordinate of any marker along the fireline. This spatial impact of the error correlations is illustrated in Fig. 6.30 and explained below.

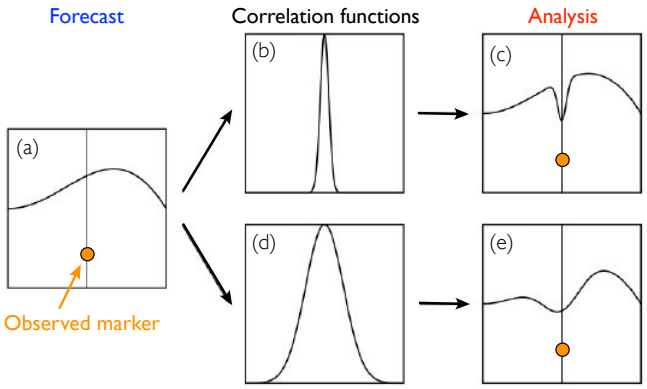


Figure 6.30: *Schematic of the behavior of the analysis with respect to the forecast error correlation functions and to the observation location (orange dot).*

- The case (a)-(b)-(c) shows that when the error correlation function is narrow (meaning that the error correlation of neighboring markers quickly decreases to 0), the analysis correction is reduced to a restricted area in the vicinity of the observed marker;

- In contrast, the case (a)-(d)-(e) shows that when the error correlation function is wider on both sides of the observed marker, the analysis correction affects a larger area along the simulated fireline.

Figure 6.31(b) presents the forecast error correlation matrix $\mathbf{C}_t^{\mathrm{f,e}}$ that is obtained in practice in comparison to the analytical solution in Fig. 6.31(a).

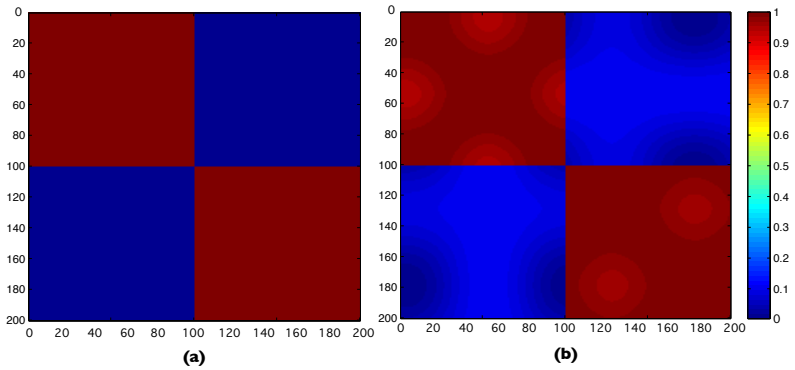


Figure 6.31: *Forecast error correlation matrix $\mathbf{C}_t^{\mathrm{f,e}}$ for the isotropic ROS case, in which (97 m, 103 m) is the ensemble mean location of fire ignition and 10 m is the error STD. (a) Analytical solution. (b) Numerical representation obtained with EnKF-*FIREFLY*. Credit: Emery et al. (2013).*

Note that the error correlation value obtained in each block is slightly different from the analytical solution in Eq. (6.21). Due to numerical inaccuracies in the isocontour algorithm (that allows to establish a correspondence between the markers along the fireline across the ensemble), there is a slight shift in the direction angle of the markers for each member and therefore, the angles $\alpha_i, i = 1, \cdots, N_{fr}$, are not identical from one member to another. This results in the slight oscillating variations on each block of $\mathbf{C}_t^{\mathrm{f,e}}$.

▷ **Assimilation cycling: Reconstruction of the initial condition.** To apply successive assimilations, the analysis fire fronts must be propagated to future observation times. However, the FIREFLY simulator evolves the progress variable c, the location of the fire front being diagnosed a posteriori. To restart FIREFLY, the two-dimensional field of the progress variable $c(x, y, t)^{(k)}$ associated with each analysis member $\mathbf{x}_t^{\mathrm{a},(k)}$, $k = 1, \cdots, N_e$, is therefore reconstructed and taken as the initial condition for the next assimilation cycle $[t, t + 1]$. This reconstruction is performed through a binarization of the two-dimensional field, i.e., $c(x, y, t)^{(k)} = 0$ in the unburnt vegetation and $c(x, y, t)^{(k)} = 1$ in the burnt vegetation. Thus, the analysis fire front $\mathbf{x}_t^{\mathrm{a},(k)}$ is the region where $c(x, y, t)$ takes values between 0 and 1. Further discussions on this reconstruction algorithm are provided in Emery et al. (2013). For demonstration purposes, a schematic of the state estimation approach is presented in Fig. 6.32.

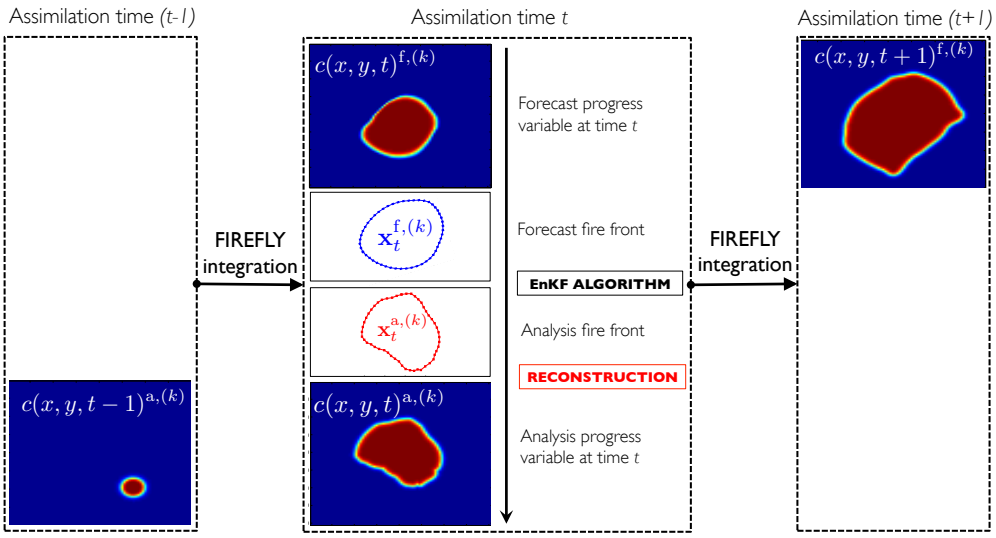


Figure 6.32: *Schematic of the sequential state estimation approach for each ensemble member k, with a focus on the reconstruction of the two-dimensional progress variable field c posterior to data assimilation over the time window $[t-1,t]$ and prior to forecast over the time window $[t,t+1]$.*

Note that this work assumes that observation errors are uncorrelated over the assimilation cycle $[t-1,t]$, i.e. the observation error covariance matrix $\mathbf{R}_t$ is treated as a diagonal matrix, in which each diagonal term is the error variance $(\sigma_t^{\mathrm{o}})^2$ associated with the x- or y-coordinate of each marker along the observed fireline $\mathbf{y}_t^{\mathrm{o}}$. This assumption can be questionable for spaceborne observations. Recent works (e.g., Brankart et al., 2009; Gorin and Tsyrulnikov, 2011) have found significant temporal and spatial correlations in the resulting observation dataset. In particular, they showed that adequately parameterizing these correlations could directly improve the quality of the assimilation updates and the accuracy of the associated error estimates. However, these issues related to observation error statistics modeling is out of the scope of this work.

6.5.5 Performance metrics of the data assimilation system

↪ Definition

For both parameter and state estimation approaches, the statistical moments of the innovation vector $\mathbf{d}_t^{\mathrm{f}} = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t^{\mathrm{f}}$ and residual vector $\mathbf{d}_t^{\mathrm{a}} = \mathbf{y}_t^{\mathrm{o}} - \mathbf{y}_t^{\mathrm{a}}$ (e.g., mean and STD) provide a convenient measure of the deviations of model predictions from observations and of the improved accuracy using data assimilation. The following terminology is used: FMO (standing for *Forecast Minus Observation*) is the distance between observed and forecast simulated fronts calculated over the assimilation time window (quantifying the error without data assimilation), while AMO (standing for *Analysis Minus Observation*) is the distance between observed and analysis simulated fronts calculated over the same assimilation time window (indicating

the error remaining with data assimilation). In the specific context of synthetic OSSE experiments (see Section 6.3), in which the true reference simulated front is known, similar metrics can be used with respect to the true simulation, instead of the observations. This results in FMT (*Forecast Minus True*) and AMT (*Analysis Minus True*) diagnostics. These distance metrics can be characterized in terms of mean and STD, denoted by the pair of parameters $(\overline{d},\ \overline{\sigma})$ and defined as follows for ensemble-based forecast/analysis estimates.

- ▷ $\overline{d}$ [m] represents the mean value (over all the N_{fr} markers along the simulated front) of the mean distance between reference and simulated fronts over the N_e ensemble members such that:

$$\overline{d_{FMT}} = \frac{1}{N_{fr}} \sum_{i=1}^{N_{fr}} \overline{d_{FMT,i}}, \tag{6.22}$$

with $\forall i = 1, \cdots, N_{fr}$,

$$\overline{d_{FMT,i}} = \frac{1}{N_e} \sum_{k=1}^{N_e} \sqrt{(x_i^{\mathrm{f},(k)} - x_i^{\mathrm{t}})^2 + (y_i^{\mathrm{f},(k)} - y_i^{\mathrm{t},(k)})^2}.$$

- ▷ $\overline{\sigma}$ [m] represents the mean value (over all the N_{fr} markers along the simulated front) of the STD of the distance between reference and simulated fronts over the N_e ensemble members such that:

$$\overline{\sigma_{FMT}} = \frac{1}{N_{fr}} \sum_{i=1}^{N_{fr}} \sigma_{FMT,i}, \tag{6.23}$$

with $\forall i = 1, \cdots, N_{fr}$,

$$\sigma_{FMT,i} = \sqrt{\frac{1}{N_e - 1} \sum_{k=1}^{N_e} \left(\sqrt{(x_i^{\mathrm{f},(k)} - x_i^{\mathrm{t}})^2 + (y_i^{\mathrm{f},(k)} - y_i^{\mathrm{t}})^2} - \overline{d_{FMT,j}} \right)^2}.$$

The definition of the AMT performance metrics is similar to FMT, with the difference that it considers $(x_\bullet^{\mathrm{a},(\bullet)}, y_\bullet^{\mathrm{a},(\bullet)})$ instead of $(x_\bullet^{\mathrm{f},(\bullet)}, y_\bullet^{\mathrm{f},(\bullet)})$. Furthermore, when the reference front is the observation (instead of the true simulation), the difference in the performance metrics is that the mean value and STD are computed over the N_{fr}^{o} observed markers instead of the N_{fr} simulated markers.

↪ Optimality conditions

Based on this performance metrics, the data assimilation system produces a good-quality analysis when:

$$\overline{d_{AMO}} < \overline{d_{FMO}}, \tag{6.24}$$

meaning that, on average, the markers from the analysis fire front are closer to the observations than the points along the forecast fronts. However, the main a posteriori diagnostics of a data assimilation system consists in verifying that the error variance of the control variables (i.e., the trace of the error covariance matrix) is reduced. Thus, the analysis $\mathbf{x}_t^{\mathrm{a}}$ must satisfy the following condition to guarantee its optimality:

$$\overline{\sigma_{AMO}} < \overline{\sigma_{FMO}}. \tag{6.25}$$

In the context of OSSE experiments, $\overline{d_{AMT}} < \overline{d_{FMT}}$ and $\overline{\sigma_{AMT}} < \overline{\sigma_{FMT}}$ are additional diagnostics.

Conclusion

The objective in Part II of this thesis is to develop a prototype data-driven wildfire simulator capable of forecasting wildfire spread dynamics at future lead-times. The prototype simulator features the following main components: a regional-scale wildfire spread model FIREFLY that treats wildfires as propagating fronts and uses a description of the local wildfire spread-rate as a function of environmental conditions based on a modified Rothermel's model (i.e., adapted to two-dimensional configurations with wind-aided fire propagation); a series of observations of the fire front location; and an ensemble-based data assimilation algorithm (e.g., standard and reduced-cost EnKF, SIR/ASIR particle filters). This prototype assumes that observations of the fire front position are available at frequent times but (possibly) provide an inaccurate and incomplete description of the fire front. To describe accurate error statistics, ensemble members are generated through variations in estimates of the fire ignition location and/or variations in the ROS model parameters. Furthermore, the data assimilation algorithm features either a parameter estimation approach (in which the control variables are the ROS input parameters), or a state estimation approach (in which the control variables are the two-dimensional coordinates of the discretized fire front). These two approaches are implemented based on the OPENPALM dynamic coupler combined with the PARASOL functionality for generating ensembles of FIREFLY fire front trajectories. The different algorithms associated with parameter estimation and state estimation approaches have been introduced in this chapter with an emphasis on their benefits for improving the accuracy of FIREFLY wildfire spread simulations.

Chapter 7

Evaluation of the data-driven wildfire spread simulator

The prototype data-driven wildfire simulator presented in this thesis features the following main components: a level-set-based fire propagation solver that adopts a regional-scale viewpoint, treats wildfires as propagating fronts, and uses a description of the local rate of spread (ROS) of the fire as a function of vegetation properties and wind conditions based on Rothermel's model; a series of observations of the fire front position; and a data assimilation algorithm featuring a parameter estimation approach (in which the control variables are the ROS model parameters) or a state estimation approach (in which the control variables are the two-dimensional coordinates of the discretized fire front). The prototype simulator is evaluated in a series of verification tests using synthetically-generated observations; these tests include representative cases of field-scale fires with spatially-varying vegetation properties and temporally-varying wind conditions, in which the ROS takes values on the order of 1 m/s. The prototype simulator is also evaluated in a validation test corresponding to a controlled grassland burning experiment.

The objective of these verification and validation tests is to demonstrate that data assimilation provides a suitable response to the challenge of accurately forecasting wildfire behavior at a computational cost that is consistent with operational framework. This demonstration features an original technical focus based on the following three steps:

▷ *a discussion on the Kalman-based analysis and forecast behavior for wildfire spread (e.g., treatment of model non-linearity, temporal variability of the errors, existence of multiplicity of solutions);*

▷ *a comparative study of different data assimilation algorithms for parameter estimation (e.g., extended and ensemble Kalman filter, particle filters, hybrid approaches based on uncertainty quantification techniques);*

▷ *a comparative study of the parameter estimation approach and the state estimation approach based on the ensemble Kalman filter, in terms of ensemble generation, analysis update performance and forecast quality at different lead-times.*

Outline

"Forecasting the behavior of existing wildfires will require a greater degree of sophistication than this model and our knowledge of fuels will permit at the present time. Variations in fuel and weather cause departures from predicted spread and intensity that pose risks unacceptable in fire suppression activities. A method for forecasting the behavior of a specific fire eventually will be developed; most likely, it will be patterned on a probability basis similar to that used for forecasting weather."

Rothermel (1972), US Department of Agriculture, Forest Service.

7.1 Preliminary insight into Kalman filtering

A preliminary version of the prototype data-driven wildfire spread simulator was proposed in Rochoux et al. (2013, PROCI) and Rochoux et al. (2013, INCA) to demonstrate the feasibility of data assimilation for wildfire spread, under simplified but relevant conditions. The initial prototype featured a data assimilation algorithm

based on an extended Kalman filter[1] (EKF) combined to a parameter estimation approach. This prototype is here useful to offer insight into the behavior of Kalman filtering. In particular,

- ▷ the sensitivity of the Kalman analysis is studied with respect to forecast uncertainty;
- ▷ the interpretation of the analysis with respect to the temporal variability of the errors on the ROS model parameters is provided and its impact on the Kalman optimal solution is highlighted for the controlled grassland burning experiment;
- ▷ the impact of the control space on the formulation of the estimation problem is highlighted through the evaluation of cost functions for different control parameters (the mapping between control space and observation space can exhibit non-linearities).

One particularity of the EKF-based data assimilation algorithm is to assume a linear observation operator $\mathcal{G}_t(\mathbf{x}_t)$ (i.e., a linear mapping between control space and observation space). The Jacobian matrix $\mathbf{G}_t$ related to the observation operator (i.e., the tangent-linear operator) is calculated by simple differentiation after perturbing each element of the control vector $\mathbf{x}_t$ and evaluating the corresponding change in the observation operator $\mathcal{G}_t(\mathbf{x}_t)$. In the context of parameter estimation, the observation operator includes the temporal integration of the regional-scale wildfire spread model Firefly and is therefore non-linear with respect to the ROS model parameters that are included in the control vector $\mathbf{x}_t$. These non-linearities are partially accounted for by using an iterative procedure at each assimilation cycle referred to as *outer-loops*. These loops consist in successive applications of the Kalman update Eq. (5.78), Chapter 5, in which the gain matrix $\mathbf{K}_t$ is updated at each iteration k through the calculation of the tangent-linear operator $\mathbf{G}_t$ around a reference vector. This reference vector is taken as the forecast vector for the first iteration ($k = 1$) and as the analysis vector taken from iteration $(k-1)$ for $k > 1$. The outer-loop procedure explained in detail in Section 5.6, Chapter 5, is repeated until convergence towards the optimal analysis $\mathbf{x}_t^{\mathrm{a}}$ is obtained at the assimilation time t; the number of iterations is typically between 1 and 10.

7.1.1 Sensitivity of the analysis to forecast uncertainty

We present first results from OSSE verification tests, in which synthetic observations are generated using specified values of the control parameters (i.e., a case in which the true vector $\mathbf{x}^{\mathrm{t}}$ exists and is known). The ROS is expressed as $\Gamma = P\,\delta_v(x, y)$ (see Eq. 6.6, Chapter 6), with δ_v the biomass fuel layer depth that is here a known randomly-distributed function of x and y over the $300\,\mathrm{m} \times 300\,\mathrm{m}$ two-dimensional domain as shown in Fig. 7.1(a). Ignition occurs at $(x, y) = (150\ \mathrm{m}, 0\ \mathrm{m})$, and the true fireline evolution is approximately semi-circular during the assimilation

[1] See Section 5.6.1, Chapter 5.

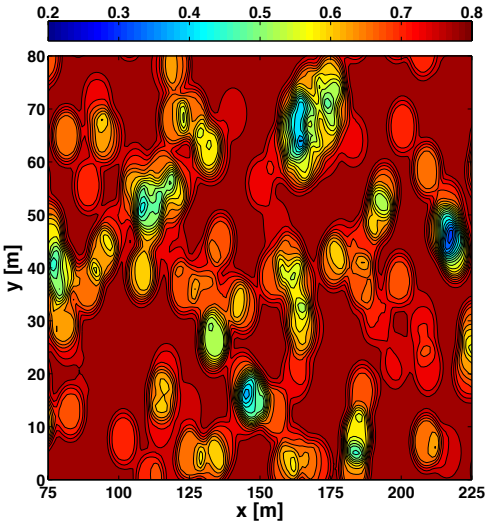
time window $[0; 800\,\text{s}]$. The control vector $\mathbf{x}$ is taken as the coefficient P such that $\mathbf{x} = P$ (the rank of the estimation problem is $n = 1$); its true value is $\mathbf{x}^{\text{t}} = P^{\text{t}} = 0.1\,\text{s}^{-1}$. Observations are made at 50 s intervals over the time window $[0; 800\,\text{s}]$, meaning that 16 observed fire fronts are available to estimate the control parameter P. Observation errors are assumed small (i.e., $\sigma^{\text{o}} = 1$ m) in order to evaluate the ability of the EKF to track the observed fire front location. Note that the time index is removed here for clarity purposes since only one assimilation cycle (including a forecast step and an analysis step) is considered.

The EKF-based data assimilation algorithm is applied to different cases corresponding to different initial values of the prior $\mathbf{x}^{\text{f}} = P^{\text{f}}$, ranging from 0.02 to $0.18\,\text{s}^{-1}$. This range of values for $\mathbf{x}^{\text{f}}$ represents different perturbations (from -80 % to $+80$ %) of the true value $\mathbf{x}^{\text{t}} = 0.1$ s^{-1}. The forecast error standard deviation (STD) σ^{f} is set according to these perturbations as $\sigma^{\text{f}} = \mathbf{x}^{\text{t}} - \mathbf{x}^{\text{f}}$, meaning that these OSSE tests study the sensitivity of the analysis $\mathbf{x}^{\text{a}}$ to forecast uncertainty. The perturbation added to the reference parameter value to compute the tangent-linear operator is taken as 5 % of the reference value at each outer-loop. Figure 7.1(b) demonstrates the efficiency of the EKF algorithm for a forecast perturbation taken as -80 % of the true value (i.e., $\mathbf{x}^{\text{f}} = 0.02$ s^{-1}) since the analysis (without and with outer-loop) perfectly coincides with the observed fire front. It is worth mentioning that $\mathbf{x}^{\text{f}}$ and $\mathbf{x}^{\text{a}}$ represent in EKF the mean value of the PDF related to the forecast and analysis control vector, respectively.

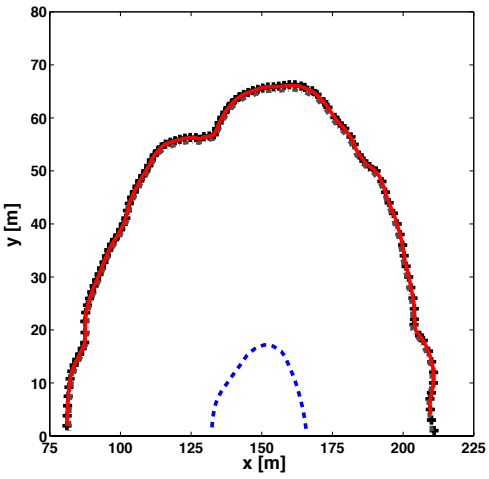
Figure 7.1(c) presents a global plot of the data assimilation experiment, showing different performance metrics as a function of the forecast value $\mathbf{x}^{\text{f}}$. $\overline{\sigma_{FMO}}$ represents the STD of the distance between observed and simulated fronts calculated over the assimilation time window $[0; 800\,\text{s}]$ and using $\mathbf{x} = \mathbf{x}^{\text{f}}$. Thus, $\overline{\sigma_{FMO}}$ provides a measure of the error that would be obtained without data assimilation. $\overline{\sigma_{AMO}}$ is the same metric but calculated using $\mathbf{x} = \mathbf{x}^{\text{a}}$, with or without outer-loop. For all cases, $\overline{\sigma_{AMO}}$ takes small values that are orders of magnitude lower than $\overline{\sigma_{FMO}}$. Thus, the EKF-based parameter estimation approach was successful both at significantly decreasing the distance between observed and simulated fire fronts and at retrieving the true value of the ROS model parameter P within less than 0.1 %. Typically, 2 or 3 outer-loops were sufficient to reach convergence and to achieve this performance.

7.1.2 Retrieval of mean conditions over the assimilation cycle

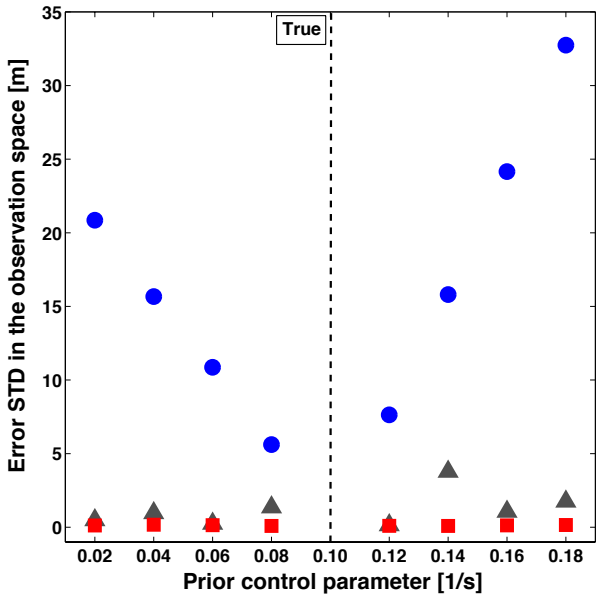
In a second series of OSSE tests, the fire configuration is reduced to a straight line propagation along the x-axis (ignited at $x = 150$ m and propagating towards decreasing x during 100 s) and with spatially-uniform ROS parameters that are representative of the short grass category in Rothermel's fuel database (see Appendix A). Thus, the fireline is characterized by a single value of the ROS. The objective of these tests is to offer insight into the meaning of the EKF-based analysis over each assimilation cycle in a basic configuration.



(a) Randomly-distributed vegetation δ_v [m].



(b) EKF example for $\mathbf{x}^{\mathrm{f}} = 0.02\ \mathrm{s}^{-1}$.



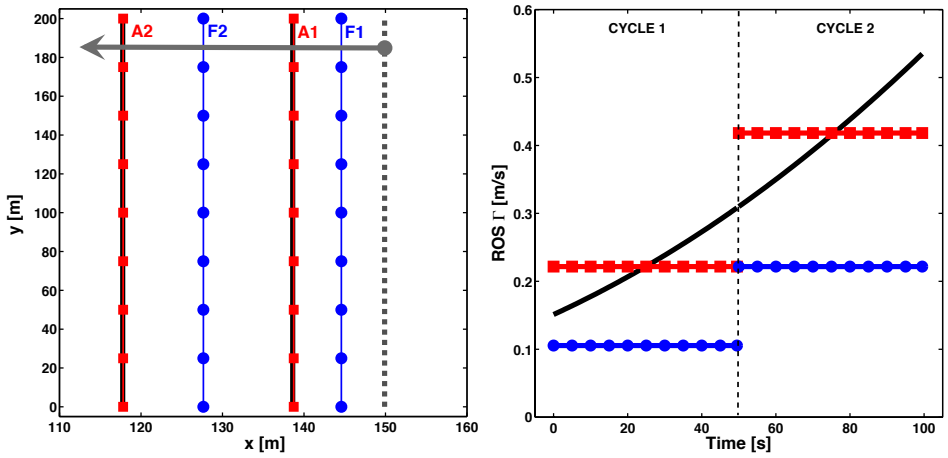
(c) EKF results for varying forecast uncertainty.

Figure 7.1: *OSSE verification tests using the EKF-based parameter estimation approach with a unique control parameter P [s^{-1}]. (a) Randomly-distributed vegetation layer depth $\delta_v(x, y)$. (b) Comparison between the forecast $\mathbf{x}^{\mathrm{f}}$ (blue dashed line), observations (black crosses), the analysis $\mathbf{x}^{\mathrm{a}}$ without outer-loop (gray dashed-dotted line) and with outer-loop (red solid line) at time $t = 800$ s for an initial parameter uncertainty equal to -80 % ($\mathbf{x}^{\mathrm{f}} = 0.02\ s^{-1}$). (c) STD between the simulated fire front and the observation over the assimilation cycle $[0; 800\,s]$ for different perturbations in the forecast P^{f} (-80 % to $+80$ % of the true value $\mathbf{x}^{\mathrm{t}} = 0.10\ s^{-1}$): $\overline{\sigma_{FMO}}$ (blue circles); $\overline{\sigma_{AMO}}$ without outer-loop (gray triangles) and $\overline{\sigma_{AMO}}$ with outer-loops (red squares).*

The behavior of the analysis $\mathbf{x}^{\mathrm{a}}$ is illustrated in Fig. 7.2 for a wind-aided fire propagation: the control vector $\mathbf{x}$ is taken as the wind magnitude u_w in the Rothermel's formulation ($\mathbf{x} = u_w$); the fire duration is divided into two assimilation cycles $[0; 50$ s] and $[50; 100$ s], with a single observed fire front per cycle at time $t_1 = 50$ s and $t_2 = 100$ s, respectively. The true wind magnitude varies linearly from 1 to 2 m/s over the whole fire duration $[0; 100$ s]. Observation errors are assumed small, implying that observations and the true fire front trajectory are equivalent.

(1) **Assimilation cycle 1.** The prior (forecast) value for the wind magnitude is taken as a 20 % perturbation of the true value at initial time $t_0 = 0$ s such that $\mathbf{x}_1^{\mathrm{f}} = 0.8$ m/s (the index 1 referring to $[0; 50$ s]). This forecast (F1) leads to a constant ROS equal to 0.1 m/s, while the true ROS varies between 0.15 and 0.30 m/s, see Fig. 7.2(b). The EKF algorithm with outer-loops is first performed at $t_1 = 50$ s; the resulting analysis $\mathbf{x}_1^{\mathrm{a}}$ (wind magnitude) is equal to 1.25 m/s and perfectly retrieves the location of the true fire front (A1) as shown in Fig. 7.2(a).

(2) **Assimilation cycle 2.** The analysis $\mathbf{x}_1^{\mathrm{a}}$ at $t_1 = 50$ s is then used as the forecast of the next assimilation cycle $[50; 100$ s] such that $\mathbf{x}_2^{\mathrm{f}} = 1.25$ m/s (the index 2 referring to $[50; 100$ s]). This forecast (F2) gives a ROS nearby 0.22 m/s and significantly underestimates the true ROS that varies between 0.30 and 0.55 m/s up to time $t_2 = 100$ s, see Fig. 7.2(b). The EKF algorithm applied at time t_2 leads to the analysis $\mathbf{x}_2^{\mathrm{a}} = 1.76$ m/s (A2), which also significantly decreases the error in the simulated fire front, see Fig. 7.2(a).



(a) Fire front locations.

(b) Time-evolution of ROS.

Figure 7.2: *EKF-based sequential assimilation with $\mathbf{x} = u_w$ [m/s] for a straight fireline propagation (from right to left); 2 assimilation cycles; time-varying true control parameter. (a) Comparison between the forecast (blue circles), the truth (black solid lines) and the analysis (red squares) at $t_1 = 50$ s and $t_2 = 100$ s. (b) Time-evolution of the ROS: the black solid line corresponds to the time-varying true ROS, blue circles (red squares) correspond to the forecast (analysis) ROS over the time window $[0; 100$ s].*

Figure 7.2(b) proves that these two values of the analysis, $\mathbf{x}_1^{\mathrm{a}}$ and $\mathbf{x}_2^{\mathrm{a}}$, are associated with a ROS that gives the mean behavior of the true fireline over each assimilation cycle. Stated differently, the EKF algorithm searches for the value of the control vector $\mathbf{x}_t$ that retrieves the mean behavior of the real system between two successive assimilation times, from time $(t-1)$ to time t. This property of KF and its extensions implies that the size of the assimilation window (for which a constant optimal parameter is searched) must be specified according to the temporal variability of the errors in the physical system (if the observation frequency is compatible).

7.1.3 Analysis of cost function evaluations

The presence of non-linearities in the observation operator is now studied for different control parameters. As explained in Section 5.6, Chapter 5, the iterative EKF algorithm (with outer-loops) is equivalent to the multi-incremental 3D-Var cost function given in Eq. (5.81). In the proposed EKF-based approach, the cost function associated with the control parameters is therefore computed at each outer-loop as posterior diagnostics of non-linearities in the observation operator $\mathcal{G}_t$ for the assimilation cycle $[t-1, t]$.

In the following tests, Firefly considers a fire propagation during 200 s with spatially-uniform vegetation (corresponding to the tall grass category in Rothermel's fuel database); ignition occurs at $(x, y) = (100 \text{ m}, 100 \text{ m})$ and in the absence of wind, the propagation of the fireline is circular over time; there is one observation at time $t = 200$ s.

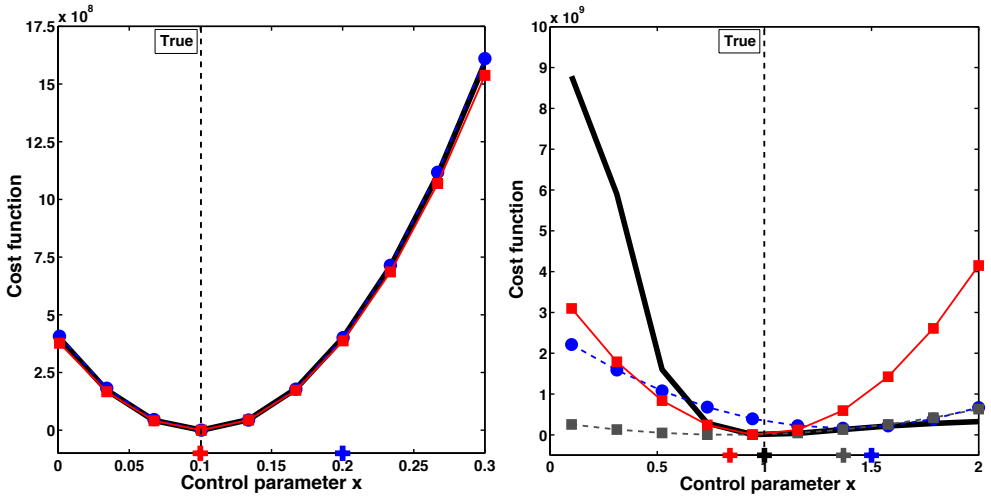
↪ Simple example

For illustration purposes, the ROS due to Rothermel is reformulated as $\Gamma = P\,(\delta_v)^{n_v}$ to introduce a non-linear dependency with respect to the biomass fuel layer depth δ_v through the exponent n_v. The objective is to highlight the effect of this dependency on data assimilation results and in particular, on the shape of the cost function to minimize.[2] Since no external flow ($u_w = 0$ m/s) is considered, the fire propagation is isotropic; the objective of the EKF is to estimate the proportionality coefficient P ($\mathbf{x} = P$) or the exponent n_v ($\mathbf{x} = n_v$) assuming negligible observation errors ($n = 1$). The true fireline trajectory is obtained for $P^{\mathrm{t}} = 0.10$ s^{-1} and $n_v^{\mathrm{t}} = 1.0$, resulting in a ROS equal to 0.10 m/s.

- ▷ **Estimation of the proportionality coefficient P.** The forecast is taken as $P^{\mathrm{f}} = 0.20$ s^{-1}. Since there is a linear relationship between the ROS and the coefficient P in the present isotropic case, the related cost function is quadratic and therefore, the EKF-based analysis retrieves the true value without outer-loop as shown in Fig. 7.3(a).

[2] The model exponent n_v is equal to 1 in the other data assimilation experiments presented in this thesis.

▷ **Estimation of the coefficient n_v.** The forecast is taken as $n_v^{\mathrm{f}} = 1.50$. In this case, the non-linear relationship between this exponent and the ROS induces a strong non-quadratic cost function with a relatively flat region for increasing values of the exponent n_v as shown in Fig. 7.3(b). Thus, the minimization of this cost function requires three successive outer-loops to converge towards the true value $n_v^{\mathrm{t}} = 1.0$, with $n_v^{\mathrm{a},(1)} = 1.369$, $n_v^{\mathrm{a},(2)} = 0.84$, and $n_v^{\mathrm{a},(3)} = 0.98$. These results highlight the crucial role of outer-loops in the determination of the optimal control parameter when the mapping between control space and observation space is non-linear.



(a) Proportionality coefficient P.

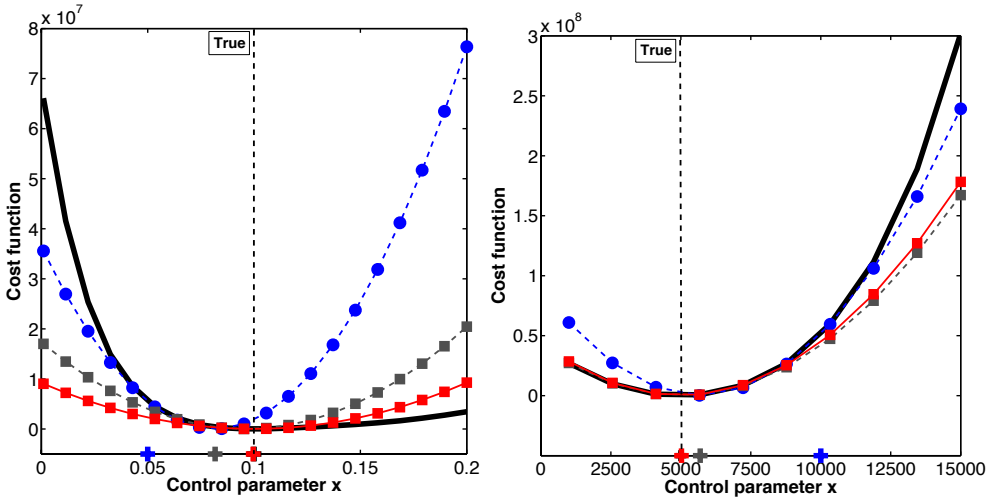
(b) ROS exponent n_v.

Figure 7.3: *Comparison between the true cost function (black solid line) of the EKF-based estimation problem and the quadratic cost functions obtained at each outer-loop for (a) the proportionality coefficient P [s^{-1}], and (b) the ROS model exponent n_v [-]. The blue circled-dashed line represents the cost function related to the first outer-loop, the gray squared-dashed line its counterpart at the second outer-loop and the red squared-solid line its counterpart at the third outer-loop. Crosses correspond to the reference control parameter (used as forecast) related to each outer-loop.*

↪ Non-linear response to environmental conditions

The same numerical configuration is used to study the effect of physical parameters on the cost function to minimize and thereby, on the EKF-based analysis. Since the fuel moisture content M_v and the fuel particle surface-area-to-volume ratio Σ_v have been identified as important sources of uncertainties as well as ROS sensitive parameters in wildfire spread modeling over a broad range of environmental conditions (see Fig. 6.24, Chapter 6, and further discussion in Appendix A), the EKF-based cost function is studied for $\mathbf{x} = M_v$ in Fig. 7.4(a) and $\mathbf{x} = \Sigma_v$ in Fig. 7.4(b), the size of the estimation problem is $n = 1$. These figures show that only 2 or 3 outer-loops are enough to make the EKF algorithm efficient at retrieving the true

value of the control parameter ($M_v^{\mathrm{t}} = 0.10$ % or $\Sigma_v^{\mathrm{t}} = 5,000$ 1/m), even though the forecast was specified for each experiment as a 100 % perturbation of the true value ($M_v^{\mathrm{f}} = 0.05$ % or $\Sigma_v^{\mathrm{f}} = 10,000$ 1/m). Note that the forecast error STD σ^{f} was specified accordingly with $\sigma^{\mathrm{f}} = \mathbf{x}^{\mathrm{t}} - \mathbf{x}^{\mathrm{f}}$. The analysis at the third outer-loop provides a consistent estimation with $M_v^{\mathrm{a},(3)} = 0.0996$ and $\Sigma_v^{\mathrm{a},(3)} = 5,003$ 1/m.



(a) Fuel moisture content M_v.

(b) Fuel particle surface/volume Σ_v.

Figure 7.4: *Cost functions for (a) the fuel moisture content M_v [−], and (b) the fuel particle surface-area-to-volume ratio Σ_v [1/m] (see caption of Fig. 7.3).*

It is worth mentioning that Fig. 7.4(a) highlights the non-linearity between the moisture content M_v and the ROS in the model due to Rothermel; the cost function was found to be non-quadratic, even for an isotropic fire propagation, while the cost function related to Σ_v is quadratic in this no-wind configuration. Thus, attention must be paid to this estimation problem when environmental conditions become heterogeneously-distributed and thereby, enhance model non-linearity.

↪ Equifinality issue in multi-parameter estimation

The size of the estimation problem is now $n = 2$ with the simultaneous estimation of the fuel moisture content M_v and the fuel surface-area-to-volume ratio Σ_v to retrieve the observed fire front at time $t = 200$ s for a no-wind or a wind-aided fire spread. The errors on these parameters are rightfully assumed to be independent. Thus, $\mathbf{x} = [M_v, \Sigma_v]$ with its true value $\mathbf{x}^{\mathrm{t}} = [0.10,\ 4,921$ 1/m$]$. The forecast is taken as a 20 % perturbation of $\mathbf{x}^{\mathrm{t}}$ such that $\mathbf{x}^{\mathrm{f}} = [0.08,\ 5,905$ 1/m$]$; these values induce an overestimation of the ROS compared to the true trajectory.

▷ **Isotropic case.** Table 7.1 compares the analysis obtained at each outer-loop with the forecast and the observations, in both parameter space and observation space, for the no-wind fire spread configuration. The analysis was found to reproduce the

observation after only 3 outer-loops since the mean distance between the observed and analysis fire fronts (AMO) is reduced to less than 1 % of the mean distance between the observed and forecast fire fronts (FMO). However, 20 outer-loops must be performed to retrieve the true value of the control parameters in this case, while the AMO statistics were converged only after 3 outer-loops. This indicates that for this no-wind simulation configuration, multiple sets of parameters can reproduce the same observed behavior of the fire spread without guarantee to obtain physical values for the control parameters. This problem of multiplicity of solutions is commonly referred to as *equifinality* (Beven and Freer, 2001). The presence of equifinality is confirmed by the parameter response surface shown in Fig. 7.5(a): this figure corresponds to the cost function evaluated for different sets of parameters, between 6 and 14 % for M_v, between 3,000 and 6,800 1/m for Σ_v. This cost function exhibits a plateau in the parameter response corresponding to the black zone in Fig. 7.5(a), which is difficult to handle in a minimization process.

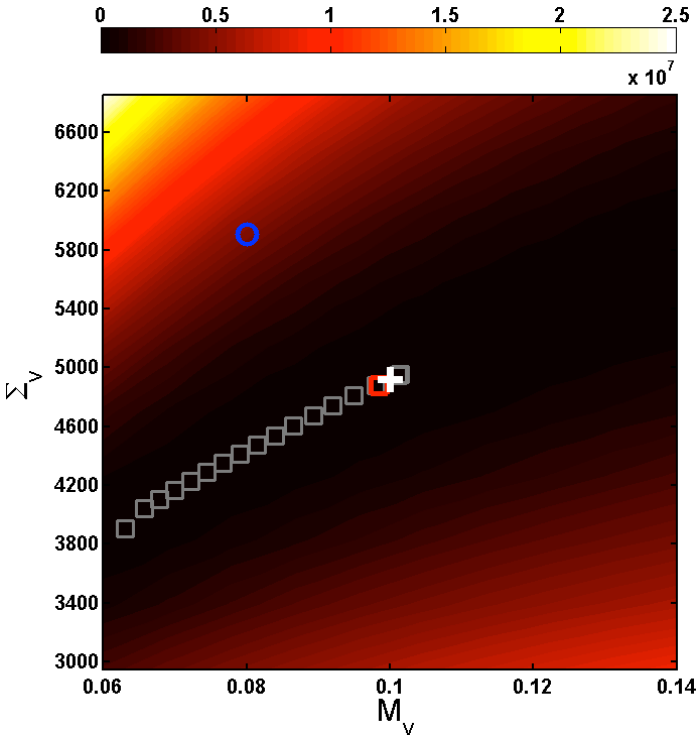
▷ **Wind-aided case.** The same data assimilation experiment is performed for a wind-aided fire spread (with a wind velocity $u_w = 0.5$ m/s blowing northward, $\alpha_w = 0°$), the topology of the cost function is significantly modified as shown in Fig. 7.5(b). This cost function is indeed characterized by a single minimum and there is no plateau in the parameter response surface (instead, the cost function exhibits a *bowl* shape) such that EKF retrieves the exact true parameters within 2 or 3 external loops (see Table 7.2). The wind velocity adds a physical constraint to the data assimilation algorithm and thereby, reduces the number of degrees of freedom for the analysis solution.

Table 7.1: *EKF-based experiment for isotropic propagation,* $\mathbf{x} = [M_v, \Sigma_v]$.

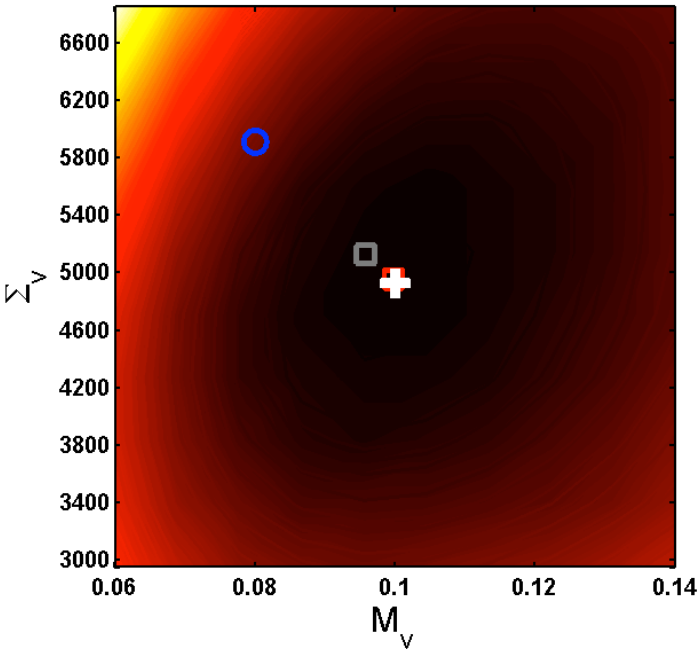
	$\mathbf{x}^{\mathrm{t}}$	$\mathbf{x}^{\mathrm{f}}$	$\mathbf{x}^{\mathrm{a},(1)}$	$\mathbf{x}^{\mathrm{a},(2)}$	$\mathbf{x}^{\mathrm{a},(3)}$	$\mathbf{x}^{\mathrm{a},(20)}$
M_v [-]	0.10	0.08	0.063	0.066	0.068	0.10
Σ_v [1/m]	4921	5905	3908	4041	4107	4959
FMO/AMO (mean) [m]	-	1.34	0.07	0.01	0.007	0.007
FMO/AMO (STD) [m]	-	2.00	0.11	0.02	0.01	0.01

Table 7.2: *EKF-based experiment for wind-aided propagation,* $\mathbf{x} = [M_v, \Sigma_v]$.

	$\mathbf{x}^{\mathrm{t}}$	$\mathbf{x}^{\mathrm{f}}$	$\mathbf{x}^{\mathrm{a},(1)}$	$\mathbf{x}^{\mathrm{a},(2)}$	$\mathbf{x}^{\mathrm{a},(3)}$
M_v [-]	0.10	0.08	0.0960	0.0998	0.10
Σ_v [1/m]	4921	5905	5120	4951	4922
FMO/AMO (mean) [m]	-	1.21	0.22	0.02	0.02
FMO/AMO (STD) [m]	-	1.86	0.33	0.03	0.03



(a) Isotropic fire spread with no external flow.



(b) Wind-aided fire spread.

Figure 7.5: *Cost function associated with EKF with respect to the control vector* $\mathbf{x} = [M_v, \Sigma_v]$. *The blue circle represents the forecast* $\mathbf{x}^{\mathrm{f}} = [0.08,\ 5,905\ 1/m]$; *gray squares correspond to the reference control vector at each outer-loop; the red square corresponds to the analysis* $\mathbf{x}^{\mathrm{a}}$; *and the white cross corresponds to the true value* $\mathbf{x}^{\mathrm{t}}$.

These EKF results for a two-parameter estimation highlight two key aspects underlying Kalman filtering:

(1) The KF-based analysis must be viewed as effective values that incorporate the effects of a number of modeling choices and that are related to the description of the sources of uncertainties. While searching for the combination of parameter values that best reproduces the observed fire behavior (when observation errors are assumed small), it does not guarantee to retrieve physical values for the parameters due to the equifinality issue as well as to knowledge gaps in fire spread uncertainties. More realistic values of the model parameters may be expected if the size n of the control vector is increased, thereby including more physical parameters and more physical constraints in the estimation problem.

(2) The equifinality issue is found to be drastically reduced when the fire propagation is anisotropic. Thus, Kalman filtering can achieve successful parameter estimation for realistic cases of wildfire spread, in which the wind conditions may vary and the vegetation may exhibit spatially-varying properties (if it is affordable to deal with a large size n of the control vector).

7.1.4 Application to the controlled grassland burning

Results from a validation test in which observations are taken from the controlled grassland burning experiment introduced in Section 6.3, Chapter 6, are now presented. This corresponds to a real reduced-scale fire occurring under moderate wind conditions, in which the ROS takes values on the order of 1 cm/s. In contrast to OSSE experiments, the true control vector $\mathbf{x}^{\mathrm{t}}$ is not known and may not exist if the forward model is not sufficiently representative.

Observations are described as full fire contours represented by $N_{fr}^{\mathrm{o}} = 20$ markers, at 28 s intervals from $t_0 = 50$ s (initial condition) to $t_1 = 78$ s (assimilation time) and $t_2 = 106$ s (forecast time). The EKF-based data-driven simulation is studied with respect to the size n of the control vector $\mathbf{x}$ for one assimilation cycle $[t_0; t_1] = [50; 78$ s$]$; $[t_1; t_2] = [78; 106$ s$]$ corresponds to the forecast time window. First, the EKF algorithm searches for the effective values of the fuel moisture content M_v and the fuel particle surface-area-to-volume ratio Σ_v (the rank of the control vector $\mathbf{x}$ is $n = 2$, i.e., $\mathbf{x} = [M_v, \Sigma_v]$), while featuring the presence of a (known) uniform wind velocity ($u_w = 1.0$ m/s, $\alpha_w = 307°$). Second, the EKF algorithm considers wind conditions as uncertain and therefore, the control vector is extended to the wind magnitude u_w and to the wind direction α_w (the rank of the control vector $\mathbf{x}$ is $n = 4$, i.e., $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$). In all tests, observation errors are assumed to be small according to the spatial resolution of the camera (i.e., $\sigma^{\mathrm{o}} = 5$ cm). The ROS parameters are assumed spatially-uniform and are optimized to reduce discrepancies between simulated and observed fire front locations. Also in all tests, the location of the origin of the fire is assumed to be known and is taken as the observed fire front at time $t_0 = 50$ s.

↪ Constant and uniform wind conditions

The fire spread model is first used over the assimilation window $[50; 78\,\text{s}]$. The ROS is treated using Rothermel's model, with $M_v^{\text{f}} = 21.7\,\%$ and $\Sigma_v^{\text{f}} = 4,921\,1/\text{m}$. These initial prior values correspond to the best educated guess available using in-situ measurements and Rothermel's fuel database; the STD of the forecast error is arbitrarily specified as a 30 % perturbation of the prior (*arbitrarily* due to knowledge gaps and lack of uncertainty quantification in the short grass properties). With these values, the model predicts a maximum ROS in the wind direction of $0.01\,\text{m/s}$ over the time period $[50; 78\ \text{s}]$. Figure 7.6(a) shows that the free run simulation significantly underestimates the ROS, while the EKF algorithm (with outer-loops) is successful at significantly decreasing the distance between observed and simulated fronts. The STD of the distance between observed and analysis fire fronts $\overline{\sigma_{AMO}}$ is reduced by a factor of 2 compared to the STD of the distance between observed and forecast fire fronts $\overline{\sigma_{FMO}}$; this improved performance is obtained with the new control parameters $\mathbf{x}^{\text{a}} = [11\ \%,\ 13,193\ 1/\text{m}]$. With these new values, the FIREFLY model predicts a maximum ROS in the wind direction of $0.04\,\text{m/s}$ over the time period $[50; 78\,\text{s}]$. As expected from the EKF-based algorithm, the STD of this corrected control vector is significantly reduced to only $3\ \%$ of the analysis values.

Figure 7.6(b) adopts a slightly different perspective and presents a comparison of different forecasts of the fire front location at time $t_2 = 106$ s using the fire spread model FIREFLY, with or without data assimilation. When using data assimilation, FIREFLY is integrated from $t_0 = 50$ s for the new control parameters $\mathbf{x}^{\text{a}} = [11\ \%,\ 13,193\ 1/\text{m}]$. Consistently with previous results, the forecast run with data assimilation at $t_1 = 78$ s significantly reduces the distance between observed and simulated fronts; $\overline{\sigma_{AMO}}$ is reduced by a factor of 2 compared to $\overline{\sigma_{FMO}}$. This result illustrates the improved prediction capability of the fire model that is achieved when uncertain environmental conditions are calibrated against past observations.

↪ Extension of the control to wind conditions

Even though the 2-parameter estimation was found to properly retrieve the mean behavior of the fire front, it was not able to track the temporal variations of the front curvature at small scales. The extension of the control vector to wind conditions is able to overcome some of these limitations in the wind direction, since the additional estimation of the wind magnitude and direction avoids to overestimate the ROS in the wind direction and retrieves a more accurate fire front location. Figures 7.6(a)-(b) show indeed that the mean analysis estimate of the fire front obtained with a 4-parameter estimation better tracks the location of the observed fire front, both at the assimilation and forecast times.

The analysis fire front trajectory is obtained for the following set of parameters, $\mathbf{x}^{\text{a}} = [7.1\ \%,\ 7,185\ 1/\text{m},\ 0.38\ \text{m/s},\ 300°]$. The new values for the fuel moisture content M_v and the fuel particle surface-area-to-volume ratio Σ_v are largely

modified compared to the values obtained with the 2-parameter estimation. This highlights that the EKF-based parameter estimation is a dynamic-learning process, in which the value of the parameters is case-dependent. Furthermore, while the analysis values of M_v and Σ_v obtained in the first test [11 %, 13,193 1/m] were realistic, they must be viewed as effective values that incorporate the effects of a number of modeling choices. With the extension of the control vector to wind conditions, more realistic values of the model parameters were obtained. These results indicate the need for a dynamic model that represents the temporal variability of the fuel moisture content M_v over the fire duration to improve prior information (the prior value was far from being predictive). They also indicate the persistent uncertainty in the modeling of the fuel particle surface-area-to-volume ratio Σ_v, which is difficult to measure and therefore crucial to control in parameter estimation due to its large impact on the wind-aided ROS.

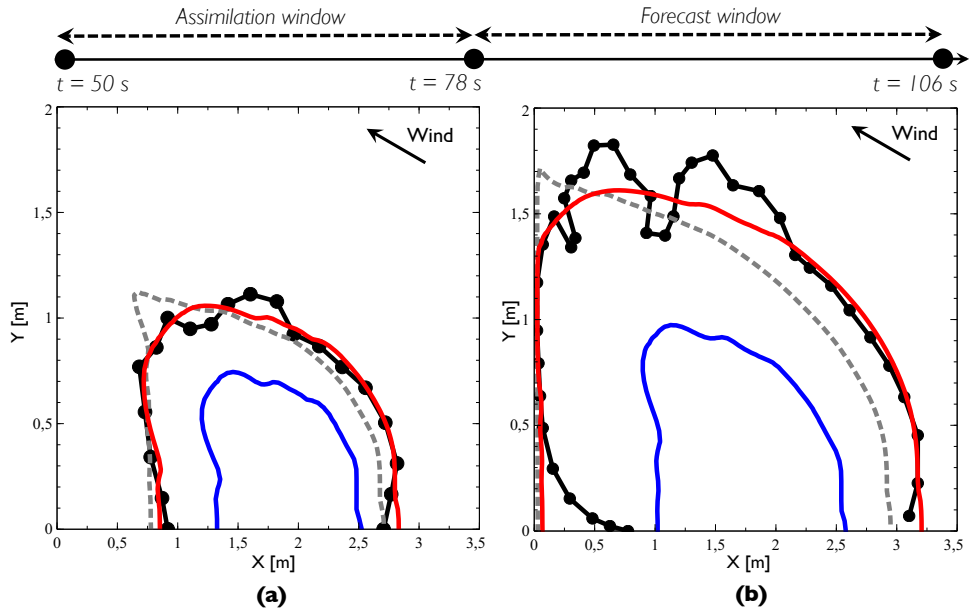


Figure 7.6: *Comparison between simulated and observed fire front positions at $t_1 = 78$ s and $t_2 = 106$ s. Blue solid lines: free run predictions (without data assimilation). Gray dashed lines: analysis predictions for 2-parameter estimation $\mathbf{x} = [M_v, \Sigma_v]$. Red solid lines: analysis predictions for 4-parameter estimation $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$. Black circled-solid lines: observations. (a) Assimilation window $[50; 78$ s$]$. (b) Forecast window $[78; 106$ s$]$.*

While these EKF-based studies, partly presented in Rochoux et al. (2013, PROCI) and Rochoux et al. (2013, INCA), produced encouraging results and confirmed the value of a data assimilation strategy for improved wildfire spread predictions, some of the design choices made during the initial development of the prototype simulator were proposed on a preliminary and temporary basis with the understanding that they would have to be re-visited in subsequent work. The choice of the EKF algorithm was considered questionable because it assumes a linear relationship between

control space and observation space (i.e., a linear relationship between changes in the parameters of the Rothermel-based ROS model and the resulting changes in the fire front positions). This linear assumption is believed to be of limited value in general wildfire problems, where the wind direction and magnitude vary and the vegetation properties are potentially strongly heterogeneous. In particular, this assumption can lead to erroneous analysis due to local minima in the cost function, which prevents the convergence of the outer-loops towards the optimal solution of the problem. To better account for non-linearity in the observation operator that relates the ROS model parameters (control space) and the fire front location (observation space), the modification to an ensemble-based data assimilation approach is now explored.

7.2 Ensemble-based parameter estimation strategies

Ensemble-based data assimilation approaches, originally developed for dynamic state estimation (Evensen, 1994), have been extended to sequential parameter estimation, for instance in the field of hydrodynamics and hydrology (Durand et al., 2008; Moradkhani et al., 2005). Following the strategy presented in Section 6.5.3, Chapter 6, the application of the EnKF-based parameter estimation approach to regional-scale wildfire spread is explored here based on work presented in Rochoux et al. (2012, CTR) and Rochoux et al. (2014a, NHESS); its advantages and drawbacks are shown in comparison to particle filters (PF) and reduced-cost strategies based on a polynomial chaos surface response model (PC-EnKF).

7.2.1 Behavior of the standard ensemble Kalman filter

↪ Simple example

As a validation step (referred to as P-OSSE-ANISO), the EnKF algorithm is applied to correct the proportionality coefficient P [s^{-1}], which accounts here for all the uncertainties in the ROS model (see Eq. 6.6, Chapter 6). In the following experiments, the fire is ignited at $(x, y) = (100\ \text{m}, 100\ \text{m})$ as a circular front with a radius of 5 m and spreads upon a random fuel distribution $\delta_v = \delta_v(x, y)$ over a $200\,\text{m} \times 200\,\text{m}$ domain. Observations (represented using $N^{\text{o}}_{fr} = 20$ front markers) are synthetically generated at 50 s intervals with Firefly and a chosen true value $\mathbf{x}^{\text{t}} = P^{\text{t}} = 0.4\ s^{-1}$. An observation error characterized by the error STD σ^{o} (assumed constant along assimilation cycles) is also introduced. The ensemble of prior values for the first cycle $\mathbf{x}^{\text{f}}_1$ is drawn from a Gaussian distribution centered in $0.2\ s^{-1}$ with an error STD $\sigma^{\text{f}} = 0.05\ s^{-1}$ (also assumed constant along assimilation cycles).

▷ **Sensitivity to observation errors**. While EKF-based results were produced with a low value of the observation error STD (relatively to the perimeter of the fireline), Fig. 7.7(a) examines the influence of this error on the EnKF performance up to $\sigma^{\text{o}} = 60$ m. Statistics (in terms of mean value and STD) of the analysis

obtained for $N_e = 48$ members over one assimilation cycle (i.e., at time $t = 50$ s) are presented as a function of the magnitude of observation errors measured by σ^o. Vertical bars give a graphical representation of the magnitude of the STD within the analysis ensemble. When the observation error STD σ^o is small, the EnKF algorithm successfully drives the analysis ensemble towards the true value of the parameter $P^t = 0.4$ s^{-1}; the resulting analysis exhibits a much reduced scatter by at least a factor 4 in comparison to the forecast STD $\sigma^f = 0.05$ s^{-1}. In contrast, when the observation error STD is large, the EnKF algorithm has reduced effects and the analysis ensemble remains close to the forecast ensemble (the analysis STD is similar to the forecast STD $\sigma^f = 0.05$ s^{-1}). For intermediate values of the observation error STD, the EnKF algorithm produces optimized predictions lying between forecast and observation. These different regimes illustrate how data assimilation combines information from both models and observations and produces better results than those that would be obtained if models or observations were considered separately.

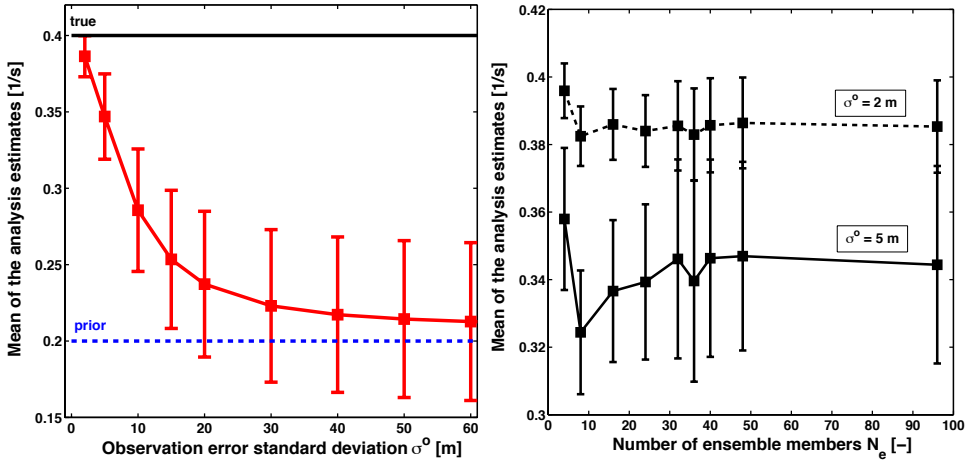
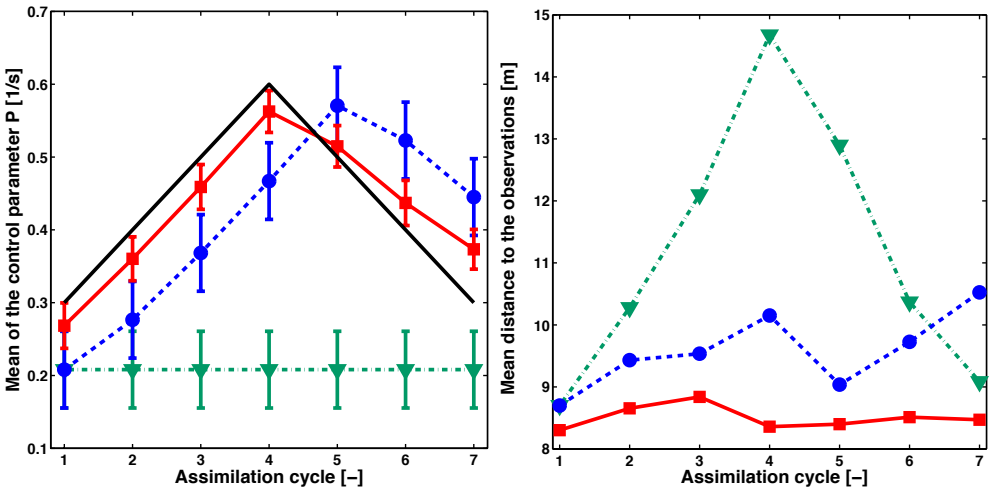


Figure 7.7: *Estimation of the coefficient P [s^{-1}] for one assimilation cycle (P-OSSE-ANISO). (a) Mean and STD of the analysis estimates as a function of the observation error STD σ^o. The black solid line corresponds to the true value 0.4 s^{-1}; the blue dashed line corresponds to the mean value of the forecast 0.2 s^{-1}; and the red squared-solid line corresponds to the mean value of the analysis. (b) Convergence of the mean analysis estimates with respect to the number of members N_e within the ensemble for $\sigma^o = 2$ m (dashed line) and $\sigma^o = 5$ m (solid line). Vertical error bars correspond to the associated error STD.*

▷ **Sensitivity to sampling errors.** Convergence properties of the EnKF-based analysis are studied in Fig. 7.7(b) with respect to the number of members N_e in the ensemble (when the EnKF ensemble contains N_e members, the regional-scale wildfire spread simulator is integrated N_e times to produce N_e fire front trajectories associated with each set of control parameters, i.e., each ensemble member, during

each assimilation cycle). The EnKF converges for a minimum of $N_e = 48$ members in the present configuration (for both $\sigma^o = 2$ m and $\sigma^o = 5$ m). The uncertainty related to the analysis is also lower for $\sigma^o = 2$ m (by nearly a factor 2 compared to $\sigma^o = 5$ m); as expected, the more accurate the observations, the more certain the analysis for a given forecast error.

▷ **Sensitivity to temporal variability in the control parameter.** Sequential application of the EnKF allows for a temporal correction of the parameter P for a case in which the time-varying profile of the true parameter was artificially set up between 0.3 and 0.6 s^{-1} over 7 assimilation cycles as illustrated in Fig. 7.8(a). Note that for this experiment, $N_e = 48$ and a constant observation error STD $\sigma^o = 5$ m is assumed. While the mean value of the prior estimates is set to 0.2 s^{-1} for the first assimilation cycle, it is set to the mean of the analysis estimates from the previous assimilation cycle otherwise.



(a) Control parameter space. (b) Observation space.

Figure 7.8: *Sequential EnKF estimation of the coefficient P [s^{-1}] over 7 assimilation cycles with $N_e = 48$ members and $\sigma^o = 5$ m; time-varying true control parameter (P-OSSE-ANISO). The green triangled-dashed-dotted line corresponds to the free run (without data assimilation); the blue circled-dashed line corresponds to the mean forecast estimate; the red squared-solid line corresponds to the mean analysis estimate; and the black solid line corresponds to the truth. Vertical error bars correspond to the associated error STD. (a) Parameter estimates. (b) Mean distance to the observed fire front.*

As shown in Fig. 7.8(a), the EnKF solution allows an optimal mean value of the parameter to be identified, which results in an ensemble of fire fronts that is coherent with the observation error statistics in Fig. 7.8(b). In contrast, the model without data assimilation (referred to as *free run*) significantly underestimates the ROS. While being not as accurate as the analysis at the assimilation time, the forecast provides a significant improvement in the prediction of wildfire spread at future lead-times compared to the free run. For instance, the mean distance to the

observations is multiplied by a factor 1.3 for the free run compared to the forecast over the third assimilation cycle. It is worth noting that there is a temporal shift between the forecast and analysis estimates in Fig. 7.8(a). The analysis estimate is obtained at the current observation and thereby, provides the most recently-updated information. In contrast, the forecast only contains information up to the previous observation time and the correction obtained at the current time through parameter estimation is only integrated in the forecast further in time. This explains the time delay of the forecast to gain information on the actual fire spread in comparison to the analysis. This illustrates that the quality of the forecast highly depends on the temporal variability of the errors in the control parameters and that the assimilation frequency must be set according to this variability in order to guarantee a high level of performance of the data-driven simulation.

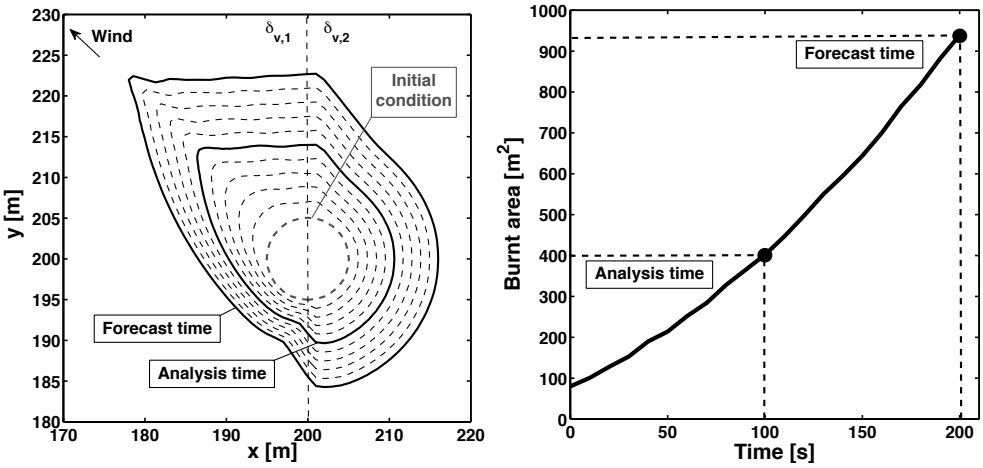
↪ Multi-parameter multi-cycle estimation

Results from OSSE tests, in which multiple parameters of the ROS model due to Rothermel are controlled (referred to as P-OSSE-ANISO-WIND), are now presented. The numerical configuration corresponds to a 400 m $\times$ 400 m domain with spatially-varying vegetation properties (short grass vegetation type) and with wind. The fuel depth is assumed to be spatially-varying taking different values in the two parts of the square-shaped computational domain ($\delta_{v,1}$ for $x \leq 200$ m and $\delta_{v,2}$ $x > 200$ m); the fuel moisture content M_v is assumed spatially-uniform. Ignition occurs at $(x, y) = (200\text{ m}, 200\text{ m})$ as a circular front with a radius of 5 m. The true trajectory of the fire front (see Table 7.3 for the related environmental conditions) is shown in Fig. 7.9(a) over the time period $[0; 200\text{ s}]$; the time-evolving burnt area is presented in Fig. 7.9(b). These figures show that the fire front propagates at an anisotropic ROS along the fireline; faster rates are simulated along the wind direction; at the flanks and at the back, the fire front propagates faster in the zone of highest fuel depth ($x \leq 200$ m).

Table 7.3: *Properties of the ensemble forecast, in terms of mean value and STD.*

Input parameter	True value	Ensemble mean	Ensemble STD
$\delta_{v,1}$ [m]	0.50	0.40	0.15
$\delta_{v,2}$ [m]	1.00	1.10	0.15
M_v [%]	16	20	4
(u_w, α_w) ([m/s], [°])	(1.0, 315)	(0.85, 280)	(0.15, 30)

In the following experiments, uncertainties in the forecast ensemble are due to variations in as many as 5 input parameters (2 values of δ_v for the 2 parts of the domain, plus values of M_v, u_w and α_w). These parameters are perturbed around mean values and with prescribed uncertainties presented in Table 7.3. The observed fire fronts are discretized using $N^{\text{o}}_{fr} = 20$ markers; the observation error STD is assumed to be small (i.e., $\sigma^{\text{o}} = 0.1$ m) in order to evaluate the ability of the EnKF to track the observed fire front location.



(a) (b)

Figure 7.9: *Time-evolving true trajectory over the time window* $[0; 200\ s]$ *for a wind-aided fire propagation (P-OSSE-ANISO-WIND). (a) True fire front location at 10 s intervals (solid lines correspond to the fire front location at analysis and forecast times). (b) Corresponding time-evolving burnt area for the true trajectory.*

▷ **Sampling errors in a multi-parameter estimation.** The EnKF-based analysis is studied with respect to the number of ensemble members N_e for different sizes n of the control vector, from 1-parameter (with the stand-alone estimation of the wind magnitude, i.e., $\mathbf{x} = u_w$) to 5-parameter estimation (i.e., $\mathbf{x} = [u_w, \alpha_w, M_v, \delta_{v,1}, \delta_{v,2}]$), corresponding to increasing sources of uncertainties. The estimation is performed over one assimilation cycle $[0; 100\ \text{s}]$.

- Figure 7.10 represents the statistics of the distance to the observations for the forecast (FMO) and analysis (AMO) ensembles, in terms of mean value and STD, for varying numbers of ensemble members N_e. For any number N_e, the EnKF-based algorithm is able to retrieve parameters that lead to an ensemble of simulations that is more consistent with the observations (with the mean value of AMO always below 1 m). The distance between predicted and observed fire fronts is drastically reduced through the data assimilation feedback, even though the performance of the forecast degrades when increasing the number of control parameters n (i.e., when more uncertainties are represented). The mean FMO is near 4 m for at least 3 control parameters (since the wind velocity magnitude and direction angle introduce the largest scatter in the ensemble), while it is nearly 0.5 m for $n = 1$ and 2 m for $n = 2$. These discrepancies in the representation of uncertainties (in terms of fire front location and behavior) are illustrated in Fig. 7.11. The analysis counterparts in the observation space are represented in Fig. 7.12; they show that despite the increasing level of uncertainties, the EnKF remains capable of improved performance compared to a stand-alone forecast since the analysis exhibits a much reduced scatter in terms of fireline location and shape.

- As illustrated in Fig. 7.12, when the number of control parameters n increases, the scatter in the analysis ensemble is larger. This is due to the increased complexity of the estimation problem: even though the distance to the observations is significantly reduced, it is more difficult to retrieve parameter values that provide the exact shape of the observed fire front. As shown in Fig. 7.13 for the 5-parameter estimation case, a large ensemble (at least $N_e = 320$ members) is indeed required to find the true values of the control parameters, in particular of the wind magnitude u_w (see Fig. 7.13(b)) and of the fuel moisture content M_v (see Fig. 7.13(d)). Consistently, the correct burnt area is retrieved for at least $N_e = 320$ members. This large ensemble is necessary to accurately describe the spatial variability in the errors and to dissociate the effects of each control parameter on the fireline location and shape, partly due to equifinality issues. Note that the threshold value for N_e (for which the error statistics are converged) increases with the number of control parameters n as shown in Table 7.4.

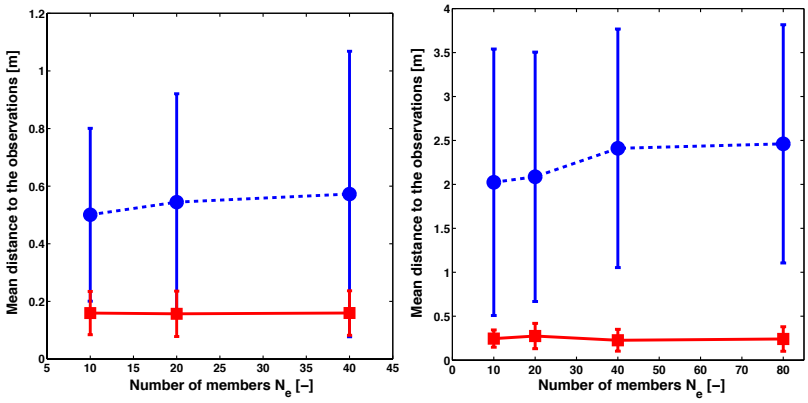
Table 7.4: *Threshold number of ensemble members N_e with respect to the size of the parameter estimation problem n.*

Size of the control vector (n)	1	2	3	4	5
Threshold number of members (N_e)	10	40	80	160	320

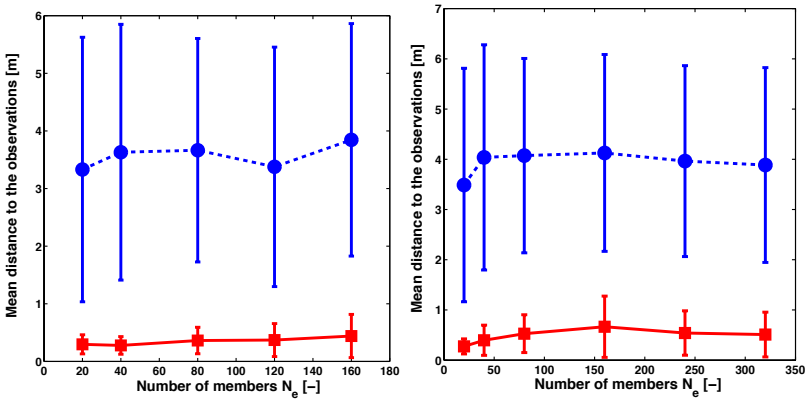
However, there is no general rule to define this threshold number of members in the ensemble; this is case-dependent. In practice, the true values of the control parameters are unknown and cannot be used as reference for the validation of the data assimilation feedback. Furthermore, the stand-alone criterion of the distance to the observations is not sufficient here to ensure that the analysis provides physical values of the control parameters (less than 100 members are sufficient to obtain converged error statistics in the observation space in Fig. 7.10 for any size of the control vector n, while N_e must be at least 320 for a 5-parameter estimation to obtain realistic values of the control parameters). It is therefore recommended for parameter estimation to include several criteria as for instance the burnt area extent shown in Fig. 7.13(a) and to generate large ensembles according to the computational time of the forward model and available computational capacities.

The linear increased computational cost with respect to N_e is shown in Fig. 7.14 for the present 5-parameter estimation experiment on a 4-processor machine. Thus, achieving a detailed representation of parameter error covariances is computationally intensive for increasing number of control parameters.

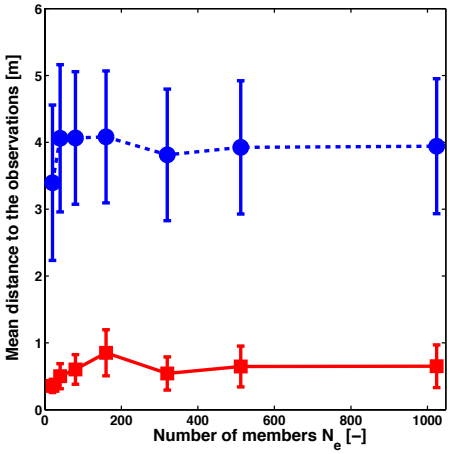
It is worth noting that the calculation of the ensemble of forecast fire front trajectories is performed using the Parasol functionality in OpenPalm (see Section 6.5.2, Chapter 6), which takes advantage of the independence between the ensemble members and which thereby, already reduces the total computational cost of the EnKF algorithm if multiple processors are available.



(a) $\mathbf{x} = [u_w],\ n = 1.$ (b) $\mathbf{x} = [u_w, \alpha_w],\ n = 2.$

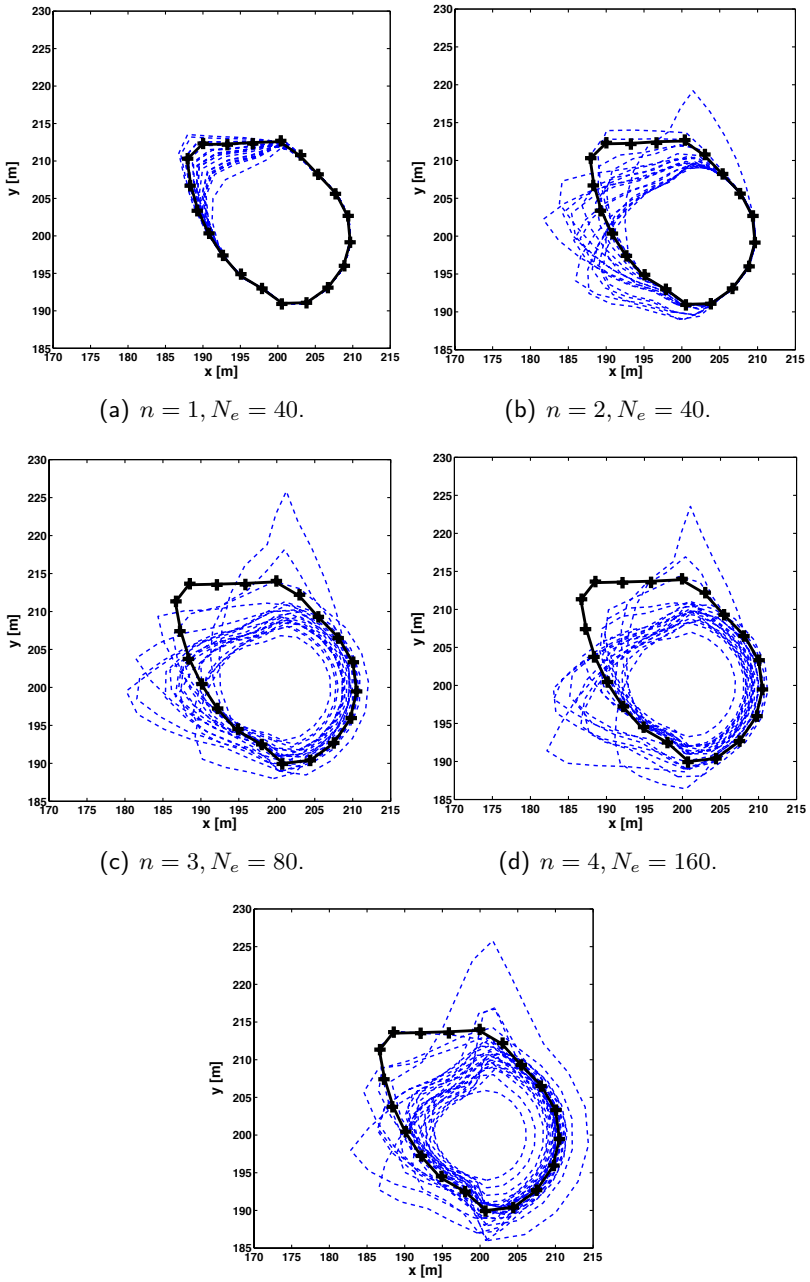


(c) $\mathbf{x} = [u_w, \alpha_w, M_v],\ n = 3.$ (d) $\mathbf{x} = [u_w, \alpha_w, M_v, \delta_{v,1}],\ n = 4.$



(e) $\mathbf{x} = [u_w, \alpha_w, M_v, \delta_{v,1}, \delta_{v,2}],\ n = 5.$

Figure 7.10: *Spatially-varying wind-aided OSSE test with respect to the number of members N_e for different sizes n of the control vector (P-OSSE-ANISO-WIND). The blue dashed line represents the mean distance between the observed and forecast front positions $\overline{d_{FMO}}$; the red solid line represents the mean distance between the observed and analysis front positions $\overline{d_{AMO}}$; error bars indicate the associated STD ($\overline{\sigma_{FMO}}$ and $\overline{\sigma_{AMO}}$).*

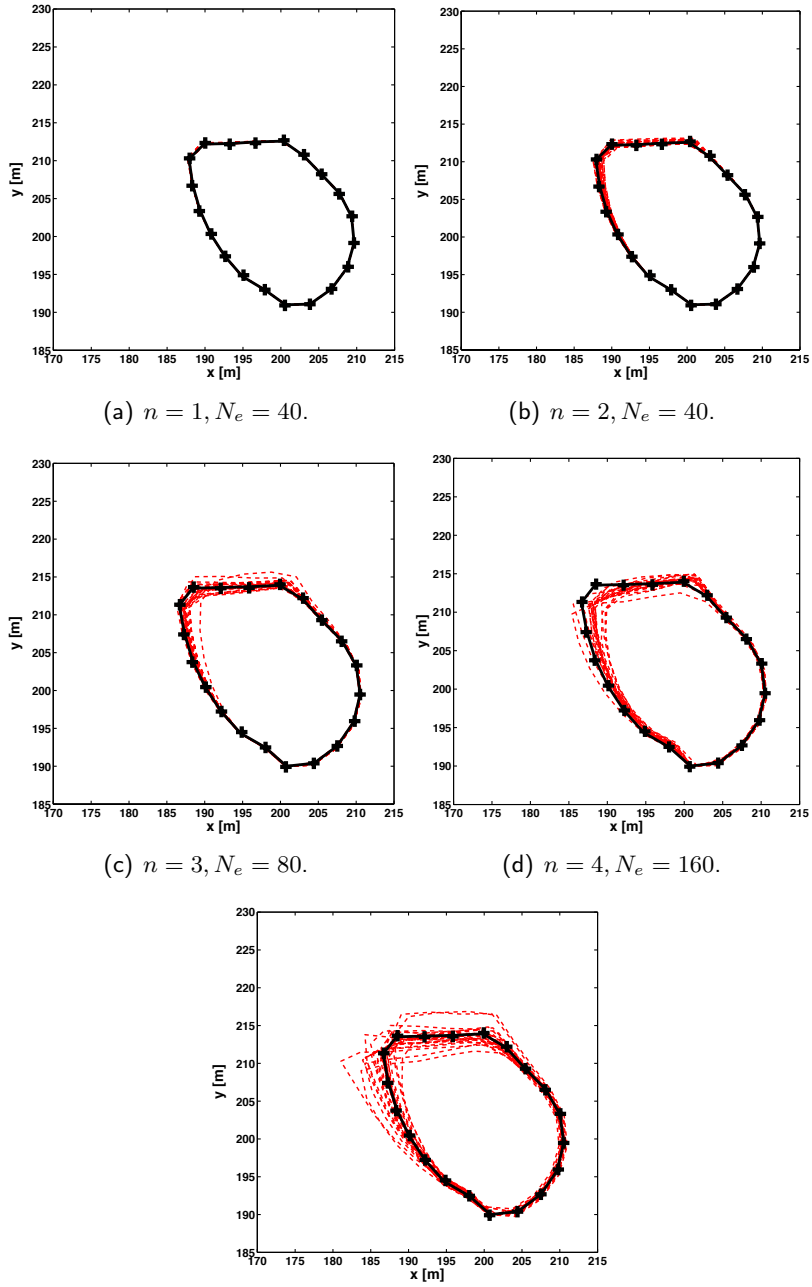


(a) $n = 1, N_e = 40$. (b) $n = 2, N_e = 40$.

(c) $n = 3, N_e = 80$. (d) $n = 4, N_e = 160$.

(e) $n = 5, N_e = 320$.

Figure 7.11: *Ensemble of forecasts for the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) with respect to the size of the control vector n. Blue dashed lines correspond to the simulated members without data assimilation for different sets of control parameters; the black solid line corresponds to the true fire front; and black crosses correspond to observations. Only a subset of the ensemble is presented for clarity purposes.*



(a) $n = 1, N_e = 40$.

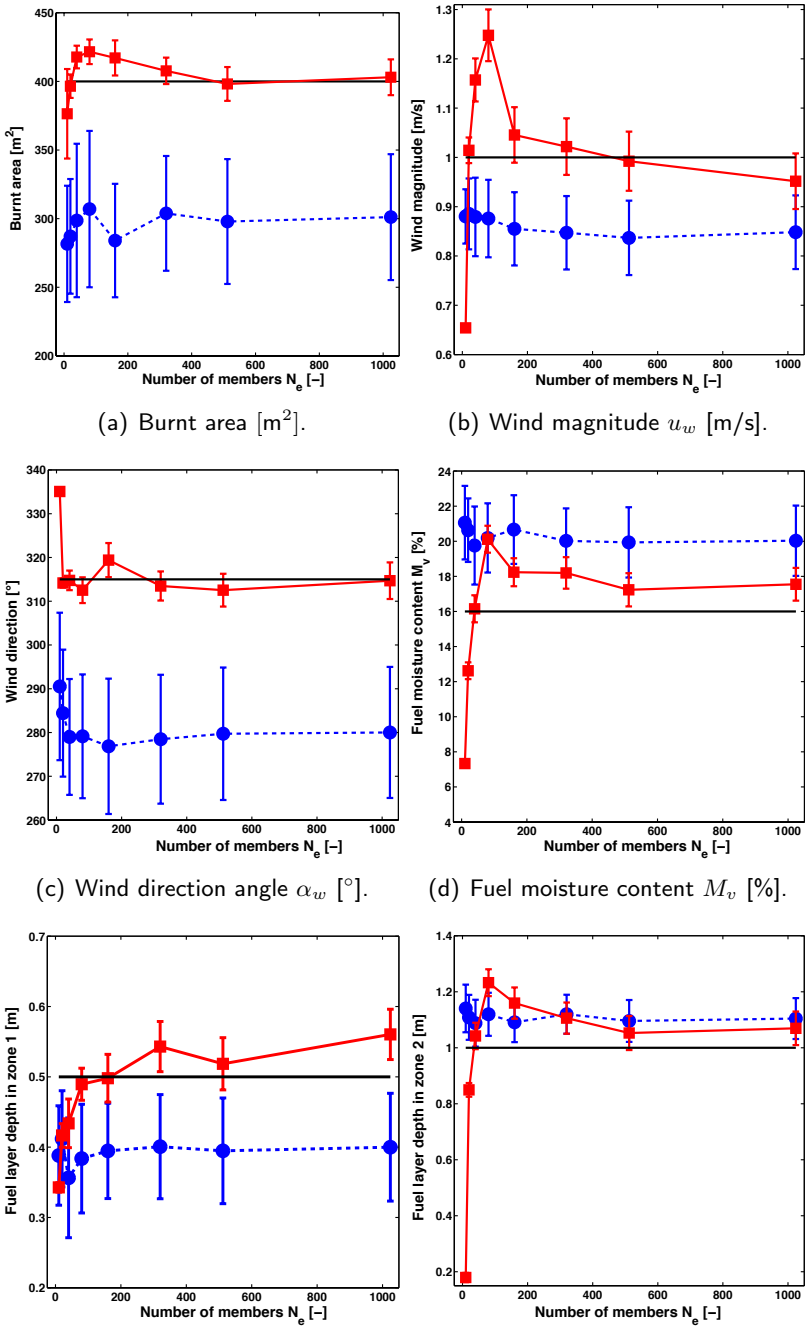
(b) $n = 2, N_e = 40$.

(c) $n = 3, N_e = 80$.

(d) $n = 4, N_e = 160$.

(e) $n = 5, N_e = 320$.

Figure 7.12: *Ensemble of analyses for the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) with respect to the size of the control vector n. Red dashed lines correspond to the simulated members with data assimilation for different sets of control parameters; the black solid line corresponds to the true fire front; and black crosses correspond to observations. Only a subset of the ensemble is presented for clarity purposes.*



(a) Burnt area $[\mathrm{m}^2]$. (b) Wind magnitude u_w [m/s].

(c) Wind direction angle α_w [°]. (d) Fuel moisture content M_v [%].

(e) Fuel layer depth in zone 1 $\delta_{v,1}$ [m]. (f) Fuel layer depth in zone 2 $\delta_{v,2}$ [m].

Figure 7.13: *Spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) with respect to the number of members N_e in the ensemble for a 5-parameter estimation. Blue circled-dashed lines represent the mean forecast estimate; red squared-solid lines represent the mean analysis estimate; black solid lines represent the true value. Error bars indicate the associated ensemble STD.*

In the present test only a few processors were used for illustration purposes, but in a real-world application it is envisioned that the number of processors devoted to forecasting wildfire spread will be high, meaning that the computational time will be significantly lower for each number of members N_e than in Fig. 7.14, allowing for real-time application.

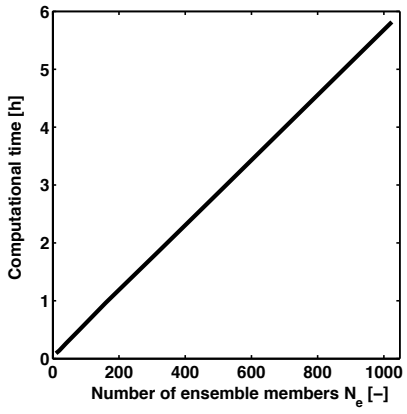


Figure 7.14: *Computational cost [h] with respect to the number of members N_e for the 5-parameter estimation approach ($n = 5$) on a multi-core processor machine (Intel-Xeon-E5520, 4 cores).*

▷ **Error statistics for multi-parameter estimation.** To highlight the key aspects of the EnKF parameter estimation approach, the error statistics in parameter space and in observation space are studied here for the 5-parameter estimation experiment with $N_e = 1024$ members in the ensemble. The Gaussian distribution of the forecast control parameters within the ensemble is compared to the distribution of the analysis control parameters and to their true value in Figs. 7.15 to 7.19. Consistently with Fig. 7.13, the probability of occurrence of the mean analysis estimate is higher than that of the mean forecast estimate (this mean analysis is closer to the true value) and the distribution STD is significantly reduced. The following aspects are important to mention to explain the complexity related to the 5-parameter estimation approach.

These normalized histograms clearly show that the wind magnitude u_w is the most difficult parameter to estimate among the 5 control parameters; its analysis distribution remains relatively scattered and is not significantly changed compared to its forecast distribution. In contrast, the wind direction angle α_w has a very specific impact on the fireline position and shape; thus, the EnKF achieves to perfectly retrieve its true value (its analysis distribution is dense around its mean value). One reason for this limited EnKF performance on the correction of the wind magnitude is that its impact on the fireline is limited to the head of the fire as shown in Fig. 7.11(a). There is a limited number of observed markers on this section of the fireline, meaning that a limited amount of information on the wind magnitude is included in the estimation problem. Thus, the observation of the head of the fire is

of primary importance to correct the wind magnitude u_w. Note that the distribution associated with the fuel moisture content M_v is not perfectly Gaussian in practice due to the moisture at extinction $M_{v,ext} = 0.30$. This induces a skewed distribution towards the decreasing moisture values (the tail on the side of lower moisture contents is longer than on the side of higher moisture contents). This shape of the distribution might lead to a non-optimal EnKF solution and could partly explained the remaining bias (with respect to the true value) in the estimation of M_v.

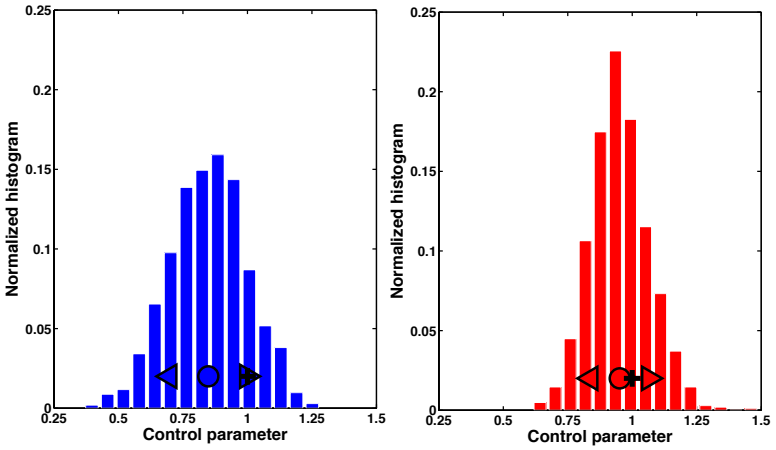


Figure 7.15: *Forecast (left) and analysis (right) normalized histograms for the wind magnitude u_w [m/s] in the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) for a 5-parameter estimation. Circles correspond to ensemble means; triangles correspond to ensemble STD; and crosses correspond to true values.*

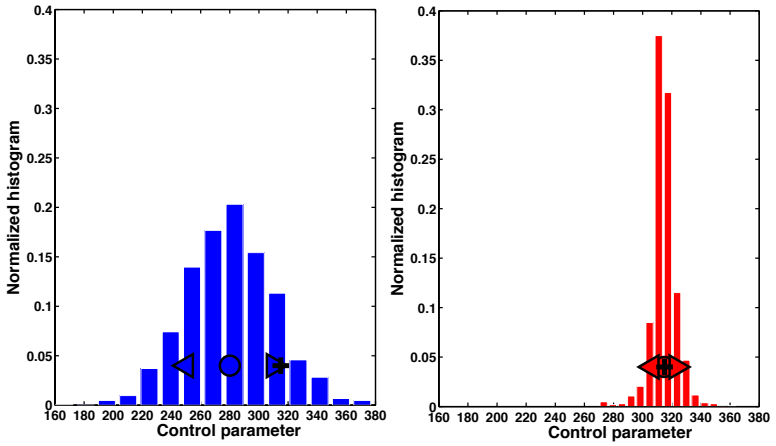


Figure 7.16: *Normalized histograms for the wind direction α_w [°] (see caption of Fig. 7.15).*

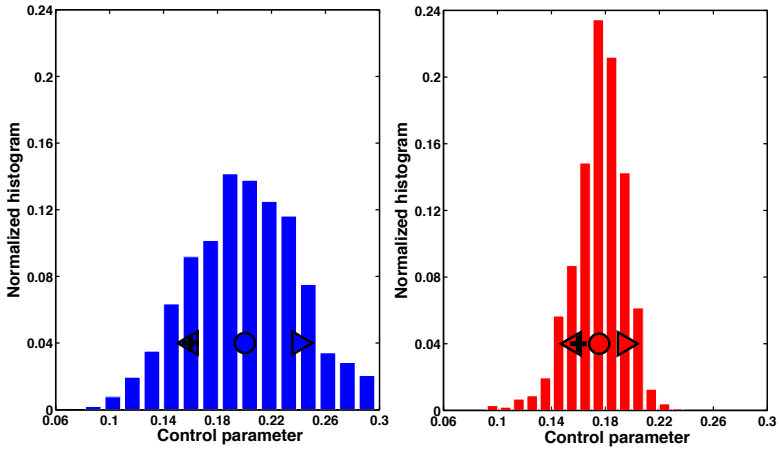


Figure 7.17: *Normalized histograms for the fuel moisture content* M_v $[-]$.

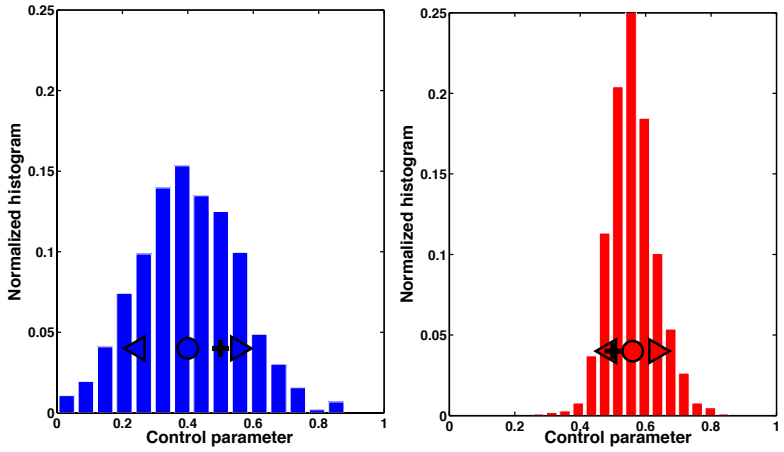


Figure 7.18: *Normalized histograms for the fuel layer depth in zone 1* $\delta_{v,1}$ $[m]$.

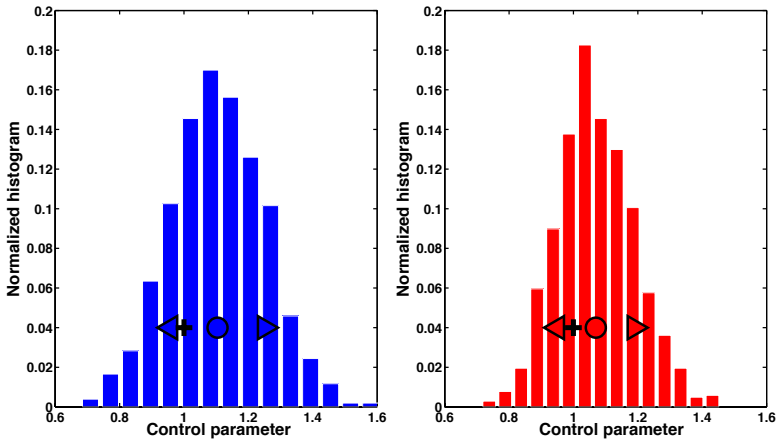
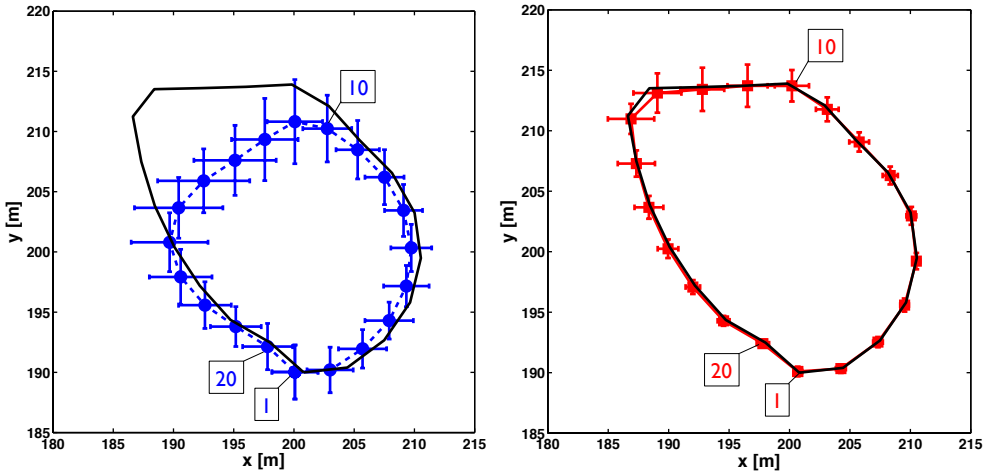


Figure 7.19: *Normalized histograms for the fuel layer depth in zone 2* $\delta_{v,2}$ $[m]$.

Figure 7.20(a) shows how uncertainties in the control parameters translate into uncertainties in the fire front marker locations (there are $N_{fr}^{o} = 20$ observed markers in this experiment). The analysis counterpart is presented in Fig. 7.20(b), which confirms the significant uncertainty reduction in the ensemble of analyses noticed in Fig. 7.12. Each coordinate of the simulated marker location is associated with a distribution represented in Figs. 7.21 and 7.22 for the ensemble of forecasts and in Figs. 7.23 and 7.24 for the ensemble of analyses.



(a) Forecast estimates. (b) Analysis estimates.

Figure 7.20: *Comparison of simulated and true fire fronts in the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) for a 5-parameter estimation. Colored symbols correspond to the ensemble mean location of the simulated fire front markers; error bars indicate the error STD related to the front marker location along x- and y-directions (horizontal and vertical error bars, respectively); and the black solid line corresponds to the true fire front.*

These figures confirm that the fire front marker positions exhibit approximate Gaussian or log-normal distributions, in particular at the head and flanks of the fire (the statistics at the rear of the fire present less Gaussian characteristics due to a very slow propagation in this section of the fireline). This justifies the application of a standard EnKF in the present parameter estimation approach (see Chapter 6). This choice of assimilated variables does not lead to bi-modal distributions, in contrast to temperature assimilation (non-burnt and burnt states) at fixed in-situ locations (Mandel et al., 2008; Beezley and Mandel, 2008). The definition of the observation space is of primary importance to obtain approximate Gaussian error statistics and thereby, make the EnKF parameter estimation approach successful.

▷ **Sensitivity to model errors.** The performance of the EnKF-based parameter estimation approach is now evaluated in a more realistic configuration, in which the observations are made with significant error (i.e., $\sigma^{o} = 2$ m per observed marker along the fireline) and in which there are significant model errors that are

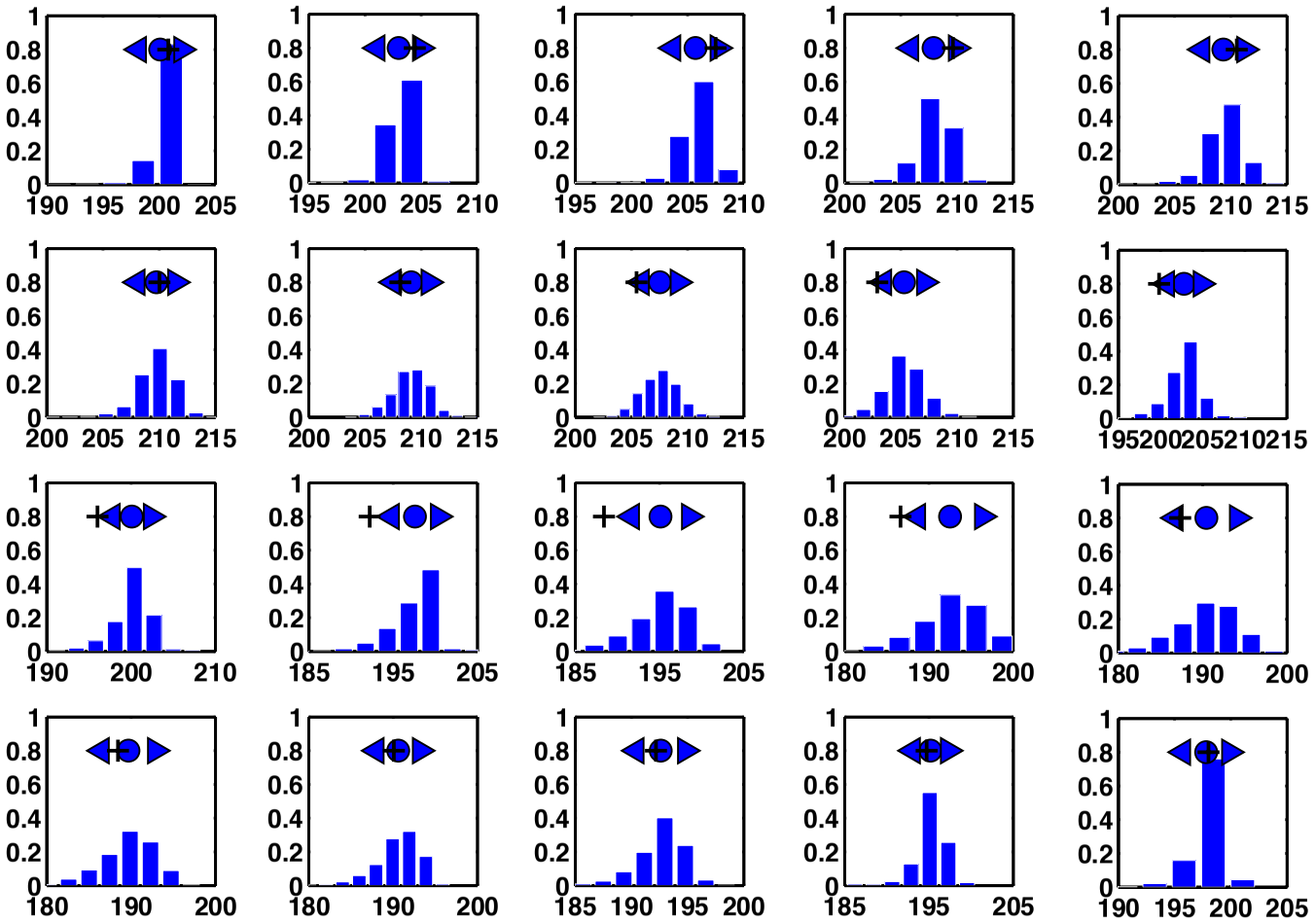


Figure 7.21: *Forecast normalized histograms for the forecast x-coordinate of each simulated fire front marker in the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) for a 5-parameter estimation. Each box corresponds to a marker along the fireline (the first box corresponds to the marker indexed by 1 in Fig. 7.20(a) and then, next boxes - from left to right and from bottom to top - correspond to the adjacent markers turning counter-clockwise along the fireline). Circles correspond to ensemble means; triangles correspond to ensemble STD; and crosses correspond to true values.*

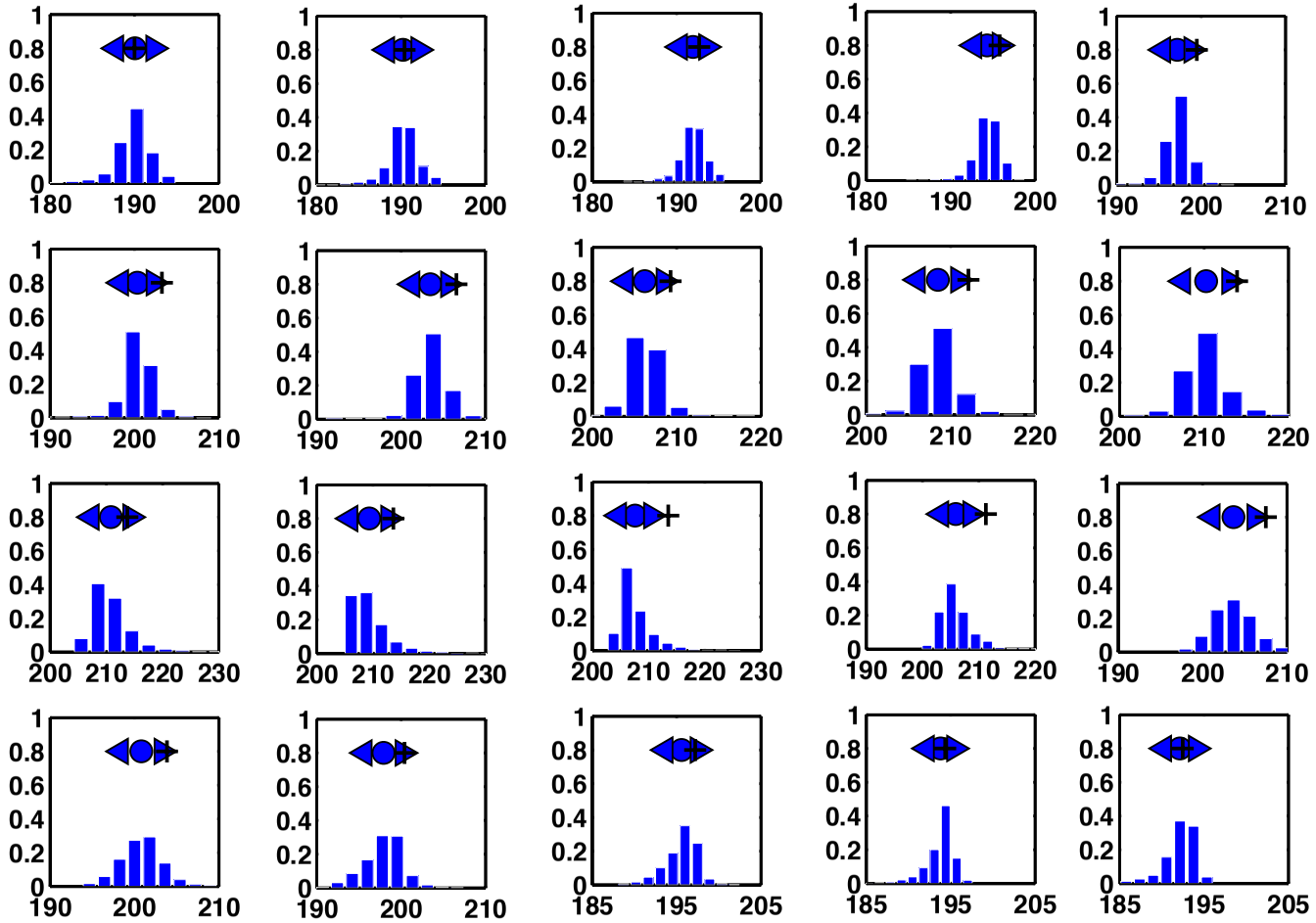


Figure 7.22: *Forecast normalized histograms for the forecast y-coordinate of each simulated fire front marker in the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) for a 5-parameter estimation. Circles correspond to ensemble means; triangles correspond to ensemble STD; and crosses correspond to true values.*

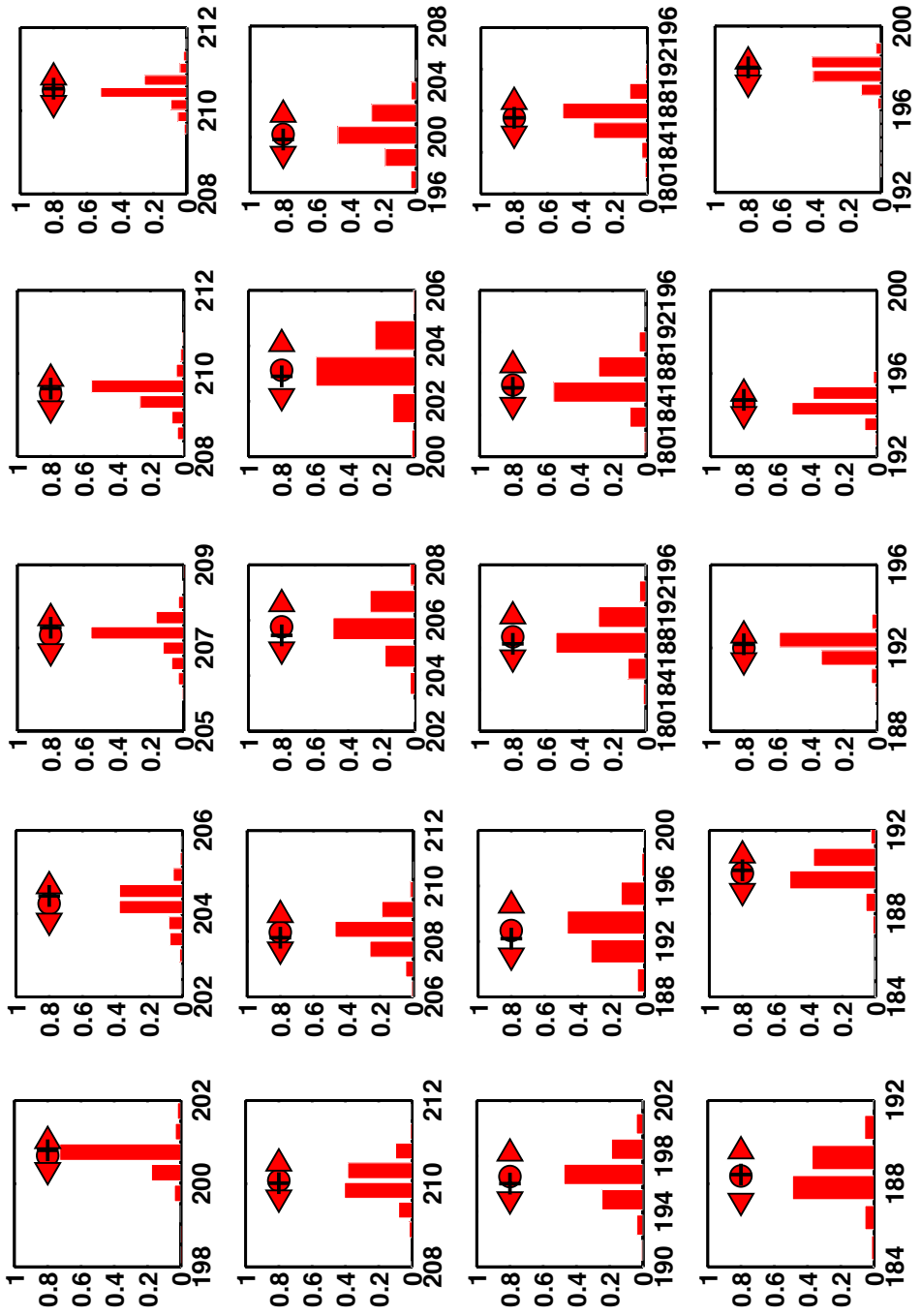


Figure 7.23: *Analysis normalized histograms for the analysis x-coordinate of each simulated fire front marker (see caption of Fig. 7.21).*

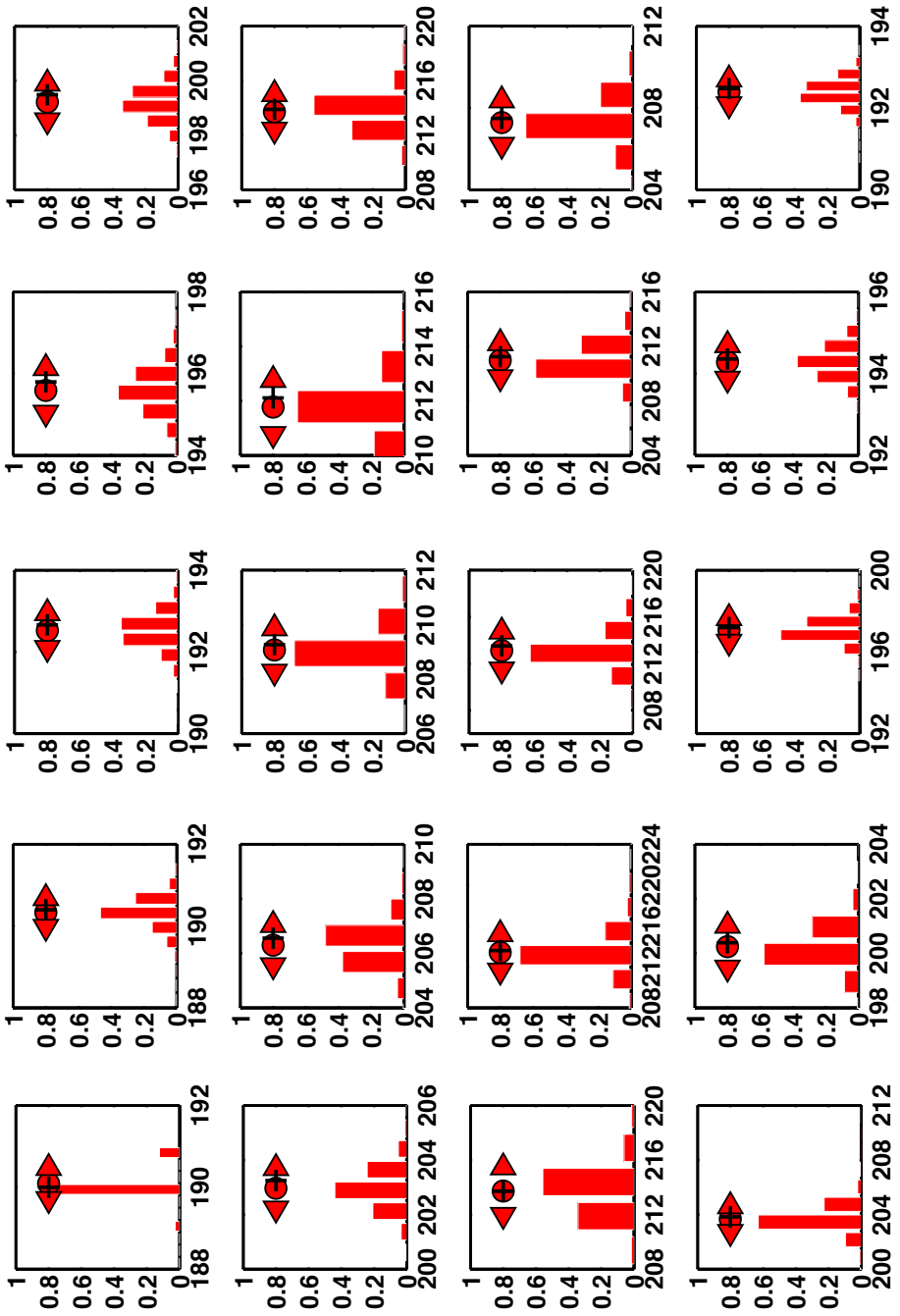


Figure 7.24: *Analysis normalized histograms for the analysis y-coordinate of each simulated fire front marker (see caption of Fig. 7.22).*

not controlled (referred to as *knowledge gaps*). This configuration deviates from the strict definition of OSSE experiments. Two different experiments listed below (corresponding to the case P-OSSE-ANISO-WIND and to a 5-parameter estimation problem with $\mathbf{x} = [u_w,\ \alpha_w,\ M_v,\ \delta_{v,1},\ \delta_{v,2}]$ and with $N_e = 512$ members) are performed to study the EnKF sensitivity to model errors. Figure 7.25 is used as reference (no model uncertainty is added to generate the ensemble of forecasts). Due to significant observation errors, the location of the observed markers does not coincide with the true fire front, in contrast to previously-mentioned results obtained in the case P-OSSE-ANISO-WIND.

(i) **Knowledge gap in ROS input parameter.** This test considers a case in which the fuel particle surface-area-to-volume ratio Σ_v (an input parameter of the Rothermel's ROS model) is not known, while not being included in the control vector $\mathbf{x}$. Observations are therefore obtained for a reference value of Σ_v chosen as $4,921$ 1/m, while the ensemble of forecast and analysis fire fronts are produced for a different value Σ_v chosen as $11,485$ 1/m (the typical short grass value in Rothermel's biomass fuel database used in the reference case). The knowledge gap in Σ_v constitutes an additional source of uncertainty that is not accounted for in the data assimilation experiment. As shown in Table 7.5, the analysis allows to slightly reduce the ensemble scatter in terms of burnt area and parameter STD. This leads to a more accurate prediction of the fire front shape, while the true head of the fire is not correctly retrieved by the analysis (see Fig. 7.26) similarly to the reference case (see Fig. 7.25). Despite this knowledge gap in Σ_v (to which the ROS is sensitive, in particular for the present wind-aided fire propagation), some information are gained through the EnKF estimation. However, the analysis values of the control parameters are not systematically more realistic: for instance, the mean moisture content is increased from 0.20 in the forecast to 0.22 in the analysis, while the true value is 0.16. The error correction due to the misspecification of Σ_v induces an over-correction of the control parameters included in $\mathbf{x}$. This illustrates the previously-mentioned discussion in Section 7.1.3 related to equifinality; the analysis must be viewed as effective values that incorporate the effects of uncertainty representation; physical values can only be retrieved if the main important sources of uncertainties are included in the control vector.

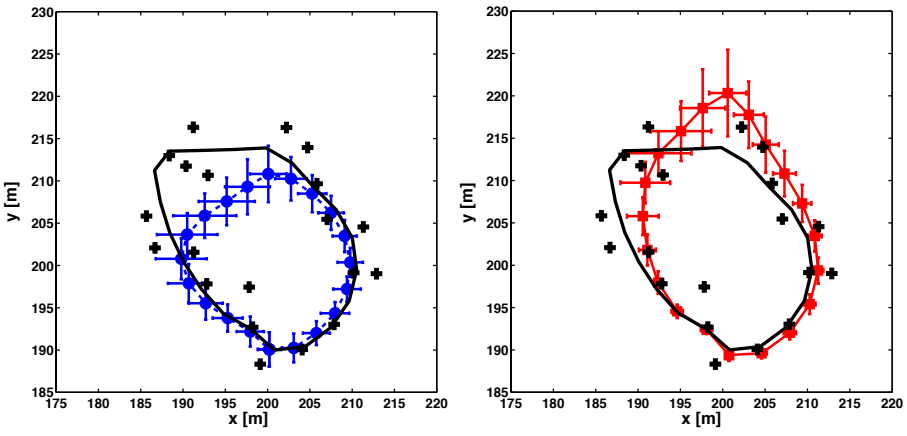
(ii) **Knowledge gap in ROS model parameter.** This test considers a case in which the ROS formulation is subject to uncertainty. As detailed in Appendix A, the wind coefficient ϕ_w^* in the ROS model due to Rothermel is expressed as follows:

$$\phi_w^* = C\, u_w^B \left(\frac{\beta_v}{\beta_{v,op}}\right)^{-E}, \tag{7.1}$$

with $B = 0.02526\, \Sigma_v^{0.54}$. In the present case, the ROS model parameter B (that is non-linearly related to the wind coefficient ϕ_w^* and thereby to the ROS) is uncertain. The true fire front is obtained for a coefficient B

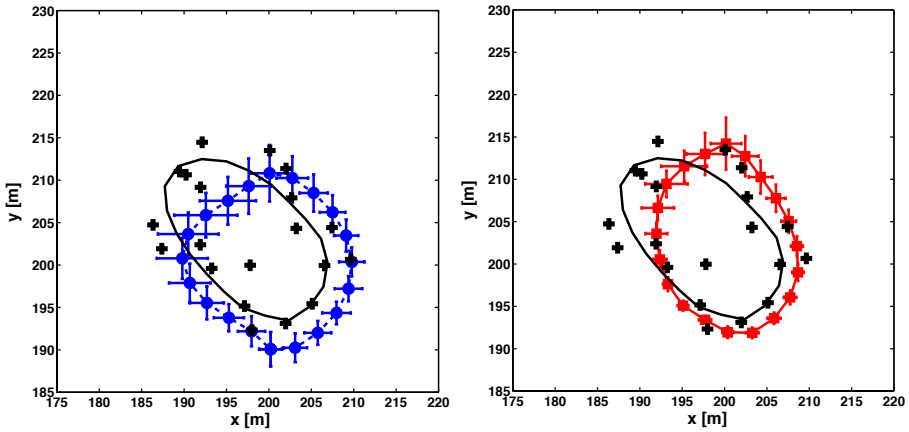
multiplied by a factor 2, while the ensemble of forecasts and analyses are obtained for the standard Rothermel's coefficient B as presented in Eq. (7.1). This increased ROS model parameter for the true trajectory intensifies the effect of the wind on the head of the fire as shown in Fig. 7.27 (this corresponds to a representative case, in which observations are obtained independently from the forward model, based on remote measurements and/or based on a more physically-detailed fire spread model). Figure 7.27(b) shows that the analysis does not retrieve an accurate location of the fireline, but there is a significant improvement in the estimation of the shape of the fireline and of the burnt area compared to the forecast. Table 7.5 also shows that the mean analysis estimates provide a good approximation of the true values (the bias between the true and simulated fire fronts is corrected), while their associated STD remain relatively large.

Even though observation errors are not negligible and uncertainties are not accurately represented in the ensemble in these tests, the EnKF parameter estimation approach is still able to provide valuable information on the fire, in particular on the shape of the fireline and on the size of the burnt area. The correction on the location of the fire front is limited and the control parameters are more difficult to estimate. While the new values of the control parameters are still realistic, they must be viewed as effective values that highly depend on the ensemble generation and on the choice of uncertainty representation. These tests highlight the potential of a data assimilation strategy for improved wildfire spread predictions in real-world applications.



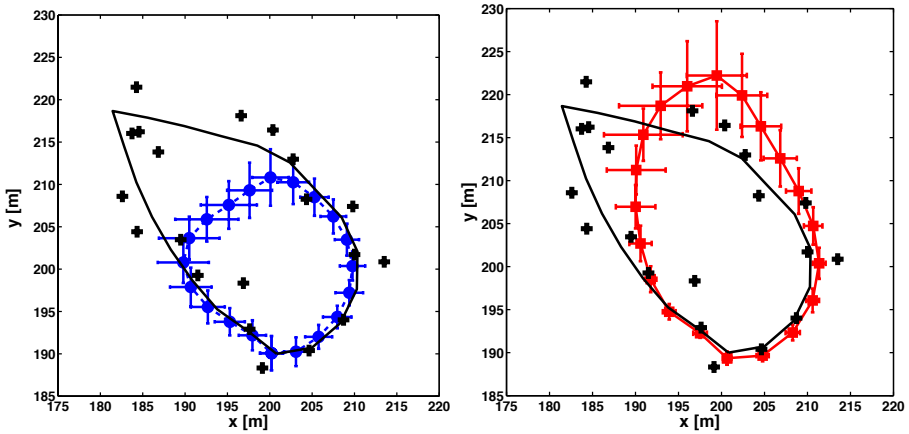
(a) Forecast estimates. (b) Analysis estimates.

Figure 7.25: *Comparison of the simulated and true fire fronts for a 5-parameter estimation with* $\mathbf{x} = [u_w, \alpha_w, M_v, \delta_{v,1}, \delta_{v,2}]$, $N_e = 512$ *and* $\sigma^o = 2$ *m (P-OSSE-ANISO-WIND); reference test. Colored symbols correspond to the ensemble mean location of the simulated fire front markers; error bars indicate the error STD related to the front marker position along* x- *and* y-*directions (horizontal and vertical error bars, respectively); the black solid line corresponds to the true fire front; and black crosses correspond to observations.*



(a) Forecast estimates. (b) Analysis estimates.

Figure 7.26: *Test 1 (knowledge gap in ROS input parameter Σ_v), see caption of Fig. 7.25.*



(a) Forecast estimates. (b) Analysis estimates.

Figure 7.27: *Test 2 (knowledge gap in ROS modeling in the wind coefficient ϕ_w^*, the coefficient B being multiplied by a factor 2 in the true trajectory), see caption of Fig. 7.25.*

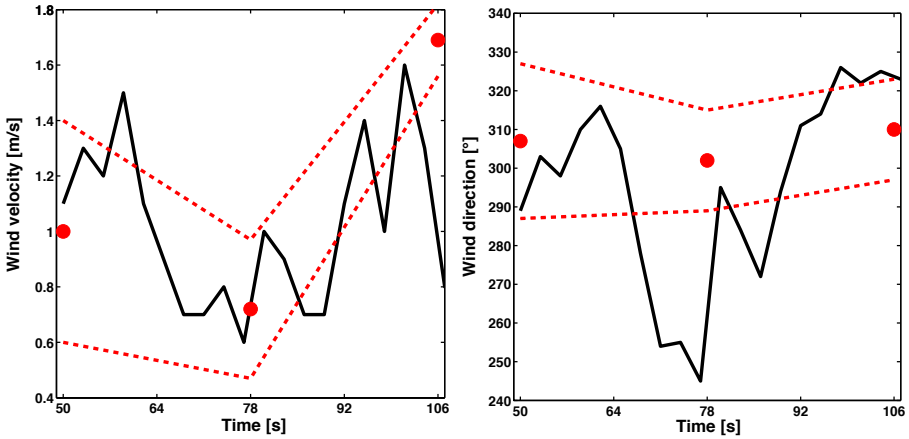
7.2.2 Application to the controlled grassland burning

The EnKF-based parameter estimation approach is applied to the real-world case used in Section 7.1.4 to evaluate the EKF approach. Here the EnKF is performed over two successive assimilation cycles, $[50; 78$ s] with an analysis update at $t_1 = 78$ s and $[78; 106$ s] with an analysis update at $t_2 = 106$ s. EKF-based results have shown the importance to account for uncertainties in wind conditions, i.e., in the wind magnitude u_w and direction angle α_w since they are subject to high-frequency variations as highlighted by measurements in Fig. 7.28. Thus, $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$ ($n = 4$). The observed fire front at $t_0 = 50$ s is taken as the initial condition.

Table 7.5: *Comparison of control parameter values and burnt areas between the true control vector, the ensemble of forecasts and the ensemble of analyses in the spatially-varying wind-aided OSSE test (P-OSSE-ANISO-WIND) for a 5-parameter estimation with* $\mathbf{x} = [u_w, \alpha_w, M_v, \delta_{v,1}, \delta_{v,2}]$ *and with model errors: test 1 corresponds to misknowledge of the fuel particle surface-area-to-volume ratio* Σ_v *(ROS input parameter) and test 2 corresponds to misknowledge of the ROS model parameter* B *in Eq. (7.1).*

Test	True		Forecast		Analysis	
	Parameter	Burnt area	Ens. mean/STD	Burnt area	Ens. mean/STD	Burnt area
Ref.	$\begin{pmatrix} 1.00 \text{ m/s} \\ 315° \\ 0.16 \\ 0.50 \text{ m} \\ 1.00 \text{ m} \end{pmatrix}$	400 m^2	$\begin{pmatrix} 0.85 \pm 0.15 \text{ m/s} \\ 280 \pm 30° \\ 0.20 \pm 0.04 \\ 0.40 \pm 0.15 \text{ m} \\ 1.10 \pm 0.15 \text{ m} \end{pmatrix}$	$298 \pm 91 \text{ m}^2$	$\begin{pmatrix} 0.90 \pm 0.15 \text{ m/s} \\ 332 \pm 16° \\ 0.16 \pm 0.023 \\ 0.49 \pm 0.10 \text{ m} \\ 1.14 \pm 0.13 \text{ m} \end{pmatrix}$	$442 \pm 48 \text{ m}^2$
1	$\begin{pmatrix} 1.00 \text{ m/s} \\ 315° \\ 0.16 \\ 0.50 \text{ m} \\ 1.00 \text{ m} \end{pmatrix}$	242 m^2	$\begin{pmatrix} 0.85 \pm 0.15 \text{ m/s} \\ 280 \pm 30° \\ 0.20 \pm 0.04 \\ 0.40 \pm 0.15 \text{ m} \\ 1.10 \pm 0.15 \text{ m} \end{pmatrix}$	$298 \pm 91 \text{ m}^2$	$\begin{pmatrix} 0.88 \pm 0.14 \text{ m/s} \\ 336 \pm 17° \\ 0.22 \pm 0.023 \\ 0.52 \pm 0.09 \text{ m} \\ 0.97 \pm 0.13 \text{ m} \end{pmatrix}$	$277 \pm 38 \text{ m}^2$
2	$\begin{pmatrix} 1.00 \text{ m/s} \\ 315° \\ 0.16 \\ 0.50 \text{ m} \\ 1.00 \text{ m} \end{pmatrix}$	480 m^2	$\begin{pmatrix} 0.85 \pm 0.15 \text{ m/s} \\ 280 \pm 30° \\ 0.20 \pm 0.04 \\ 0.40 \pm 0.15 \text{ m} \\ 1.10 \pm 0.15 \text{ m} \end{pmatrix}$	$298 \pm 91 \text{ m}^2$	$\begin{pmatrix} 1.01 \pm 0.14 \text{ m/s} \\ 335 \pm 16° \\ 0.15 \pm 0.024 \\ 0.54 \pm 0.10 \text{ m} \\ 1.11 \pm 0.13 \text{ m} \end{pmatrix}$	$501 \pm 55 \text{ m}^2$

Prior estimates of the control parameters are described in Table 7.6 along with the associated STD. They are based on field measurements (wind conditions are specified as mean values of the wind magnitude and direction angle before the fire ignition, in order to mimic real-time data assimilation conditions; measurements over the fire duration are used for validation), except for Σ_v that is initially guessed based on Rothermel's fuel database (short grass category).



(a) Wind magnitude u_w [m/s]. (b) Wind direction angle α_w [°].

Figure 7.28: *Comparison of in-situ measurements during the controlled grassland fire experiment (black solid line) to EnKF estimation of wind conditions: red circles correspond to mean values (forecast at $t_0 = 50$ s, analysis at $t_1 = 78$ s and $t_2 = 106$ s), and red dashed lines correspond to associated STD.*

Table 7.6: *Controlled grassland fire experiment using EnKF-based parameter estimation: forecast and analysis ensemble statistics for $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$, $n = 4$.*

EnKF	Forecast control vector		Analysis control vector	
	Ens. mean	Ens. STD	Ens. mean	Ens. STD
Cycle 1	0.220	0.060	0.089	0.035
	11500 1/m	4000 1/m	20995 1/m	3000 1/m
	1.00 m/s	0.40 m/s	0.72 m/s	0.25 m/s
	307°	20°	302°	13°
Cycle 2	0.089	0.06	0.112	0.04
	20995 1/m	4000 1/m	18959 1/m	2725 1/m
	0.72 m/s	0.40 m/s	1.69 m/s	0.18 m/s
	302°	20°	310°	13°

A prior ensemble of $N_e = 1,000$ members is generated and corrected by assimilating the fire front at time $t_1 = 78$ s. The new values of the control parameters are used

to produce the forecast over the next assimilation cycle up to $t_2 = 106$ s: they are taken as the analysis at time $t_1 = 78$ s, while the associated error STD are taken as those of the prior estimates (random walk model) as shown in Table 7.6. Again, the new values of the control parameters obtained at $t_2 = 106$ s can be used to forecast the fire behavior at future lead-times, for instance at $t_3 = 134$ s. Results for the different temporal sequences are presented in Fig. 7.29: for each map, black crosses represent the current observed fire front (discretized with 40 markers) that is compared to the spread of the forward model simulations, without (blue dashed lines) or with (red solid lines) the EnKF update at the current time. The EnKF algorithm is found to significantly decrease the distance between observations and simulated fronts at both assimilation times $t_1 = 78$ s and $t_2 = 106$ s. The scatter of the fire spread trajectories within the ensemble is also reduced, in particular the STD of the control parameters is smaller (by at least 30 %) for the analysis than for the forecast in Table 7.6. It was also found that optimized wind conditions are consistent with field measurements (see Fig. 7.28) and this confirms that the EnKF provides realistic values for the control parameters. Note however that the EnKF is not able to retrieve the exact shape of the observed fireline (in particular at time $t_2 = 106$ s), even though the observation error is small ($\sigma^o = 5$ cm). The challenges found in the EnKF update are two-fold.

- ▷ First, there is the classical data assimilation challenge associated with providing accurate representations of uncertainties in the control variables. In this parameter estimation approach, these control variables are assumed spatially-uniform over the grassland lot. This assumption introduces some limitations in the way to correct the location of the simulated fireline; the deformations at the head of the fire cannot be represented with the current definition of the control vector $\mathbf{x}$ and need to be handled with an algorithm addressing the spatial variability of the uncertainties without increasing the computational cost of the data-driven strategy. This aspect is discussed in Section 7.3.
- ▷ Second, there is the challenge associated with accounting for the limited accuracy of the semi-empirical ROS model. Rothermel's model was specifically designed to represent the propagation at the head of the fire and is therefore not able to accurately represent the fire propagation at the flanks. In the present study, since the wind is blowing westward, the fire spread model tends to overestimate the ROS in the wind direction to be able to map the observed fire front on the eastward flank. This behavior is particularly obvious for the analysis at time $t_2 = 106$ s and for the forecast at time $t_3 = 134$ s when the fire only propagates on its flanks (since it has reached the field boundary on the West flank). This behavior is also observed for alternative data assimilation algorithms such as particle filters.[3]

  For instance, Fig. 7.30 compares the analysis obtained at time $t_1 = 78$ s for the EnKF with $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$, $n = 4$, and for the SIR/ASIR particles

[3] See Section 5.7, Chapter 5, for details on the principles underlying particle filters.

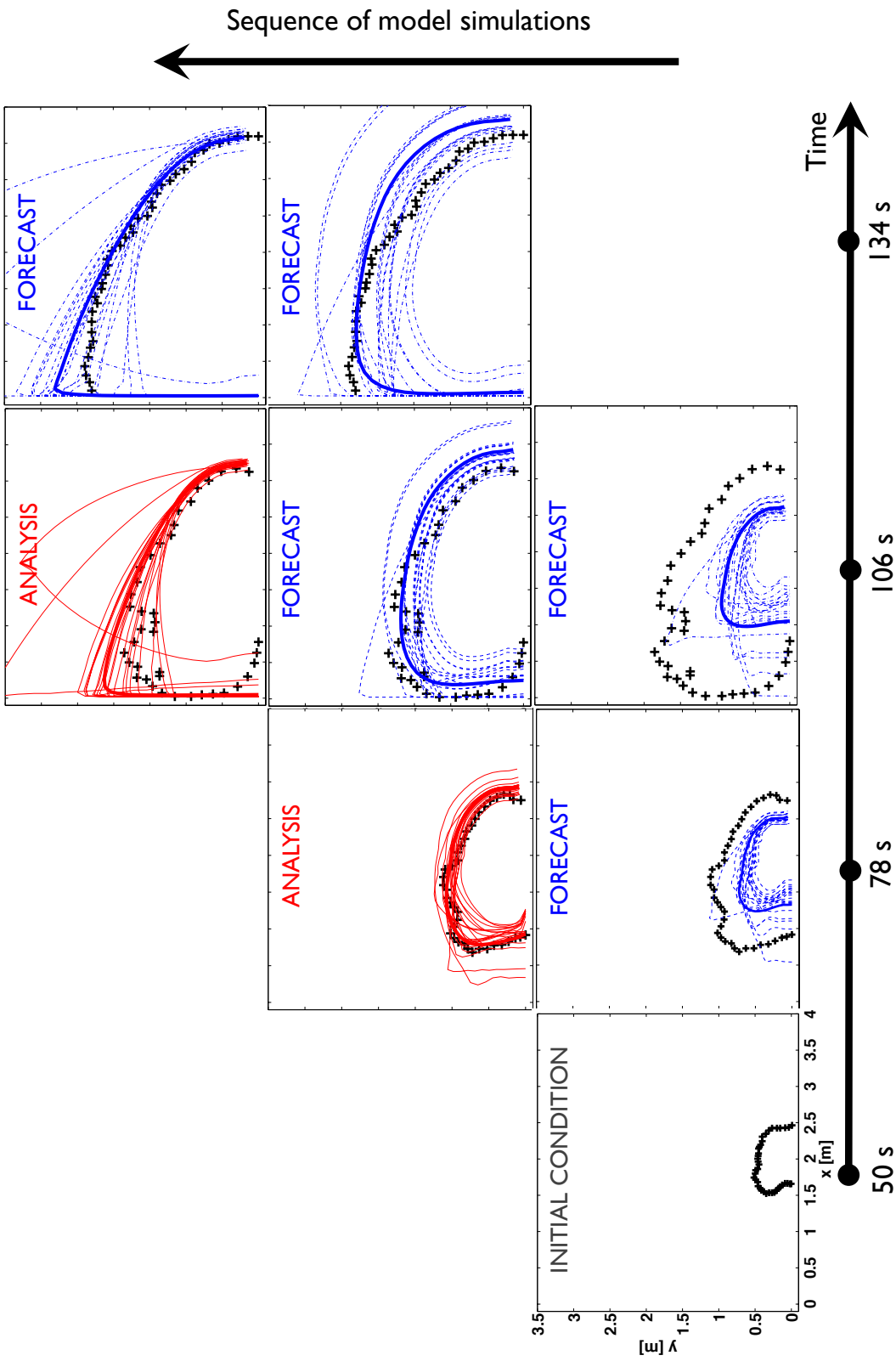


Figure 7.29: *Comparison of EnKF results (lines) to observations (black crosses) for data assimilation at 28 s time intervals from 50 s to 134 s. Each column corresponds to a model simulation snapshot at a given time; each line corresponds to a model simulation sequence, starting either from the initial condition at $t_0 = 50$ s or from the EnKF update. Only a subset of the ensemble is shown for clarity purposes.*

filters with $\mathbf{x} = [M_v, \Sigma_v, u_w]$, $n = 3$. This comparison shows that the optimized behavior of the fire spread model is independent from the choice of the data assimilation algorithm. SIR and ASIR particle filters as the EnKF are not able to retrieve the exact shape of the fireline: either the ROS is overestimated in the wind direction to exactly locate the observed fire front on the eastward flank, or the location of the head of the observed fire in the wind direction is correctly retrieved but the eastward flank is not. This choice is made through the estimation of the wind magnitude u_w and of the fuel particle surface-area-to-volume ratio Σ_v, which significantly affect the wind contribution to the ROS in the model due to Rothermel. As shown in Table 7.7, the main differences between the EnKF and particle filters are indeed in the analysis values for u_w and Σ_v. This discussion shows that ROS modeling limitations affect the shape of the optimized fireline in spite of the application of parameter estimation. While parameter estimation can provide insight into the temporal evolution of the environmental parameters, a complementary approach is therefore required to overcome ROS modeling limitations. This aspect is also discussed in Section 7.3 through the application of a state estimation approach. Note that the application of SIR and ASIR particle filters to wildfire spread is explained in detail in da Silva et al. (2014, HTHP) and da Silva et al. (2013, IPDO); a summary is provided in Appendix H.

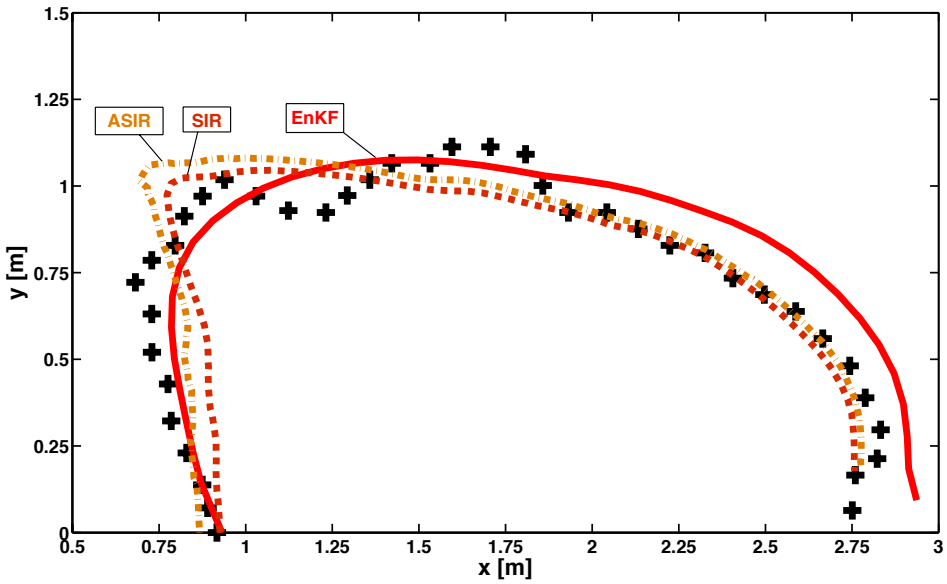


Figure 7.30: *Parameter estimation approaches applied to the controlled grassland fire experiment at $t_1 = 78$ s: comparison of the EnKF mean analysis obtained for a 4-parameter estimation (red solid line) to the results provided by SIR ($N_e = 200$) and ASIR ($N_e = 50$) particle filters for a 3-parameter estimation with $\mathbf{x} = [M_v, \Sigma_v, u_w]$: the orange dashed line corresponds to the SIR mean analysis estimate and the yellow dashed-dotted line corresponds to the ASIR mean analysis estimate. Black crosses correspond to observations.*

Table 7.7: *Validation of the EnKF for the controlled grassland fire experiment: forecast/analysis mean and STD of the control parameters using SIR/ASIR particle filters for* $\mathbf{x} = [M_v, \Sigma_v, u_w]\,,\ n = 3.$

SIR	**Forecast control vector**		**Analysis control vector**	
$N_e = 200$	Ens. mean	Ens. STD	Ens. mean	Ens. STD
Cycle 1	0.220	0.066	0.102	0.027
	11500 1/m	3450 1/m	15094 1/m	2596 1/m
	1.000 m/s	0.150 m/s	0.982 m/s	0.064 m/s
Cycle 2	0.102	0.031	0.104	0.028
	15094 1/m	4528 1/m	13263 1/m	4038 1/m
	0.982 m/s	0.147 m/s	1.095 m/s	0.082 m/s
ASIR	**Forecast control vector**		**Analysis control vector**	
$N_e = 50$	Ens. mean	Ens. STD	Ens. mean	Ens. STD
Cycle 1	0.220	0.066	0.089	0.035
	11500 1/m	3450 1/m	14914 1/m	2902 1/m
	1.000 m/s	0.150 m/s	0.986 m/s	0.050 m/s
Cycle 2	0.089	0.027	0.096	0.013
	14914 1/m	4474 1/m	14230 1/m	1942 1/m
	0.986 m/s	0.148 m/s	1.094 m/s	0.057 m/s

Even though the corrected shape of the fireline is not perfectly retrieved along assimilation cycles, the EnKF algorithm is able to provide more accurate forecasts of the fire spread than the stand-alone model, with a significant improvement compared to the free run (without data assimilation) as shown in Fig. 7.29. This result illustrates the improved accuracy of the simulation and forecast using the EnKF on a real-world controlled grassland burn. However, this estimation is made at the expense of heavy computational cost ($N_e = 1,000$ members for 4 control parameters and thereby, $N_e = 1,000$ forward model integrations for the forecast estimates as well as for the analysis estimates in observation space).

7.2.3 Reduced-cost ensemble-based data assimilation strategy

↪ Motivation

The classical EnKF algorithm is based on a Monte Carlo (MC) sampling to stochastically represent the forecast error covariance matrix using the members $\mathbf{x}_t^{\mathrm{f}}$ and their associated fire front location $\mathbf{y}_t^{\mathrm{f}}$. While this sampling technique is generic for the simulation of stochastic models and provides accurate access to the full statistics of modeling uncertainties, it is however computationally expensive due to the slow convergence rate of the MC technique (Fishman, 1996; Rubinstein and Kroese,

2008; Li and Xiu, 2008). The computational cost of one realization may be already expensive itself (Lucor et al., 2007). In particular, the large number of realizations required by the EnKF-based parameter estimation approach to obtain satisfactory results may prove computationally burdensome within an operational framework (especially considering that a fire spread model including spatially-distributed information on vegetation, terrain topography and atmospheric conditions such as Forefire/Méso-NH would be used in future work). The required size of the sample significantly increases with the complexity of the physics (i.e., multi-parameter estimation) and the non-linearity in the model (i.e., complex physics) even if the MC convergence rate is independent from these factors. These aspects emphasize the need for a reduced-cost EnKF. Efforts have therefore been devoted to designing ad-hoc variance reduction techniques (Boyaval, 2012) and more efficient EnKF schemes by reducing sampling errors (Saad, 2007; Szunyogh et al., 2008; Li and Xiu, 2008, 2009; Blanchard et al., 2010; Xiu, 2010; Rosic et al., 2013).

A large number of sampling methods have been developed to achieve the same level of accuracy with fewer model simulations than MC-based techniques; in particular, polynomial chaos (PC) non-intrusive methods issued from spectral-based representations and introduced by Wiener (1938) are efficient in terms of precision and cost (Ghanem and Spanos, 1991; Le Maître and Knio, 2010). For this purpose and following work from Li and Xiu (2009), a hybrid EnKF strategy based on a PC expansion and referred to as PC-EnKF is proposed in this thesis. Its feasibility was initially explored in Rochoux et al. (2012, CTR).

The key idea underlying this PC-EnKF approach is to build a polynomial representation of the Firefly forward model response (referred to as the *surrogate model*) to varying input parameters of the ROS model (Glimm and Sharp, 1999; Stern et al., 2001). Within the ensemble-based data assimilation framework, the use of the PC-based surrogate model instead of Firefly significantly reduces the cost of one realization and thereby, a large number of model simulation members can be produced to accurately characterize modeling uncertainties, at a lower computational cost than the standard MC-based sampling approach.

↪ Formulation of the reduced-cost ensemble-based strategy

▷ **General formulation of the surrogate model.** The PC-EnKF hybrid strategy is presented for the assimilation time window $[t-1,t]$; its flowchart is provided in Fig. 7.32. The PC-based surrogate model approximates the generalized observation operator $\mathcal{G}_t$ at time t, denoted by $\mathcal{G}_{pc,t}$. It is parameterized with respect to the multi-dimensional control vector $\mathbf{x}_t^{\mathrm{f}} \in \mathbb{R}^n$ following the forecast PDF $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t)$. The random vector $\mathbf{x}_t^{\mathrm{f}}$ may be regarded as a set of second-order random variables (with finite variance) expressed in terms of a random event ω such that $\mathbf{x}_t^{\mathrm{f}} = \mathbf{x}_t^{\mathrm{f}}(\omega)$. It can be projected onto a stochastic space spanned by orthogonal PC functions of

independent Gaussian random variables $\boldsymbol{\zeta}(\omega)$ as follows:

$$\mathbf{x}_t^{\mathrm{f}}(\omega) = \Big[\mathrm{x}_{1,t}^{\mathrm{f}}, \mathrm{x}_{2,t}^{\mathrm{f}}, \cdots, \mathrm{x}_{n,t}^{\mathrm{f}}\Big] = \sum_{q=0}^{\infty} \hat{\mathbf{x}}_q \, \varphi_q\Big(\boldsymbol{\zeta}(\omega)\Big). \tag{7.2}$$

The simulated positions of the fire front $\mathbf{y}_t^{\mathrm{f}} = \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}}(\boldsymbol{\zeta}))$ can also be viewed as a random variable and therefore, they can be projected onto a stochastic space spanned by orthogonal PC functions as follows:

$$\mathbf{y}_t^{\mathrm{f}} = \mathcal{G}_{pc,t}\Big(\mathbf{x}_t^{\mathrm{f}}(\boldsymbol{\zeta})\Big) = \sum_{q=0}^{\infty} \hat{\mathbf{y}}_q \, \varphi_q(\boldsymbol{\zeta}), \tag{7.3}$$

where $\hat{\mathbf{y}}_q \equiv \hat{\mathbf{y}}_q(t)$ are time-dependent coefficients, and where $(\varphi_q)_{q=0,\cdots,\infty}$ designate the multi-dimensional approximating polynomial functions forming an orthogonal basis with respect to the joint PDF $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t) = \mathrm{p}^{\mathrm{f}}(\mathrm{x}_{1,t}, \mathrm{x}_{2,t}, \cdots, \mathrm{x}_{n,t})$. The choice for the basis functions may depend on the type of random variable functions (Xiu and Karniadakis, 2002).[4] Since the control vector $\mathbf{x}_t^{\mathrm{f}}$ is assumed to follow a Gaussian PDF $\mathrm{p}^{\mathrm{f}}(\mathbf{x}_t)$ within the framework of Kalman filtering and in particular in EnKF, the surrogate model of the observation operator $\mathcal{G}_{pc,t}$ is built upon the basis of the Hermite polynomials (Ghanem and Spanos, 1991). Stated differently, the Hermite polynomials form the optimal basis for random variables following multi-variate Gaussian PDF. Note that the model outputs $\mathbf{y}_t^{\mathrm{f}}$ are represented in terms of the same random event ω as the model inputs $\mathbf{x}_t^{\mathrm{f}}$, since the uncertainty in the model outputs is assumed to be mainly due to the uncertainty in the ROS model parameters in the context of parameter estimation.

▷ **Truncated expression.** In practice, a truncated expansion of Eq. (7.3) is used; there are several ways of constructing the approximation space. The most common choice is to constrain the number of terms N_{pc} in the PC expansion by the number of control parameters n and by the maximum order of the polynomial basis Q_{po} such that:

$$N_{pc} = \frac{(n + Q_{po})!}{(n!\; Q_{po}!)}. \tag{7.4}$$

This choice of N_{pc} ensures that the PC approximation is of highest order Q_{po}. Note that Q_{po} is a user-defined quantity that must be chosen carefully, according to the model non-linearity, in order to obtain an accurate representation of the model outputs $\mathbf{y}_t^{\mathrm{f}}$ with a high-order convergence rate. Theoretically, $Q_{po} = 1$ (i.e., only two terms for $n = 1$ corresponding to the mean and STD of the control variable) is enough to approximate exactly a Gaussian random variable. Note also that N_{pc} rapidly grows with n and Q_{po}, implying that a balance between accuracy and computational cost must be found. For instance, if $n = 2$ and $Q_{po} = 2$, there

[4] For non-Gaussian PDF, this stochastic approach is referred to as *generalized polynomial chaos* (gPC) in literature.

are $N_{pc} = 6$ terms retained in the PC expansion as illustrated in Fig. 7.31. Using this formalism, the surrogate model $\mathcal{G}_{pc,t}$ can be formulated as follows:

$$\mathbf{y}_t^{\mathrm{f}} \cong \mathcal{G}_{pc,t}\Big(\mathbf{x}_t^{\mathrm{f}}(\boldsymbol{\zeta})\Big) = \sum_{q=0}^{N_{pc}} \hat{\mathbf{y}}_q \, \varphi_q(\boldsymbol{\zeta}), \tag{7.5}$$

where the unknowns are the following time-dependent vectors:

$$\hat{\mathbf{y}}_q \equiv \hat{\mathbf{y}}_q(t) = \Big[(\hat{x}_1, \hat{y}_1)_q, \ldots, (\hat{x}_{N_{fr}^{\mathrm{o}}}, \hat{y}_{N_{fr}^{\mathrm{o}}})_q\Big]_t,$$

q varying between 1 and N_{pc}, with N_{fr}^{o} the number of markers along the observed fire front at time t. Note that the size of the q-th vector $\hat{\mathbf{y}}_q$ is $2N_{fr}^{\mathrm{o}}$ (each marker location being represented with both the x- and y-coordinate on the horizontal plane) and thereby, the computation of $(2\,N_{fr}^{\mathrm{o}}\,N_{pc})$ coefficients (also referred to as the *PC modes*) is necessary to build the surrogate model $\mathcal{G}_{pc,t}$.

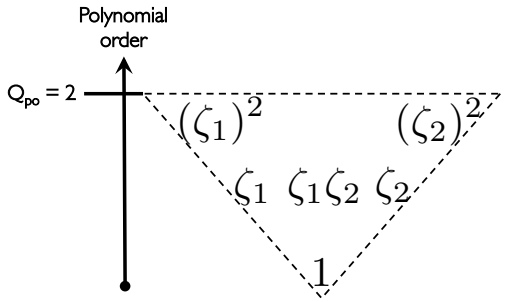


Figure 7.31: *Schematic of the $N_{pc} = 6$ terms retained in the truncated PC expansion for $n = 2$ ($\boldsymbol{\zeta} = [\zeta_1, \zeta_2]$) and $Q_{po} = 2$.*

▷ **Calculation of the PC modes.** Due to the orthogonality of the PC basis, it can be shown that the q-th PC coefficients $\hat{\mathbf{y}}_q$ are given by:

$$\hat{\mathbf{y}}_q = \frac{\mathbb{E}\Big[\mathcal{G}_{pc}(\mathbf{x}_t^{\mathrm{f}})\,\varphi_q(\boldsymbol{\zeta})\Big]}{\mathbb{E}\Big[\varphi_q(\boldsymbol{\zeta})^2\Big]}, \tag{7.6}$$

where:

- $\mathbb{E}[\,\cdot\,]$ refers to the expectation operator satisfying $\mathbb{E}[\varphi_q(\boldsymbol{\zeta})\,\varphi_l(\boldsymbol{\zeta})] = 0$ if $q \neq l$, with the following definition for the inner product:

$$\mathbb{E}\Big[\varphi_q(\boldsymbol{\zeta})\,\varphi_l(\boldsymbol{\zeta})\Big] = \int_{\mathbb{R}^n} \varphi_q(\boldsymbol{\zeta})\,\varphi_l(\boldsymbol{\zeta})\,\mathrm{p}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta} = \delta_{ql}\,\Big[\varphi_q^2\Big], \tag{7.7}$$

  with δ_{ql} the Kronecker delta-function;

- $\mathbb{E}[\varphi_q(\boldsymbol{\zeta})^2]$ is a normalization factor equal to 1 if the basis is constructed orthonormal;

- $\mathbb{E}[\mathcal{G}_{pc,t}(\mathbf{x}_t^{\mathrm{f}})\,\varphi_q(\mathbf{x}_t^{\mathrm{f}})]$ is computed using a Gauss-Hermite quadrature rule, with $[\mathbf{x}_t^{\mathrm{f},(j)}]$ (j varying between 1 and N_{quad}) the quadrature roots vector of size $(N_{quad})^n$ constrained by the maximum order of the polynomial basis Q_{po} such that $2Q_{po} < 2(N_{quad}-1)$. Thus, this term is computed as follows:

$$\begin{aligned}\mathbb{E}\Big[\mathcal{G}_{pc,t}(\mathbf{x}_t^{\mathrm{f}})\,\varphi_q(\boldsymbol{\zeta})\Big] &= \int_{\mathbb{R}^n} \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f}})\,\varphi_q(\boldsymbol{\zeta})\,\mathrm{dp}(\boldsymbol{\zeta}) \\ &\cong \sum_{j=1}^{(N_{quad})^n} \mathcal{G}_t(\mathbf{x}_t^{\mathrm{f},(j)})\,\varphi_q(\boldsymbol{\zeta})\,w^{(j)},\end{aligned} \tag{7.8}$$

where $\mathbf{y}_t^{\mathrm{f},(j)} = \mathcal{G}(\mathbf{x}_t^{\mathrm{f},(j)})$ corresponds to the FIREFLY forward model integration evaluated at the j-th quadrature root $\mathbf{x}_t^{\mathrm{f},(j)}$ with its associated weight $w^{(j)}$, and where $\varphi_q(\zeta)$ corresponds to the q-th multi-dimensional basis function formulated as tensor products of one-dimensional polynomial bases:

$$\varphi_q \equiv \varphi_q(\boldsymbol{\zeta}) = \prod_{l=1}^{n} \varphi_{i(l)}^{1D}\Big(\zeta_l\Big), \tag{7.9}$$

with $\varphi_{i(l)}^{1D}$ the one-dimensional polynomial basis and its multi-index $i(l)$ varying between 0 and Q_{po} to determine the proper term in the multi-variable space illustrated in Fig. 7.31.

The mean values of the marker locations along the simulated fire fronts $\mathbf{y}_t^{\mathrm{f}}$ are provided by the first mode of the PC expansion $\hat{\mathbf{y}}_0$, while the covariance between one marker location $y_{i,t}^{\mathrm{f}}$ (indexed by $i = 1, \cdots, N_{fr}^{\mathrm{o}}$) and the control parameter $\mathrm{x}_{j,t}^{\mathrm{f}}$ (indexed by $j = 1, \cdots, n$) is given by:

$$\mathrm{cov}(y_{i,t}^{\mathrm{f}}, \mathrm{x}_{j,t}^{\mathrm{f}}) = \sum_{q=1}^{N_{pc}} \sum_{l=1}^{N_{pc}} \hat{y}_{i,q}\,\hat{\mathrm{x}}_{j,l} \int_{\mathbb{R}^n} \varphi_q(\boldsymbol{\zeta})\,\varphi_l(\boldsymbol{\zeta})\,\mathrm{p}(\boldsymbol{\zeta})\mathrm{d}\boldsymbol{\zeta},$$

$$\Rightarrow \begin{cases} \mathrm{cov}(x_{i,t}^{\mathrm{f}}, \mathrm{x}_{j,t}^{\mathrm{f}}) = \displaystyle\sum_{q=1}^{N_{pc}} \hat{x}_{i,q}\,\hat{\mathrm{x}}_{j,q}\,\mathbb{E}[\varphi_q(\boldsymbol{\zeta})^2] \\ \mathrm{cov}(y_{i,t}^{\mathrm{f}}, \mathrm{x}_{j,t}^{\mathrm{f}}) = \displaystyle\sum_{q=1}^{N_{pc}} \hat{y}_{i,q}\,\hat{\mathrm{x}}_{j,q}\,\mathbb{E}[\varphi_q(\boldsymbol{\zeta})^2], \end{cases} \tag{7.10}$$

assuming the control parameters and the front marker locations can be expressed in the same PC basis.

Based on this formulation, the construction of the surrogate model $\mathcal{G}_{pc,t}$ over the assimilation window $[t-1, t]$ requires a limited number of $(N_{quad})^n$ forward model integrations (see the first step in Fig. 7.32). The polynomial approximation $\mathcal{G}_{pc,t}$

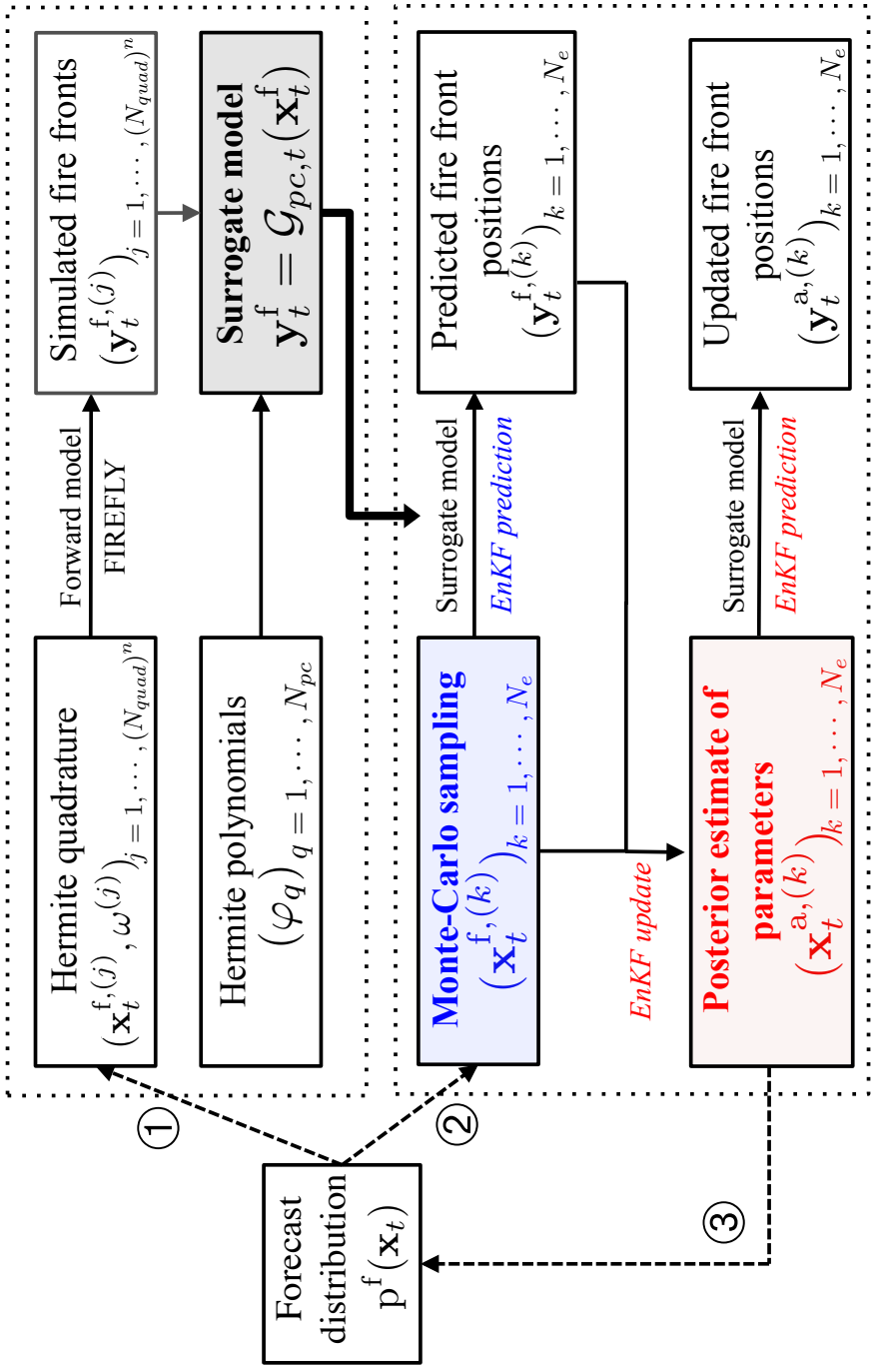


Figure 7.32: *Flowchart of the hybrid PC-based EnKF algorithm (PC-EnKF) during the assimilation cycle $[t-1,t]$ decomposed into 3 steps: (1) Construction of the PC expansion of the generalized observation operator; (2) EnKF prediction and update for the assimilation cycle $[t-1,t]$; and (3) Parameter evolution to the next assimilation cycle $[t,t+1]$..*

calculated in Eq. (7.5) is then used in the prediction step of the EnKF algorithm instead of the observation operator $\mathcal{G}_t$, to compute the predictions of the time-evolving fire front locations $\mathbf{y}^{\mathrm{f}}_{pc,t}$ for a large number of ensemble members N_e (see the second step in Fig. 7.32). This ensemble of forecasts is used to accurately estimate the error covariance matrices $\mathbf{C_{xy}} \in \mathbb{R}^{n\times 2N^{\mathrm{o}}_{fr}}$ and $\mathbf{C_{yy}} \in \mathbb{R}^{2N^{\mathrm{o}}_{fr}\times 2N^{\mathrm{o}}_{fr}}$ that are required in the Kalman update Eq. (6.15), Chapter 6. Thus, the EnKF update can be performed with reliable covariance matrices at a reduced computational cost, compared to the standard EnKF algorithm based on a MC sampling. This approach leads to analysis estimates of the control parameters $\mathbf{x}^{\mathrm{a},(k)}_t$ and to accurate PDF of the fire front locations $\mathbf{y}^{\mathrm{a},(k)}_t$ $(k = 1, \cdots, N_e)$ using the same surrogate model as for the forecast estimates.

In order to reduce the computational cost of the EnKF algorithm, a surrogate model based on a PC expansion is used in place of the forward model, i.e., the FIREFLY regional-scale wildfire spread model. The performance of the resulting PC-EnKF algorithm is assessed on synthetically-generated fire spread cases based on preliminary work presented in Rochoux et al. (2012, CTR) as well as on the controlled grassland fire experiment.

↪ Simple examples

▷ **Convergence of the PC-EnKF algorithm.** As for the standard EnKF (see the first test case named P-OSSE-ANISO in Section 7.2.1), the PC-EnKF algorithm is applied to correct the proportionality coefficient P [s^{-1}], linearly related to the ROS Γ. In this test case, the fire is ignited as a circular front at $(x, y) = (100\ \mathrm{m}, 100\ \mathrm{m})$ and the fuel layer depth δ_v is randomly-distributed over the $200\ \mathrm{m} \times 200\ \mathrm{m}$ computational domain; the anisotropic fire spread is simulated during 50 s, at which time the EnKF update is performed. The true value of the control parameter is $\mathbf{x}^{\mathrm{t}} = P^{\mathrm{t}} = 0.4\ \mathrm{s}^{-1}$.

A PC approximation (with a polynomial order $Q_{po} = 4$ and subsequently a quadrature order $N_{quad} = 5$) is used to build the model response surface to the control parameter $\mathbf{x} = P$ corresponding to a forecast taken as $\mathbf{x}^{\mathrm{f}} = 0.2\ \mathrm{s}^{-1}$ and $\sigma^{\mathrm{f}} = 0.05\ \mathrm{s}^{-1}$.

- **Sensitivity to sampling errors.** Convergence properties of the EnKF-based analysis estimates are studied in Fig. 7.33(a) with respect to the number of ensemble members N_e for a fixed observation error STD $\sigma^{\mathrm{o}} = 2$ m and for one assimilation cycle. Since there is no analytical solution of the problem, the convergence of the EnKF is assumed to be achieved if the mean value of the control parameter and its STD remain constant when increasing N_e. The performance of the PC-EnKF algorithm is compared to that of the standard EnKF algorithm (black squares) for different PC polynomial orders, $Q_{po} = 2$ (orange triangled-dashed line) and $Q_{po} = 4$ (red circled-dashed line). Figure 7.33(a) shows that in the present configuration, the EnKF algorithm

converges for a minimum of $N_e = 48$ members (meaning that Firefly is integrated 48 times to produce 48 fire front trajectories associated with each realization of the control parameter). It is shown that the PC-EnKF algorithm provides a comparable result as EnKF (in terms of mean and STD) above $N_e = 40$ members for a polynomial order $Q_{po} = 4$. However, the results achieved with PC-EnKF are obtained for a lower number of Firefly time-integrations (i.e., 5 Firefly model integrations only since $N_{quad} = 5$ quadrature points are used to build the model surface response $\mathcal{G}_{pc}$) than the standard EnKF, while considering the same number of members N_e to generate the forecast/analysis estimates. Thus, the PC-EnKF algorithm provides a solution that reproduces the converged solution of the EnKF for a computational cost that is reduced by a factor of at least 8. This implies that for more complex fire spread cases where more members are required to track spatial variations in wind and vegetation conditions, the PC-EnKF algorithm appears as a promising alternative to obtain accurate simulations of fire spread at a reasonable computational cost. Additionally, the PC-EnKF algorithm provides a mean estimate that is less fluctuating than the EnKF algorithm, with a slightly reduced scatter for low values of N_e, indicating that the PC-EnKF strategy requires less ensemble members N_e to reach convergence.

Figure 7.33(a) also illustrates the sensitivity of the PC-EnKF-based analysis to the choice of the PC polynomial order Q_{po} for a varying number of ensemble members N_e. While $Q_{po} = 2$ (i.e., $N_{quad} = 3$) provides a reasonable approximation of the mean analysis estimate when considering the standard EnKF as reference, $Q_{po} = 4$ (i.e., $N_{quad} = 5$) leads to a more accurate estimate without loss of accuracy. Even though the fire front marker locations exhibit approximate Gaussian PDF and in theory $n = 1$ is sufficient to characterize their distributions, a high polynomial order is required in this case since the true value ($P^{\mathrm{t}} = 0.4$ s^{-1}) is not in the zone of high probability occurrence of the forecast estimates ($P^{\mathrm{f}} = 0.2$ s^{-1} with $\sigma^{\mathrm{f}} = 0.05$ s^{-1}). Indeed, the true fire front locations are at the tail of the forecast PDF, which makes the estimation of the fire front locations more difficult. This difficulty shows the ability of the PC-EnKF procedure to retrieve accurate estimates of the fire spread at a low computational cost and without loss of accuracy, even though prior information is very uncertain.

- **Example of PC-based surface response.** Figure 7.33(b) provides a comparison in the observation space between the observed fire front and the forecast/analysis estimates obtained through the PC-EnKF algorithm for an observation error STD $\sigma^{\mathrm{o}} = 2$ m, a PC polynomial order $Q_{po} = 4$ and a number of ensemble members $N_e = 1,000$. As expected, the analysis estimates provide a more accurate approximation of the observed fire front location than the forecast estimates.

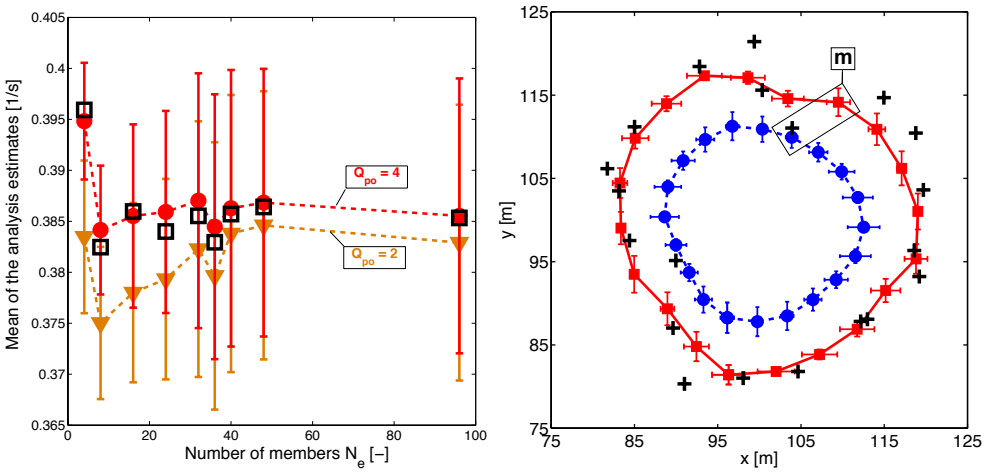
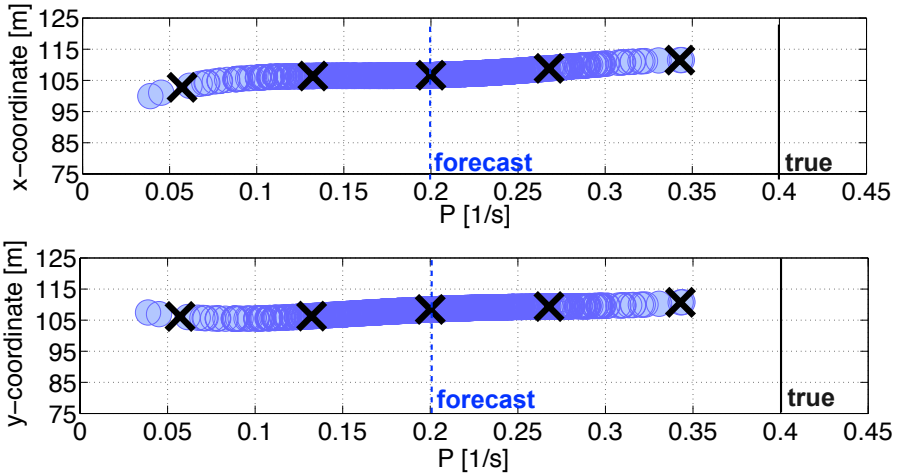


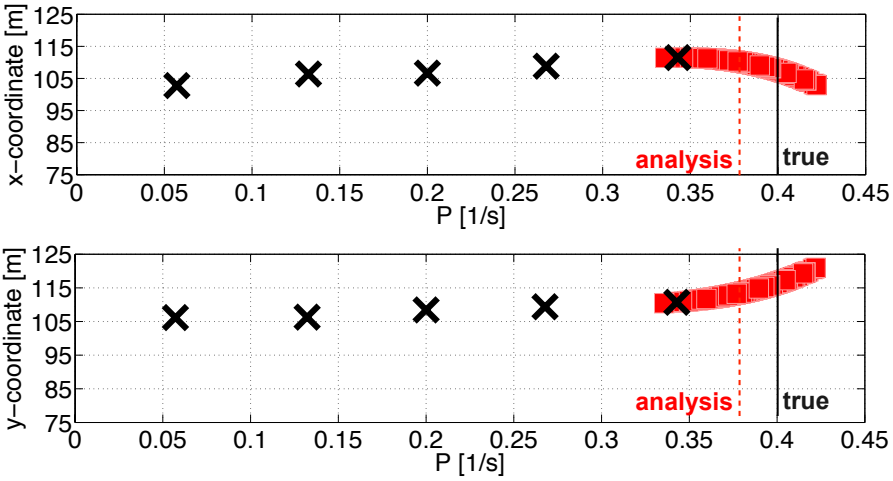
Figure 7.33: *Comparison of the analysis estimates between EnKF and PC-EnKF in the parameter/observation space; one assimilation cycle (P-OSSE-ANISO); fixed observation error STD $\sigma^{\mathrm{o}} = 2$ m. (a) Convergence of the mean analysis estimates of the control parameter P $[\mathrm{s}^{-1}]$ with respect to the number of members N_e for varying polynomial orders Q_{po}. The orange triangled-dashed line corresponds to $Q_{po} = 2$; the red circled-dashed line corresponds to $Q_{po} = 4$. Vertical error bars correspond to the associated error STD. EnKF results (black squares) are indicated as reference. (b) Comparison of fire front locations using the PC-EnKF approach with an EnKF update at 50 s for $Q_{po} = 4$ and $N_e = 1,000$; all fronts correspond to time 50 s. Black crosses correspond to observations; the blue circled-dashed line corresponds to the mean forecast estimate of the fire front and the red squared-solid line corresponds to the analysis counterpart. Horizontal and vertical error bars correspond to the associated error STD along the x- and y-directions, respectively.*

To offer insight into the main ideas underlying the PC-EnKF algorithm, Fig. 7.34 illustrates the mapping between control space and observation space for one marker of the fireline; its position on the forecast/analysis fireline is indicated (indexed by m) in Fig. 7.33(b). The variations in the x- and y-coordinates of this marker are represented with respect to variations in the control parameter P: black crosses indicate the simulated marker positions associated with the $N_{quad} = 5$ quadrature roots (i.e., FIREFLY model integrations) corresponding to the first step of the PC-EnKF algorithm; and blue circles indicate the forecast estimates obtained through the surrogate model evaluation combined with MC sampling ($N_e = 1,000$) corresponding to the second step of the PC-EnKF algorithm (see Fig. 7.32). These fire front estimates are associated with the forecast control parameter $P^{\mathrm{f}} = 0.20$ s^{-1} and its error STD $\sigma^{\mathrm{f}} = 0.05$ s^{-1}. In contrast, red squares are produced by the EnKF update applied for any of the 1,000 ensemble members, they correspond to the analysis estimates related to $P^{\mathrm{a}} = 0.38$ s^{-1} and $\sigma^{\mathrm{a}} = 0.01$ s^{-1}. The scatter of the ensemble is significantly reduced in the analysis, around the

true value $P^{\mathrm{t}} = 0.40\ \mathrm{s}^{-1}$, highlighting the uncertainty reduction achieved through data assimilation. This result is obtained all along the fireline as confirmed by the PDF of the front marker x- and y-coordinates in Fig. 7.35: the location of high probability occurrence (the red zone) is continuous and thinner in the analysis than in the forecast for all observed front markers.



(a) Forecast estimates.



(b) Analysis estimates.

Figure 7.34: *Model surface response (or surrogate model) of the x- and y-coordinates of the front marker indexed by m on the fireline in Fig. 7.33(b), with respect to the control parameter P $[s^{-1}]$; one assimilation cycle with an analysis update at 50 s (P-OSSE-ANISO). Black crosses correspond to quadrature roots (forward model integrations with* Firefly*); blue circles correspond to (a) forecast estimates, and red squares to (b) analysis estimates obtained through the PC-EnKF algorithm at time 50 s. The vertical solid line indicates the true value $P^{\mathrm{t}} = 0.4\ s^{-1}$; the vertical dashed lines indicate the mean forecast and analysis estimates of P.*

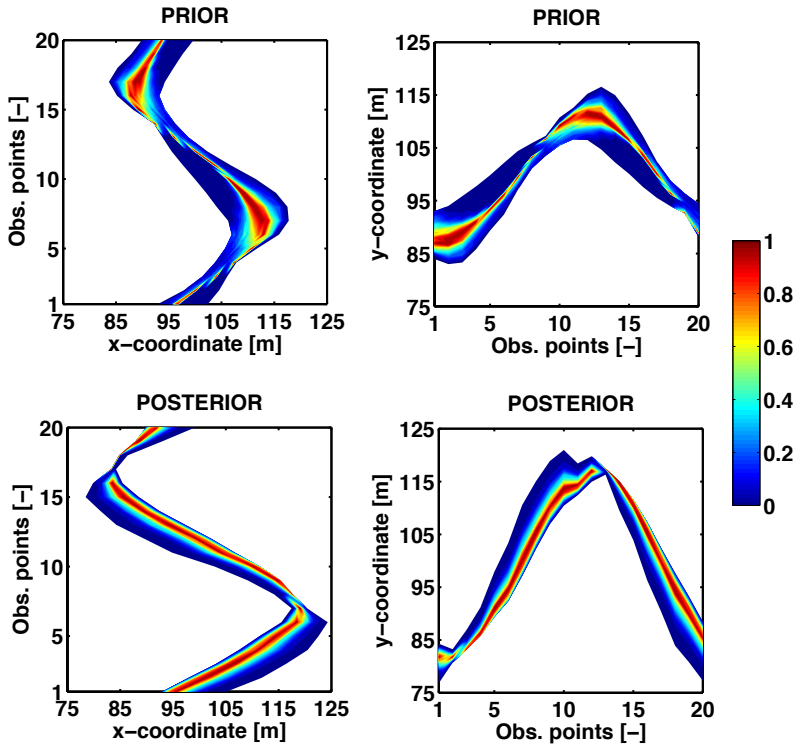


Figure 7.35: *PDF in observation space for the estimation of P using PC-EnKF for $\sigma^{\mathrm{o}} = 2$ m and $N_e = 1,000$; one assimilation cycle with an analysis update at 50 s (P-OSSE-ANISO). PDF of x- (left) and y- (right) coordinates of each fire front marker location for the forecast (top) and the analysis (bottom). The forecast corresponds to $P^{\mathrm{f}} = 0.20$ s^{-1} with its error STD $\sigma^{\mathrm{f}} = 0.05$ s^{-1}; the analysis corresponds to $P^{\mathrm{a}} = 0.38$ s^{-1} with its error STD $\sigma^{\mathrm{a}} = 0.01$ s^{-1} in parameter space.*

- **Sensitivity to observation errors.** For verification purposes on the behavior of the PC-EnKF algorithm, Fig. 7.36 examines the influence of the observation error on the performance of EnKF and PC-EnKF (the EnKF algorithm is used as reference). Statistics (in terms of mean value and STD) of the analysis obtained for $N_e = 48$ members over one assimilation cycle, at time $t = 50$ s, are presented as a function of the magnitude of the observation errors measured by σ^{o} (up to $\sigma^{\mathrm{o}} = 30$ m); vertical bars give a graphical representation of the magnitude of the STD within the analysis ensemble. The results show the consistency of PC-EnKF with EnKF in retrieving realistic values for the control parameter, even though the observation error is significant. When the observation error STD σ^{o} is small, the PC-EnKF algorithm successfully drives the analysis ensemble towards the true value of the parameter $P^{\mathrm{t}} = 0.4$ s^{-1}; the resulting analysis exhibits a much reduced scatter by at least a factor 4 in comparison to the forecast STD $\sigma^{\mathrm{f}} = 0.05$ s^{-1}. In contrast, when σ^{o} is large, the PC-EnKF algorithm has reduced effects and the analysis ensemble remains close to the forecast ensemble (the analysis

STD is similar to the forecast STD $\sigma^{\mathrm{f}} = 0.05$ s^{-1}). For intermediate values of σ^{o}, the PC-EnKF algorithm produces optimized analyses lying between forecast and observation; as expected, the more accurate the observations, the more certain the analysis for a given forecast error.

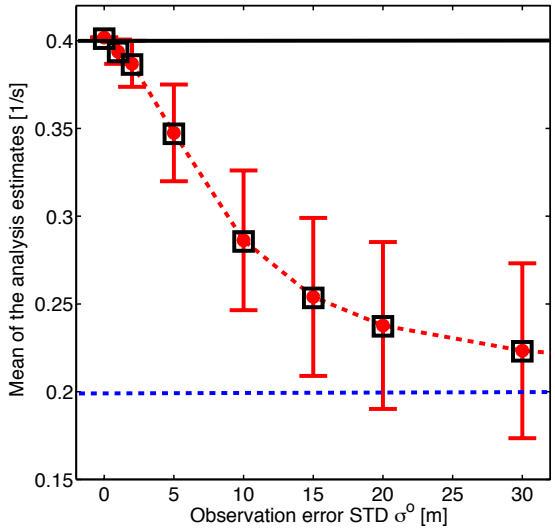
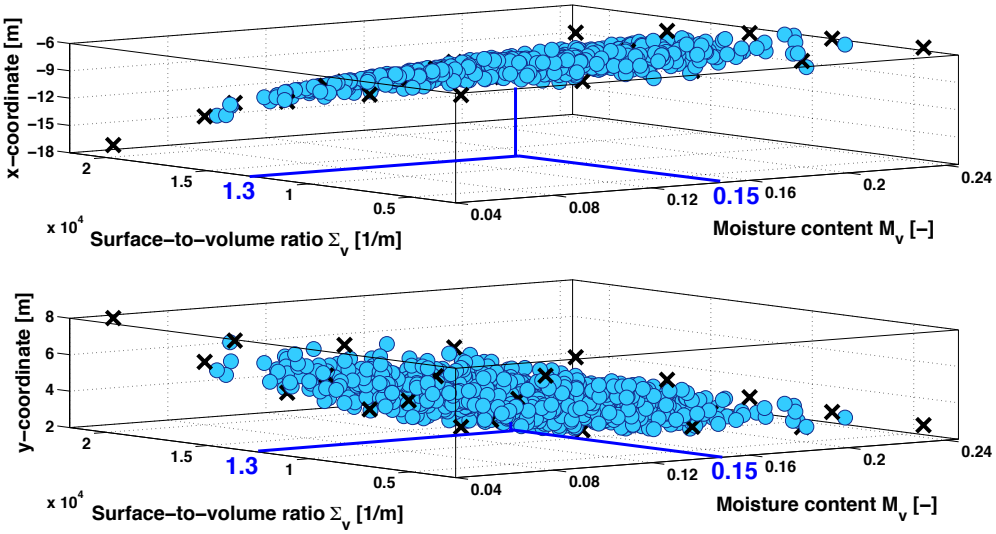


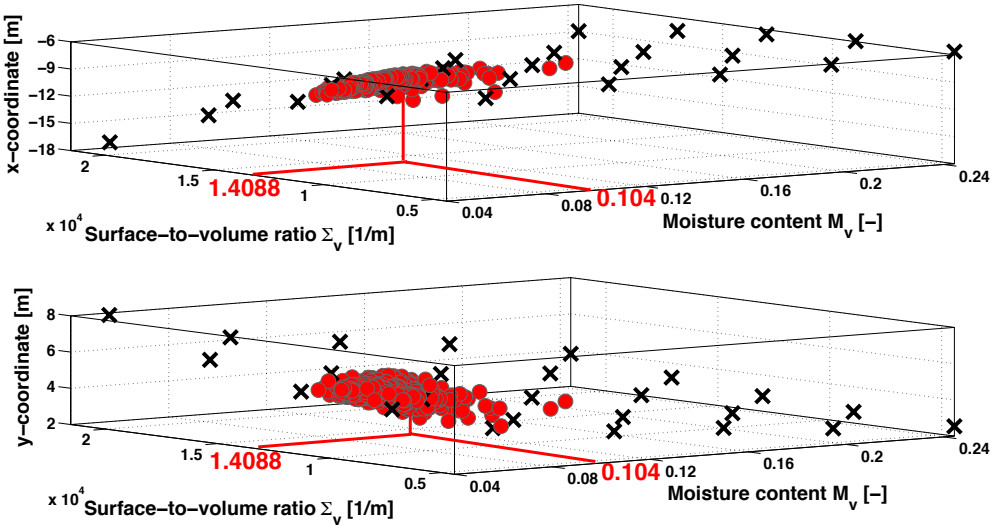
Figure 7.36: *Mean and STD of the analysis estimates of the control parameter P [s^{-1}] as a function of the observation error STD σ^{o} for a fixed number of members, $N_e = 48$; comparison between EnKF and PC-EnKF; one assimilation cycle (P-OSSE-ANISO). The black solid line corresponds to the true value 0.4 s^{-1}; the blue dashed line corresponds to the mean value of the forecast 0.2 s^{-1}; the red circled-dashed line corresponds to the PC-EnKF-based mean analysis estimate; and vertical error bars correspond to the associated error STD. EnKF results (black squares) are indicated as reference.*

▷ **Wind-aided propagation.** The application of the PC-EnKF algorithm is illustrated for a wind-aided grassland fire propagation during 50 s and directed northward with $u_w = 0.8$ m/s and $\alpha_w = 0°$ (referred to as P-OSSE-ISO-WIND). The control parameters are the fuel moisture content M_v and the fuel particle surface-area-to-volume ratio Σ_v such that $\mathbf{x} = [M_v, \Sigma_v]$ with $n = 2$. The observed fire front is synthetically-generated at time $t = 50$ s by adding an artificial noise $\sigma^{\mathrm{o}} = 1$ m to the true trajectory obtained for $M_v^{\mathrm{t}} = 10$ % and $\Sigma_v^{\mathrm{t}} = 14,500$ 1/m. The mean values of the forecast estimates are 15 % and $13,000$ 1/m with 3.3 % and $3,000$ 1/m error STD, respectively. A PC approximation (with a polynomial order $Q_{po} = 4$ and a quadrature order $N_{quad} = 5$) is used to build the model response surface to the two control parameters. Similarly to the 1-parameter estimation case presented in Fig. 7.34, Fig. 7.37(a) illustrates the observation operator mapping onto the space spanned by the control parameters (for x- and y-coordinates of the front marker positions, top and bottom panels, respectively). Since the size of the control vector is $n = 2$, this mapping now features a two-dimensional surface, indicating the Firefly model response to varying values for M_v and Σ_v. Results are shown for the fifth marker located on the fireline starting from the westward flank

(see Fig. 7.38); black crosses represent the simulated front positions associated with the 25 quadrature roots (i.e., 5 quadrature roots for each control parameter).



(a) Forecast estimates.



(b) Analysis estimates.

Figure 7.37: *Model surface response (or surrogate model) of the front marker coordinates with respect to the control vector* $\mathbf{x} = [M_v, \Sigma_v]$, $n = 2$; *the true control vector is* $\mathbf{x}^{\mathrm{t}} = [10\ \%, 14,500\ 1/\mathrm{m}]$ *(P-OSSE-ISO-WIND). Forecast and analysis estimates (colored circles) of the* x- *(top) and* y-*coordinates (bottom) of the fire front positions mapped onto the PC-based model surface response (black crosses).*

The surrogate model of the observation operator $\mathcal{G}_{pc}$ is then used with $N_e = 1,000$ members in the ensemble to evaluate the forecast front positions shown in

Fig. 7.37(a) for the 1,000 forecast estimates; the mean value of the forecast estimates $[15\ \%, 13,000\ 1/\text{m}]$ is indicated. Note that the ensemble members evaluated with the surrogate model are contained within the surface response constructed with Firefly, meaning that the PC decomposition $\mathcal{G}_{pc}$ properly approximates the observation operator (including Firefly model integration for parameter estimation).

Figure 7.37(b) shows the 1,000 analysis estimates of the control parameters obtained using the PC-EnKF update; the mean value of the analysis estimates is $[10.4\ \%,\ 14,088\ 1/\text{m}]$ that is more consistent with the true value of the control vector $\mathbf{x}^{\text{t}} = [10\ \%,\ 14,500\ 1/\text{m}]$. The scatter of the analysis estimates is also significantly reduced compared to that of the forecast ensemble. For this example, the computational time is significantly reduced compared to the classical EnKF parameter estimation since only 25 forward model integrations were used to perform the EnKF update. In contrast, nearly 40 members were necessary on the previously-mentioned wind-aided test case (see Table 7.4), while it could not provide converged statistics of the fire front position due to the limited number of ensemble members (for this purpose, 1,000 members are used in the PC-EnKF algorithm for the evaluation of error statistics).

Figure 7.38 compares the simulated fire front using the mean value of the analysis estimates with the most probable fire front derived from the PDF of the analysis estimates. These fronts are found to feature a similar topology, indicating that the Gaussian assumption on the error statistics made in EnKF does not degrade the quality of the analysis for the present tests.

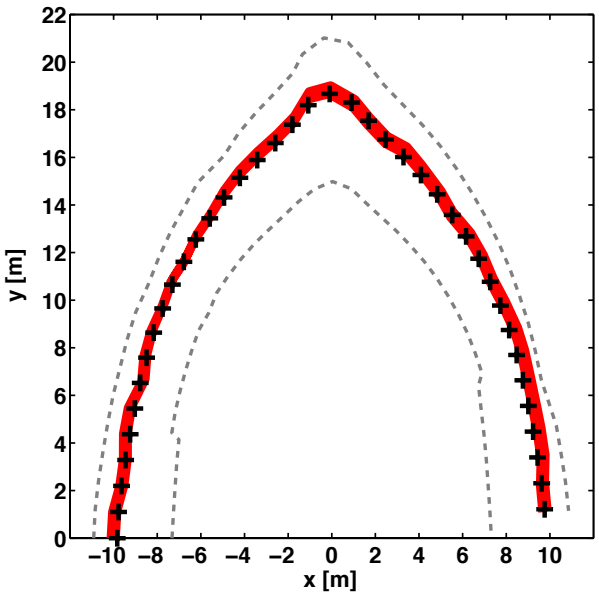


Figure 7.38: *Comparison of the mean analysis estimate (black crosses) with the most probable front (red zone) at 50 s (P-OSSE-ISO-WIND); the scatter of the ensemble (measured by STD) is delimited by gray dashed lines.*

↪ Application to the controlled grassland burning

The PC-EnKF algorithm is now applied to the controlled grassland burning experiment over one assimilation cycle $[50; 78$ s$]$ with an EnKF update at time $t_1 = 78$ s (analysis mode) as well as over a forecast time period $[78; 106$ s$]$ with an EnKF forecast at time $t_2 = 106$ s (forecast mode); the initial condition at $t_0 = 50$ s is taken as the observed fire front at t_0. The control space includes two biomass fuel parameters, the moisture content M_v and the particle surface-area-to-volume ratio Σ_v such that $\mathbf{x} = [M_v, \Sigma_v]$ with $n = 2$. Wind conditions are assumed constant over the simulated fire duration from $t_0 = 50$ s to $t_2 = 106$ s. Observations at times $t_1 = 78$ s and $t_2 = 106$ s are discretized with $N_{fr}^{\text{o}} = 40$ front markers with an observation error STD $\sigma^{\text{o}} = 5$ cm (consistently with previous EKF and EnKF studies).

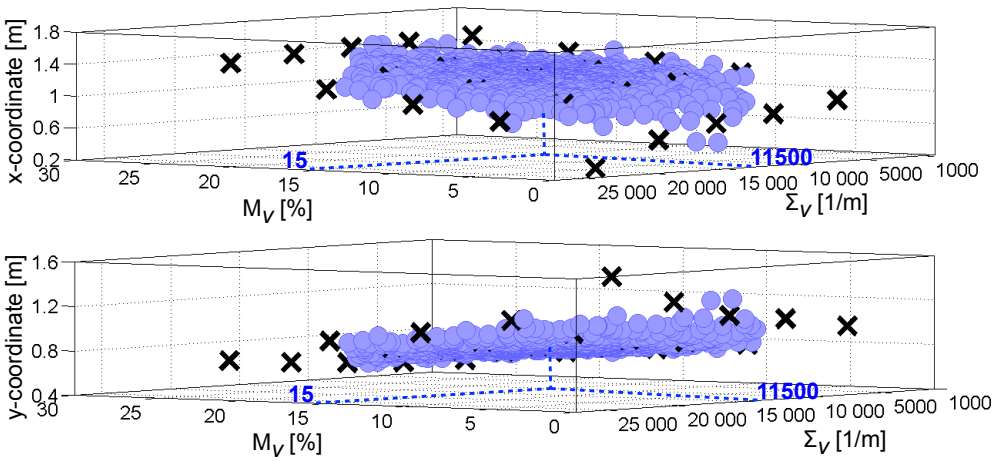
A PC approximation (with a polynomial order $Q_{po} = 4$ and subsequently a quadrature order $N_{quad} = 5$) is used to build the model response surface to the two control parameters M_v and Σ_v; this response surface is shown for one particular simulated front marker in Fig. 7.39(a) using black crosses. A forecast ensemble of $N_e = 1,000$ members (blue circles) is generated at no cost using the PC-based surrogate model assuming uncertainties in M_v and Σ_v; the forecast estimates of these control parameters are described in Table 7.8 along with the associated STD. Note that the blue circles are contained within the surface response described by the black crosses that represent the $(N_{quad})^2 = 25$ FIREFLY model integrations performed to build the PC-based surface response. The forecast ensemble is corrected by assimilating the fire front at time $t_1 = 78$ s.

Table 7.8: *PC-EnKF-based experiment for the controlled grassland burning experiment: error statistics of the forecast and analysis ensemble estimates for* $\mathbf{x} = [M_v, \Sigma_v]$, $n = 2$. *The number of* FIREFLY *model integrations is also presented as indicator of the computational cost.*

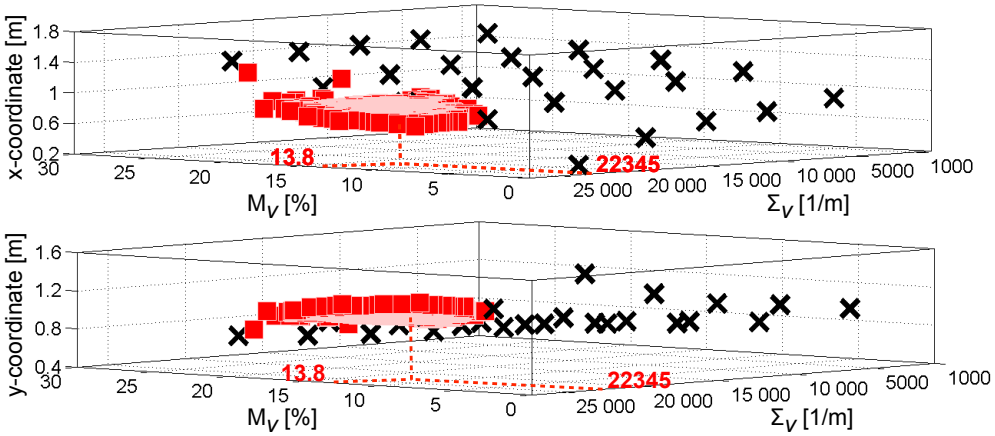
	Cost	Forecast control vector		Analysis control vector	
		Ens. mean	Ens. STD	Ens. mean	Ens. STD
PC-EnKF	25	0.150 11500 1/m	0.040 3000 1/m	0.138 22583 1/m	0.014 1157 1/m
EnKF	1000	0.150 11500 1/m	0.040 3000 1/m	0.135 22345 1/m	0.014 1170 1/m

▷ **Analysis mode.** The forecast ensemble is corrected by assimilating the fire front at time $t_1 = 78$ s. The comparison between the observations (black crosses), the forecast estimates (blue circled-dashed line) and the PC-EnKF-based analysis estimates (red squared-solid line) obtained at time $t_1 = 78$ s are presented in Fig. 7.40(a). The forecast trajectory represents the ensemble mean of the surrogate model simulations obtained without data assimilation (i.e., using the forecast

estimates of the control parameters), while the analysis trajectory derives from an EnKF update at t_1 using the analysis estimates in the surrogate model integrations.



(a) Forecast estimates.

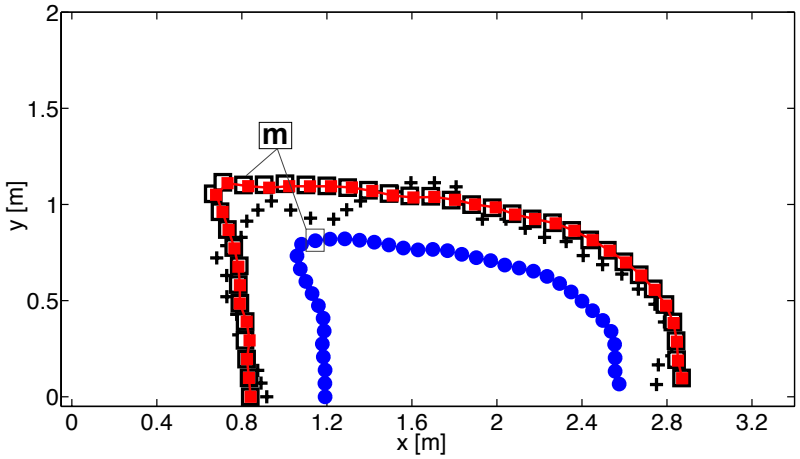


(b) Analysis estimates.

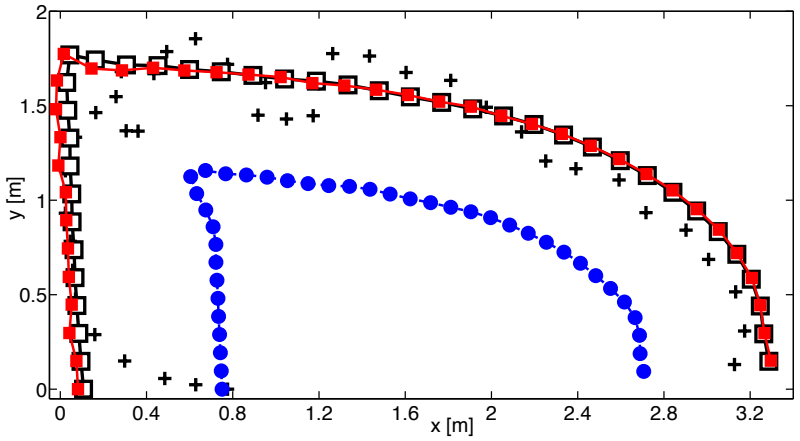
Figure 7.39: *Model surface response (or surrogate model) of the x- and y-coordinates of the front marker indexed by m on the fireline (see Fig. 7.40) with respect to the control vector $\mathbf{x} = [M_v, \Sigma_v]$ ($n = 2$). Black crosses correspond to quadrature roots (FIREFLY). (a) Forecast estimates (blue circles), and (b) analysis estimates (red squares) of the x- (top) and y-coordinates (bottom) of the fire front positions are mapped onto the PC-based model surface response.*

It is found that the PC-EnKF strategy allows to significantly decrease the distance between the observations and the simulated fronts with the same level of accuracy as the standard EnKF algorithm (the PC-EnKF algorithm provides similar analysis mean and STD, see Table 7.8). As illustrated in Fig. 7.39(b), the uncertainty in the fire front positions is significantly reduced in comparison to the forecast since the STD related to the analysis estimates is much smaller than that of the forecast

estimates. This indicates that the PC-EnKF algorithm allows reliable statistical information to be retrieved for only 25 FIREFLY model integrations (in contrast, the standard EnKF algorithm requires 1,000 members to correct $n = 2$ control parameters and thereby, $2N_{fr}^{o} = 80$ fire front marker coordinates).



(a) Analysis time, $t_1 = 78$ s.

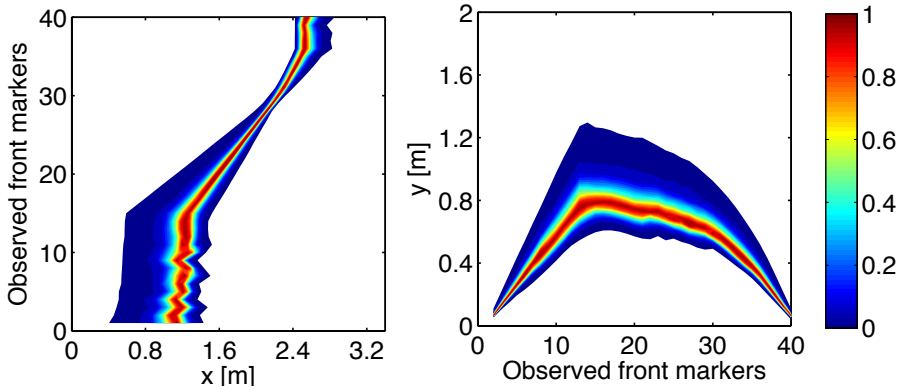


(b) Forecast time, $t_2 = 106$ s.

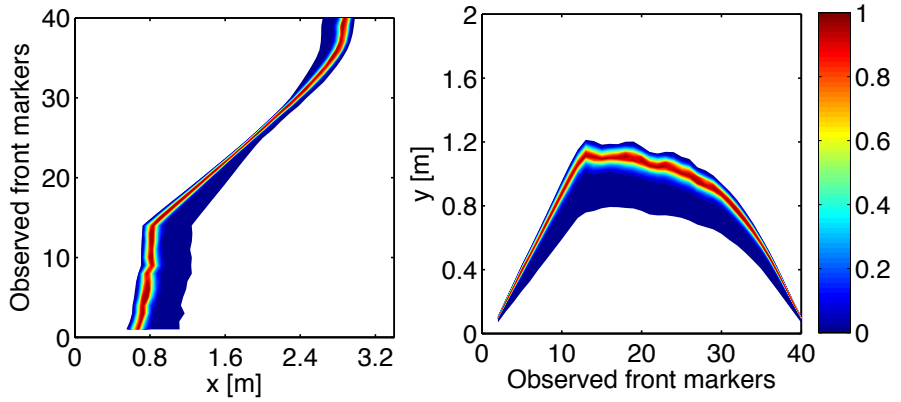
Figure 7.40: *Comparison between simulated and measured fire front positions for the controlled grassland fire experiment: black crosses correspond to observations, the blue circled-dashed line corresponds to the mean forecast estimate constructed through the PC-based surrogate model; the red squared-solid line corresponds to the mean analysis estimate obtained by the PC-EnKF procedure applied at time $t_1 = 78$ s. Black squares correspond to the standard EnKF used as reference.*

Consistently, Fig. 7.41 shows that the support of the analysis PDF is significantly reduced compared to the forecast PDF for the x- and y-coordinates of the $N_{fr}^{o} = 40$ observed front markers. The topology of the PDF along the observed fire front is found to be overall preserved through the EnKF update, implying that the as-

sumption of Gaussian error statistics for the modeling error statistics seems not to deteriorate the performance of the ensemble-based data assimilation algorithms. Some regions of the PDF related to the x-coordinates of the front marker locations (nearby $x = 2$ m) are not sensitive to variations in the fuel moisture content M_v and Σ_v. These regions correspond to the flank of the fire, meaning that the x-coordinates of the surrounding front markers do not vary and the growth of the burning area only induces variations in the y-coordinates.



(a) Forecast PDF with respect to the x- (left) and y- (right) coordinates of the observed fire front markers.



(b) Analysis PDF with respect to the x- (left) and y- (right) coordinates of the observed fire front markers.

Figure 7.41: *Colormap of the PDF of the fire front marker locations (in terms of x and y-coordinates) for the controlled grassland fire experiment at the analysis time $t_1 = 78$ s. (a) PDF related to the ensemble of forecast estimates. (b) PDF related to the ensemble of analysis estimates.*

As discussed for the OSSE test cases, the non-linear response of the observation operator to the control parameters induces a slightly non-Gaussian PDF for the forecast estimates: it is indeed found that the mode of the PDF does not exactly coincide with the mean value. Note that the PDF exhibits a relatively flat tail for

decreasing x- and increasing y-coordinates of the observed fire front markers: this is due to a sharp ROS acceleration when decreasing the fuel moisture content M_v or alternatively, when increasing the fuel particle surface-to-volume ratio.

▷ **Forecast mode.** Figure 7.40(b) compares the fire front position at $t_2 = 106$ s obtained using the forecast estimates (without data assimilation) and the analysis estimates derived from an assimilation update at $t_1 = 78$ s. The PC-EnKF algorithm appears to properly represent the forecast trajectory at $t_2 = 106$ s in comparison to the standard EnKF. This result illustrates that a PDF sampling based on PC (instead of MC in the standard EnKF) can significantly reduce the computational cost of the EnKF prediction/update steps (in terms of number of FIREFLY model integrations that constitute the most time-consuming task in PC-EnKF) and thereby, provide accurate error statistics on the inputs and outputs of FIREFLY. For instance, $N_e = 1,000$ forward model integrations were used in EnKF to accurately represent the error statistics, in contrast to only 25 forward model integrations in PC-EnKF. Thus, the number of FIREFLY model integrations is here divided by a factor of 40.

Additionally, Fig. 7.40(b) shows that the errors in the control parameters do not significantly change in-between the two observation times (i.e., at $t_1 = 78$ s and $t_2 = 106$ s), meaning that an observation time period of 28 s seems appropriate for applying data assimilation (relatively to the temporal variability of the errors in the control vector $\mathbf{x}$).

While the improved accuracy of EnKF-based data-driven simulations is obtained at the expense of heavy computational cost (in the context of multi-parameter estimation, see Section 7.2.1), the PC-EnKF strategy appears as a promising strategy for solving Bayesian filtering problems at a low computational cost that is a requirement of operational frameworks. Future plans include to extend this approach to a larger size n of the estimation problem (case studies were limited to $n = 2$ here) and to investigate how this approach can be further optimized in terms of computational cost to meet operational requirements.

7.3 Contributions of a state estimation strategy

While the previously-presented parameter estimation approaches produced encouraging results and confirmed the value of an EnKF strategy for improved predictions of the fire front location, their application to highly-anisotropic cases in which the fire front is locally deformed (due partly to unknown spatial variations in local biomass fuel properties and temporal variations in wind conditions) showed some limitations. In particular, applying a spatially-uniform correction of the control parameters to the grassland controlled burning experiment only led to a global correction of the fire front location and did not retrieve, locally, a topology of the fire front that is consistent with observations.

For this purpose, the choice of a stand-alone parameter estimation approach is considered questionable for more general wildfire problems in which the vegetation, topographical and possibly meteorological properties exhibit arbitrary spatial variations. The extension to the estimation of spatialized vegetation and wind ROS parameters would be indeed computationally prohibitive in the context of real-time forecast of the fire behavior, since it would dramatically increase the size n of the control vector. Besides, high-resolution distributions for the ROS parameters are not available and thereby, there would be no means of assessing if the analysis values are consistent with the physics of wildfire spread (for validation purposes).

In this context, an extension of the data assimilation strategy is proposed, based on a change from a parameter estimation approach to a state estimation approach (see Fig. 7.42). This change was inspired by data assimilation applications in numerical weather prediction and operational oceanography as well as by previous studies by Beezley and Mandel (2008) and Beezley (2009). In these studies, the control variable is the temperature field that is characterized by a bimodal PDF in the fire region (i.e., burning state or not-burning state). In order to satisfy the Gaussian assumption on the error statistics in EnKF, the idea of morphing from image processing was introduced; however, this choice led to technical difficulties in the EnKF implementation. In the present work, the control variable is the fire front position and its uncertainty is approximated by a Gaussian PDF, which allows for a straightforward application of EnKF. In the following,

- ▷ OSSE tests cases are performed to highlight the key aspects of the proposed state estimation approach presented in Rochoux et al. (2014, IAFSS) and Rochoux et al. (2014b, NHESS);
- ▷ a comparative study of the state estimation approach and the parameter estimation approach (based on the EnKF algorithm) is presented for the controlled grassland burning experiment, in order to discuss the advantages and limitations of each approach, in terms of ensemble generation, analysis update performance and forecast quality at different lead-times. This comparative study is proposed in Rochoux et al. (2014, PROCI).

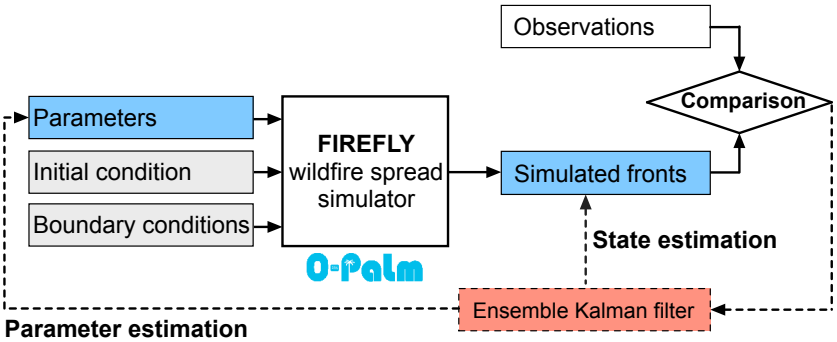


Figure 7.42: *Data assimilation flowchart for parameter estimation and state estimation approaches (the control variables are colored in blue).*

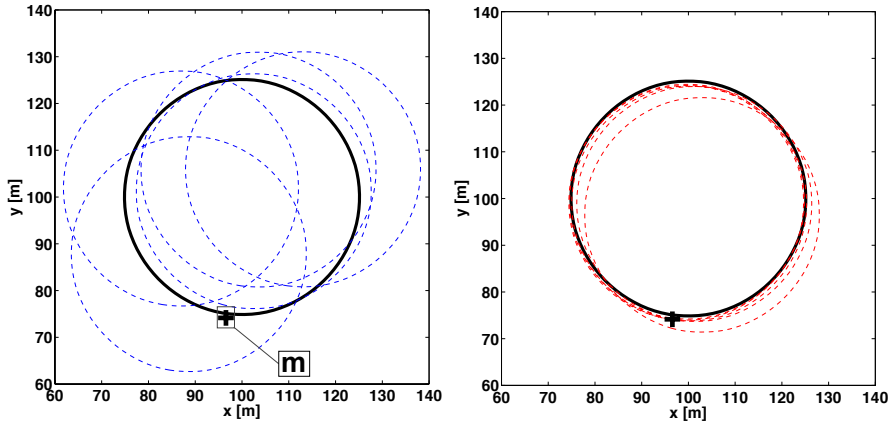
7.3.1 Analysis behavior with respect to ensemble generation

Results from OSSE experiments representative of field-scale fires are presented. Synthetic observations are generated from a reference FIREFLY simulation using chosen values of the ROS model parameters and of the ignition location (i.e., the true state is known and can be represented by the model).

↪ Isotropic case of wildfire spread

An isotropic case (referred to as S-OSSE-ISO), corresponding to a 200 m $\times$ 200 m domain with uniform vegetation properties and no wind, is considered. The ROS is constant and uniform and is set to 0.2 m/s. The true fire front is initialized as a circular front centered at (x_{ign}, y_{ign}) = (100 m, 100 m), with a radius of 5 m. FIREFLY is first integrated in time in order to produce at the analysis time (chosen to be $t = 200$ s) the true fire front location. A forecast ensemble of $N_e = 25$ members is then produced based on spatial variations of the ignition location (x_{ign}, y_{ign}) around a mean value (97 m, 103 m) and with a 10-m STD for both x- and y-directions. Uncertainties in the forecast ensemble are only due to the initial location of the fire. The observed fire front is described by a stand-alone marker ($N_{fr}^{o} = 1$), while simulated fire fronts are discretized using $N_{fr} = 100$ markers. The observation error STD is assumed small (relatively to the fireline perimeter), with $\sigma^{o} = 1$ m.

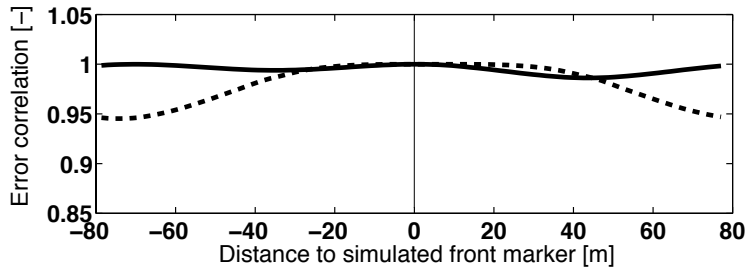
Figure 7.43(a) presents a comparison between the true and forecast fire front positions at time $t = 200$ s. This figure shows that due to uncertainties in the ignition location of the fire, the predicted front positions are scattered over a large area.



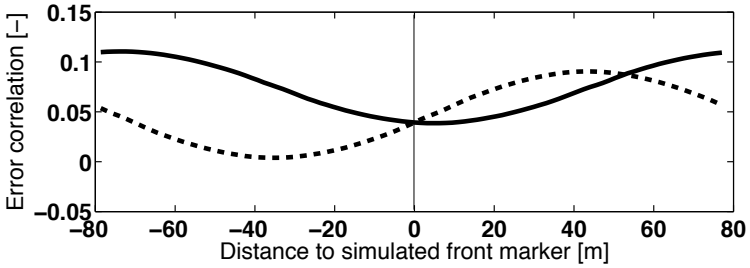
(a) Forecast. (b) Analysis with $\sigma^{o} = 1$ m.

Figure 7.43: *Spatially-uniform OSSE test (S-OSSE-ISO) with constant ROS but uncertain ignition location (x_{ign}, y_{ign}); single assimilation cycle; all figures correspond to time $t = 200$ s. (a) Comparison between true (black solid line) and forecast (blue dashed lines) fire front positions; the cross symbol is the only observation available. (b) Similar comparison between true (black solid line) and analysis (red dashed lines) fire front positions.*

Since in this test, uncertainties in the distribution of the vegetation properties are not accounted for in the ensemble generation, the propagation of the fire front is isotropic (i.e., simulated fire fronts remain circular). Thus, errors in the position of the N_{fr} simulated fire front markers are highly correlated within the ensemble. Figure 7.44 presents the error correlation functions along the fire front related to one simulated marker. Note that these error correlations translate the area of influence of each marker location error on the other marker location errors (that are distributed along the fire front) as explained in Fig. 6.30, Chapter 6. This figure shows that the univariate correlations are almost equal to one, while the multi-variate correlations are nearly zero.[5] This means that information on the propagation isotropy are mostly contained in the univariate correlations, while the non-zero multi-variate correlations compensate for the non-unity univariate correlations (Emery et al., 2013). As a result, the data assimilation algorithm translates the information observed at one point into a uniform correction along the fire front.



(a) Univariate error correlations along the fireline.



(b) Multi-variate error correlations along the fireline.

Figure 7.44: *Error correlation functions along the fireline associated to one marker location (the vertical bar represents its reference marker location indexed by m in Fig. 7.43(a)) for the isotropic propagation case (S-OSSE-ISO) with constant ROS but uncertain ignition location (x_{ign}, y_{ign}). (a) Univariate correlations: the dashed (solid) line indicates the error correlation of the reference marker x-coordinate (y-coordinate) with respect to the errors in the x-coordinates (y-coordinates) of the other markers along the fireline. (b) Multi-variate correlations: the dashed (solid) line indicates the error correlation of the reference marker x-coordinate (y-coordinate) with respect to the errors in the y-coordinates (x-coordinates) of the other markers along the fireline.*

[5] A detailed explanation is provided in Section 6.5, Chapter 6.

Figure 7.43(b) presents the comparison between the true and analysis fire front positions. Here the analysis ensemble corresponds to the updated front positions that are produced by the EnKF algorithm at the end of the analysis cycle (i.e., at time $t = 200$ s), when only one observation is available. As expected, the analysis estimates of the fire front positions feature a much reduced scatter; they are located close to the true front position and the EnKF correction is isotropic (due to the high correlations along the fireline), implying that analyses exhibit the same circular shape as the forecast members. It is worth mentioning that with this uniform definition of the forecast ensemble, deforming the fire front through the ensemble-based analysis is impossible. To be able to stochastically represent more complex fire front shapes that are representative of the multiple sources of uncertainties in wildfire spread and of their spatial distributions, it is therefore of primary importance to consider non-uniform environmental conditions when generating the forecast ensemble of fire fronts.

↪ Extension to anisotropic cases of wildfire spread

An anisotropic case of wildfire spread (referred to as S-OSSE-ANISO-WIND), subject to spatially-varying vegetation properties and wind-aided propagation, is now considered. In the previous spatially-uniform OSSE test (S-OSSE-ISO), only one observation was required to produce an optimal analysis. Uncertainties in the forecast ensemble were only due to the initial location of the fire (x_{ign},y_{ign}), meaning that the variety of the uncertainties was very limited in the ensemble. If this type of ensemble forecast were used for the present spatially-varying OSSE test, the EnKF would not be able to provide an anisotropic correction of the fire front location. To allow for a spatialized correction, it is necessary to account for uncertainties on the spatial distribution of the environmental conditions when generating the ensemble forecast. An ensemble of $N_e = 20$ forecasts is therefore produced based on assumed uncertainties in:

(1) the ROS input parameters, specifically in the fuel depth δ_v, the moisture content M_v, the fuel particle surface-area-to-volume ratio Σ_v and the wind velocity vector (magnitude u_w and direction angle α_w);

(2) the ignition location (x_{ign}, y_{ign}).

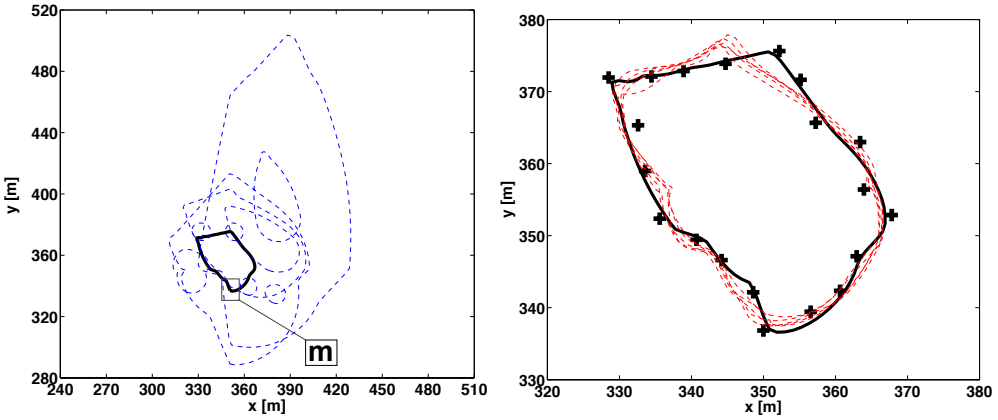
In addition, the fuel depth δ_v is assumed to be spatially-varying, taking different values in the 4 quadrants of the square-shaped 700 m $\times$ 700 m computational domain. Thus, in this configuration, uncertainties in the forecast ensemble are due to spatial variations in as many as 10 input parameters, i.e., 4 values of δ_v for the 4 quadrants as well as values of M_v, Σ_v, u_w, α_w and (x_{ign},y_{ign}). Corresponding values for the mean and STD are presented in Table 7.9.

Figure 7.45(a) presents a comparison between the true and forecast fire front positions at time $t = 150$ s. Due to uncertainties in the ROS model parameters, the propagation is now anisotropic and the present fire fronts are character-

ized by stronger irregularities and more complex shapes than results presented in Fig. 7.43(a), corresponding to uncertainties only in the ignition location. The features observed in Fig. 7.45(a) are the consequence of both the presence of wind and the spatial variations in fuel depth. The observed fire front is described by $N_{fr}^{o} = 20$ markers, uniformly-distributed along the true fire front (while the simulated fire fronts are discretized using $N_{fr} = 100$ markers), and with an error STD $\sigma^{o} = 1$ m. Figure 7.45(b) presents a similar comparison between the true and analysis fire front positions at time $t = 150$ s. As expected, the analysis ensemble features a much reduced scatter, with fire fronts located close to the true one.

Table 7.9: *Properties of the ensemble forecast, in terms of mean value and STD, in the spatially-varying OSSE test (S-OSSE-ANISO-WIND).*

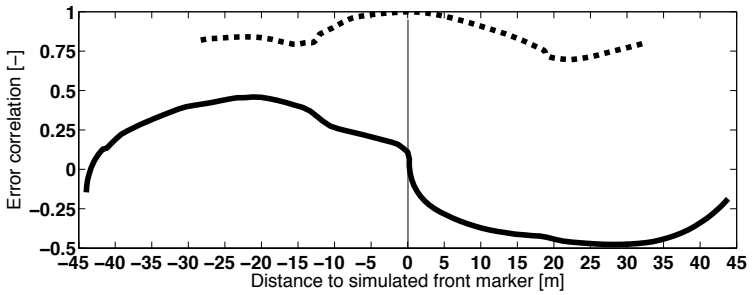
Input parameter	True value	Ensemble mean	Ensemble STD
$\delta_{v,1}$ [m]	0.25	0.25	0.10
$\delta_{v,2}$ [m]	1.25	1.25	0.10
$\delta_{v,3}$ [m]	0.75	0.75	0.10
$\delta_{v,4}$ [m]	1.75	1.75	0.10
M_v [%]	20	20	10
Σ_v [1/m]	11500	11500	4000
(u_w, α_w) ([m/s], [°])	(1.0, 315)	(0.75, 315)	(0.15, 45)
x_{ign} [m]	350	350	20
y_{ign} [m]	350	350	20



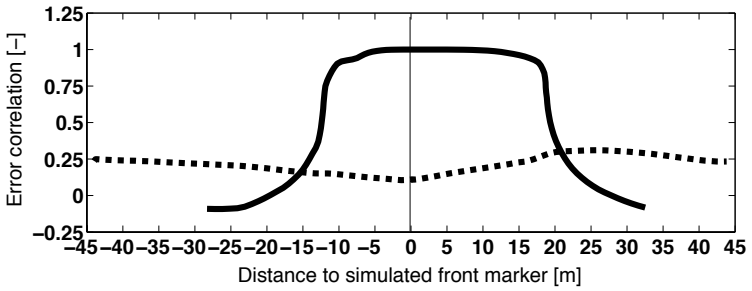
(a) Forecast. (b) Analysis with $\sigma^{o} = 1$ m.

Figure 7.45: *Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location; single analysis cycle; all figures correspond to time $t = 150$ s (S-OSSE-ANISO-WIND). (a) Comparison between true (black solid line) and forecast (blue dashed lines) fire front positions. (b) Similar comparison between true (black solid line) and analysis (red dashed lines) fire front positions. Cross symbols are the observations.*

Since the anisotropy of the wildfire propagation is now represented by a wide range of uncertainties in the ensemble, the error in the location of one observed point is only correlated with the error in the location of the other adjacent points along the fireline as shown in Fig. 7.46 for the forecast ensemble.



(a) Univariate error correlations along the fireline.



(b) Multi-variate error correlations along the fireline.

Figure 7.46: *Error correlation functions along the fireline associated to one marker location (the vertical bar represents its reference marker location indexed by m in Fig. 7.45(a)) for the anisotropic propagation case (S-OSSE-ANISO-WIND). (a) Univariate correlations: the dashed (solid) line indicates the error correlation of the reference marker x-coordinate (y-coordinate) with respect to the errors in the x-coordinates (y-coordinates) of the other markers along the fireline. (b) Multi-variate correlations: the dashed (solid) line indicates the error correlation of the reference marker x-coordinate (y-coordinate) with respect to the errors in the y-coordinates (x-coordinates) of the other markers along the fireline.*

The estimation problem translates the information coming from one observation marker into a local correction restricted to the closest neighbours only. The distance over which the observation marker affects the correction of the simulated front marker locations is referred to as *correlation length-scale* (Daley, 1991; Pannekoucke et al., 2008; Weaver and Mirouze, 2012). This length-scale depends on the spatial variability of the errors in the ensemble generation. Figure 7.46 shows that the length-scale associated with the univariate error correlations of the x-coordinate typically takes values on the order of 15 m on both sides of the considered simulated marker. This means that if this marker is observed during data assimilation, the correction of its location modifies the shape of the fire front along 15 m on both sides of this marker. Figure 7.47(a) maps the univariate error correlations onto the

mean forecast front and confirms that the section of the fireline over which the error correlation related to the x-coordinate is above 0.75 is very limited. Consistently, Fig. 7.48(a) demonstrates that the forecast error correlation matrix is not constant by block, in contrast to the isotropic case (see Fig. 6.31, Chapter 6).

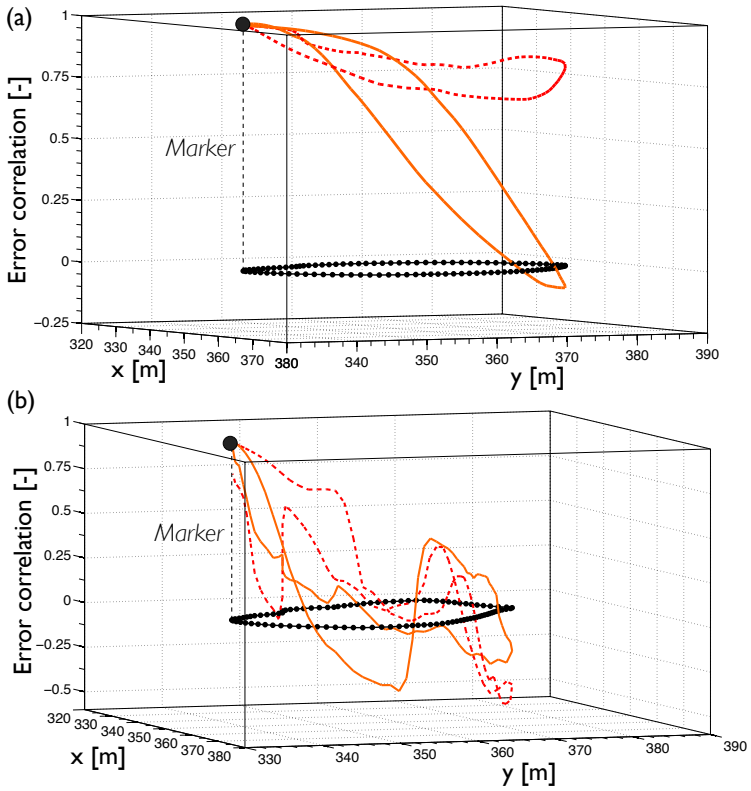
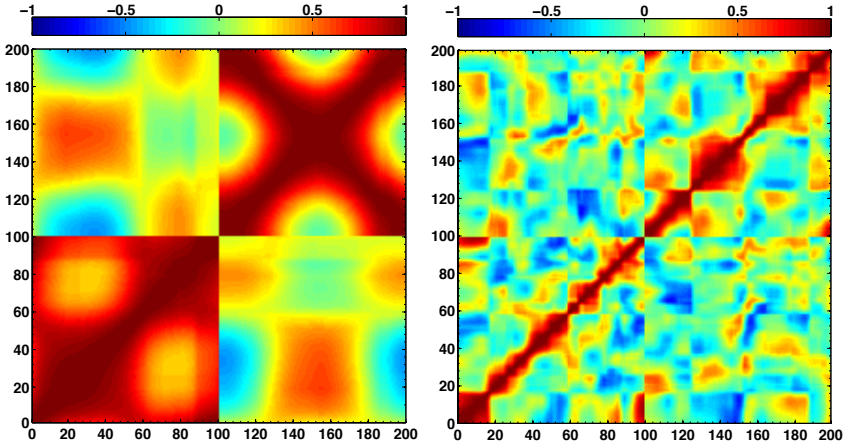


Figure 7.47: *Error univariate correlation functions along the fireline associated to one fire front marker (the vertical dashed bar represents this marker) for anisotropic propagation (S-OSSE-ANISO-WIND). Red dashed (orange solid) lines indicate the error correlation of the reference marker x-coordinate (y-coordinate) with respect to the x-coordinates (y-coordinates) of the other markers along the fireline. (a) 3-D view of the forecast univariate correlations (the black circled line represents the mean forecast front). (b) 3-D view of the analysis univariate correlations (the black circled line represents the mean analysis front).*

The update step induces narrower error correlation length-scales for the analysis ensemble than for the forecast ensemble as shown in Figs. 7.47(a)-(b). The correction of the front marker locations (which results from a linear combination of the error correlation length-scales of each observed marker along the fireline) modifies the spatial structure of the error correlations as shown by the analysis error correlation matrix in Fig. 7.48(b). As a consequence, when several observations are available, a non-uniform correction is obtained and error variances associated with each simulated front marker are significantly reduced. Thus, the data assimilation algorithm is able to realistically modify the shape of the fireline based on observation information.

However, even though accounting for a wide range of uncertainties in the ROS parameters provides a wide range of possible fire front shapes, the EnKF does not guarantee an optimal analysis. Sampling errors can indeed degrade the representation of the error statistics by the forecast ensemble (if N_e is not large enough) and thereby, the analysis solution. Figure 7.49 examines the impact of the number of members N_e on the EnKF performance, in terms of mean and STD of the distances between the true and forecast front positions (FMT) as well as between the true and analysis front positions (AMT). These results show that for the present anisotropic propagation, $N_e = 20$ members are sufficient to obtain converged error statistics. Below this threshold value, sampling noise induces errors in the representation of the error covariances.



(a) Forecast. (b) Analysis with $\sigma^{\mathrm{o}} = 1$ m.

Figure 7.48: *Error correlation matrices for the anisotropic propagation case (S-OSSE-ANISO-WIND) with $N_{fr} = 100$ simulated front markers.*

7.3.2 Sensitivity of the analysis to the observation network

Because in real-world applications measurements will be sparse and imperfect, we study in the framework of OSSE experiments the sensitivity of the data-driven solution to the STD of the observation error (σ^{o}), to the number of observation markers along the fireline (N^{o}_{fr}), and to the level of completeness of the observations (i.e., a case in which only a limited section of the fireline is effectively observed).

↪ Sensitivity to observation errors

While results in Fig. 7.43(b) were produced with a low value of the observation error STD ($\sigma^{\mathrm{o}} = 1$ m) relatively to the fireline perimeter, Fig. 7.50(a) examines the influence of this error on the EnKF performance up to $\sigma^{\mathrm{o}} = 20$ m (similarly to Fig. 7.7(a) for the EnKF parameter estimation approach). The statistics (in terms of mean value and STD) of the distance between the true and forecast fire

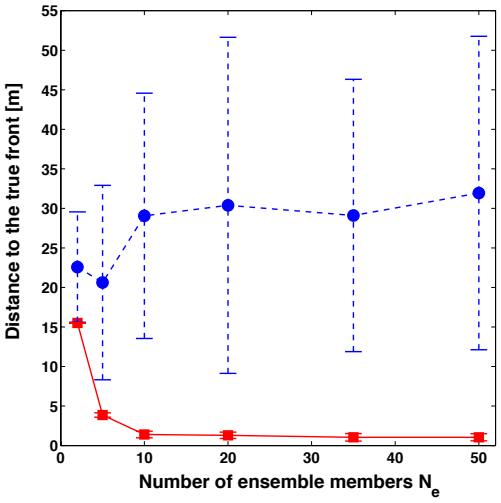


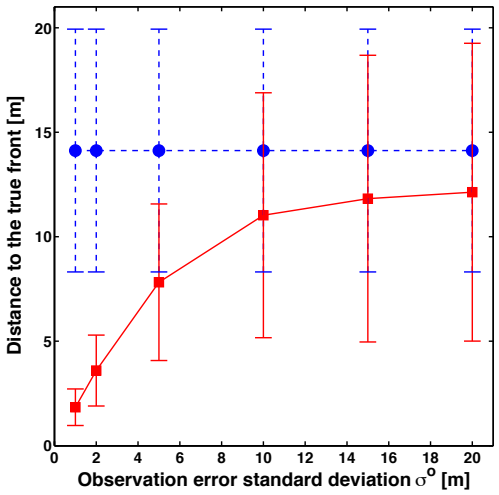
Figure 7.49: *Spatially-varying OSSE test as a function of the number of members N_e (S-OSSE-ANISO-WIND). The blue dashed line represents the mean distance between the true and forecast front positions $\overline{d_{FMT}}$ (error bars indicate the associated STD $\overline{\sigma_{FMT}}$); the red solid line represents the mean distance between the true and analysis front positions $\overline{d_{AMT}}$ (error bars indicate the associated STD $\overline{\sigma_{AMT}}$).*

front positions (FMT) as well as between the true and analysis fire front positions (AMT) are presented as a function of the observation error STD σ^{o}. Vertical bars give a graphical representation of the magnitude of the STD in the forecast and analysis ensembles, noted $\overline{\sigma_{FMT}}$ and $\overline{\sigma_{AMT}}$, respectively. The figure shows that when the observation error STD σ^{o} is small, the EnKF algorithm successfully drives the analysis ensemble towards the true state. In contrast, when the observation error STD is large, the analysis ensemble remains close to the forecast ensemble. For intermediate values of σ^{o}, the EnKF algorithm produces optimized predictions lying between forecast and observation. In the following tests, observation errors are assumed to be small in order to evaluate the EnKF ability to track the observed fire front location, in terms of behavior and shape.

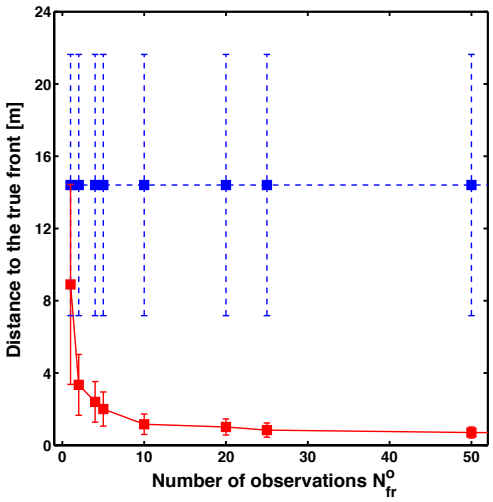
↪ Sensitivity to the number of observed markers

Figure 7.50(b) examines the influence of the number of uniformly-distributed observation markers N_{fr}^{o} along the fireline on the EnKF performance. In particular, it presents the statistics (in terms of mean value and STD) of the distance between the true and forecast fire front positions (FMT) as well as between the true and analysis fire front positions (AMT) as a function of the ratio $r = N_{fr}^{\text{o}}/N_{fr}$ varying between 0.01 and 0.5 (since the number of simulated markers is $N_{fr} = 100$). For this study, a forecast ensemble of $N_e = 25$ members is produced based on spatial variations of the ignition location (x_{ign}, y_{ign}) around a mean value (97 m, 103 m) and with a 10-m STD in both x- and y-directions. However, in contrast to the uniform case (S-OSSE-ISO), this case exhibits non-uniform ROS due to the spatially-distributed fuel

layer depth $\delta_v = \delta_v(x, y)$; there is no external flow u_w. This spatially-distributed case (referred to as S-OSSE-ANISO) leads to anisotropic propagation of the fire; the corresponding forecast estimates are presented in Fig. 7.51(a).



(a)

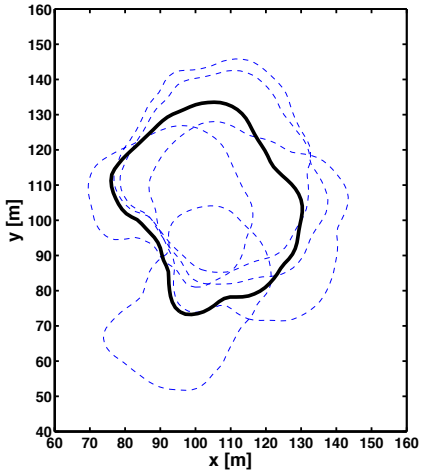


(b)

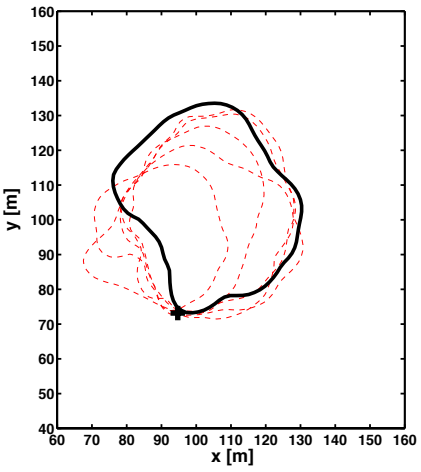
Figure 7.50: *EnKF performance with respect to (a) the observation error STD σ° (S-OSSE-ISO) and (b) the number of observed markers N°_{fr} (S-OSSE-ANISO). The blue circled-dashed line represents the mean FMT $\overline{d_{FMT}}$ and the red squared-solid line represents the mean AMT. Error bars indicate the corresponding STD, $\overline{\sigma_{FMT}}$ and $\overline{\sigma_{AMT}}$.*

When the number of observed markers N^{o}_{fr} is large, the EnKF algorithm successfully drives the analysis ensemble towards the true state as confirmed by the analysis ensemble presented in Fig. 7.51(c) for $N^{\circ}_{fr} = 25$. In contrast, when N°_{fr} is small, the EnKF algorithm has reduced effects as illustrated in Fig. 7.51(b) for $N^{\circ}_{fr} = 1$.

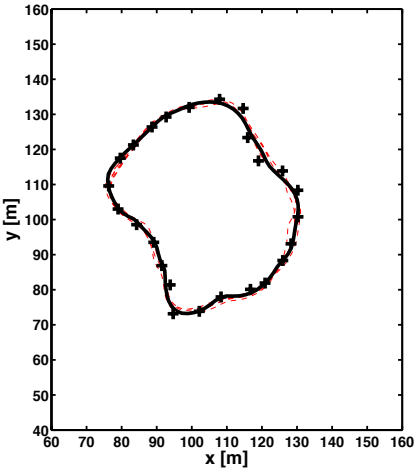
In this case, the analysis remains close to the forecast presented in Fig. 7.51(a). Besides, above $N_{fr}^{o} = 25$, the statistics of the distance for the analysis remains stable, implying that the added observed markers do not contribute to the analysis solution and constitute repetitive information; they are therefore unnecessary here.



(a) Forecast.



(b) Analysis with $N_{fr}^{o} = 1$.



(c) Analysis with $N_{fr}^{o} = 25$.

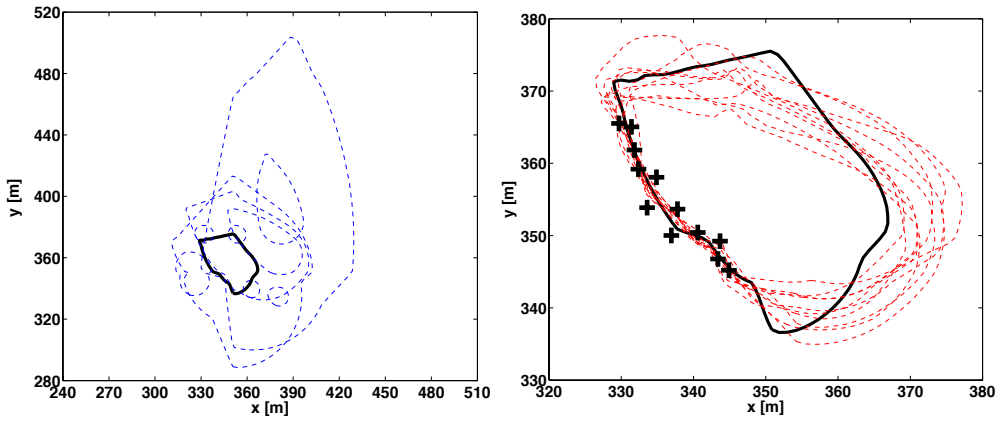
Figure 7.51: *Spatially-varying OSSE test with anisotropic no-wind ROS due to spatially-distributed vegetation and uncertain ignition location (S-OSSE-ANISO); single analysis cycle; all figures correspond to time $t = 200$ s. (a) Comparison between true (solid line) and forecast (dashed lines) fire front positions. (b) Similar comparison between true (solid line) and analysis (dashed lines) fire front positions for $N_{fr}^{o} = 1$. (c) See caption (b) for $N_{fr}^{o} = 25$ uniformly-distributed markers. Crosses correspond to observation.*

In summary, the performance of the state estimation approach and its ability to capture the high-resolution features of the fire front strongly depend on both the spatial variability of the errors in the simulated fire front locations and the density

of the observation network. If this spatial variability is significant, a large number of observations with a wide coverage of the fire front perimeter is required. In contrast, if the errors in the fire front positions do not vary spatially, only a few observed markers are necessary to produce a consistent analysis.

↪ Sensitivity to the location of the observed markers

While Figs. 7.45(a)-(b) show that the direct observation of the fire front position can overcome various uncertainties in the ROS model parameters, Fig. 7.52(b) illustrates for the same anisotropic configuration (S-OSSE-ANISO-WIND) that the spatial distribution of the observations along the fireline has a significant impact on the analysis.



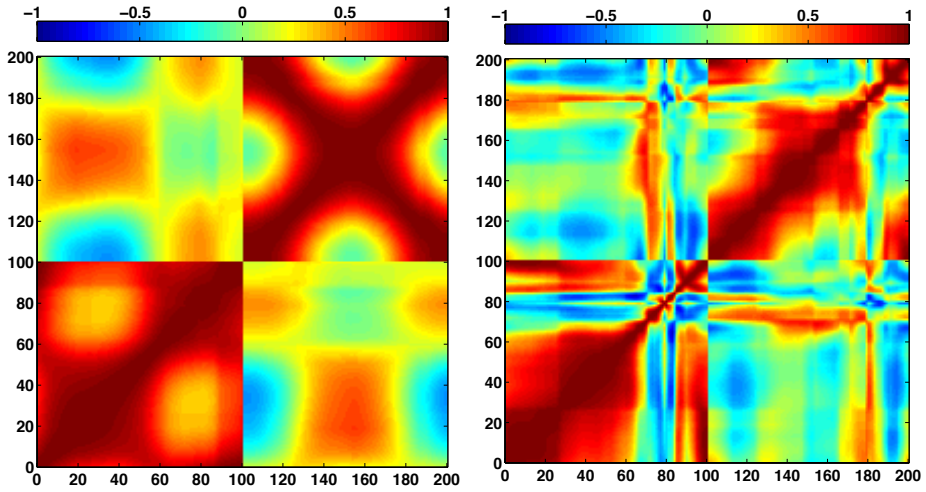
(a) Forecast. (b) Analysis with $N^{o}_{fr} = 12$.

Figure 7.52: *Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location; single analysis cycle; all figures correspond to time $t = 150$ s (S-OSSE-ANISO-WIND). (a) Comparison between true (black solid line) and forecast (blue dashed lines) fire front positions. (b) Comparison between true (black solid line) and analysis (red dashed lines) fire front positions with $N^{o}_{fr} = 12$ and with an incomplete set of observations (non-uniformly distributed along the fireline). Cross symbols are the observations.*

Figure 7.52(b) considers a practically-relevant situation, in which observations are limited to a certain section of the fire front (i.e., the informed section) and therefore, provide an incomplete picture of the real situation. Such a situation could occur for instance due to the opacity of the fire plume standing in the way of the remote sensor line-of-sight. In this situation, while the EnKF algorithm produces an analysis that is close to the true state in the informed section, the benefits of data assimilation are reduced in the non-informed sections.

However, despite a reduced level of performance, the EnKF informed by incomplete observations remains capable of improved performance compared to a free run. In particular, the analysis ensemble exhibits errors with lower spatial variability than

the forecast ensemble presented in Fig. 7.52(a). This improvement is due to the error correlations of the simulated front markers that are observed. They constrain (to a certain extent) the shape of the fire front when the position of these markers is modified through the EnKF update. This constraint is imposed by the correlation length-scales, which depend on the sources of uncertainties that are represented during the generation of the forecast estimates. The more spatial variabilities in the errors are considered, the thinner the length-scales, meaning that local effects on the behavior and shape of the fire front are represented (in contrast to the isotropic propagation case S-OSSE-ISO). This is confirmed by the structure of the analysis error correlation matrix in Fig. 7.53(b), showing that the reduction of the spatial extent of high error correlations nearby the observed markers is limited to the closest simulated markers. For comparison, see Fig. 7.48(b), in which this reduction of the error correlations systematically occurs for all simulated front markers since the whole front is observed and assimilated.



(a) Forecast. (b) Analysis with $\sigma^{o} = 1$ m.

Figure 7.53: *Error correlation matrices for the anisotropic propagation case (S-OSSE-ANISO-WIND) with $N_{fr} = 100$ simulated markers and $N_{fr}^{o} = 12$ non-uniformly distributed assimilated markers.*

7.3.3 Performance of wildfire spread forecast

Multiple analysis cycles for the spatially-varying OSSE test (S-OSSE-ANISO-WIND) are now considered. The behavior of the forecast between successive observations or after the last observation is examined. The true fire front spread is simulated for time-varying wind conditions presented in Table 7.10 (observations are representative of the true fire front since the observation error STD is small, i.e., $\sigma^{o} = 1$ m), whereas the forecast ensemble is simulated for constant wind conditions using the parameters presented in Table 7.9.

The performance of the forecast is expected to deteriorate in time for two reasons. First, because the impact of the fire front correction at a previous time decreases as the forecast lead-time increases. Second, because the present implementation of the EnKF does not provide any correction for ROS model errors or for uncertainties in the ROS model parameters (including the incorrect assumption of a constant wind). The correction of the ROS model parameters can be addressed through the parameter estimation approach presented in Section 7.2. Accounting for model errors is out of the scope of this work.

Table 7.10: *Time-varying true wind conditions (in terms of magnitude u_w and direction angle α_w) in the spatially-varying OSSE test (S-OSSE-ANISO-WIND).*

Assimilation cycle	**Wind magnitude** [m/s]	**Wind direction** [°]
1	1.0	315
2	0.75	290
3	0.83	257
4	1.20	232

In this test, the EnKF-based state estimation approach is applied over four successive assimilation cycles: the EnKF update is performed at times $t_1 = 150$ s, $t_2 = 300$ s, $t_3 = 450$ s, and $t_4 = 600$ s. Each assimilation cycle includes a prediction step that integrates the wildfire spread model until the observation time t_n ($n = 1, \cdots, 4$) and an update step providing the corrected fire front location at time t_n. To move to the next assimilation cycle, the two-dimensional progress variable field c associated with each analysis member must be reconstructed through a binarization procedure explained in Section 6.5.4, Chapter 6. This field is then used as initial condition of the next assimilation cycle to further integrate FIREFLY during the prediction step up to time t_{n+1}. Note that the perturbation of the ignition location (x_{ign}, y_{ign}) is only introduced during the first assimilation cycle (as a means to account for uncertainties in the fire ignition location before remote sensing detection).

Figure 7.54 presents the successive comparison between the averaged free run (i.e., the stand-alone FIREFLY simulation without data assimilation), the average fire front location related to the forecast and the analysis, and the observations (considered to be close to the true state) from $t_1 = 150$ s to $t_4 = 600$ s. The free run simulation does not accurately estimate the fire spread ROS and direction, and the distance to the observations (and thereby to the true front) is drastically reduced through the EnKF update. The forecast, resulting from the integration of FIREFLY starting from the analysis at the previous assimilation time (for instance, the forecast at time $t_2 = 300\ s$ is derived from the analysis at time $t_1 = 150$ s), provides a better prediction of the fire front location and a more physically-consistent front shape than the free run at each assimilation time. Still, its prediction quality is significantly lower than that exhibited by the analysis update, similarly to the results obtained in previous OSSE tests.

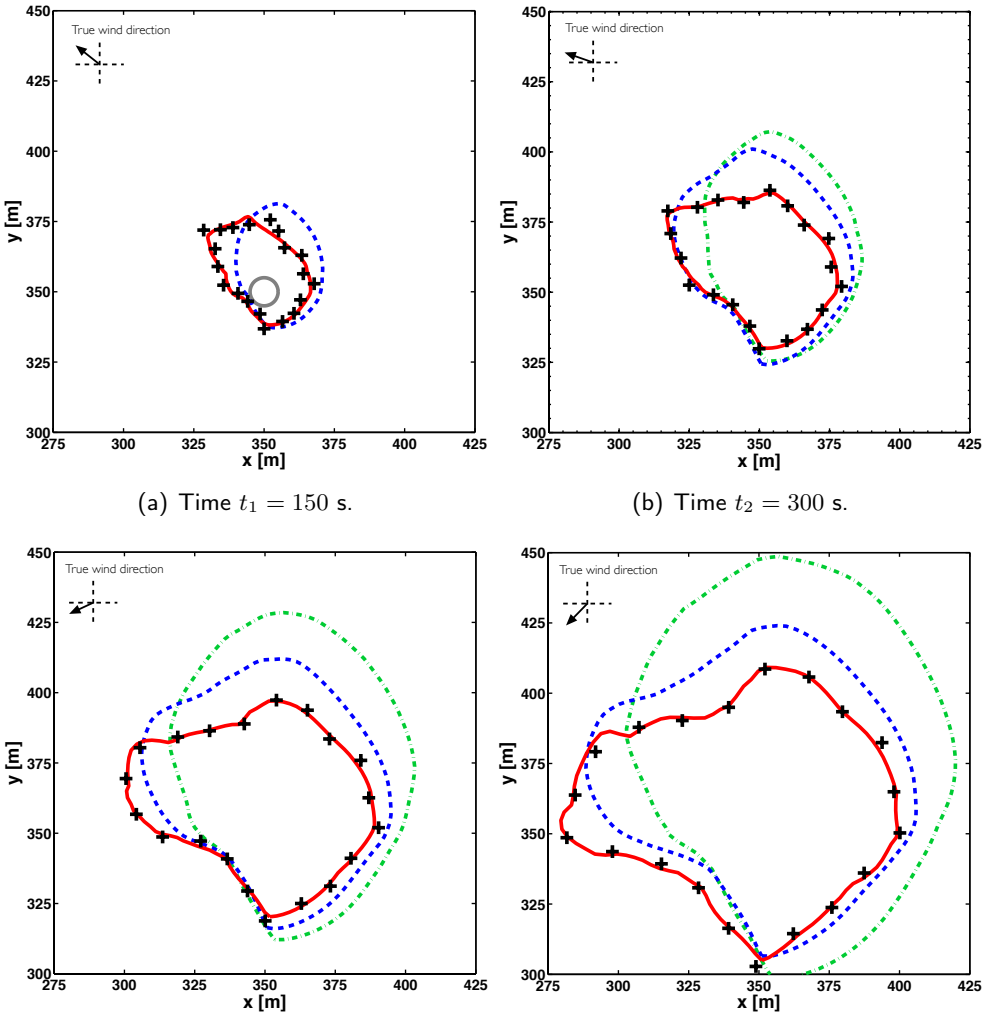
Figure 7.55 compares the forecast and analysis estimates to the true fire front for the 4 successive assimilation cycles. The forecast estimates provide a wide range of fire fronts (in terms of location and shape) due to the prediction step, including uncertainties in the ROS model parameters to cover all the potential fire front candidates for the analysis and to foresee any change in the behavior of the wildfire spread over each assimilation cycle. In contrast, the analysis estimates exhibit a much reduced scatter due to the EnKF update, thereby drastically reducing uncertainties in the wildfire spread.

This performance of the EnKF update is confirmed by the error statistics presented in Fig. 7.56(a). The typical cyclic evolution of the deviations of model predictions from observations is presented for 5 successive assimilation cycles. During the update step of the assimilation cycle n, the analysis (An) provides a correction to the front position and the distance between the true state and the forecast (Fn-1) is drastically reduced. The ensemble of analyses at the end of cycle n provides the initial conditions for the next cycle $(n+1)$; during the prediction step of cycle $(n+1)$, the wildfire spread model simulates the fire evolution but, without additional observations, this forecast deviates from the true state. For instance, the mean distance between the true state and the forecast derived from the analysis performed at time t_1 (A1) is approximately 80 m at time $t_5 = 750$ s. During the update step of the analysis cycle $(n+1)$, the distance between the true state and the forecast is again reduced. For instance, the mean distance is significantly reduced by the EnKF algorithm, from 20 m for the forecast (F1) to less than 1 m for the analysis (A2). Then, the assimilation cycle may be repeated.

Still, EnKF-based data-driven simulations bring valuable information on the wildfire spread behavior at short lead-times, even though data assimilation is not applied systematically. This is illustrated in Fig. 7.56(b), which presents a comparison of different forecasts of the fire front location at time $t_4 = 600$ s using Firefly, with or without data assimilation. The analysis at $t_4 = 600$ s (A4) provides the most accurate tracking of the observations, while the forecast runs with data assimilation performed at previous times (i.e., $t_1 = 150$ s for (F1), $t_2 = 300$ s for (F2) and $t_3 = 450$ s for (F3)) deviate from the observations. However, the closer the assimilation time to $t_4 = 600$ s, the lower the forecast deviation from the observations. Consistently, the free run provides the less accurate prediction with a mean distance to the observations approximately equal to 70 m; in contrast, this mean distance is reduced by a factor of nearly 2 for (F2) and by a factor of 70 for (A4). This result illustrates the improved prediction capability of the wildfire spread model, achieved when calibrated against past observations.

In summary, these results on OSSE test cases show that in a state estimation approach, EnKF updates provide valuable information on wildfire spread and lead to more accurate forecasts on the wildfire spread behavior at short lead-times (consistently with the persistence of the model initial condition). Thus, the best strategy to produce accurate data-driven simulations with Firefly consists in performing

the EnKF update at regular short-term time intervals. If this condition is satisfied in the data assimilation system, the data-driven prototype simulator based on a state estimation approach is capable of efficiently tracking the actual fire front and shape. Note that this condition is feasible for airborne and/or UAVs systems that can continuously observe a fire over a certain time period.



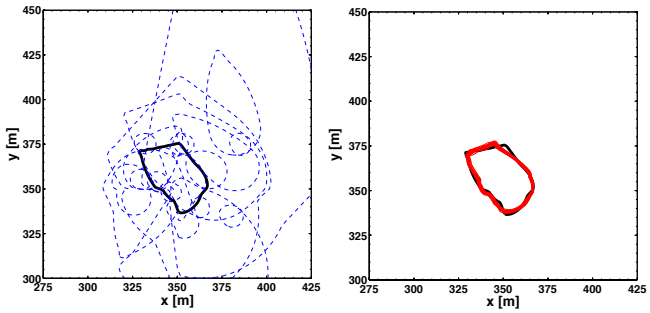
(a) Time $t_1 = 150$ s.

(b) Time $t_2 = 300$ s.

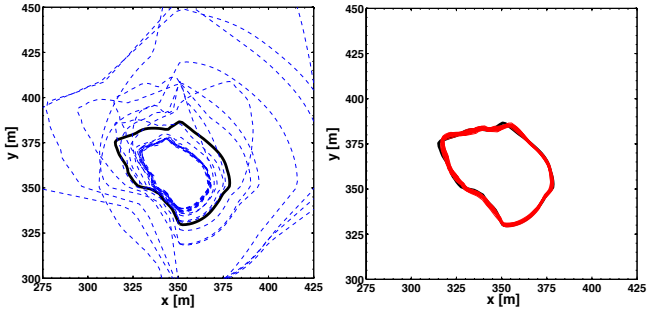
(c) Time $t_3 = 450$ s

(d) Time $t_4 = 600$ s

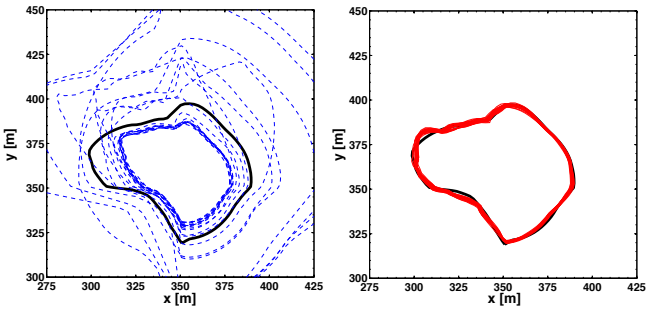
Figure 7.54: *Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location (S-OSSE-ANISO-WIND); multiple analysis cycles from 0 to 600 s. Comparison between the free run (green dashed-dotted lines), the mean forecast estimate (blue dashed lines), the observations (black crosses) and the mean analysis estimate (red solid lines) at the 4 successive analysis times: (a) $t_1 = 150$ s (the gray circle corresponds to initial condition); (b) $t_2 = 300$ s; (c) $t_3 = 450$ s; and (d) $t_4 = 600$ s.*



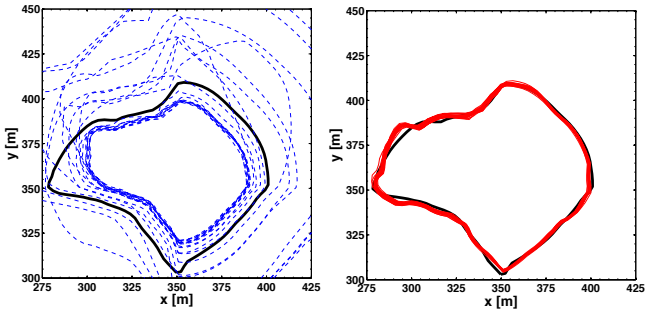
(a) Time $t_1 = 150$ s.



(b) Time $t_2 = 300$ s.

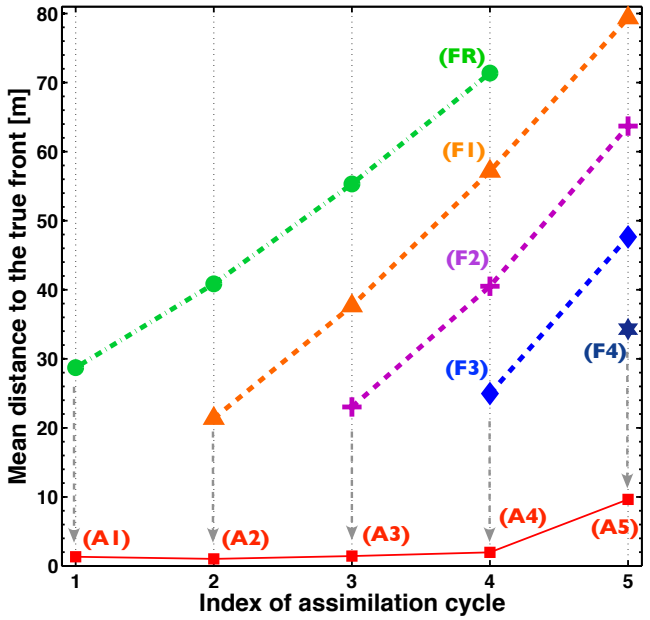


(c) Time $t_3 = 450$ s.

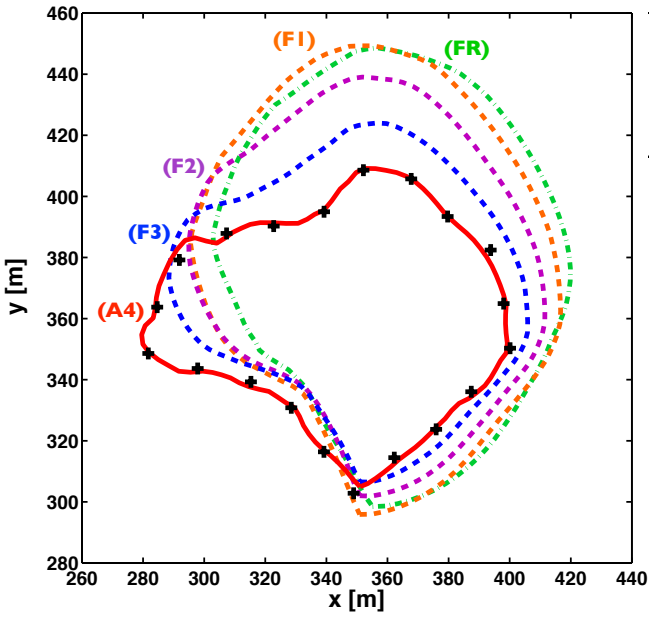


(d) Time $t_4 = 600$ s.

Figure 7.55: *Spatially-varying OSSE test with uncertain ROS model parameters and uncertain ignition location (S-OSSE-ANISO-WIND); multiple analysis cycles from 150 to 600 s. Left: Comparison between the true (black solid line) and forecast estimates (blue dashed lines) of fire fronts. Right: Similar comparison between the true (black solid line) and analysis estimates (red solid lines) of fire fronts.*



(a)



(b)

Figure 7.56: *(a) Average distance between the true and forecast/analysis fire front positions (FMT/AMT) as a function of the assimilation cycle index (S-OSSE-ANISO-WIND); green circles correspond to the free run (FR); triangles, crosses, diamonds and stars correspond to a forecast with an analysis update at $t_1 = 150$ s (F1), $t_2 = 300$ s (F2), $t_3 = 450$ s (F3) and $t_4 = 600$ s (F4), respectively; square symbols correspond to an analysis performed at times t_1 (A1), t_2 (A2), t_3 (A3), t_4 (A4) and t_5 (A5). (b) Comparison between the free run (FR) in green dashed-dotted line, the mean forecast estimate (F1, F2, F3) in dashed lines, the observations in black crosses, and the mean analysis estimate (A4) in red solid line at time $t_4 = 600$ s.*

7.3.4 Comparative study of state and parameter estimations

EnKF-based parameter and state estimation approaches are evaluated by comparison with data taken from the controlled grassland fire experiment presented in Fig. 6.13, Chapter 6. The objective is to highlight the benefits and drawbacks of each approach. This study considers 4 successive, 14-seconds-long, assimilation cycles with initialization at $t_0 = 50$ s and updates at times $t_1 = 64$ s, $t_2 = 78$ s, $t_3 = 92$ s and $t_4 = 106$ s. Observations are described as full fire contours represented by $N^{\text{o}}_{fr} = 40$ markers with an uncertain location $\sigma^{\text{o}} = 5$ cm along the x- and y-directions. Note that previous studies (EKF, EnKF and PC-EnKF for parameter estimation) assimilate observations at 28 s intervals; the objective here is to investigate if assimilating at a higher temporal frequency leads to more accurate analysis and forecast estimates, in terms of fire front location and shape.

↪ Ensemble generation

In the parameter estimation approach, 4 parameters are used as control variables with $\mathbf{x} = [M_v, \Sigma_v, u_w, \alpha_w]$. These parameters are perturbed around mean values and with prescribed uncertainties (according to assumed levels of uncertainty), but remain spatially-uniform for each ensemble member. The EnKF ensemble contains $N_e = 1,000$ members, meaning that during each assimilation cycle, Firefly produces 1,000 fire front trajectories associated with each set of control parameters.

In the state estimation approach, the control variables are the spatial coordinates of the discretized fire front. The EnKF ensemble now contains $N_e = 50$ members corresponding to different choices of the ROS model parameters and of the fire initial location. In addition to M_v, Σ_v, u_w and α_w, the fuel depth δ_v is varied in the ensemble in 4 zones. The intent here is to generate a rich forecast ensemble featuring a wide range of fire front shapes and locations.

The statistical properties (mean and STD) of the variations in the ROS model parameters are reported in Table 7.11.

Table 7.11: *Statistical properties of the ROS model parameters, in terms of mean value and STD, being treated as random variables to generate the forecast ensemble (the parameter estimation approach only uses perturbations in* $[M_v, \Sigma_v, u_w, \alpha_w]$*).*

ROS parameter	Ensemble mean	Ensemble STD
$\delta_{v,1}$ [m]	0.06	0.04
$\delta_{v,2}$ [m]	0.08	0.04
$\delta_{v,3}$ [m]	0.1	0.04
$\delta_{v,4}$ [m]	0.12	0.04
M_v [%]	22	6
Σ_v [1/m]	11500	4000
(u_w, α_w) [(m/s, °)]	(1, 307)	(0.4, 45)
x_{ign} [m]	2	0.65

↪ Update performance

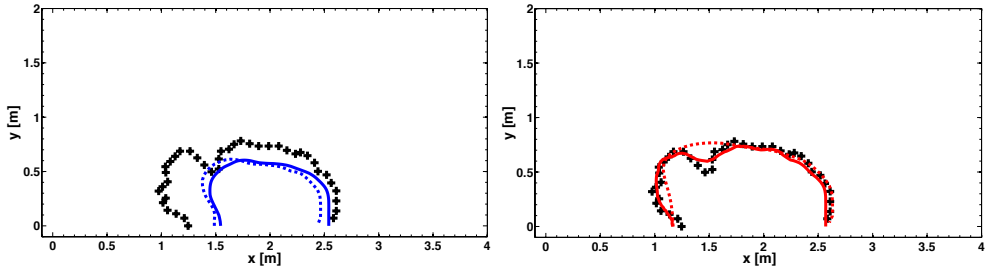
Figures 7.57(a)-(b) present the mean (ensemble-averaged) fire front location at time $t_1 = 64$ s, as predicted by FIREFLY starting from initial conditions specified at time $t_0 = 50$ s. The mean predictions are compared to experimental observations in Fig. 7.57(a). It is shown that the mean free forecast (without data assimilation) significantly underestimates the observed ROS of the fire. In contrast, the predictions made using parameter estimation and state estimation ensembles after an EnKF update performed at time t_1 successfully reduce the distance between simulations and observations in Fig. 7.57(b). This improvement is achieved by an adjustment of the ROS model parameters in the parameter estimation approach or by a direct adjustment of the fire front location in the state estimation approach. The state estimation does not correct the sources of uncertainties (for instance, the uncertainties in the ROS model parameters) but provides a new initial condition for the next prediction step.

Similar comparisons between the mean forecast/analysis and observations at times $t_2 = 78$ s, $t_3 = 92$ s and $t_4 = 106$ s are presented in Fig. 7.57. Both parameter estimation and state estimation approaches provide accurate estimates of the fire front position. Figure 7.58 quantifies this statement and presents the averaged distance between the observations and the mean fire front position produced by the forecast (FMO) and the analysis (AMO). The mean distance to the observations is reduced by a factor 2 in the parameter estimation approach and by a factor of at least 5 in the state estimation approach. Thus, the agreement between predicted and observed front positions is significantly better than the level of agreement that would be obtained in the absence of data assimilation, see Fig. 7.57(a).

The performance of the parameter estimation approach is degraded at times t_1 and t_4 due to irregularities in the fire shape. These irregularities cannot be captured in the parameter estimation approach because its EnKF ensemble relies on spatially-uniform parameters. In contrast, the state estimation approach uses a richer ensemble and its performance is very good: the mean distance between observed and simulated fronts remains below 0.1 m for all assimilation cycles and the scatter of the analysis ensemble (measured by the STD) is significantly lower than that obtained in the parameter estimation approach, see Fig. 7.58(b). This result suggests that the state estimation approach is able to provide an accurate and non-uniform correction of the fire front location; it also highlights the importance of accounting for spatial variations of the environmental conditions to obtain more accurate fire front shapes at update times.

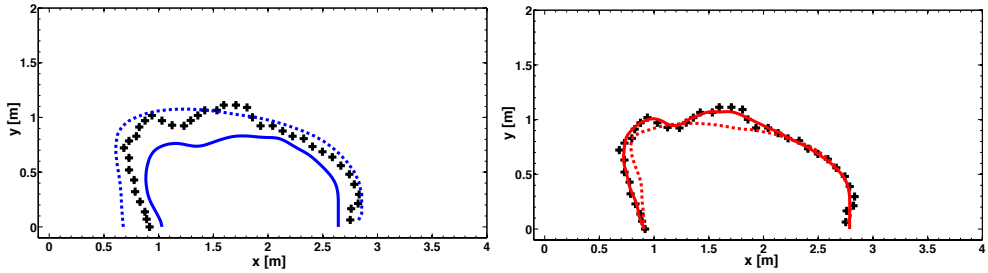
↪ Forecast performance

In spite of the overall quality of the correction provided by both EnKF estimation approaches, the accuracy of the forecast remains limited to short-term predictions, while still significantly more accurate than the free run (as illustrated for the state estimation approach in Fig. 7.59).



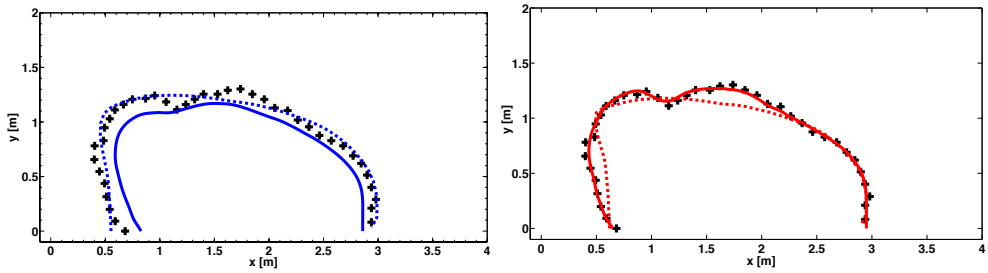
(a) Forecast at $t_1 = 64$ s.

(b) Analysis at $t_1 = 64$ s.



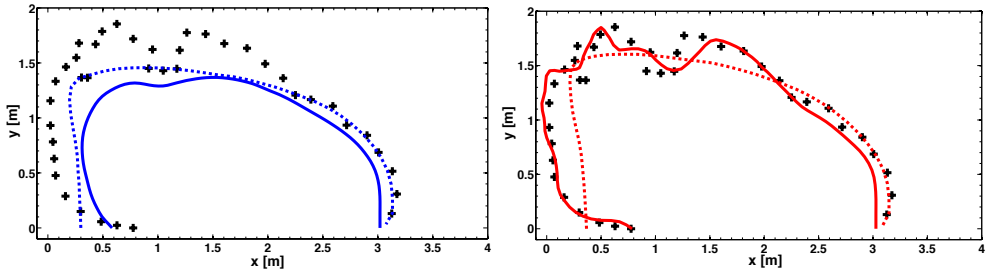
(c) Forecast at $t_2 = 78$ s.

(d) Analysis at $t_2 = 78$ s.



(e) Forecast at $t_3 = 92$ s.

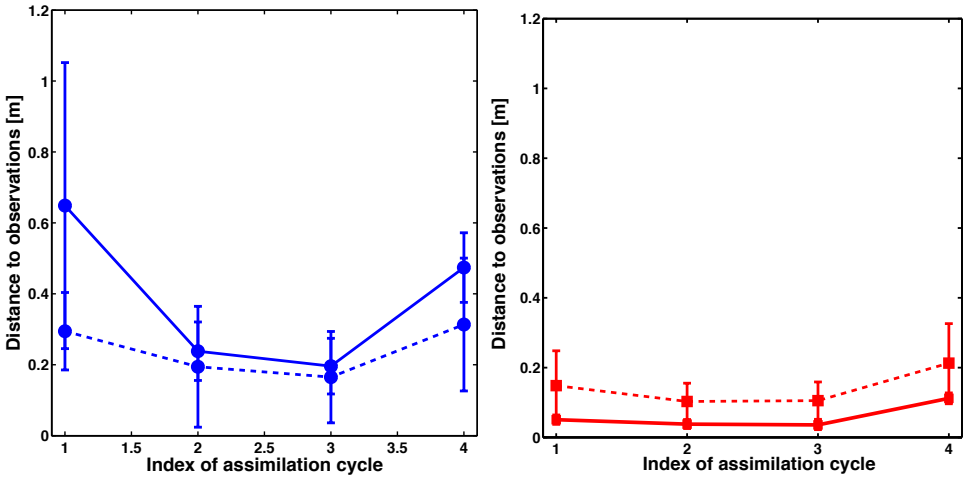
(f) Analysis at $t_3 = 92$ s.



(g) Forecast at $t_4 = 106$ s.

(h) Analysis at $t_4 = 106$ s.

Figure 7.57: *Comparison between simulated (lines) and observed (black crosses) fire front positions at $t_1 = 64$ s, $t_2 = 78$ s, $t_3 = 92$ s and $t_4 = 106$ s. The simulated fire front position is the mean position calculated as the average of the EnKF ensemble. Solid lines correspond to the simulation based on state estimation, and dashed lines correspond to the simulation based on parameter estimation. (a)-(c)-(e)-(g) Mean forecast (without data assimilation for the first observation time or with data assimilation at the previous observation time). (b)-(d)-(f) Mean analysis (with a data assimilation update at the current observation time).*



(a) Forecast. (b) Analysis.

Figure 7.58: *Evolution over successive assimilation cycles of the averaged distance between the observations and the mean fire front position produced by the (a) forecast ($\overline{d_{FMO}}$) and (b) analysis ($\overline{d_{AMO}}$) estimates. Dashed (solid) lines represent parameter (state) estimation results; error bars indicate the STD of the distance to the observations, noted $\overline{\sigma_{FMO}}$ and $\overline{\sigma_{AMO}}$, respectively.*

These results suggest that while a data assimilation approach provides excellent forecast performance at short lead-times, this level of performance may not be persistent and needs to be renewed by frequent observations with an assimilation frequency that is high enough to track the temporal variability of the errors on the control variables, especially when the wildfire behavior is strongly time-dependent. In the present configuration, due to the increased knowledge in the ROS model parameters along the assimilation cycles, the parameter estimation approach provides better forecasts of the fire front position than the state estimation approach (which only updates the initial condition of the wildfire spread model for a given assimilation cycle). This improved forecast performance of the parameter estimation approach is illustrated in Fig. 7.58(a) at the successive assimilation times. The adjustment of the ROS model parameters allows for a correction of inaccuracies in initial guesses as well as an adaptation to time-dependent properties; the statistical properties of the EnKF ensemble are dynamically-evolving. In contrast, the state estimation approach is limited to adjusting the fire front location; in that case, the statistical properties of the EnKF ensemble are constant. Figure 7.60 shows that the initial values of the moisture content M_v and the fuel particle surface-to-volume ratio Σ_v are not adequate. These values are corrected in the parameter estimation approach: the mean (ensemble-averaged) value of M_v is decreased from 0.22 to approximately 0.10; the mean (ensemble-averaged) value of Σ_v is increased from $11,500$ 1/m to approximately $19,000$ 1/m. These values are not corrected in the state estimation approach and therefore induce a significant bias in the wildfire spread simulations.

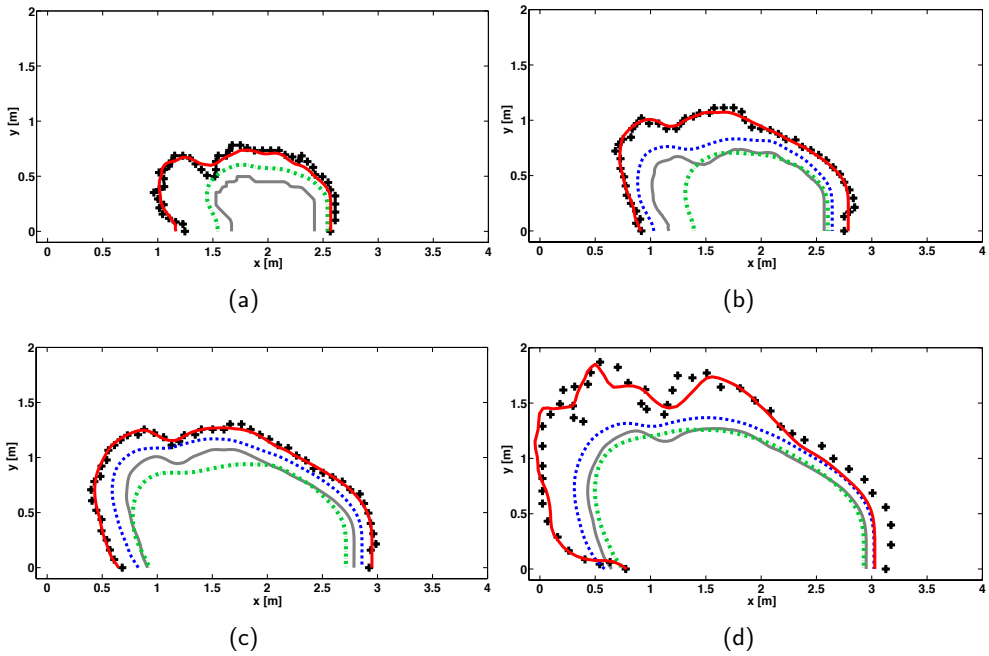


Figure 7.59: *Sequential state estimation results at (a) $t_1 = 64$ s, (b) $t_2 = 78$ s, (c) $t_3 = 92$ s and (d) $t_4 = 106$ s. Black crosses correspond to observations, the gray solid line corresponds to the initial condition of the assimilation cycle, the green dashed-dotted line corresponds to the free run, the blue dashed line corresponds to the mean forecast (without data assimilation for the first observation time or with data assimilation at the previous observation time), and the red solid line corresponds to the mean analysis (with a data assimilation update at the current observation time).*

↪ Discussion

Figure 7.58 shows that the parameter estimation approach provides better results after a prediction step (i.e. a better forecast), while the state estimation approach provides better results after an update step (i.e. a better analysis). This ranking between these EnKF estimation approaches is problem-dependent and may not hold in cases where the assimilation cycle is longer (due to a lower observation frequency) and the values of the control parameters vary significantly during an assimilation cycle (i.e., the temporal variability of the errors on the control parameters is higher than the assimilation frequency). The duration of the assimilation cycle is therefore of primary importance in the success of the proposed data assimilation approaches.

A parameter estimation approach searching for uniform parameters over the assimilation cycle may not be an efficient data assimilation strategy if the assimilation is not renewed by frequent observations and if this observation frequency does not match at least the frequency at which environmental conditions (ROS model parameters) evolve. To allow for a spatially-varying correction of the fire front position, the generation of the EnKF ensemble must represent the anisotropy in wildfire

spread that results from spatial variations in biomass properties and from the presence of temporally-fluctuating wind. This anisotropy was implicitly introduced in EnKF-based state estimation by selecting spatially-dependent biomass properties and different wind conditions between the members.

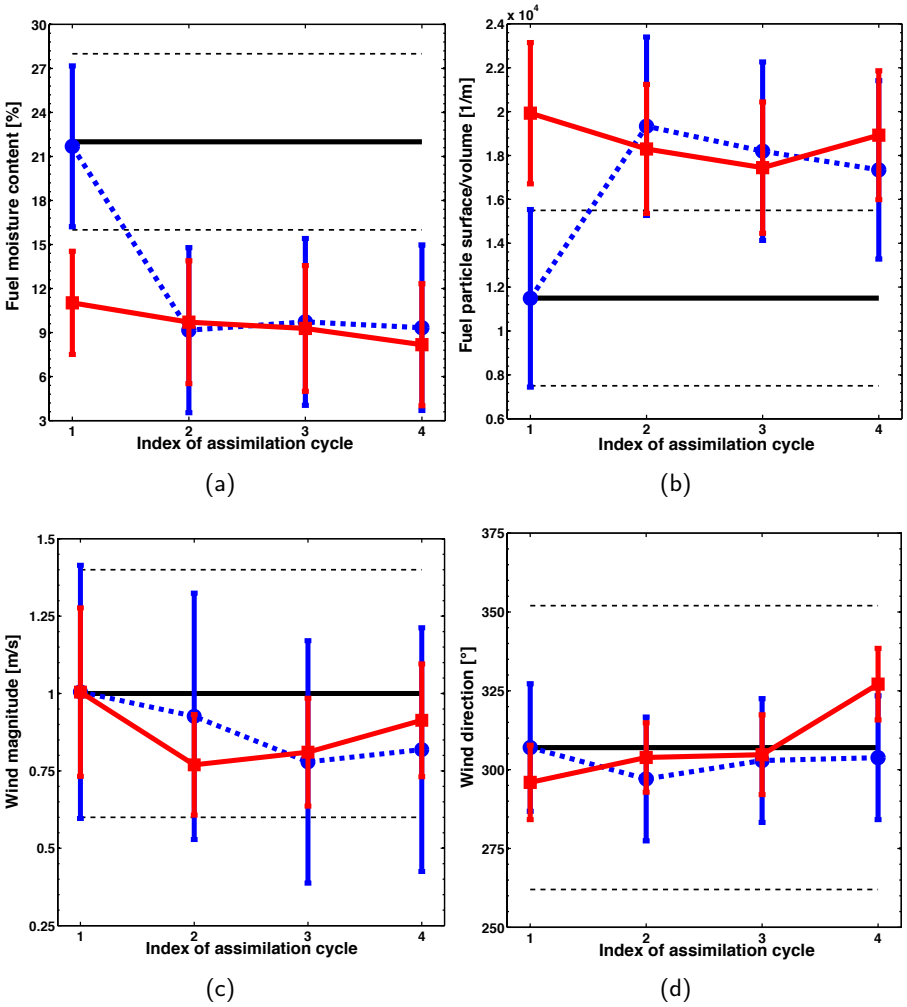


Figure 7.60: *Evolution of ROS model parameters over 4 successive assimilation cycles. Parameter estimation-based ensemble: circle symbols (connected by a dashed line) indicate mean values in the forecast ensemble; square symbols (connected by a solid line) indicate mean values in the analysis ensemble; vertical bars indicate the STD. State estimation-based ensemble: horizontal solid lines indicate mean values; horizontal dashed lines indicate STD. (a) Fuel moisture content M_v [%]. (b) Fuel particle surface-area-to-volume ratio Σ_v [1/m]. (c) Wind magnitude u_w [m/s]. (d) Wind direction α_w [°].*

However, accounting for the detailed spatial variations of these environmental conditions in a parameter estimation approach would significantly increase its computational cost, with no means of assessing the consistency of the analysis update to

in-situ measurements (since those are usually provided with a much coarser spatial resolution). The state estimation approach appears as the most suitable approach to retrieve and simulate the real shape of the fire front in strong anisotropic propagation conditions, since it allows for a non-uniform correction of the front location.

Furthermore, the performance of the state estimation approach for long-term predictions could be improved based on a more realistic prior knowledge of the control parameters (see Fig. 7.60). Thus, these results indicate that a dual state estimation/parameter estimation approach may overcome some of the limitations seen in each approach applied separately. In such dual approach, parameter estimation could be extended to the case of coarse-resolution spatial variations of the ROS model parameters.

Still, in all tests performed to date (these tests correspond to cases in which the observation error is small), the agreement between the analysis and observed fire front positions is very good and significantly better than the level of agreement that would be obtained in the absence of data assimilation.

Conclusion

This chapter presents an original application of data assimilation methodologies to the problem of simulating wildfire spread at regional scales and forecasting its behavior at future lead-times. It is assumed that airborne and/or space-borne observations of the fire front location are available at frequent times but provide an inaccurate and (possibly) incomplete description of the fire front. Thus, data assimilation provides an attractive framework for integrating (incomplete) remote sensor observations into the front-tracking wildfire spread simulator Firefly (limited to flat terrain configurations here). The prototype data-driven simulator features a parameter estimation approach or a state estimation approach based on the ensemble Kalman filter (EnKF). In the parameter estimation approach, the control variables are the input parameters of the spread-rate model due to Rothermel (e.g., wind conditions, vegetation properties) in order to provide feedback to Firefly on the future wildfire behavior. This approach is limited to spatially-uniform distribution for the control parameters due to the absence of high-resolution information on these parameters and to remain consistent with operational forecast. In the state estimation approach, the control variables are the locations of the fire front markers (these markers being the chosen representation of the simulated fire front). This approach aims at providing spatially-distributed corrections of the fire front location and at accounting for modeling uncertainties that cannot be assessed by uncertainties in the spread-rate model parameters, for instance due to the chaotic nature of wildfire spread or due to knowledge gaps in the spread-rate model formulation itself.

The prototype data-driven simulator was evaluated in a series of synthetic cases, including configurations with spatially-varying vegetation properties and temporally-varying wind conditions, and in a validation test corresponding to a controlled grassland burning experiment. In all test cases, data-driven simulations were successful at significantly decreasing the distance between observed and simulated fire fronts and at providing an optimized forecast of the wildfire behavior. In particular, the following aspects are important to mention.

- ▷ **Ensemble-based data assimilation.** While used as a preliminary approach to wildfire spread forecast in Rochoux et al. (2013, PROCI) and Rochoux et al. (2013, INCA), the choice of the extended Kalman filter (EKF) algorithm was considered questionable because it assumes a linear relationship between control space and observation space; this linear assumption is believed to be of limited value in general wildfire problems in which the wind conditions may vary and the vegetation properties are potentially strongly heterogenous. The modification to an EnKF ap-

proach was explored in Rochoux et al. (2012, CTR), in which some of the observation model non-linearities were accounted for.

- ▷ **Comparison to particle filters.** Despite of its linear combination of ensemble members as well as of its Gaussian assumption on the modeling and observation error statistics, the optimality of the EnKF was demonstrated based on comparative studies to particle filters that are more general Bayesian filters and that still produced equivalent results to the EnKF, see da Silva et al. (2014, HTHP) and da Silva et al. (2013, IPDO).
- ▷ **Representation of the error spatial variability.** This work also shows that in order to allow for a spatially-varying correction of the front position, the generation of the EnKF-based ensemble should represent the anisotropy in fire propagation, due to spatial variations in vegetation properties and due to the presence of temporally-fluctuating wind. This anisotropy was implicitly introduced in the EnKF-based state estimation by selecting spatially-dependent vegetation properties and different wind conditions between the members, see Rochoux et al. (2014, IAFSS) and Rochoux et al. (2014, PROCI). A significant gain in accuracy resulted in the data-driven simulations compared to the proposed parameter estimation approaches, which were limited to a uniform description of the input parameters of the spread-rate model to remain consistent with operational framework. The benefits of the state estimation approach are even more evidenced in cases in which the observations are made with significant error and/or cases in which the observations are incomplete, e.g., when only a fraction of the fireline perimeter is observed.
- ▷ **Reduced-cost parameter estimation.** Realistic values for the control parameters could be accurately inferred, with limited equifinality issues, if the number of members in the ensemble is large enough to describe properly the error correlations between the members. However, there is no criterion to assess the accuracy of parameter estimation in real-world applications. It is therefore important to produce large ensembles. However, accounting for the spatial variability of the spread-rate model parameters would increase the size of the control vector in a parameter estimation approach and the required size of the sample. Consistently, it would drastically raise the computational cost of the data assimilation procedure and devote computational time to correct the control parameters in zones of no interest at the assimilation time (i.e., far away from the front), where no information could be obtained on the wildfire behavior. Even though the correction of the control parameters could be restricted to the flame zone, spatially-distributed parameters are difficult to validate since high-resolution spatial distributions of environmental conditions are

not yet available. For this purpose, a reduced-cost EnKF strategy based on polynomial chaos (PC-EnKF) was designed in Rochoux et al. (2012, CTR), the polynomial-based surrogate model being used in place of the forward model Firefly during the EnKF prediction step to generate a large number of model simulation members at no cost and without loss of accuracy. The benefits of this PC-EnKF strategy are explored in Rochoux et al. (2014a, NHESS), with an extensive comparison to the EnKF performance.

- ▷ **Forecast performance.** The resulting data-driven simulator can in turn be used to provide a near-future forecast of the wildfire location. Results indicate that the forecast performance of a parameter estimation approach or a state estimation approach may be limited to near-term predictions (i.e., at short lead-times). The forecast performance highly depends on the consistency between the assimilation frequency and the spatial/temporal variability of all the errors that are corrected in the data assimilation procedure (either the ROS model parameters or the locations of the fire front markers). In particular, the variability of these errors is determined by the persistence of the initial condition of the wildfire spread model (i.e., the time period over which memory effects induced by the dependence on the initial condition affect the simulation) as well as by the temporal and spatial variability of the environmental conditions (e.g., wind conditions, vegetation properties) as shown in Rochoux et al. (2014, PROCI) and Rochoux et al. (2014b, NHESS). For instance, if the error variability in the parameters do not change over time, then a parameter estimation approach exhibits a high persistence. The size of the assimilation window (for which constant control variables are searched) must be specified adequately (if the observation frequency is compatible) to capture the sudden changes in wildfire behavior that are the most critical to predict.

To guarantee a high level of performance in wildfire spread forecasts, the data assimilation procedure (for the state estimation approach or for the parameter estimation approaches) needs to be renewed by frequent observations with a frequency in adequation to the temporal variability of the errors on the estimation targets.

Conclusion

Real-time predictions of the direction and speed of a propagating wildfire (typical of Fig. 7.61) have been identified as a valuable research objective with direct applications in fire emergency management. While much progress has been achieved over the past few decades in the basic understanding of wildfire dynamics, while also much progress has been achieved in the mathematical formulation and numerical simulation of wildfire spread, forecasting reliable scenarios of wildfire spread at an operational level remains a challenging task because the problem involves both multi-physics and multi-scales. In order to overcome some of the current limitations of regional-scale wildfire spread modeling, uncertainties in the mathematical representation of the wildfire spread as well as in the input parameters or external forcing required by the models need to be quantified and reduced. For this purpose, this thesis explored two complementary strategies, model evaluation and data assimilation. They include the study of the multi-physical processes occurring at flame scales to address knowledge gaps in fire spread modeling as well as the development of a prototype data-driven simulator to improve the ability to forecast both regional-scale wildfire dynamics and plume emissions at future lead-times.

Figure 7.61: *Airborne snapshot of Perthus regional-scale wildland fire (Pyrénées-Orientales, France, July 2012) © Pauline Crombette.*

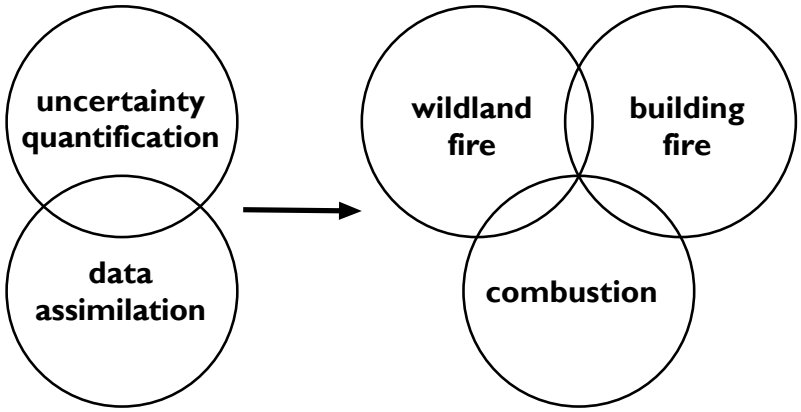
In Part I (Insight from multi-physics flame-scale large-eddy simulations), multi-physics large-eddy simulations of fire spread were performed at laboratory flame scale and compared to measurements to provide a comprehensive understanding of the mechanisms underlying natural fires. In particular, a coupling strategy based on high-performance computing and a multi-model multi-solver approach (based on a coupling between a large-eddy simulation solver, a radiation solver and a biomass thermal degradation solver) was developed to simulate propagating buoyant diffusion flames, corresponding to a limit case, in which no air is premixed with pyrolysis gases before flame ignition. From our perspective, this study constitutes a novel and original application of a multi-physics computational fluid dynamics strategy to natural fire propagation, solving for the detailed flame structure (at a scale on the order of 1 mm) and its interaction with turbulence, including radiation heat transfer, detailed combustion chemistry and biomass pyrolysis. Valuable information were obtained on the characteristic flame features of laboratory-scale fires, in terms of chemical structure, burnt gas products, puffing behavior, slope-induced effects as well as spatially-distributed flame radiation received by vegetation. While further validation studies are required, the proposed coupling strategy constitutes a preliminary step towards fully physical simulations of fire spread at flame scale, allowing for future improvement of each model component and for the exploration of the wildfire response to varying environmental conditions. It also constitutes a reference to evaluate and improve semi-empirical modeling approaches used in current regional-scale wildfire spread simulators.

Part II (Data assimilation for regional-scale wildfire spread forecast) presents a novel analysis of the potential benefits of ensemble-based data assimilation techniques for wildfire spread. Indeed, uncertainties in wildfire spread modeling go beyond the limits of deterministic forecast abilities with the dynamical model and recent progress made in airborne remote sensing provides new ways to monitor real-time fire front locations. These thermal-infrared measurements provide an incomplete description of the fire spread (in particular due to the opacity of the thermal plume induced by the fire) and are subject to instrumental errors as well as representativeness errors. A forecast capability must therefore rely on a computer simulation tool that is compatible with operational forecast to provide an accurate prediction of the wildfire behavior, even in the zones where the flame is not observed.

All data assimilation approaches share the idea of merging measurements and computer models, while accounting for both observation and modeling errors. However, a wide range of sophistication exists between the different existing algorithms. Thus, the main challenge in this work was to determine the most suitable data assimilation procedure to wildfire spread, i.e., a procedure that accounts for non-linearity in wildfire behavior, that handles spatial and temporal variability in the errors and that reliably delivers an accurate forecast of the time-evolving location of the fire front with a positive lead-time (i.e., with simulations achieved faster than real-time). The ensemble Kalman filter is therefore retained in this work, in order

to stochastically characterize the non-linear response of the specifically-developed regional-scale wildfire spread simulator FIREFLY to variations in the input environmental conditions (e.g., wind conditions, vegetation properties) and in the fire front locations. This work emphasizes the potential of data assimilation to dramatically increase fire simulation accuracy and produce high-fidelity data-driven simulations of wildfire spread. It also emphasizes the need for a reduced-cost strategy to remain consistent with operational forecast. We propose therefore data-driven modeling as one of the two cornerstones of a fire spread forecast capability. The other cornerstone corresponds to the integration of a variety of in-situ and remote sensors providing (real-time) information on fire location, vegetation, terrain topography and atmospheric conditions.

While fire spread forecast capabilities are still at an early stage of development, it is envisioned that they will be similar to current weather forecasting capabilities and that the general ability to predict the evolution of wildfires will rely on the continuous assimilation of remote sensing observations into a multi-physics fire model (accounting for fire surface propagation and atmospheric dynamics). It is also envisioned that these future capabilities for forecasting wildfire spread scenarios will not uniquely rely on an unique spread-rate model but instead on a variety of spread-rate models that are characterized by different validity ranges and whose prediction capacity can thereby vary for different fire regimes. As a complement to data assimilation, these spread-rate models could be drastically improved using the critical analysis derived from multi-physics flame-scale large-eddy simulations (currently restricted to research projects since they are not compatible with real-time forecast operational objectives). Still, data assimilation has already shown great potential to relate comprehensively computational fire modeling and fire sensor technology, which is not yet available in fire research and on a broader level, in combustion.



Contributions of this thesis

The main contributions of this thesis are summarized below:

- ▷ the **development of the regional-scale wildfire spread simulator** FIREFLY, including a spread-rate model based on Rothermel's formulation and a level-set-based front-tracking solver able to deal with heterogeneous biomass fuel properties, time-varying wind conditions and complex terrain topography.

- ▷ the **development of a biomass thermal degradation model** PYROWO, based on a 0-D formulation of the solid phase of porous biomass fuel subject to an external radiation heating and on a calibration of model parameters. This approach is cost-effective to allow for multi-physics multi-model coupling.

- ▷ the **use of a generic platform** for these developments, based on the dynamic code coupler OPENPALM, in particular:
  - the development of a coupling strategy based on high-performance computing and message passing interface, including the AVBP large-eddy simulation solver, the PRISSMA radiation solver and the PYROWO biomass thermal degradation solver, to explicitly solve for the strong coupling between the gas and solid phases due to mass, momentum and energy transfers;
  - the development of a generic scheme for applying the ensemble Kalman filter (EnKF) to wildfire spread, in which the fire spread simulator FIREFLY can be easily modified and which can serve as a good starting point for future applications of data assimilation (e.g., biomechanics, Peyrounette, 2013).
  - the validation and application of the PARASOL functionality of OPENPALM, recently developed at CERFACS and particularly attractive for Monte-Carlo simulations, for uncertainty quantification as well as for data assimilation purposes. In particular, the reduced-cost Kalman filtering strategy (PC-EnKF) developed in this thesis can be integrated as a new generic functionality in OPENPALM;

- ▷ the demonstration of the **feasibility of particle image velocimetry** for inferring the flame-induced air entrainment.

- ▷ the **introduction of a paradigm-shift in wildfire modeling**, including:
  - a novel and original description of the observation operator for wildfire spread, which relates the input parameters of the spread-rate model to the model counterparts of the observed fire front. The definition of the observations as a discretized fire front is able to capture much of the observation information content and allows for approximate Gaussian

error statistics and thereby, a straightforward application of ensemble-based data assimilation;

- the comparison of benefits and drawbacks between a multi-parameter multi-cycle estimation approach (well-suited for current parallel computing platforms) and a state estimation strategy (able to deal with high anisotropy in the rate and direction of wildfire spread);
- the development and validation of a new reduced-cost data assimilation strategy based on the construction of a surface response model compatible with the operational framework;
- an original application of the data-driven simulation capability to a real controlled burning experiment;
- the study of flame-scale processes governing fire spread such as flame-induced radiation heat transfer, combustion of pyrolysis gases and puffing.

▷ the **emergence of collaborations** including:

- the multi-disciplinary, international collaboration between Ecole Centrale Paris (France), CERFACS (France) and the University of Maryland (USA) bridging the gap between the different related but separate research communities that are involved in this thesis (compartment fire, wildfire, data assimilation and combustion);
- the collaboration between CERFACS (France), Mines Albi (France) and the Federal University of Rio de Janeiro (Brazil) for the application of particle filters to wildfire spread forecast and the comparison to data assimilation methodologies, with a common research project with Wellington da Silva;
- the collaboration between CERFACS (France) and Institut d'Alembert (France) for the application of polynomial chaos to wildfire spread forecast, with a common research project with Didier Lucor;
- the active participation in the research training of master students for engineering and scientific fire and data assimilation applications (Blaise Delmotte, Charlotte Emery, Clément Doche, Myriam Peyrounette).

Future lines of approach

> *After climbing a great hill, one only finds*
> *that there are many more hills to climb.*
> Nelson Rolihlahla Mandela (1918-2013).

▷ Towards more realistic flame-scale large-eddy simulations

Since the coupling approach for flame-scale large-eddy simulations is modular, it constitutes a valuable starting point to improve the underlying models based on systematic quantitative comparison to laboratory-scale measurements. In particular, the following studies could be useful to gain more physical insight into the physical mechanisms governing wildfire spread:

- **Coupling capability.** A comparative study of physics-based fire spread simulators (e.g., WFDS, FIRETEC) with the coupling capability proposed in this work could be performed to further assess the validity of our approach and to investigate its benefits for furthering our understanding of wildfire controlling processes. The validity of the underlying models could also be analyzed for varying fire spread conditions, in conjunction with the simultaneous development of adapted experimental diagnostics to reach reliable quantitative comparisons between simulation and experiment.
- **Pyrolysis.** The formulation of the PYROWO biomass thermal degradation model could be extended to a one-dimensional model accounting for the spatial (vertical) profile of flame-induced radiation in the vegetation layer and for radiation heat transfer due to the biomass solid particles within the vegetation layer. Also, the models for the different physical processes underlying biomass thermal degradation (i.e., drying, pyrolysis, char oxidation) could be evaluated against more complex physics-based models (that are not affordable for large-eddy simulations) and against a more complete experimental dataset based on thermogravimetry but also on cone calorimeter experiments. This evaluation could be performed for varying burning conditions, in terms of flame-based radiation and biomass properties.
- **Combustion.** Flame-scale simulations need to be performed for varying composition of the pyrolysis gases to mimic the behavior of partially-premixed flames and analyze the differences with the limit case of diffusion flames tested in this work. The fuel/oxidizer equivalence ratio is indeed unknown and its impact on the flame structure needs to be quantified to improve knowledge on the effective burning conditions in real-world fire cases. Such characterization is useful to choose the most suitable chemical kinetic scheme to describe the combustion of pyrolysis gases as well as to design an adequate turbulent

combustion model that accurately accounts for subgrid-scale buoyancy and strain effects.

- **Radiation.** The present work focuses on the study of the gas phase, but a multi-scale analysis needs to be performed to quantify the amount of soot that is produced during biomass fuel pyrolysis and that is convected to the gas phase, where its contribution to radiation (in addition to $\mathrm{CO_2}$ and $\mathrm{H_2O}$) may be significant and essential to obtain consistent rate of spread values.
- **Wildfire emission.** Detailed simulations of wildfire spread are useful to describe the production of burnt gases at flame scale and especially to track the formation and oxidation of pollutants such as CO. Quantifying the amount of CO that is released from the flame zone into the atmosphere could be valuable to improve wildfire air prediction tools, for instance through the improvement of emission factors.

▷ Towards operational application of the data-driven strategy

This work demonstrated that in order to allow for a spatially-varying correction of the front position, the generation of the forecast estimates must represent the anisotropy in fire propagation that results from spatial variations in vegetation properties and from temporally-fluctuating wind conditions. A spatially-distributed correction along the fireline was obtained for a state estimation approach, while a parameter estimation approach was necessary to obtain persistent predictions at future lead-times. Future plans to build an efficient strategy to track regional-scale fire spread are based on the following components:

- **Dual state/parameter estimation approach.** A dual state/parameter estimation (Moradkhani et al., 2005) would overcome the limitations illustrated in the present and past studies. The parameter estimation approach could be extended to the case of weak spatial variations of the spread-rate model parameters. Assuming that the errors on the parameters vary slowly in time, the correction provided by data assimilation could reasonably be used for forecast, thus allowing for mid- to long-term forecast. In addition, the state estimation approach could be used for short-term forecast in order to locally correct the shape of the fire front.
- **Extensive validation for regional-scale wildfire spread.** The wildfire spread model FIREFLY was extended to treat configurations with complex topography as explained in Appendix G (see Section G.3). While featuring realistic fire spread, an extensive validation study including representative field-scale wildfires needs to be performed to evaluate its performance for complex terrain topography. Once this complex terrain capability is available, the proposed data-driven strategy could be evaluated for regional-scale wildfire spread for which in-situ, airborne and/or spaceborne data are available at high temporal and spatial resolution.

- **Fire/atmosphere interactions.** Future plans also include the integration of the data assimilation algorithm into a CFD atmospheric solver in order to describe the interactions between the fire and the atmosphere. One example of a coupled fire/atmosphere model is ForeFire/Méso-NH (Filippi et al., 2013), see Section 1.3.3, Chapter 1. The developments of data-driven modeling for coupled fire/atmosphere models require a significant effort as the atmospheric state must be corrected in coherence with the correction of the fire front location. Furthermore, a high spatial resolution description of vegetation, meteorological conditions and terrain topography is required to improve forecast accuracy at short lead-times. Such methodology could be useful to better quantify pollutant/aerosol emissions by wildfires (Filippi et al., 2011; Strada et al., 2012). For instance, this is the objective of the proposed FireCaster project between SPE, LA (CNRS/Paul Sabatier University, Toulouse), CERFACS, EM2C, CNRM (CNRS/Météo-France), INRIA, which is aimed (based on the ANR-IDEA project) at developing a crisis probabilistic model that can quickly evaluate multiple fire spread scenarios in case of emergency.
- **Improved data-driven strategy.** From a technical perspective, the relationship between control space and observation space (i.e., the observation operator) in the data-driven strategy could be re-visited. In the present work, the observations are the time-evolving location of the fire front that is obtained by a post-processing of mid-infrared imaging (FRP products, see Section 1.4, Chapter 1). One possible strategy currently investigated by the wildfire research group at the Department of Geography of King's College London[6] is to directly define the observation as the FRP map to avoid introducing post-processing errors. The data-driven strategy would then rely on data assimilation of images, which appear as the next step towards real-world applications of data assimilation for wildfire spread.
- **Assimilation of spaceborne and/or airborne measurements.** New observing systems are now available for wildfire spread monitoring from space, for instance aerial thermal imagery or remote sensing at high resolution with the Pléiades constellation for instance, from which the position of the fireline can be detected. These data could be combined with the data assimilation strategy proposed in this work, to improve the representation of wildfire dynamics. The SERTIT[7] for instance provides regional-scale geo-localized image of fires from a rich combination of remote sensing data.

The ultimate goal of this research is to provide real-time fire forecasts using thermal-infrared imaging data including a description of both wildfire dynamics and fire plume emissions.

[6] wildfire.geog.kcl.ac.uk/.

[7] Service Régional de Traitement d'Image et de Télédétection, Univ. de Strasbourg (France), sertit.u-strasbg.fr/RMS/.

▷ Extension of data assimilation to combustion applications

In the context of gas turbines, where CFD numerical models have substantive sources of uncertainties in initial/boundary conditions, numerical and physical parameters, such a data assimilation methodology may be used for example to improve the prediction of flame ignition and propagation by a quantitative analysis of images and measurements. Until now, diagnostics and CFD simulations have been mostly developed independently, with the same goal of quantifiable accuracy to study reacting flows in a time-resolved, simultaneous, multi-dimensional framework. Introducing a data assimilation approach in combustion would mean to use measurements (images or probe signals) to reduce uncertainties on, for example, the turbulent flame speed or the heat transfer to the walls. This could also be applied to unsteady processes such as burner ignition or extinction. It may directly improve simulation results, through improved initial and/or boundary conditions (e.g., spray injection, wall temperatures), or improve physical models when applied to parameters (e.g., soot concentration, flame structure in terms of wrinkling and thickening). By combining experiments and large-eddy simulations along with their uncertainties using mathematical tools such as data assimilation or uncertainty quantification, the accuracy of predictions of turbulent flames could be improved and uncertainties could be controlled. For example, flame front observations in a burner, obtained with thermal-infrared visualizations or Laser-Induced Fluorescence (LIF) of the OH radical, could be used for a data assimilation approach with large-eddy simulation of the burner, to improve the accuracy of the results or the turbulent combustion model. Thus, data assimilation could be a powerful framework to better understand the relationship between measurements and simulated reacting flow variables as well as to increase the accuracy of both experimental and numerical techniques.

▷ Further developments of data assimilation based on polynomial chaos

A reduced-cost EnKF strategy based on polynomial chaos (PC-EnKF) was proposed in this thesis, the polynomial-based surrogate model being used in place of the forward model FIREFLY during the EnKF prediction step to generate a large number of model simulation members at no cost and without loss of accuracy. As highlighted by Li and Xiu (2009) and Rosic et al. (2013), alternative polynomial chaos strategies can be explored to reduce the cost of Kalman filtering and to better address model non-linearity. This is equivalent to project the Kalman update equation onto the polynomial chaos basis. For instance, the estimation targets could be directly the modes of the polynomial chaos decomposition in parameter space or in model state space. One advantage of this alternative approach is to avoid Monte-Carlo sampling, which introduces sampling errors during each assimilation cycle. The benefits of such alternative polynomial chaos approach need to be further investigated.

Appendix A

Models of rate of spread: Rothermel versus Balbi

Inspired by past analysis studies from Fons (1946), Byram and Fons (1952), Emmons (1964) and Frandsen (1971), the rate of spread (ROS) model due to Rothermel (1972) laid the foundations of the US operational fire danger rating system (NFDRS) and of the fire behavior simulators Behave *(Andrews, 1986) as well as* Farsite *(Finney, 1998). Due to its breakthrough in fire modeling, Rothermel's model is still a reference in the wildfire research field. New ROS formulations have been proposed in the literature to overcome some of the uncertainties and knowledge gaps present in the original semi-empirical approaches, e.g., Cheney et al. (1998), Balbi et al. (2009). In this appendix, the ROS models due to Rothermel and Balbi are reviewed along with a sensitivity study to highlight the crucial input parameters to control in a data-driven simulation capability.*

A.1 Rothermel's model of rate of spread

A.1.1 Historical background

Early work in wildfire spread research conducted by the US Forest Service aimed at quantifying the effect of measurable variables (e.g., wind velocity, fuel moisture content, fuel density, fuel compacity) on the fire ROS using both laboratory-based and field-based burning experiments (see Fig. A.1). Fons (1946) laid the first stones of a mathematical model for wildfire spread with a focus on the head of the fire. This theoretical analysis of wildfire spread states that a significant amount of heat is required ahead of the flame to bring the surrounding vegetation to its ignition temperature and that a wildfire can be regarded as a series of successive ignitions in the vegetation, or as the displacement of the biomass pyrolysis zone towards unburnt vegetation. Thus, the ROS is primarily controlled by the ignition time and

the properties of the biomass solid particles as illustrated in Fig. A.2.

Figure A.1: *Wind-tunnel experiments of fire spread: the test section is* 9.2 *m long with a cross section of* 1.8 *m* $\times$ 1.8 *m and pine needles are used to model typical Californian vegetations. Credit: Fons (1946), US Forest Service.*

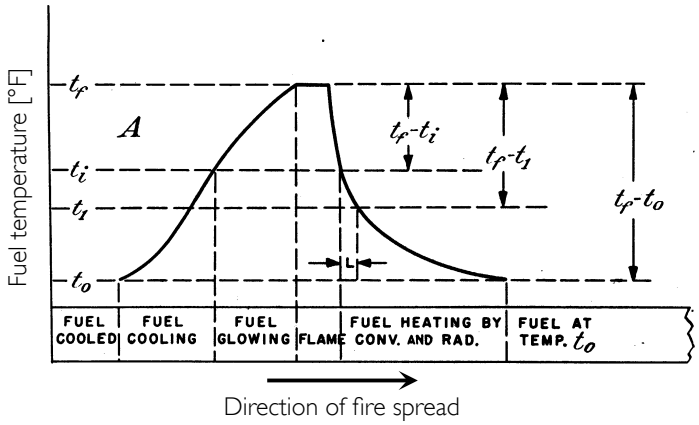


Figure A.2: *Schematic of temperature gradients at/near the flame region. Credit: Fons (1946).*

Preliminary studies performed by Fons (1946) provided the key elements to understand the mechanisms underlying wildfire spread and led to a model that mathematically translates these physical insights. In particular, Frandsen (1971) proposed to apply an energy balance equation within a control volume of the vegetation located ahead of the fire front as illustrated in Fig. A.3. Frandsen's analysis assumed an infinite straight fireline along the y-direction that spreads at a constant rate Γ [m/s] (i.e., quasi-steady state assumption) and distinguished the components along x- and z-directions of the propagating heat flux I_p received in the control volume. The steady-state ROS can be formulated as the ratio of the flame-induced heat flux (primarily by radiation) to the heat required for ignition of the biomass

fuel, i.e.,

$$\Gamma = \frac{I_p}{\rho_b \, \varepsilon \, Q_{ig}} = \frac{1}{\rho_b \, \varepsilon \, Q_{ig}} \left[I_{x,ig} - \int_0^{+\infty} \left(\frac{\partial I_z}{\partial z} \right)_{z=z_c} \mathrm{d}x \right], \tag{A.1}$$

with:

- $\rho_b \, \varepsilon$ [kg/m^3] the effective fuel density (i.e., the amount of vegetation per unit volume of the fuel bed raised to ignition ahead of the advancing fire);
- Q_{ig} [J/kg] the heat of pre-ignition (i.e., the heat required to bring a unit weight of fuel to ignition);
- $I_{x,ig}$ [W/m^2] the horizontal heat flux along the x-direction received by the control volume at the time of ignition;
- $\left(\frac{\partial I_z}{\partial z} \right)_{z=z_c}$ [W/m^3] the gradient of the vertical intensity along the z-direction evaluated at a plane at a constant depth z_c of the vegetation bed.

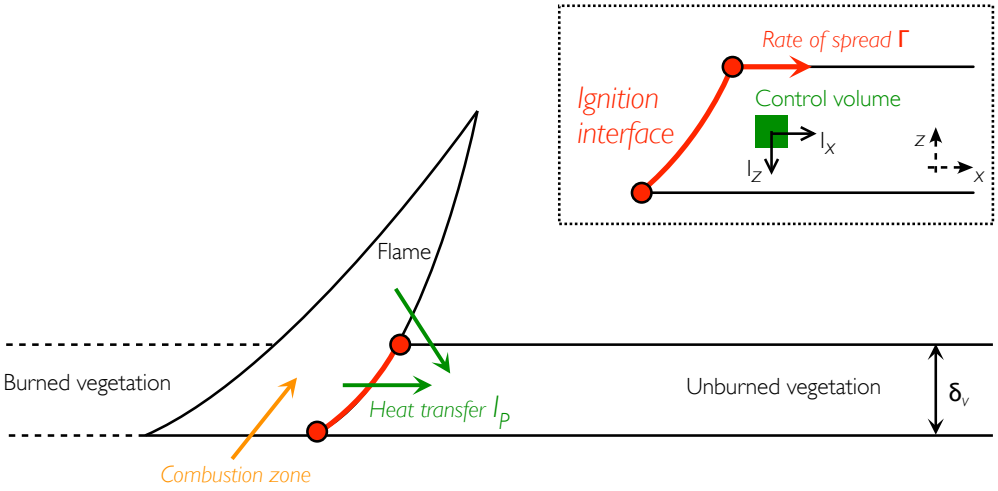


Figure A.3: *Schematic of the wildfire spread mechanism. Credit: Dupuy and Valette (1997).*

Based on the mathematical formulation proposed by Frandsen (1971), Rothermel (1972) demonstrated that the vertical component of the propagating heat flux I_p only matters in the case of a wind-aided or up-slope fire propagation due to the higher flame tilt angle towards the unburned vegetation, which enhances radiation and convection heat transfer (see Fig. A.4). To the contrary, for a no-slope no-wind fire propagation, the contribution to the propagating heat flux is mainly horizontal within the vegetation layer; this horizontal heat flux is noted $I_{p,0}$. In the Rothermel's ROS formulation, $I_{p,0}$ is a constant heat flux to which additional effects of wind and slope are combined as follows:

$$I_p = I_{p,0} \, (1 + \phi_w^* + \phi_{sl}^*) \, . \tag{A.2}$$

This expression for the propagating heat flux I_p means that the wind and slope effects on the ROS are assumed additive and proportional to the no-wind no-slope propagating heat flux $I_{p,0}$ through the positive coefficients ϕ_w^* and ϕ_{sl}^*, respectively. This also implies that radiation is the dominant external heating received by the unburnt vegetation ahead of the fire front.

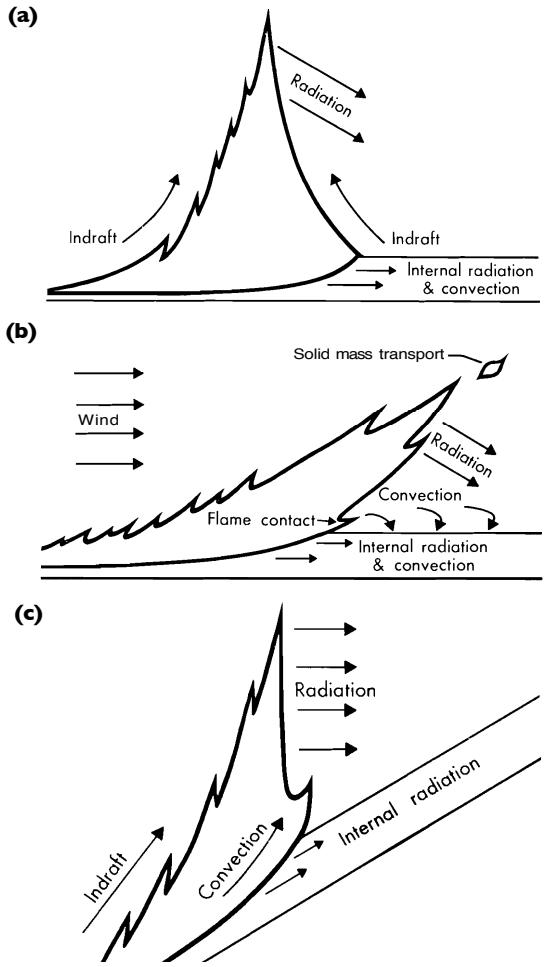


Figure A.4: *Schematic of the wildfire spread for different environmental conditions. (a) No-wind no-slope fire. (b) Wind-driven fire. (c) Up-slope fire. Credit: Rothermel (1972).*

Rothermel added two points to Frandsen (1971) analysis:

(1) The no-wind no-slope propagating heat flux $I_{p,0}$ is directly related to the energy release rate of the combustion I_r through the following relationship:

$$I_{p,0} = \chi\, I_r, \tag{A.3}$$

with χ the dimensionless propagating flux ratio, which describes the proportion of the flame heat release transferred to the vegetation in the non-flaming zone.

(2) A heterogeneous formulation for the vegetation is included to represent a natural fuel bed composed of particles varying in size, shape, density and specific heat. This heterogeneous formulation is based on the weighting concept, meaning that each ROS model parameter can be estimated by properly weighting the variations in this parameter in the heterogeneous vegetation layer. For this purpose, biomass fuels are grouped into categories of similar properties (e.g., dead/living cells, size of solid particles).

As a result, the ROS $\mathit{\Gamma}$ [m/s] proposed by Rothermel reads:

$$\mathit{\Gamma} = \mathit{\Gamma}_0 \, (1 + \phi_w^* + \phi_{sl}^*) = \frac{\chi \, I_r}{\rho_b \, \varepsilon \, Q_{ig}} \, (1 + \phi_w^* + \phi_{sl}^*) \, , \tag{A.4}$$

where $\mathit{\Gamma}_0$ represents the minimal value for the ROS achieved for no-wind no-slope conditions, while ϕ_w^* and ϕ_{sl}^* represent the additive wind and slope effects on $\mathit{\Gamma}$.

A.1.2 List of input/model parameters

The one-dimensional formulation of the ROS $\mathit{\Gamma}$ proposed by Rothermel (1972) requires 11 input parameters described in Table A.1. The physical quantities involved in Eq. (A.4) such as the combustion-induced energy release rate I_r or the wind/slope correction coefficients ϕ_w^* and ϕ_{sl}^* were parameterized with respect to these parameters using statistics derived from wind-tunnel experiments. Thus, Eq. (A.4) can be expressed as follows:

$$\mathit{\Gamma} = \mathit{\Gamma}\Big(\delta_v, \beta_v, M_v, M_{v,ext}, \Sigma_v, m_v^{''}, \rho_p, \Delta h_c, s_t, s_e, u_w, \alpha_{sl}\Big). \tag{A.5}$$

Among this list of input parameters, the moisture content at extinction $M_{v,ext}$, the fuel particle mass density ρ_p, the fuel low heat of combustion Δh_c, the fuel particle total mineral content s_t and the fuel particle effective mineral content s_e are assumed to be independent of the biomass fuel type; their *nominal value* in the international system (SI) of units is given in Table A.1 and is partly justified below:

- ▷ As explained in Chapter 3, the vegetation moisture content at extinction $M_{v,ext}$ represents the threshold value of the moisture content M_v, over which the fire stops propagating and combustion cannot sustain itself due to the prohibitive amount of energy required to evaporate the biomass moisture. The nominal value $M_{v,ext} = 30 \ \%$ was determined experimentally using the fiber saturation point of common biomass fuels.
- ▷ The fuel particle total mineral content s_t, evaluated on average at $5.55 \ \%$, represents the ratio of mineral mass to oven-dry-wood mass, i.e., the non-combustible mass content of the vegetation that is responsible for ash formation during the thermal degradation process.

The biomass fuel database related to Rothermel's ROS model is presented in Table A.2; only the properties of the fine dead particles are mentioned since they are the main particles involved in the propagation of the fire front.

Table A.1: *Input parameters of the Rothermel's ROS model.*

Name	Symbol	Unit	Nominal value
Fuel depth (vertical thickness of the vegetation layer)	δ_v	m	-
Fuel packing ratio	β_v	%	-
Fuel moisture (mass of water divided by mass of dry vegetation)	M_v	%	-
Fuel moisture at extinction	$M_{v,ext}$	%	30.0
Fuel particle surface-to-volume ratio	Σ_v	1/m	-
Fuel loading	$m_v^{''}$	kg/m^2	-
Fuel particle mass density	ρ_p	kg/m^3	512.4
Fuel low heat of combustion	Δh_c	J/kg	18.608×10^6
Fuel particle total mineral content	s_t	%	5.55
Fuel particle effective mineral content	s_e	%	1.0
Wind velocity at mid-flame height (projected into horizontal plane)	u_w	m/s	-
Terrain slope angle	α_{sl}	$\circ$	-

A.1.3 Submodels underlying Rothermel's formulation

Physical quantities of interest are defined in the original Rothermel's ROS formulation using the units of the Bristish Imperial system (instead of the SI units). In the following presentation of the ROS submodels, these quantities are defined in the SI units for clarity purposes.

▷ Reaction intensity I_r [W/m^2]

$$I_r = \gamma \, m_n^{''} \, \Delta h_c \, n_m \, n_s. \tag{A.6}$$

▷ Optimum reaction velocity γ [s^{-1}]

$$\gamma = \gamma_{\max} \left(\frac{\beta_v}{\beta_{v,op}} \right)^A \exp \left[A \left(1 - \frac{\beta_v}{\beta_{v,op}} \right) \right], \tag{A.7}$$

with $A = \left(4.774 \, (\Sigma_v)^{0.1} - 7.27 \right)^{-1}$.

▷ Maximum reaction velocity $\gamma_{\max}$ [s^{-1}]

$$\gamma_{\max} = \Sigma_v^{1.5} \left(495 + 0.0594 \, \Sigma_v^{1.5} \right)^{-1}. \tag{A.8}$$

Table A.2: *Rothermel's database for 11 fuel models ($\overline{\beta_v}$ represents the mean packing ratio over all particle sizes).*

Fuel name	Fine particles			δ_v [m]	$\overline{\beta_v}$ [%]
	Σ_v [1/m]	m_v'' [kg/m^2]	β_v [%]		
(1) Grass (short)	11483	0.166	0.106	0.305	0.106
(2) Grass (tall)	4921	0.674	0.173	0.762	0.172
(3) Brush	6562	0.225	0.072	0.610	0.252
(4) Chaparral	6562	1.123	0.120	1.829	0.383
(5) Timber (grass)	9843	0.449	0.192	0.457	0.383
(6) Timber (litter)	6562	0.337	1.078	0.061	3.594
(7) Timber (understory)	6562	0.674	0.431	0.305	1.725
(8) Hardwood (litter)	8202	0.654	2.094	0.061	2.500
(9) Slash (light)	4921	0.337	0.216	0.305	1.653
(10) Slash (medium)	4921	0.8984	0.25	0.701	2.156
(11) Slash (heavy)	4921	1.5721	0.335	0.914	2.779

▷ Optimum packing ratio $\beta_{v,op}$ [-]

$$\beta_{v,op} = 3.348\,\Sigma_v^{-0.8189}. \tag{A.9}$$

▷ Bulk mass density ρ_b [kg/m^3]

$$\rho_b = \beta_v\,\rho_p. \tag{A.10}$$

▷ Fuel loading m_v'' [kg/m^2]

$$m_v'' = \frac{\rho_b\,\delta_v}{1+s_t}. \tag{A.11}$$

▷ Net fuel loading m_n'' [kg/m^2]

$$m_n'' = \frac{m_v''}{1+s_t}. \tag{A.12}$$

▷ Moisture damping coefficient n_m [-]

$$n_m = 1 - 2.59\left(\frac{M_v}{M_{v,ext}}\right) + 5.11\left(\frac{M_v}{M_{v,ext}}\right)^2 - 3.52\left(\frac{M_v}{M_{v,ext}}\right)^3. \tag{A.13}$$

▷ Mineral damping coefficient n_s [-]

$$n_s = 0.174\,s_e^{-0.19}. \tag{A.14}$$

▷ Propagating heat flux χ [-]

$$\chi = (192 + 0.2595\,\Sigma_v)^{-1}\,\exp\left[\left(0.792 + 0.681\,\Sigma_v^{0.5}\right)\,(\beta_v + 0.1)\right]. \quad \text{(A.15)}$$

▷ Wind coefficient ϕ_w^* [-]

$$\phi_w^* = C\,u_w^B\,\left(\frac{\beta_v}{\beta_{v,op}}\right)^{-E}, \quad \text{(A.16)}$$

with:

$$C = 7.47\,\exp\left[-0.133\,\Sigma_v^{0.55}\right],$$
$$B = 0.02526\,\Sigma_v^{0.54},$$
$$E = 0.715\,\exp\left[-3.59\times10^{-4}\,\Sigma_v\right]$$

▷ Slope coefficient ϕ_{sl}^* [-]

$$\phi_{sl}^* = 5.275\,\beta_v^{-0.3}\,(\tan\,\alpha_{sl})^2. \quad \text{(A.17)}$$

▷ Effective heating number ε [-]

$$\varepsilon = \exp\left[-\frac{138}{\Sigma_v}\right]. \quad \text{(A.18)}$$

▷ Heat of pre-ignition Q_{ig} [J/kg]

$$Q_{ig} = 250 + 1.116\,M_v. \quad \text{(A.19)}$$

Note that the formulation presented here is slightly different from the original formulation proposed by Rothermel. The difference lays in the determination of the fuel loading m_v'': while m_v'' was an input parameter provided in the biomass fuel database for each fuel type in the original model, it is now derived from the packing ratio β_v and the fuel particle mass density ρ_p [kg/m^3] since:

$$m_v'' = \frac{\rho_b\,\delta_v}{1+s_t} = \frac{(\beta\,\rho_p)\,\delta_v}{1+s_t}.$$

When the fuel layer thickness δ_v increases, the same packing ratio β_v is maintained in the vegetation bed, in contrast to the original Rothermel's model that strains the distribution of the solid fuel particles in the increasing volume of vegetation. Thus, the proposed formulation provides a linear dependence of the ROS Γ to the vegetation layer depth δ_v.

A.1.4 Sensitivity study to input parameters

To illustrate the sensitivity of the Rothermel's ROS to varying input parameters (associated with the short grass fuel type), the theoretical values of the no-wind ROS Γ_0 and the wind-aided ROS $\Gamma = \Gamma_0 \left(1 + \phi_w^*\right)$ are calculated for different wind conditions u_w (varying between 0 and 5 m/s) as well as for different properties of the solid particles summarized in Table A.3.

Table A.3: *Range of variations of the vegetation properties in the sensitivity study.*

M_v	β_v	Σ_v	δ_v	Δh_c	s_e
0-30 %	0.1-5 %	4000-20000 1/m	0-1 m	9-28 MJ/kg	1-5.55 %

Figure A.5 shows the overall range of values that can take the (a) no-wind and (b) wind-aided ROS due to variations in the vegetation moisture content M_v. In particular, Fig. A.5(a) shows that the no-wind ROS Γ_0 depends non-linearly on M_v and rapidly increases when the moisture content becomes lower than 5 %. Besides, Fig. A.5(b) shows the variations of the wind-aided ROS Γ with respect to the wind velocity u_w, varying between 0 and 5 m/s. Each color line represents the ROS variation for a specific value of the moisture content M_v; for instance, the red line corresponds to the case of a 0-% moisture content (i.e., the red dot in Fig. A.5(a)) and shows that under this condition of moisture content, the ROS can be multiplied by a factor 10 if the wind starts blowing up to 4 m/s.

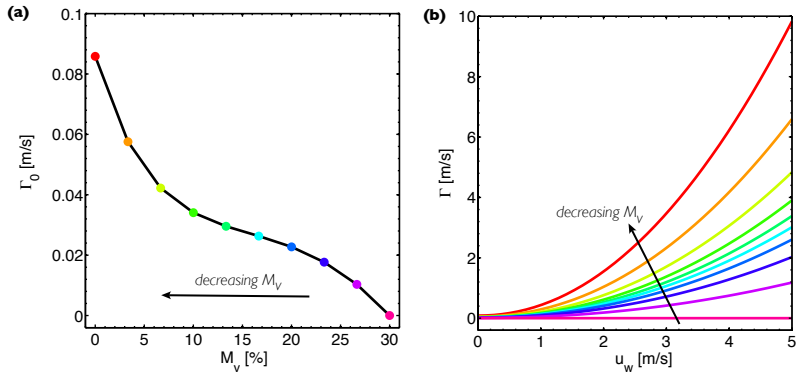


Figure A.5: *Sensitivity of the Rothermel's ROS to the fuel moisture content M_v [%] varying between 0 % and 30 % (i.e., the fuel moisture at extinction). (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of M_v).*

Similar plots can be produced for the fuel packing ratio β_v (Fig. A.6), the fuel particle surface-area-to-volume ratio Σ_v (Fig. A.7), the fuel layer depth δ_v (Fig. A.8), the fuel low heat of combustion Δh_c (Fig. A.9) and the fuel effective mineral content s_e (see Fig. A.10). It was found that the wind strongly enhances their effects on the ROS Γ since the rate of spread is multiplied by a factor 10 when the wind

blows up to 4 or 5 m/s. In particular, the effect of the fuel particle surface-area-to-volume ratio Σ_v on the ROS is severely amplified by the wind with ROS reaching up to 12 m/s for a 5-m/s wind and a 20000-1/m fuel particle surface-area-to-volume ratio. This is due to the formulation of the wind correction coefficient ϕ_w^*, whose parameters B, C and E explicitly depends on Σ_v. Σ_v represents the available surface for releasing combustion gas reactants per unit volume for a biomass solid particle; the higher Σ_v, the larger amount of gas reactants can be released per unit time to sustain and enhance combustion-related processes.

The ROS values feature a much wide scatter for the fuel particle surface-area-to-volume ratio Σ_v and the fuel moisture content M_v than for the effective mineral content s_e and the fuel packing ratio β_v, indicating that a lack of information in Σ_v and M_v results in a significant uncertainty range in the ROS prediction, which is critical to reduce for improving fire simulation accuracy. This pair of parameters is therefore important to include in the control vector $\mathbf{x}$ for parameter estimation.

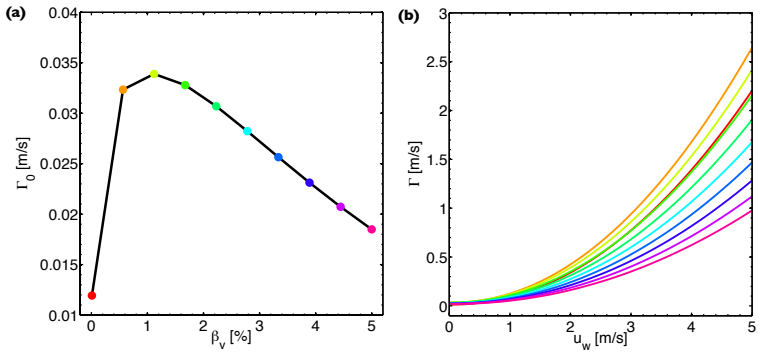


Figure A.6: *Sensitivity of the Rothermel's ROS to the fuel packing ratio β_v [%]. (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of β_v).*

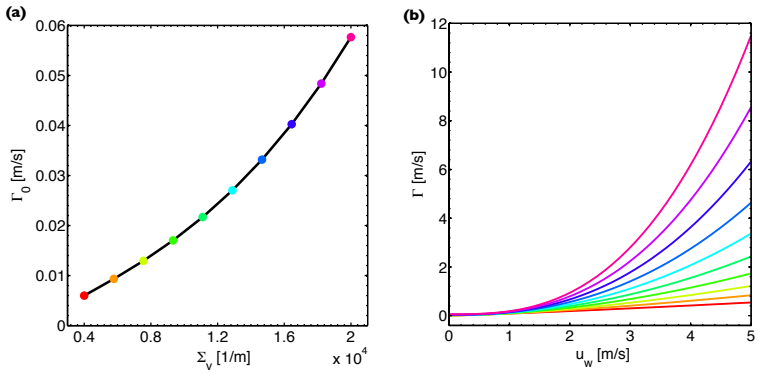


Figure A.7: *Sensitivity of the Rothermel's ROS to the fuel particle surface-area-to-volume ratio Σ_v [1/m]. (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of Σ_v).*

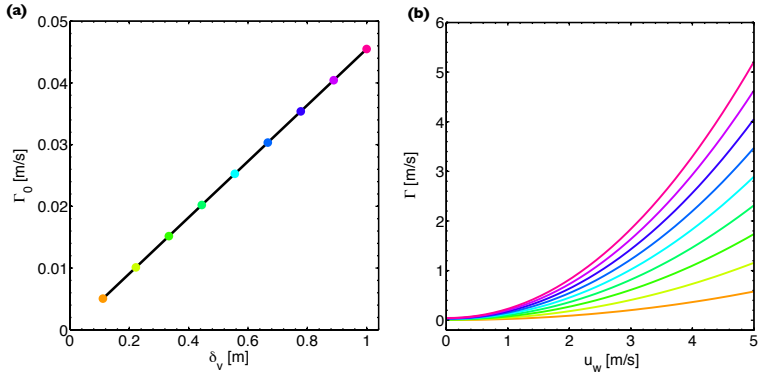


Figure A.8: *Sensitivity of the Rothermel's ROS to the fuel layer depth δ_v [m]. (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of δ_v).*

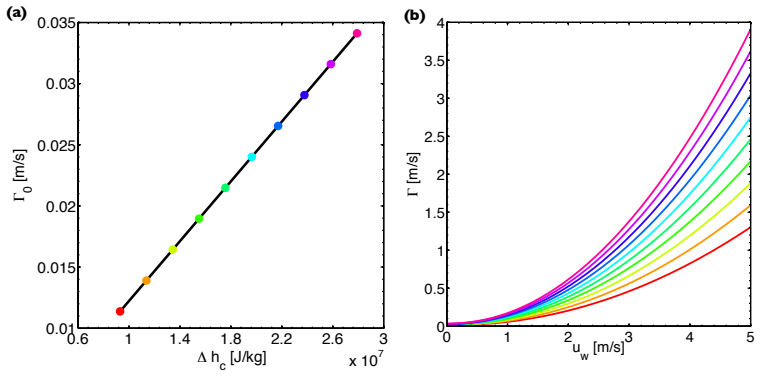


Figure A.9: *Sensitivity of the Rothermel's ROS to the fuel heat of combustion Δh_c [J/kg]. (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of Δh_c).*

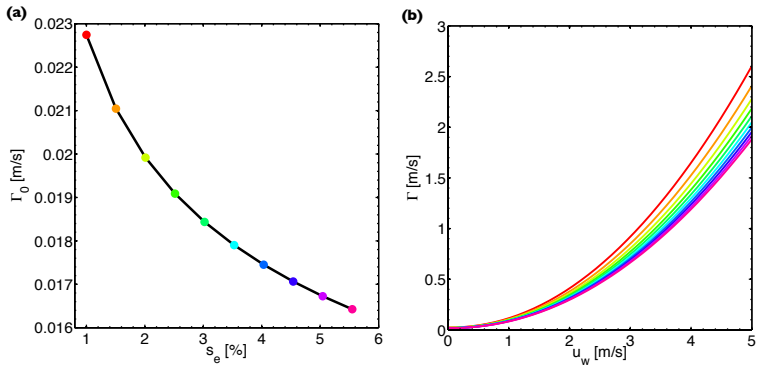


Figure A.10: *Sensitivity of the Rothermel's ROS to the fuel effective mineral content s_e [%]. (a) No-wind ROS Γ_0 [m/s]. (b) Wind-aided ROS Γ [m/s] with respect to the wind velocity u_w [m/s] (each color representing a specific value of s_e).*

In contrast to the vegetation depth δ_v and the heat of combustion Δh_c, the ROS is shown to depend non-linearly on the pair of parameters M_v and Σ_v. It is worth noting that these non-linearities will be more important when the wind magnitude fluctuates over time or when the vegetation is heterogeneously distributed. This highlights the importance of applying a data assimilation methodology able to handle multiple sources of non-linearity in the wildfire spread model.

A.2 Comparison to Balbi's model

Alternative ROS formulations exist in the literature as for instance the model due to Balbi et al. (2009). This model relies on mass, momentum and energy balance equations, but introduces some approximations to account for geometrical and thermodynamical characteristics of the flame region and to make the ROS model cost-effective in practice (in contrast, Rothermel's semi-empirical formulation is only based on the energy balance equation). From this viewpoint, Balbi's approach still belongs to the semi-empirical category of wildfire spread. One advantage of the Balbi's approach is to include a radiation submodel to estimate the amount of energy transferred to the vegetation ahead of the flame front, which is largely responsible for its thermal degradation until ignition.

A.2.1 Balbi's formulation of rate of spread

↪ **Assumptions**

With a focus on surface fires, Balbi's formulation relies on the following set of assumptions:

(1) The flame zone is assumed to exhibit on average a triangular shape, which is consistent with observed results and convenient to reduce the number of geometrical parameters required for the description of heat and mass fluxes.

(2) Thermal radiation is considered as the dominating heat transfer mechanism in the pre-heated vegetation zone under the flame (as long as the flame is not too tilted toward the ground, in which cases convection becomes the dominating heat transfer mechanism), while convection plays an essential role beyond the zone over which the flame is projected since the flame-induced flow of fresh air towards the flame has a convective cooling effect on the vegetation. In this context, the flame is supposed to behave as a radiant plane.

(3) The radiation factor (i.e., the fraction of the total energy released by the flame that is emitted by radiation), denoted by ξ_{rad} in the following, is assumed to be a decreasing function of the surface-area-to-volume ratio of the flame (denoted by Σ_{fr} in the following).

(4) The flame can be described using a uniformly-distributed average temperature (denoted by T_{fr} in the following).

(5) The combustion chemical reactions are reduced to a single reaction occurring at stoichiometry, i.e., $C + O_2 \rightarrow CO_2$. The associated stoichiometric coefficient is denoted by s and is set to $s = 9$ (meaning that 1 kg of pyrolysis gases is completely consumed for 9 kg of air).

(6) The vegetation is assumed to be made of solid particles of homogeneous properties (in terms of moisture content M_v, surface-area-to-volume ratio Σ_v, temperature T_v, etc.).

(7) A constant mass loss rate (denoted by $\dot{m}''_v$) is supposed for the vegetation as soon as the gas temperature reaches the assumed biomass ignition temperature T_{ign}.

↪ Input parameters

Nine parameters are common to Rothermel's and Balbi's ROS models, namely the fuel loading m''_v, the fuel layer depth δ_v, the fuel particle surface-area-to-volume ratio Σ_v, the fuel heat of combustion Δh_c, the fuel particle mass density ρ_p, the fuel moisture content M_v, the wind velocity at mid-flame height u_w and the terrain slope angle α_{sl}. Balbi's model requires additional parameters related to radiation and convection heat transfer described in Table A.4.

↪ Equations underlying Balbi's spread-rate model

For clarity purposes, we present here the main equations involved in Balbi's formulation for a fire that propagates in wind and up-slope directions (a schematic of the flame situation is presented in Fig. A.11). See Balbi et al. (2009) for more details on the general equations.

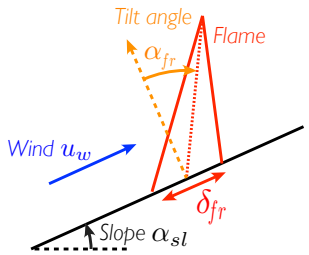


Figure A.11: *Schematic of wind- and slope-induced tilt angle of the flame noted α_{fr}, with u_w the wind magnitude in the up-slope direction and α_{sl} the slope angle. The parameter δ_{fr} represents the fire front depth.*

(E1) Mass budget. Based on the mass balance equation and on the assumption of stoichiometric mixture, the vertical velocity component u_b [m/s] (also referred to as the *buoyant velocity*) satisfies:

$$u_b = \frac{u_{b,0}}{\cos(\alpha_{sl})} = \frac{2\,\dot{m}''_v\,(s+1)}{\rho_{fr}\,\cos(\alpha_{sl})}, \tag{A.20}$$

with $u_{b,0}$ [m/s] the no-slope vertical velocity, s the stoichiometric coefficient, ρ_{fr} [kg/m^3] the flame mass density (assumed constant within the flame region) and $\dot{m}_v^{''}$ [kg/m^2/s] the vegetation mass loss rate.

Table A.4: *Additional input parameters of the ROS Balbi model related to radiation and convection heat transfer.*

Name	Symbol	Unit
Flame emissivity	ε_{fr}	-
Vegetation emissivity	ε_v	-
Radiation fraction (ratio of the radiation heat to the total heat received by the vegetation)	χ_{rad}	-
Buoyancy velocity (upward velocity of the gas reactants for no-wind no-slope conditions)	u_b	m/s
Moisture evaporation enthalpy (amount of energy required to evaporate moisture within the vegetation)	Δh_v	J/kg
Fuel calorific capacity (specific heat for vegetation at constant pressure)	$c_{p,v}$	J/kg/K
Fuel ignition temperature	T_{ign}	K
Ambient air temperature	T_{air}	K

(E2) Momentum budget. Assuming that the gas velocity field in the flame $\mathbf{u}_{fr}$ is a composition of the buoyancy velocity field $\mathbf{u}_b = (0, u_b)^{\mathrm{T}}$ and the wind velocity $\mathbf{u}_w = (u_w \cos(\alpha_{sl}), u_w \sin(\alpha_{sl}))^{\mathrm{T}}$, and assuming that the normal component of $\mathbf{u}_{fr}$ is responsible for tilting the flame, the tilt angle of the flame α_{fr} reads:

$$\tan(\alpha_{fr}) = \frac{u_w}{u_b \cos(\alpha_{sl})} + \tan(\alpha_{sl}), \tag{A.21}$$

with:

$$\cos(\alpha_{fr}) = \frac{u_b \cos(\alpha_{sl})}{\left(\sqrt{u_w^2 + u_b^2 + 2u_w u_p \sin(\alpha_{sl})}\right)},$$

and:

$$\sin(\alpha_{fr}) = \frac{u_w + u_b \sin(\alpha_{sl})}{\left(\sqrt{u_w^2 + u_b^2 + 2u_w u_p \sin(\alpha_{sl})}\right)}.$$

(E3) Energy budget. The heat release rate by gaseous combustion along the fireline noted Q_{fr} $[W/m]$ can be expressed as $Q_{fr} = \Delta h_c \, \delta_{fr} \, \dot{m}_v^{''}$, with δ_{fr} the

fire front depth and Δh_c $[J/kg]$ the low heat of combustion. Assuming that radiation occurs from the flame region above the fuel and inside the vegetation in the flaming part, and assuming that out of the flame there is a compensation between the cooling-induced airflow and the long-range radiation effect, the radiation heat release rate is given by the term $\chi_{rad}\, Q_{fr}$. The bulk flame temperature (assumed constant within the flame and noted T_{fr} [K]) then reads:

$$T_{fr} = T_{air} + (1 - \chi_{rad})\, \frac{\Delta h_c}{(1+s)\, c_p}, \tag{A.22}$$

with T_{air} [K] the ambient air temperature and c_p [J/kg/K] the gas calorific capacity.

(E4) Flame height. Since fire is a buoyancy-induced mechanism, the buoyant velocity u_b is classically calculated as a function of the mid-flame height (the flame height being noted H_{fr}) such as:

$$u_b = \sqrt{g \left(\frac{\rho_{air} - \rho_{fr}}{\rho_{fr}} \right) H_{fr}} = \sqrt{g \left(\frac{T_{air} - T_{fr}}{T_{fr}} \right) H_{fr}}, \tag{A.23}$$

with ρ_{air} [kg/m^3] the mass density of the ambient air. Thus, the flame height H_{fr} reads:

$$H_{fr} = \frac{u_b^2}{g \left(\frac{T_{air} - T_{fr}}{T_{fr}} \right)}. \tag{A.24}$$

(E5) Radiation submodel. The flame region above the vegetation and the flaming and smoldering process within the vegetation layer, respectively noted R_v and R_{fr}, are considered as primary sources of radiation in a wildfire. Under the assumption of a gray-body, R_v is expressed as follows:

$$\begin{array}{lll} R_v & = \varepsilon_v\, \sigma_{sb}\, T_{fr}^4 \left(1 - \frac{x}{\delta_{v,opt}}\right) \frac{\delta_v}{\delta_{v,opt}}, & x \leq \delta_{v,opt}, \\ & = 0, & x > \delta_{opt}, \end{array}$$

with ε_v the emissivity of the vegetation, σ_{sb} [W/m^2/K^4] the Stefan-Boltzmann constant, δ_v [m] the vegetation layer thickness, $\delta_{v,opt}$ [m] the optical length-scale satisfying:

$$\delta_{v,opt} = \frac{4}{\Sigma_v\, \beta_v}. \tag{A.25}$$

The flame contribution above the vegetation layer R_{fr} reads:

$$R_{fr} = \varepsilon_{fr}\, \sigma_{sb}\, T_{fr}^4 \left(\frac{1 - \cos(\alpha_{fr,o})}{2} \right), \tag{A.26}$$

with ε_{fr} the emissivity of the flame and $\alpha_{fr,o}$ the view-angle of the flame.

(E6) Pre-heating submodel. The heat budget in a control volume of the pre-heated zone of the vegetation is expressed as follows:

$$m_v^{''}\, c_p\, \frac{\mathrm{d}T_v}{\mathrm{d}t} + \Delta h_m\, \frac{\mathrm{d}m_{H_2O}^{''}}{\mathrm{d}t} = R_v + \nu_v\, R_{fr}, \tag{A.27}$$

with T_v [K] the mean temperature of the vegetation, Δh_m [J/kg] the moisture evaporation enthalpy, $m_{H_2O}^{''}$ [kg/m^2] the vegetation moisture loading (in contrast to $m_v^{''}$ [kg/m^2] the vegetation loading), and ν_v the fraction of the flame radiation absorbed by the fuel satisfying:

$$\nu_v = \min\left(\frac{\delta_v}{\delta_{v,opt}}, 1\right), \tag{A.28}$$

and:

$$\frac{\delta_v}{\delta_{v,opt}} = \frac{\delta_v\, \Sigma_v\, \beta_v}{4} = \frac{\Sigma_v}{4}\,(\delta_v\, \beta_v) = \frac{\Sigma_v}{4}\left(\frac{m_v^{''}}{\rho_b}\right). \tag{A.29}$$

Thus, Eq. (A.27) can be rewritten as follows:

$$\begin{aligned} &\Gamma\, \Sigma_v \left[c_p\,(T_{ign} - T_{air}) + m_{H_2O}^{''}\, \Delta h_m\right] \\ &= \int_0^{\delta_{opt}} R_v\, dx + \nu_v \int_0^{l_{fr}\,\sin(\alpha fr)} R_{fr}\, dx \qquad \text{(A.30)} \\ &= \varepsilon_v\, \sigma_{sb}\, T_{fr}^4\, \frac{\delta_v}{2} + \nu_v\, \varepsilon_{fr}\, \sigma_{sb}\, T_{fr}^4\, l_{fr} \left(\frac{1 + \sin(\alpha_{fr}) - \cos(\alpha_{fr})}{2}\right), \qquad \text{(A.31)} \end{aligned}$$

with Γ [m/s] the ROS of the flame front, T_{ign} [K] the ignition temperature of the gas (in contrast to T_{air} [K], the ambient air temperature) and l_{fr} [m] the flame length-scale.

(E7) Radiation fraction. The fraction of the heat release rate that is due to radiation χ_{rad} depends on the spread-rate Γ such that:

$$\chi_{rad} = \frac{\chi_{rad,0}}{\left(1 + \frac{\Gamma}{12\,\Gamma_0}\,\cos(\alpha_{fr})\right)}, \tag{A.32}$$

with $\chi_{rad,0}$ the radiation fraction when the flame volume-to-surface area converges to zero and Γ_0 [m/s] the no-wind no-slope ROS.

(E8) ROS formulation. As in the ROS due to Rothermel (1972), Balbi model assumes that the fire front propagates at least at the no-wind no-slope ROS Γ_0:

$$\Gamma = \Gamma_0 = \frac{\varepsilon_v\, \sigma_{sb}\, T_{fr}^4\, \delta_v}{2\, m_v^{''}\, \left(c_p\,(T_{ign} - T_{amb}) + m_{H_2O}^{''}\, \Delta h_m\right)}, \quad \alpha_{fr} \leq 0. \tag{A.33}$$

Note that a backing fire (i.e., $\alpha_{fr} \leq 0$) propagates at least at a minimal speed Γ_0. In this configuration, the minimal speed Γ_0 is not decreased by a counter-wind and is obtained by assuming that the radiation contribution of the flame above the vegetation R_{fr} is zero. To the contrary, if the flame is tilted towards the unburned vegetation (i.e., $\alpha_{fr} > 0$), the heat flux absorbed by the pre-heated vegetation is stronger and thereby, the fire propagates at a faster rate given by:

$$\Gamma = \Gamma_0 + \frac{\nu_v \, \varepsilon_{fr} \, \sigma_{sb} \, T_{fr}^4 \, l_{fr}}{2 \, m_v'' \left(c_p \, (T_{ign} - T_{amb}) + m''_{H_2O} \, \Delta h_m \right)} \left(1 + \sin(\alpha_{fr}) - \cos(\alpha_{fr})\right). \tag{A.34}$$

Equivalently, this ROS Γ reads:

$$\Gamma = \frac{1}{2} \left(Ra + \sqrt{Ra^2 + \frac{4\Gamma_0 \, (12 \, \Gamma_0)}{\cos(\alpha_{fr})}} \right), \tag{A.35}$$

with:

$$Ra = \Gamma_0 \left\{ 1 - \frac{12}{\cos(\alpha_{fr})} + \frac{12 \, \nu_v \, \chi_{rad,0} \, \Delta h_c}{4 \left(c_p \, (T_{ign} - T_{amb}) + m''_{H_2O} \, \Delta h_m \right)} \mathcal{A}_{fr} \right\},$$

and:

$$\mathcal{A}_{fr} = \frac{1 + \sin(\alpha_{fr}) - \cos(\alpha_{fr})}{\cos(\alpha_{fr})}.$$

This set of equations highlights the differences with the formulation due to Rothermel (1972). The latter relies on the stand-alone energy balance equation, in which the important quantities of interest are parameterized using wind-tunnel experiments, while the Balbi's model includes several submodels (in particular to describe the radiation heat transfer to the vegetation and the pre-heating of the vegetation ahead of the fire front) and minimizing the number of parameters to calibrate from experimental data.

A.2.2 Comparative study

Balbi et al. (2009) compared the ROS values predicted by the Balbi's model to the reference Rothermel's semi-empirical model in wind-tunnel laboratory-scale experiments (Weise and Biging, 1997), in which the vegetation was uniform and composed of wood sticks. The properties of the vegetation are described as follows: $M_v = 3$ %, $\delta_v = 0.1$ m, $\rho_p = 512.6$ kg/m^3, and $\beta_v = 0.98$ %. This comparison for a wide range of wind magnitude u_w is presented in Fig. A.12.

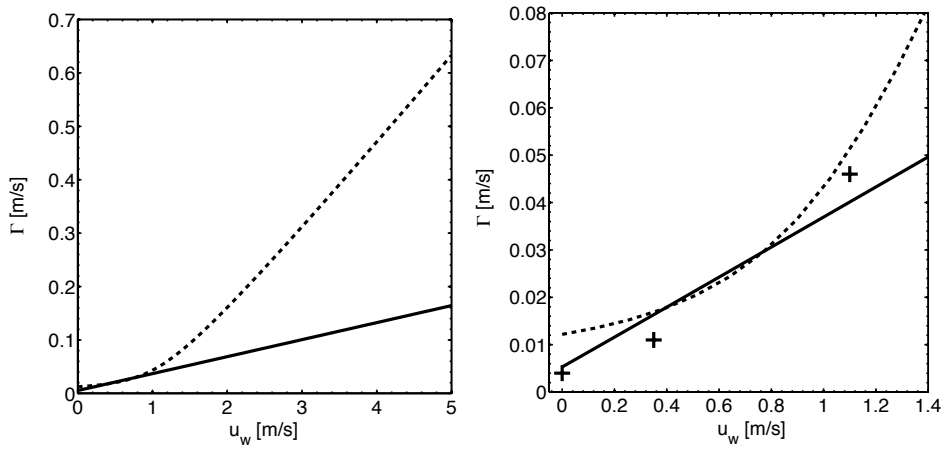


Figure A.12: *Comparison of ROS Γ by Balbi (dashed line) and Rothermel (solid line) models with respect to the wind magnitude u_w. Left: ROS variations for a wind magnitude u_w varying between 0 and 5 m/s. Right: Comparison to laboratory-scale data due to Weise and Biging (1997) represented in black crosses.*

The two models predict similar ROS values in the wind velocity range $[0, 1$ m/s$]$, in agreement with experimental data. However, their behavior drastically differs for stronger winds: while Rothermel's model predicts a quasi-linear increase of the ROS Γ with the wind velocity u_w, Balbi's model exhibits an exponential increase. Thus, Balbi's model is much more sensitive to the wind velocity than the Rothermel's model.

A comparative sensitivity study between Balbi's and Rothermel's ROS models is presented here to highlight their behavior discrepancies with respect to environmental conditions. This sensitivity study still relies on the previous configuration due to Weise and Biging (1997) experiments; the objective is to perturb the parameters (e.g., the fuel moisture content M_v, the fuel surface-area-to-volume ratio Σ_v, the fuel layer depth δ_v, the fuel packing ratio β_v) and to study the scatter of the resulting ROS values. Figure A.13 provides the ensemble of no-wind ROS Γ_0 and wind-aided ROS Γ as a function of the wind magnitude u_w for the fuel moisture content M_v. The Balbi's ROS model is not sensitive to M_v, even for the no-wind case, in contrast to Rothermel's model. To the contrary, the fuel packing ratio β_v and the fuel surface-area-to-volume ratio Σ_v have strong effects on the ROS (in particular when the fire propagation is induced by wind) as shown in Fig. A.14 and Fig. A.15, respectively. It is worth noting that Σ_v has no effect on the no-wind ROS and that its contribution to the ROS is highly related to the wind flow as in the Rothermel's model. Besides, uncertainties in the fuel layer depth δ_v can also change drastically the value of the ROS, with or without wind. In particular, Fig. A.16 shows that similarly to Rothermel's model, Balbi's formulation exhibits a linear dependence on δ_v.

While the values for the input parameters M_v, β_v, Σ_v and δ_v are realistic in the

present comparative study, their effects on the ROS significantly differ between the Rothermel's and Balbi's ROS formulations. The behavior of the Balbi's model changes significantly compared to the Rothermel's model. Further studies based on comparisons to laboratory-scale and field-scale experiments are required to diagnose the validity range of each model.

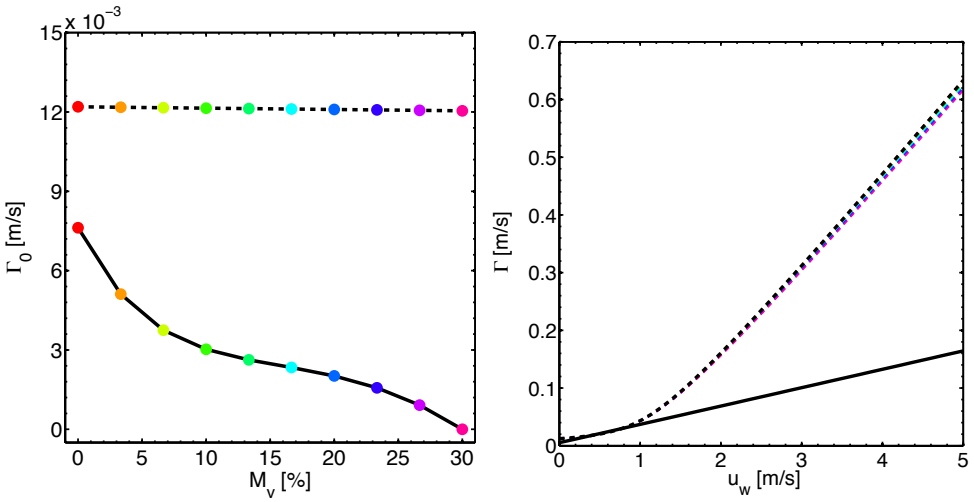


Figure A.13: *Sensitivity of the Balbi's ROS (dashed line) to the fuel moisture content M_v [%] varying between 0 % and 30 %, compared to Rothermel's ROS (solid line). Each color represents a specific value of M_v. Left: No-wind ROS Γ_0 [m/s]. Right: Wind-aided ROS Γ [m/s] as a function of the wind velocity u_w [m/s].*

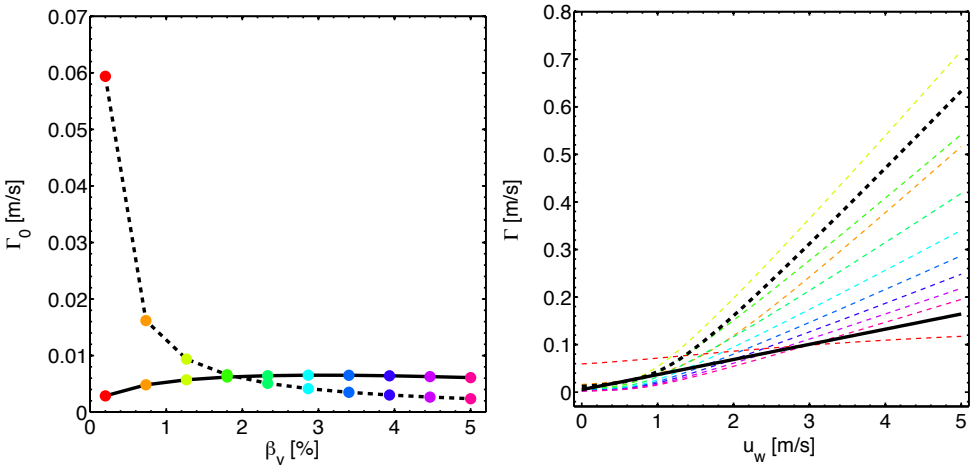


Figure A.14: *Sensitivity of the Balbi's ROS (dashed line) to the fuel packing ratio β_v [%] compared to Rothermel's ROS (solid line). Each color represents a specific value of β_v. Left: No-wind ROS Γ_0 [m/s]. Right: Wind-aided ROS Γ [m/s] as a function of the wind velocity u_w [m/s].*

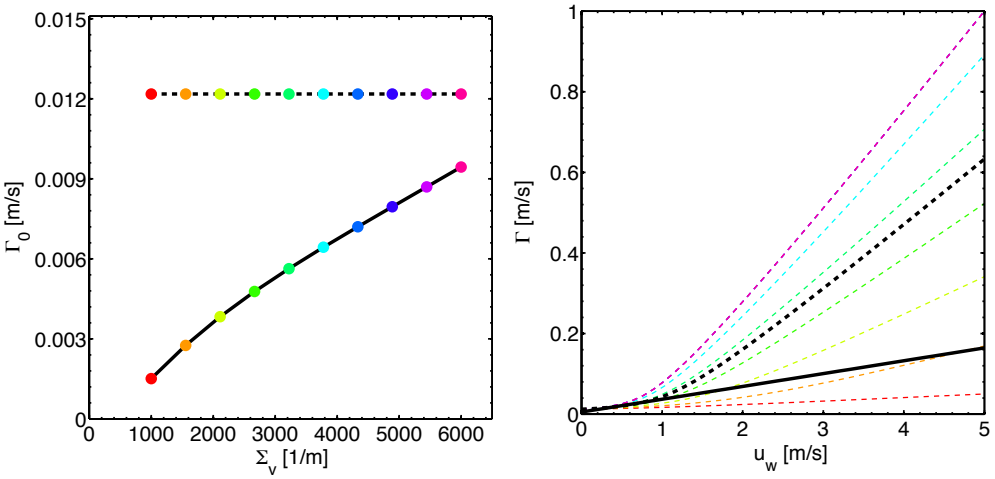


Figure A.15: *Sensitivity of the Balbi's ROS (dashed line) to the fuel particle surface-area-to-volume ratio Σ_v [1/m], compared to Rothermel's ROS (solid line). Each color represents a specific value of Σ_v. Left: No-wind ROS Γ_0 [m/s]. Right: Wind-aided ROS Γ [m/s] as a function of the wind velocity u_w [m/s].*

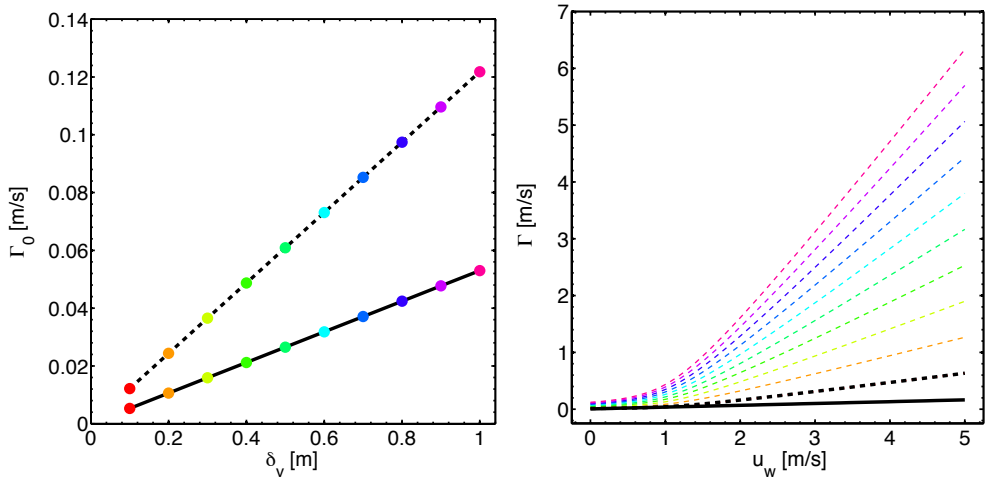


Figure A.16: *Sensitivity of the Balbi's ROS (dashed line) to the fuel layer depth δ_v [m], compared to Rothermel's ROS (solid line). Each color represents a specific value of Σ_v. Left: No-wind ROS Γ_0 [m/s]. Right: Wind-aided ROS Γ [m/s] as a function of the wind velocity u_w [m/s].*

This example illustrates the current status of semi-empirical models in the fire research field: a wide range of ROS formulations has been developed, but their domain of validity is restricted to the conditions of the experiments used during their original development and calibration. This domain of validity (in terms of wind velocity range for instance) must to be properly assessed in the perspective of ensemble-based data assimilation.

Appendix B

The OpenPALM dynamic code coupler

The OpenPALM software is a flexible and powerful dynamic code coupler that has been co-developed at CERFACS and ONERA since 1996 (open-source under lesser general public license) and that has reached a high-degree of maturity and stability, with applications ranging from operational data assimilation to multi-physics modeling, from climate change impact assessment to fluid/structure interactions. It is well-suited for the evolution of the current coupling technology towards the exaflop machines of next generations.

In this thesis, the OpenPALM coupler is used for two different applications illustrated in Fig. B.1, (1) the development of multi-physics flame-scale large-eddy simulations based on a high-performance coupling of parallel codes (data parallelism), and (2) the development of a data assimilation prototype based on an advanced scheduling of tasks of different natures (task parallelism). A general introduction of this code coupler along with a focus on both data and task parallelisms is provided in this appendix.

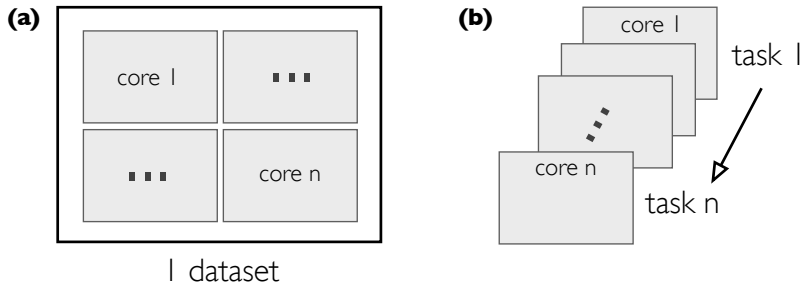


Figure B.1: *Different forms of parallelism. (a) Data parallelism. (b) Task parallelism.*

B.1 Overview of the OpenPALM code coupler

Code coupling is an appealing method to develop multi-physics and multi-component applications. For this purpose, the OpenPalm[1] software is a library of functionalities that handles the scheduling of existing components execution, sequentially or concurrently, as well as data exchange between these components. OpenPalm allows in a flexible and evolutive way, for the coupling of independent code components with a high-level of modularity in the data exchanges and treatment, while providing a straightforward parallelization environment, striking a balance between performance, software portability/flexibility and numerical accuracy. This coupling interface is able to deal with different forms of parallelism across multiple processors in parallel computing environments, data parallelism and task parallelism, as schematized in Fig. B.1.

- ▷ **Data parallelism** corresponds to the simultaneous execution on multiple cores of the same function/task across the elements of a unique dataset.
- ▷ **Task parallelism** corresponds to the simultaneous execution on multiple cores of many different functions/tasks across the same or different datasets.

OpenPalm is mainly composed of three complementary components, (1) the Palm library; (2) the Cwipi library; and (3) the graphical interface PrePalm. As the application programming interface is available in Fortran and C/C++, OpenPalm can couple codes written in different languages (e.g., F77, F90, C, C++ compiled codes, main interpreted languages such as Python, Perl, Java, Tcl/Tk, Octave, or black-box codes such as Fluent and Abaqus).

B.1.1 The PALM library

Palm[2] has been originally designed for oceanographic data assimilation algorithms (Fouilloux and Piacentini, 1999; Ferrya et al., 2007) within the MERCATOR global ocean operational forecast system, in order to couple two different numerical physical models (e.g. an ocean model with an atmospheric model for climate modeling) or to couple a physical model with a given mathematical algorithm (e.g., optimization, post-treatment algorithm). Currently, its application domain extends to a wide range of scientific applications (Buis et al., 2006; Piacentini et al., 2011). For instance, it is widely used for data assimilation applications such as the French project ADOMOCA[3] in atmospheric chemistry data assimilation (Massart et al., 2009) or for flood forecasting using the Saint-Venant solver Mascaret at SCHAPI[4] (Habert et al., 2012).

Palm has started to be used as a code coupling interface by the CFD[5] team at

[1] http://www.cerfacs.fr/globc/PALMWEB/.
[2] Projet d'Assimilation par Logiciel Multi-méthodes.
[3] Assimilation de DOnnées pour les MOdèles de Chimie Atmosphérique.
[4] Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations.
[5] Computational Fluid Dynamics.

CERFACS to perform multi-physics large-eddy simulations on gas turbines[6] (Amaya et al., 2010; Poitou et al., 2011, 2012). Although the simulations of specific components gas turbines become accessible, these stand-alone simulations are not sufficient to allow for a high-fidelity characterization of the physical system and thus, coupled systems combining a detailed description of the reactive flow (e.g. the fluid solver AVBP) as well as the radiation and conduction heat transfers at the wall of combustion chambers (e.g. the radiation solver PRISSMA, the heat transfer solver in solids AVTP) have been developed within the PALM framework (Duchaine et al., 2009; Gicquel et al., 2011; Jauré et al., 2011; Maheu et al., 2012). An example of a fluid-structure coupling is presented in Fig. B.2; this simulation of a cooled turbine blade results from the coupling between the AVBP fluid large-eddy simulation solver and the AVTP heat transfer solver, and combines data and task parallelisms.

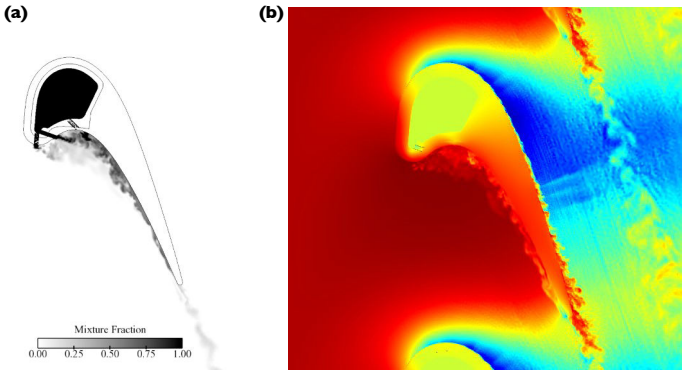


Figure B.2: *Fluid/structure coupling using the* PALM *software, to predict the cooling of a turbine blade on an unstructured mesh, composed of 6.5 millions of tetrahedral elements for the fluid zone and 600,000 elements within the solid. (a) Instantaneous snapshot of cooling air distribution. (b) Temperature distribution in the fluid and solid domains. Credit: Piacentini et al. (2011).*

Within the framework of PALM, applications are split into elementary components that can exchange data through message passing interface (MPI) communications. The main features of PALM are:

- ▷ the dynamic launching of the coupled components;
- ▷ the full independence of the components from the application algorithm;
- ▷ the parallel data exchanges with redistribution;
- ▷ the separation of the physics from the algebraic manipulations performed by the PALM algebra toolbox.

PALM offers the option to merge into a single executable the coupled components that are started in a sequence. Besides, computing resources such as the required memory and the number of concurrent processors are handled by the PALM driver.

[6] cerfacs.fr/coupling/.

This optimized communication scheme handles very complex communication patterns with some very practical features, such as the remapping of objects exchanged by parallel codes with different distributions, the selection of object subsets entirely from the user interface, the presence of an explicitly managed permanent repository for objects to be repeatedly received. Thus, PALM can be defined as a dynamic coupler for its ability to deal with situations, where the component execution scheduling and the data exchange patterns cannot be entirely defined before execution.

B.1.2 The CWIPI library

Based on the byzantine fault tolerance (BFT) and finite volume method (FVM) libraries (Fournier et al., 2011), CWIPI[7] (Refloch et al., 2011) provides a fully parallel communication layer for mesh-based coupling between different parallel solvers with MPI communications. As in most existing coupling libraries for multi-executables paradigm (DeCecchis et al., 2011; Joppich and Kürschner, 2006; Jauré et al., 2011; Valcke, 2012), CWIPI is a static coupler, in the sense that all the components of the simulation are started at the beginning, exchange data during the run phase and terminate simultaneously. Coupling is performed through one-dimensional, two-dimensional or three-dimensional exchange zones, which can be discretized differently in each coupled code as schematized in Fig. B.3. This library is able to deal with different types of geometrical elements (e.g. polygon, polyhedral) with an unstructured description. The functionalities in CWIPI involve:

- ▷ construction of the communication graph between distributed geometric interfaces through geometrical localization;
- ▷ interpolation on non-coincident meshes;
- ▷ exchange of coupling fields for massively parallel applications;
- ▷ generation of visualization file.

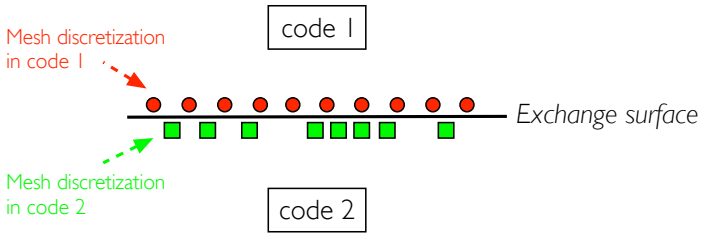


Figure B.3: *Schematic of the* CWIPI *exchange zone for two codes with different mesh discretizations.*

[7] Coupling With Interpolation Parallel Interface.

B.1.3 The graphical interface PrePALM

The graphical user interface, called PrePalm, is a portable Tcl/Tk application. The relevant features of the coupled components are described in identity cards for each component, which do not depend on the specific coupling algorithms. These identity cards are loaded by PrePalm, in order to construct the coupled applications. While the user describes the execution scheduling, the parallel sections, the data exchange patterns and the algebraic treatments, entirely from the user interface, PrePalm produces the input file for the coupler executable and the source code for the wrappers of the coupled component that manage the set-up of the communication framework with no need of change in the component sources. The same graphic tool can be used at run-time to monitor the simulation status and to provide post-mortem some statistics on the memory and CPU time resources used by the different components.

Figure B.4 gives a simple example of the PrePalm interface and highglights that an OpenPalm application can be described as a set of computational units arranged in a coupling algorithm.

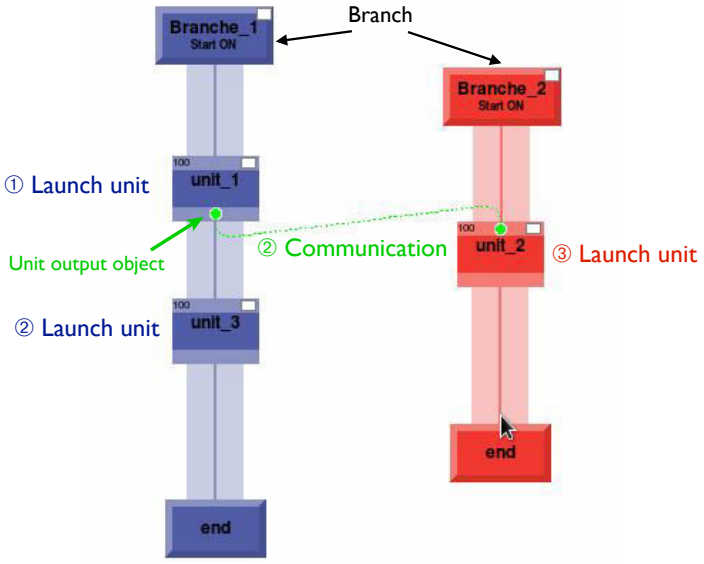


Figure B.4: *The* PrePalm *graphical user interface.*

The different units are controlled by conditional and iterative constructions and belong to algorithmic sequences called *computational branches*. A branch is structured as a sequential program in a high-level programming language. Within a branch, the coupled independent programs (referred to as *units*) can be regarded as subroutines of the branch program. Communications can be defined between different branches to exchange, send and receive data; they are represented as markers at the top (received data) or at the bottom (sent data) of the unit box. Figure B.5 presents the PrePalm interface when using OpenPalm to exchange

data at the flame/vegetation interface between the AVBP large-eddy simulation solver, the PRISSMA radiation heat transfer solver and the PYROWO vegetation thermal degradation solver (see Chapter 3).

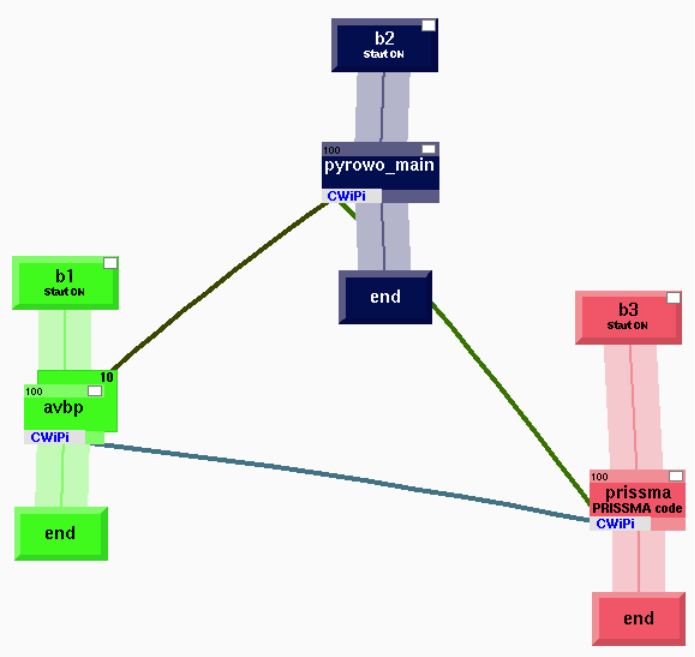


Figure B.5: *Example of the* PREPALM *graphical user interface for the coupling between* AVBP, PRISSMA *and* PYROWO *through the* CWIPI *library.*

B.2 PARASOL functionality

PARASOL is a novel functionality of the OPENPALM software (implemented in Tcl language) allowing to automatically launch, in parallel, a certain number of instances of the same computer program. Its objective is to spawn N_e instances of the same computer program with different inputs and consequently, different output variables of interest, according to the number of allocated processors specified by the user. Thus, PARASOL is particularly adapted to perform ensemble-based simulations in the context of uncertainty quantification or data assimilation, with an optimized use of the available computational ressources.

In practice, PARASOL creates two types of subroutine, a *slave* subroutine in charge of executing the function to *parasol-ize* on the one hand, and a *master* subroutine which calls N_e instances of the slave subroutines and manage input/output data (i.e., in terms of memory allocation and data exchange). In this thesis, the PARASOL functionality is applied to the wildfire spread simulator FIREFLY (i.e., FIREFLY is the subroutine to *parasol-ize*), which simulates the time-evolving location of the fire front given environmental conditions. Thus, different sets of ROS input pa-

rameters (e.g., vegetation properties, wind conditions) are provided as inputs to PARASOL and then, the corresponding fire front locations are calculated at the same simulation time (see Chapter 6). The OPENPALM-PARASOL functionality for the wildfire spread application is illustrated in Fig. B.6.

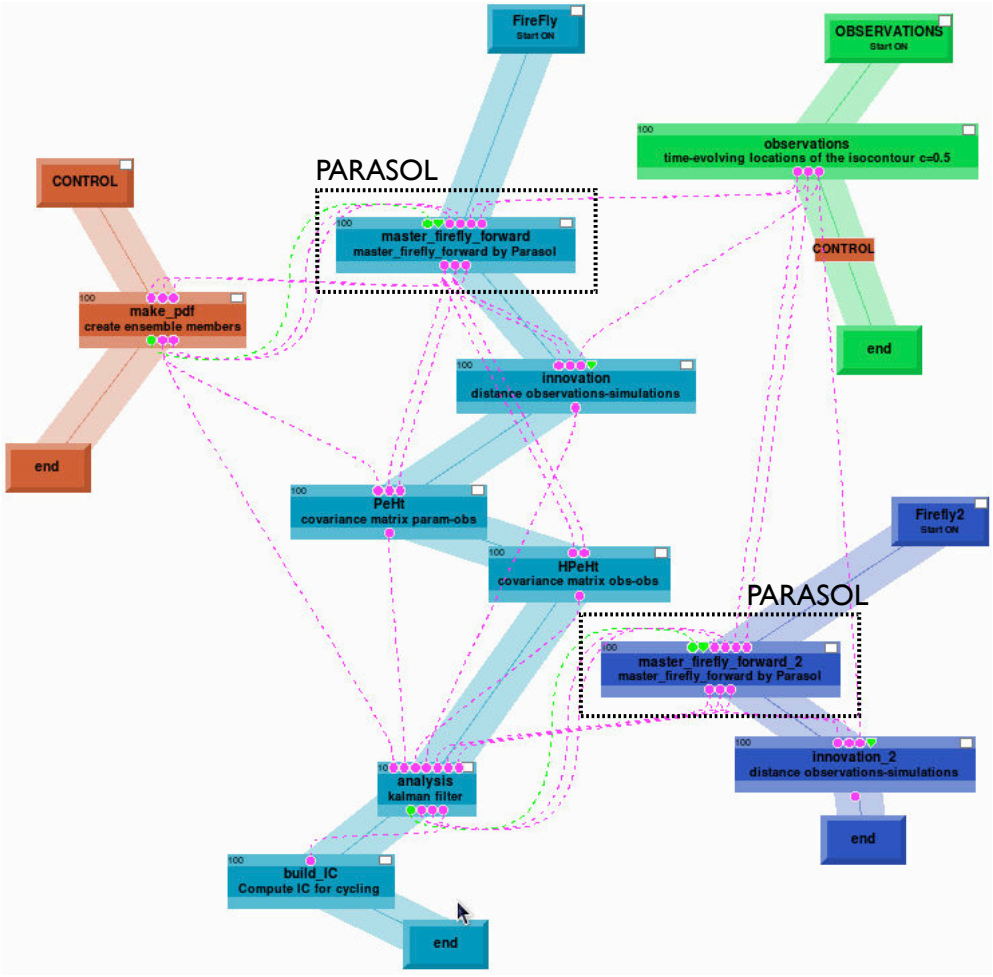


Figure B.6: *Example of the* PREPALM *graphical user interface for the application of the ensemble Kalman filter (EnKF) to wildfire spread through the parameter estimation approach. Black boxes represent the master subroutine encapsulating (*parasol-izing*)* FIREFLY*, for generating (1) the ensemble of fire front forecasts (i.e., without data assimilation) and (2) the ensemble of fire front analyses (i.e., with data assimilation). The branch named OBSERVATIONS provides the location of the fire front at the assimilation time; the branch named CONTROL corresponds to the generation of the forecast parameter estimates; the branch named FIREFLY corresponds to the different calculations (e.g., innovation vector, covariance matrices) leading to the computation of the analysis parameter estimates and to the new initial condition for the next assimilation cycle; and the branch named FIREFLY2 corresponds to the integration of* FIREFLY *leading to the analysis estimates of the fire front positions.*

Appendix C

The PYROWO vegetation thermal degradation model

The Pyrowo *vegetation thermal degradation model is designed to characterize the pyrolysis that vegetation undergoes ahead of the flame front, when subject to an external heat flux due to the flame-induced radiation heat transfer. In this appendix, it is shown that* Pyrowo *can be directly derived from the state-of-the-art multiphase formulations of flame-scale fire propagation under some assumptions.*

C.1 Multiphase model for vegetation thermal degradation

C.1.1 Description of the wildland vegetation

Multiphase formulations of the vegetation thermal degradation at macroscopic scales are proposed in literature (Grishin, 1997; Larini et al., 1998; Morvan and Dupuy, 2004; Séro-Guillaume and Margerit, 2002; Linn et al., 2002). In these approaches, vegetation is regarded as a porous medium with at least three different length-scales, from macroscopic to microscopic scales as illustrated in Fig. C.1. The forest canopy is usually modeled as a diphasic medium composed of a gas phase (i.e., mainly air) and a vegetation phase (e.g., branch, leaf). The latter can in-turn be considered as a porous medium composed of a gas phase (characterized by the index g) and a vegetation porous phase (characterized by the index p); this porous p-phase of the vegetation being itself associated with a solid phase (i.e., made of wood indexed by w and char indexed by c), a liquid phase (mainly made of moisture ($\mathrm{H_2O}_l$)) and a gas phase (i.e., made of air, water vapor $(\mathrm{H_2O})_g$ and flammable gases).

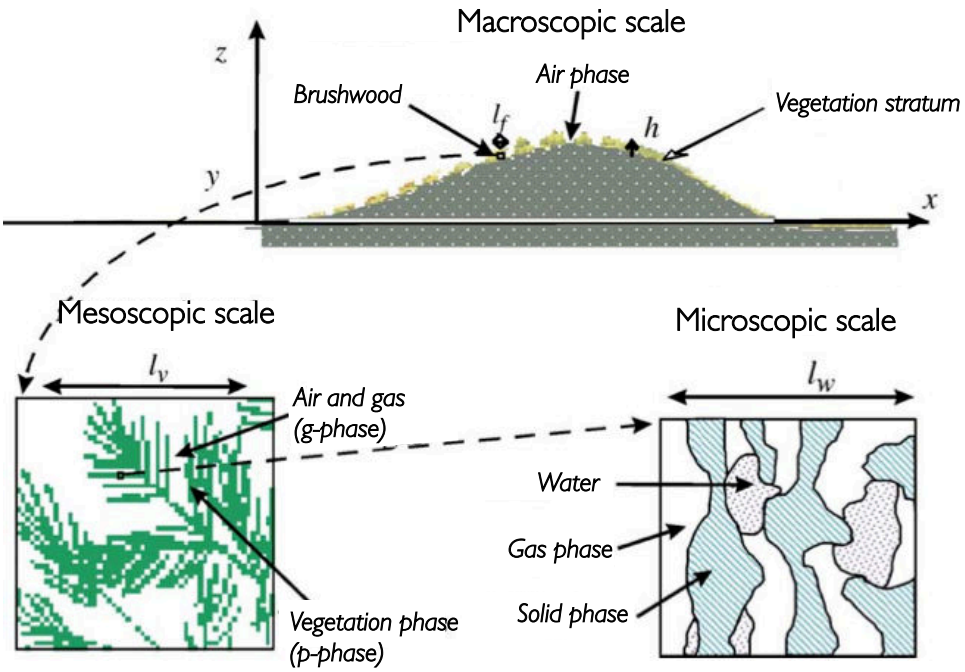


Figure C.1: *Porous representation of wildland fuels, from macroscopic to microscopic scales. Credit: Séro-Guillaume and Margerit (2002)*

Multiphase approaches rely on a closed set of equations that characterizes the successive steps occurring during the vegetation thermal degradation at macroscopic scales, namely the drying and pyrolysis of the vegetation phase as well as the combustion of the gas compounds released by the pyrolysis in the gas phase. For instance, the multiphase approach due to Séro-Guillaume and Margerit (2002) relies on the volume averaging method developed by Marle (1982) to account for the multi-scales of vegetation (i.e., from microscopic scales to macroscopic scales) in the formulation of the macroscopic vegetation properties. This volume averaging method derives an equivalent medium at macroscopic scales, with macroscopic properties.

C.1.2 Qualitative description of the vegetation pyrolysis

Di Blasi (1993) demonstrated that the chemical reactions occurring during the vegetation thermal degradation and breaking up the polymers (e.g., cellulose, hemicellulose, lignin) can be divided into two main steps:

- a series of primary reactions, in which the vegetation transforms into char (i.e., carbon-based non-volatile residue), flammable gas compounds and tar (i.e., high-molecular-weight products behaving as a gas at pyrolysis temperatures but condensing at ambient temperatures) in proportions that depend on the type of vegetation (it is known that cellulose and hemicellulose mostly

release flammable gas compounds, while lignin enhances char formation).

- ▷ a series of secondary reactions, in which intermediate tar materials degrade into flammable gas compounds and char.

However, pyrolysis-related processes are only partially known, implying necessary simplifications in biomass pyrolysis modeling. For instance, the model parameters that are required to describe the kinetics of char, flammable gas compounds and tar production are usually determined empirically. In the fire research field, fuel thermal degradation models mainly neglect secondary reactions and thereby, the intermediate tar production. This implies that the solid phase of the ongoing thermal degradation produces flammable gas compounds and char solid residue from a modeling viewpoint. Thus, the vegetation thermal degradation can be schematized as the sequence of the three following processes, which are active for different temperature ranges that can overlap:

(1) the drying of vegetation, in which moisture evaporates from the biomass fuel solid particles, leading to a production of water vapor in the g-phase (mesoscopic-scale representation);

(2) the release of flammable gas compounds (referred to as *pyrolysis gases*) from the biomass fuel solid particles in the gas phase (at microscopic scales), which are subsequently transported to the g-phase (mesoscopic-scale representation) due to buoyancy effects and flame-induced air entrainment;

(3) a partial oxidation at the surface of the remaining carbon-based materials (referred to as *char oxidation*) at high temperatures.

It is worth mentioning that the variations in time- and length-scales of pyrolysis can lead to different modes of combustion. If the mass rate of pyrolysis gases (i.e., the combustion gas reactants) is sufficiently high to supply the flame in reactants within the flammability limits (defined in terms of temperature and fuel/oxidizer equivalence ratio), this process leads to flaming combustion and sustains wildfire propagation in diffusion flame-type configurations. The features of this diffusion flame depend on the time delay between the production of the pyrolysis gases and their ignition (see Chapter 2). To the contrary, when air largely diffuses within the porous vegetation, the flame can be absent and instead, a slow oxidation of the remaining solid materials can proceed. This corresponds to the smoldering mode of combustion. This description highlights the complexity in vegetation thermal degradation modeling and its high dependencies on the flow and heat transfer conditions within the vegetation layer.

C.1.3 Mathematical formulation of the vegetation pyrolysis

In practice, a wide range of modeling approaches exist to describe the solid phase of the vegetation at microscopic scales. In the fire research field, this solid phase is commonly represented as a single homogeneous material in which primary and

secondary reactions can be distinguished. However, the specific behavior of each polymer (i.e., cellulose, hemicellulose, lignin) is not accounted for, implying that the potential interactions and bonds between polymers are neglected (Shafizadeh, 1982; Di Blasi, 1993). For instance, Séro-Guillaume and Margerit (2002) represented the vegetation thermal degradation as the combination of drying and pyrolysis, while neglecting char oxidation:

$$\begin{array}{ccc} (H_2O)_l & \longrightarrow & (H_2O)_g \\ \text{virgin solid vegetation} & \longrightarrow & \text{flammable gases} + \text{char.} \end{array}$$

↪ Microscopic scales

At microscopic scales, the p-phase of the porous vegetation is defined by the porosity β_p, which corresponds to the ratio of the volume occupied by the fluid phase V_{f_p} (i.e., gas and liquid phases) to the total volume of the p-phase V_p. β_p reads:

$$\beta_p = \frac{V_{f_p}}{V_p}. \tag{C.1}$$

The rest of the volume occupied by the solid phase (noted Vs_p) is made of char and wood, indexed by c and w respectively. Thus, the following relation holds:

$$V_p = \underbrace{V_{g_p} + V_{l_p}}_{V_{f_p}} + \underbrace{V_{w_p} + V_{c_p}}_{V_{s_p}}, \tag{C.2}$$

in which V_{g_p}, V_{l_p}, V_{w_p} and V_{c_p} correspond to the volumes occupied by the gas, the liquid, the (solid) wood and the (solid) char phases within the p-phase. The saturation of these components in the fluid and solid phases are defined as follows:

$$\beta_{g_p} = \frac{V_{g_p}}{V_{f_p}}, \quad \beta_{l_p} = \frac{V_{l_p}}{V_{f_p}}, \quad \beta_{w_p} = \frac{V_{w_p}}{V_{s_p}}, \quad \beta_{c_p} = \frac{V_{c_p}}{V_{s_p}}, \tag{C.3}$$

satisfying $\beta_{g_p} + \beta_{l_p} = 1$ for the fluid phase, and $\beta_{w_p} + \beta_{c_p} = 1$ for the solid phase. This leads to the definition of the mass density of each phase within the p-phase and thereby, to the total density of the p-phase noted ρ_p. Indeed,

$$\rho_{g_p} = \frac{M_{g_p}}{V_{g_p}}, \quad \rho_{l_p} = \frac{M_{l_p}}{V_{l_p}}, \quad \rho_{w_p} = \frac{M_{w_p}}{V_{w_p}}, \quad \rho_{c_p} = \frac{M_{c_p}}{V_{c_p}}, \tag{C.4}$$

with M_{g_p} the mass of gas contained within the volume V_{g_p}, M_{l_p} the mass of liquid within V_{l_p}, M_{w_p} the mass of wood contained within V_{w_p}, and M_{c_p} the mass of char within V_{c_p}. The mass densities of the solid and fluid phases, denoted respectively by ρ_{s_p} and ρ_{f_p}, yield:

$$\begin{aligned} \rho_{s_p} &= \beta_{w_p}\,\rho_{w_p} + \beta_{c_p}\,\rho_{c_p} = \tfrac{1}{V_{s_p}}\left(M_{w_p} + M_{c_p}\right), \\ \rho_{f_p} &= \beta_{g_p}\,\rho_{g_p} + \beta_{l_p}\,\rho_{l_p} = \tfrac{1}{V_{f_p}}\left(M_{g_p} + M_{l_p}\right). \end{aligned} \tag{C.5}$$

This leads to the following definition for the mass density of the p-phase ρ_p:

$$\rho_p = \beta_p \, \rho_{f_p} + (1 - \beta_p) \, \rho_{s_p}. \tag{C.6}$$

↪ Mesoscopic scales

At mesoscopic scales, the multiphase model due to Séro-Guillaume and Margerit (2002) assumes that the p-phase of the porous vegetation is in thermal equilibrium with the gas phase. This yields a unique energy balance equation for the vegetation mesoscopic description. The structure of the vegetation at these scales is characterized by the porosity β_g, defined as follows:

$$\beta_g = \frac{V_f}{V}, \quad \text{with} \quad V = V_p + V_g, \tag{C.7}$$

with V_p the volume occupied by the p-phase and V_g the volume occupied by the g-phase. The equation describing the evolution of the temperature T_p of both phases yields:

$$\begin{aligned} &\left[\beta_g \, \rho_g \, c_{p,g} + (1 - \beta_g) \, \rho_p \, c_{p,p}\right] \frac{\mathrm{d}T_p}{\mathrm{d}t} + \\ &\beta_g \, \rho_g \, c_{p,g} \, \mathbf{u}_g \cdot \nabla T_p + \nabla \cdot \left(-\lambda_{eq} \, \nabla T_p + \dot{s}_{p,rad} + \dot{s}_{g,rad}\right) = \dot{q}_{p,c} + \dot{q}_{g,c}, \end{aligned} \tag{C.8}$$

where:

- ▷ $\mathbf{u}_g$ is the velocity field in the g-phase, commonly modeled based on Darcy's law;
- ▷ λ_{eq} is the equivalent conductivity of the medium at mesoscopic scales;
- ▷ $c_{p,p}$ is the specific heat of the p-phase and $c_{p,g}$ its counterpart for the g-phase;
- ▷ $\dot{s}_{p,rad}$ and $\dot{s}_{g,rad}$ are the radiation heat fluxes in the p-phase and g-phase, respectively;
- ▷ $\dot{q}_{p,c}$ and $\dot{q}_{g,c}$ are the volumetric heat exchanges resulting, respectively, from the combustion of flammable gases in the g-phase and from the chemical reactions in the p-phase. The latter reads:

$$\dot{q}_{p,c} = -\left(1 - \beta_p\right) \left(\beta_{w_p} \, \rho_{w_p} \, \dot{q}_{w_p} + \beta_{l_p} \, \rho_{l_p} \, \dot{q}_{l_p}\right), \tag{C.9}$$

  with $\dot{q}_{w_p}$ the heat flux due to wood pyrolysis and $\dot{q}_{l_p}$ the heat flux due to moisture evaporation.

Since the multiphase model due to Séro-Guillaume and Margerit (2002) neglects the porosity variations within the p-phase, its mass balance equations for wood,

char and moisture can be expressed as follows:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left[(1-\beta_g)(1-\beta_p)\,\beta_{w_p}\,\rho_{w_p}\right] &= -(1-\beta_g)(1-\beta_p)\,\beta_{w_p}\,\rho_{w_p}\,k_{w_p}(T_p),\\
\frac{\mathrm{d}}{\mathrm{d}t}\left[(1-\beta_g)(1-\beta_p)\,\beta_{c_p}\,\rho_{c_p}\right] &= (1-\beta_g)(1-\beta_p)\,\beta_{w_p}\,\rho_{w_p}\,k_{c_p}(T_p),\\
\frac{\mathrm{d}}{\mathrm{d}t}\left[(1-\beta_g)\,\beta_p\,\beta_{l_p}\,\rho_{l_p}\right] &= -(1-\beta_g)\,\beta_p\,\beta_{l_p}\,\rho_{l_p}\,k_{l_p}(T_p),
\end{aligned} \tag{C.10}$$

where k_{w_p} corresponds to the mass transfer from virgin solid wood to gas compounds and char as a function of the macroscopic temperature T_p, k_{c_p} corresponds to the rate of char formation, and k_{l_p} corresponds to the rate of moisture evaporation at temperature T_p. These equations represent the time-evolving mass density of each compound of the vegetation that is identified at microscopic scales, i.e., wood, char and liquid moisture, based on the porosity at mesoscopic scales (β_p) and macroscopic scales (β_g).

C.2 Derivation of a macroscopic vegetation thermal degradation model

Within the Pyrowo thermal degradation model developed in this thesis, since the porosity of the p-phase β_p is typically small (below 5 %), it is assumed that the gas compounds emitted by the pyrolysis are directly entrained toward the g-phase. There is no residence of the emitted gases within the p-phase, implying that the fluid phase of the p-phase is saturated in liquid (i.e., $V_{f_p} = V_{l_p}$). Thus, the p-phase is reduced to dry wood, char material and liquid water, each component being characterized by a mass fraction noted $Y_{p,dry}$, $Y_{p,char}$ and $Y_{p,vap}$, respectively.

To describe the mass fractions $Y_{p,dry}$, $Y_{p,char}$ and $Y_{p,vap}$, the bulk density of the p-phase for each of these three components is introduced through the variables $\hat{\rho}_{w_p}$, $\hat{\rho}_{c_p}$ and $\hat{\rho}_{l_p}$, respectively. These quantities represent the mass of each component contained within the volume of the p-phase V_p, instead of the volume occupied by the specific component in the classical definition of the mass density. Thus, the following relations hold for wood, char and moisture contents:

$$\begin{aligned}
\hat{\rho}_{w_p} = \frac{M_{w_p}}{V_p} &= \left(\frac{M_{w_p}}{V_{w_p}}\right) \times \left(\frac{V_{w_p}}{V_p}\right)\\
&= \rho_{w_p}\left(\frac{V_{w_p}}{V_{s_p}} \times \frac{V_{s_p}}{V_p}\right)\\
&= \rho_{w_p}\,\beta_{w_p}\,(1-\beta_p)\\
\Rightarrow \hat{\rho}_{w_p} &= \rho_p\,Y_{p,dry},
\end{aligned} \tag{C.11}$$

$$\begin{aligned}\hat{\rho}_{c_p} = \frac{M_{c_p}}{V_p} &= \left(\frac{M_{c_p}}{V_{c_p}}\right)\left(\frac{V_{c_p}}{V_{s_p}} \times \frac{V_{s_p}}{V_p}\right) \\ &= \rho_{c_p}\,\beta_{c_p}\,(1-\beta_p) \\ \Rightarrow \hat{\rho}_{c_p} &= \rho_p\,Y_{p,char},\end{aligned} \tag{C.12}$$

$$\begin{aligned}\hat{\rho}_{l_p} = \frac{M_{l_p}}{V_p} &= \left(\frac{M_{l_p}}{V_{l_p}}\right)\left(\frac{V_{l_p}}{V_{f_p}} \times \frac{V_{f_p}}{V_p}\right) \\ &= \rho_{l_p}\,\beta_{l_p}\,\beta_p \\ \Rightarrow \hat{\rho}_{l_p} &= \rho_p\,Y_{p,vap}.\end{aligned} \tag{C.13}$$

Thus, the total density of the p-phase defined in Eq. (C.6) can be rewritten in terms of the bulk densities $\hat{\rho}_{w_p}$, $\hat{\rho}_{c_p}$ and $\hat{\rho}_{l_p}$ as well as in terms of the mass fractions $Y_{p,dry}$, $Y_{p,char}$ and $Y_{p,vap}$ as follows:

$$\begin{aligned}\rho_p &= \beta_p\,\rho_{f_p} + (1-\beta_p)\,\rho_{s_p} \\ &= \beta_p\,\rho_{l_p} + (1-\beta_p)\,\left(\beta_{w_p}\,\rho_{w_p} + \beta_{c_p}\,\rho_{c_p}\right) \\ &= \frac{(M_{l_p} + M_{w_p} + M_{c_p})}{V_p} \\ \Longrightarrow \rho_p &= \hat{\rho}_{l_p} + \hat{\rho}_{w_p} + \hat{\rho}_{c_p} = \rho_p\,Y_{p,dry} + \rho_p\,Y_{p,char} + \rho_p\,Y_{p,vap},\end{aligned} \tag{C.14}$$

with $Y_{p,dry} + Y_{p,char} + Y_{p,vap} = 1$. Similarly, the total specific heat $c_{p,p}$ of the p-phase yields:

$$c_{p,p} = Y_{p,dry}\,c_{p,w_p} + Y_{p,char}\,c_{p,c_p} + Y_{p,vap}\,c_{p,l_p}. \tag{C.15}$$

If β_v is defined as the vegetation packing ratio at macroscopic scales such that $\beta_v = 1 - \beta_g$, Eq. (C.10) can be expressed as follows:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\beta_v\,\rho_p\,Y_{p,dry}\right] &= -\left(\beta_v\,\rho_p\,Y_{p,dry}\right)\,k_{w_p}(T_p), \\ \frac{\mathrm{d}}{\mathrm{d}t}\left[\beta_v\,\rho_p\,Y_{p,char}\right] &= \left(\beta_v\,\rho_p\,Y_{p,char}\right)\,k_{c_p}(T_p), \\ \frac{\mathrm{d}}{\mathrm{d}t}\left[\beta_v\,\rho_p\,Y_{p,vap}\right] &= -\left(\beta_v\,\rho_p\,Y_{p,vap}\right)\,k_{l_p}(T_p).\end{aligned} \tag{C.16}$$

Since the term $(\beta_g\,\rho_g\,c_{p,g})$ is negligible in Eq. (C.8) due partly to the low density of the pyrolysis gases, the equation for the vegetation temperature T_p becomes:

$$\begin{aligned}&\beta_v\,\rho_p\,c_{p,p}\,\frac{\mathrm{d}T_p}{\mathrm{d}t} + (1-\beta_v)\,\rho_g\,c_{p,g}\,\mathbf{u}_g\cdot\nabla T_p + \nabla\cdot\left(\dot{q}''_{p,rad} + \dot{q}''_{g,rad}\right) \\ &= \dot{q}'''_{p,c} + \dot{q}'''_{g,c}.\end{aligned} \tag{C.17}$$

This derivation demonstrates that the macroscopic formulation of the vegetation thermal degradation considered by Séro-Guillaume and Margerit (2002) defined in Eq. (C.14)-(C.16)-(C.17) is equivalent (under some assumptions) to the multiphase model proposed by Morvan and Dupuy (2004) and thereby, present similarities to the description of the vegetation involved in Pyrowo, except that the latter (1) represents chemical source terms using Arrhenius-type kinetic law, (2) describes char production as a fixed amount of the rate of pyrolysis, and (3) accounts for char oxidation (i.e., surface oxidation of the remaining solid biomass particles once the drying and pyrolysis stages are achieved). Besides, the flow velocity field $\mathbf{u}_g$ within the vegetation layer is not accounted for since the transport of the pyrolysis gases is not explicitly solved, and a thermal equilibrium is assumed between the g-phase and the p-phase at mesoscopic scales (i.e., $T_v = T_p$). As a conclusion, the vegetation thermal degradation model proposed in this thesis and named Pyrowo considers an equivalent medium of the different vegetation sub-scales by including average macroscopic properties of the vegetation. Instead of relying on volume averaging methods (typical of the porous medium field), the macroscopic parameters are obtained in Pyrowo through a calibration procedure against thermogravimetric analysis (TGA) data. The presentation of this calibration procedure named Calwo follows.

C.3 Calibration procedure for designing an equivalent vegetation

C.3.1 Key ideas of the Friedmann kinetic analysis

The current mass of vegetation subject to thermal heating is noted m, its initial counterpart is noted m_0 and the mass content of inorganic compounds (representing the mass of remaining materials when the vegetation thermal degradation is finished) is noted m_{ash}. From the previous mathematical analysis, this mass decreases, while the temperature of the solid vegetation increases since there is an ongoing phase change from solid-phase to gas-phase due to pyrolysis and char oxidation (in particular for vegetation temperature above 500 K). Each mechanism responsible for the mass loss of the pre-heated vegetation can be generally described using the following relation:

$$-\frac{\mathrm{d}m}{\mathrm{d}t} = \dot{\omega}_r = F_r\, T_v\, (m - m_{ash})^{n_r}, \tag{C.18}$$

in which the reaction term F_r is expressed using an Arrhenius-type equation as follows:

$$F_r = k_r\, \exp\left[\frac{-T_{r,a}}{T_v}\right], \tag{C.19}$$

with T_v the temperature of the vegetation and the triplet $(k_r, T_{r,a}, n_r)$ the Arrhenius-type parameters associated with the ongoing reaction referred to as *index r*.

It is worth mentioning that the description of the calibration procedure focuses here on the pyrolysis reaction that occurs within the flame region (i.e., between 550 K and 650 K). However, drying and char oxidation can be similarly interpreted. Assuming the reaction index n_r is equal to 1, Eq. (C.18) can be rewritten as:

$$\begin{aligned} -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0-m}{m_0-m_{ash}}\right) &= (k_r\,T_v)\,\exp\left[\frac{-T_{r,a}}{T_v}\right]\left(\frac{m-m_{ash}}{m_0-m_{ash}}\right) \\ \Rightarrow -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0-m}{m_0-m_{ash}}\right) &= (k_r\,T_v)\,\exp\left[\frac{-T_{r,a}}{T_v}\right]\left(1-\frac{m_0-m}{m_0-m_{ash}}\right). \end{aligned} \tag{C.20}$$

The extent of reaction ξ_r is defined as the percent of reaction progress such that:

$$\xi_r = \frac{m-m_{ash}}{m_0-m_{ash}}. \tag{C.21}$$

Using this notation, Eq. (C.20) becomes:

$$\frac{\mathrm{d}\xi_r}{\mathrm{d}t} = (k_r\,T_v)\,\exp\left[\frac{-T_{r,a}}{T_v}\right](1-\xi_r). \tag{C.22}$$

As explained in Chapter 4, TGA experiments aim at retrieving the weight loss curve of a material given its temperature T_v for a constant heating rate $\mathrm{d}T_v/\mathrm{d}t$. This implies that the time-evolving extent of reaction can be decomposed as follows:

$$\frac{\mathrm{d}\xi_r}{\mathrm{d}t} = \left(\frac{\mathrm{d}\xi_r}{\mathrm{d}T_v}\right)\left(\frac{\mathrm{d}T_v}{\mathrm{d}t}\right). \tag{C.23}$$

Incorporating Eq. (C.23) into Eq. (C.22) and integrating the resulting equation from the initial state to the current state of the vegetation yield the following equation:

$$\int_{\xi_{r,0}}^{\xi_r}\frac{\mathrm{d}\xi_r}{(1-\xi_r)} = \int_{T_{v,0}}^{T_v}\left\{\frac{T_v}{(\mathrm{d}T_v/\mathrm{d}t)}\,(k_r\,T_v)\,\exp\left[\frac{-T_{r,a}}{T_v}\right]\right\}\mathrm{d}T_v, \tag{C.24}$$

in which the initial state of the vegetation is represented in terms of the initial reaction extent $\xi_{r,0}$ and the related vegetation temperature $T_{v,0}$, while the current state is represented in terms of the current reaction extent ξ_r and temperature T_v. Assuming $\xi_{r,0}=0$ at temperature $T_v=T_{v,0}$, the extent of reaction ξ_r is predicted using the following analytical formulation:

$$\xi_r = 1-\exp\left[-\int_{T_{v,0}}^{T_v}\left\{\frac{T_v}{(\mathrm{d}T_v/\mathrm{d}t)}\,(k_r\,T_v)\,\exp\left[\frac{-T_{r,a}}{T_v}\right]\right\}\mathrm{d}T_v\right]. \tag{C.25}$$

This methodology for retrieving the extent of reaction ξ_r for a constant heating rate $\mathrm{d}T_v/\mathrm{d}t$ is commonly referred to as the *Friedmann kinetic analysis* in the literature (Trick et al., 1997).

C.3.2 Calibration procedure CALWO

Based on the kinetic analysis method due to Friedmann, the following procedure is useful to calibrate the kinetic parameters $(k_r, T_{a,r})$ of reaction r, which best reproduce available TGA measurements:

(1) Isolate and evaluate separately the extent of reaction ξ_r based on the TGA mass loss curve;

(2) Compute and fit each mass loss rate curve $\mathrm{d}\xi_r/\mathrm{d}t$ with Gaussian relation (the area under the curve represents the average percent of vegetal mass loss);

(3) Calculate the reaction term F_r at temperature intervals using Eq. (C.22) as follows:

$$F_r = F_r(v) = \frac{\frac{\mathrm{d}\xi_r}{\mathrm{d}t}}{T_v\,(1-\xi_r)}. \tag{C.26}$$

(4) Compute the apparent activation temperature $T_{r,a}$ and apparent pre-exponential term k_r satisfying

$$\ln\, F_r(T_v) = \ln\, k_r - T_{r,a}\left(\frac{1}{T_v}\right). \tag{C.27}$$

The activation temperature $T_{r,a}$ is computed using linear regression methods, since it corresponds to the slope of the function $\ln\, F_r = f(1/T_v)$, while $\ln\, k_r$ corresponds to the intercept of the regression line.

Appendix D

Evaluation of reduced kinetic schemes for pyrolysis gas combustion

During the pyrolysis of wildland fuels, large amounts of flammable compounds (referred to as pyrolysis gases) are released and convected towards the flame by buoyancy. Provided their temperature and the amount of air in contact with pyrolysis gases are large enough, flaming combustion can self-sustain and thereby, enhance fire propagation. Combustion between pyrolysis gases and oxidizer requires a chemical kinetic scheme suitable for a compressible large-eddy simulation (LES) capability such as AVBP*, in terms of computational cost and chemical stiffness. It is therefore essential to rely on reduced kinetic schemes to predict realistic biomass combustion conditions in LES. While 2S-CH4-BFER (see Section 2.4.3, Chapter 2) is the reduced kinetic scheme retained in this thesis, the possibility of using a 5-step reduced scheme specifically dedicated to pyrolysis gases of Mediterranean biomass fuels (Pérez-Ramirez et al., 2012) and referred to as 5S-GLO-pyr was also explored. See Section 3.3.1 in Chapter 3 for the complete analysis of the flame thermo-chemical features corresponding to biomass combustion.*

In this appendix, the objective is to present the 5S-GLO-pyr reduced chemical kinetic scheme and to propose an additional calibration named 5S-GLO-pyr, in order to evaluate the possibility of including 5S-GLO-pyr* in more complex flame configurations than perfectly-stirred reactor for future quantification of wildfire emissions.*

D.1 Original formulation

5S-GLO-pyr is a 5-step kinetic scheme derived from the mechanism due to Revel et al. (1994) for CH_4 oxidation, which was shown to be consistent with the chemical pathways for methane combustion in wildfires (Leroy, 2007). 5S-GLO-pyr includes additional species compared to 2S-CH4-BFER (i.e., CH_3, CH_2O, H_2) through the following reactions:

$$\begin{aligned}
&\text{(R1)} \quad 2\,CH_4 \longrightarrow 2\,CH_3 + H_2 \\
&\text{(R2)} \quad 2\,CH_3 + O_2 \longrightarrow 2\,CH_2O + H_2 \\
&\text{(R3)} \quad 2\,CH_2O + O_2 \longrightarrow 2\,CO + 2\,H_2O \\
&\text{(R4)} \quad 2\,H_2 + O_2 \longleftrightarrow 2\,H_2O \\
&\text{(R5)} \quad 2\,CO + O_2 \longleftrightarrow CO_2.
\end{aligned} \tag{D.1}$$

While (R2), (R3) and (R4) correspond to oxidation reactions, the first reaction (R1) describes a dissociation mechanism of CH_4 into methyl radical CH_3 and hydrogen H_2. The final reaction (R5) represents the equilibrium between CO and CO_2. Reaction rates $\dot{q}_r$ are given in Table D.1.

Pre-exponential coefficients of reactions (R1) and (R5f) depend on the equivalence ratio Φ, similarly to the pre-exponential adjustment (PEA) technique retained in the 2S-CH4-BFER scheme. The calibration and validation of the reduced scheme 5S-GLO-pyr were originally performed from PSR simulations (Pérez-Ramirez et al., 2012), in which combustion is characterized by the residence time (1.3 s) as well as for the initial temperature and composition of the gas mixture (since reaction rates are kinetically-controlled). The main objective of the calibration was to provide predictions of CO that are consistent with PSR-based experimental data (Leroy, 2007; Leroy et al., 2008), since this species is essential in the determination of kinetics pathways related to pyrolysis gases and in the assessment of emissions.

It is worth mentioning that this calibration of 5S-GLO-pyr does not satisfy the equilibrium balance for reverse reactions (R4) and (R5); stated differently, the calibration of the related activation energies and pre-exponential factors was performed independently between the forward and backward reactions. Still, Pérez-Ramirez et al. (2012) showed that 5S-GLO-pyr follows the overall trend provided by experimental data in terms of the composition of the burnt gas mixture.

D.2 Proposed calibration

The activation temperature of the CO/CO_2 equilibrium reaction (R5f) is corrected to improve the prediction capability of 5S-GLO-pyr as this reaction significantly affects the burnt gas temperature. This termination reaction modifies the final

Table D.1: *Reaction rates $\dot{q}_r$ of the 5-step kinetic scheme 5S-GLO-pyr (the index f stands for* forward*, the index b stands for* backward*) using CGS units ($\dot{q}_r$ [mol/cm^3/s], $E_{a,r}$ [cal/mol]). Credit: Pérez-Ramirez et al. (2012).*

Reaction	Reaction rate [mol/cm^3/s]	Arrhenius parameters
(R1)	$\dot{q}_1 = k_1\,[\mathsf{CH_4}]^{-0.33}\,[\mathsf{O_2}]^{1.0}\,([\mathsf{CH_3}]+[\mathsf{CH_2O}])^{0.85}\,\exp\left[-E_{a,1}/(R_g\,T)\right]$	$k_1 = \exp\left(27.85 + 0.25\,\Phi\right)$ [CGS] $E_{a,1} = 41670$ cal/mol
(R2)	$\dot{q}_2 = k_2\,[\mathsf{CH_3}]^{0.94}\,[\mathsf{O_2}]^{0.66}\,\exp\left[-E_{a,2}/(R_g\,T)\right]$	$k_2 = 1.07 \times 10^{12}$ [CGS] $E_{a,2} = 36002$ cal/mol
(R3)	$\dot{q}_3 = k_3\,[\mathsf{CH_2O}]^{1.11}\,[\mathsf{O_2}]^{0.38}\,\exp\left[-E_{a,3}/(R_g\,T)\right]$	$k_3 = 1.06 \times 10^{13}$ [CGS] $E_{a,3} = 41976$ cal/mol
(R4f)	$\dot{q}_{4f} = k_{4f}\,[\mathsf{H_2}]^{1.00}\,[\mathsf{O_2}]^{0.50}\,\exp\left[-E_{a,4f}/(R_g\,T)\right]$	$k_{4f} = 2.90 \times 10^{13}$ [CGS] $E_{a,4f} = 48484$ cal/mol
(R4r)	$\dot{q}_{4r} = k_{4r}\,[\mathsf{H_2O}]^{1.00}\,\exp\left[-E_{a,4r}/(R_g\,T)\right]$	$k_{4r} = 3.93 \times 10^{12}$ [CGS] $E_{a,4r} = 106058$ cal/mol
(R5f)	$\dot{q}_{5f} = k_{5f}\,[\mathsf{CO}]^{1.00}\,[\mathsf{O_2}]^{0.50}\,\exp\left[-E_{a,5f}/(R_g\,T)\right]$	$k_{5f} = \exp\left(33.40 - 3.50\,\Phi\right)$ [CGS] $E_{a,5f} = 47773$ cal/mol
(R5r)	$\dot{q}_{5r} = k_{5r}\,[\mathsf{CO_2}]^{1.00}\,\exp\left[-E_{a,5r}/(R_g\,T)\right]$	$k_{5r} = 2.90 \times 10^{13}$ [CGS] $E_{a,5r} = 112042$ cal/mol

mixture composition and does not satisfy the equilibrium balance. The correction criterion is based on the deviation from GRI-Mech3.0 predictions of the adiabatic burnt gas temperature for all considered initial mixture temperatures (i.e., from 900 K to 1200 K) and equivalence ratios (i.e., from 0.5 to 2.0). Based on this criterion, the activation energy $E_{a,5f}$ is modified from 47773 to 60773 cal/mol. As shown in Fig. D.1 for a 1000 K initial gas temperature, $E_{a,5\mathrm{f}} = 60773$ cal/mol minimizes the deviation from GRI-Mech3.0 burnt gas temperature predictions for all considered values of the equivalence ratio Φ.

Since auto-ignition delay times predicted by 5S-GLO-pyr also significantly deviate from GRI-Mech3.0 predictions, an additional correction step is proposed to predict more accurate values. Since the dissociation reaction (R1) is mainly responsible for the initiation of the combustion process, its pre-exponential factor k_1 is corrected as a function of the equivalence ratio Φ based on the kinetic scheme presented in Table D.1 with $E_{a,5\mathrm{f}} = 60773$ cal/mol. Thus, the multiplication factor $c_1(\Phi)$ is introduced such that the corrected pre-exponential factor reads $k_{1,cor} = k_1\, c_1(\Phi)$. The variations of the auto-ignition delay time with respect to this multiplication factor $c_1(\Phi)$ are presented in Fig. D.2(a); the value $c_1(\Phi) = 1$ corresponds to the standard version of the global scheme 5S-GLO-pyr.

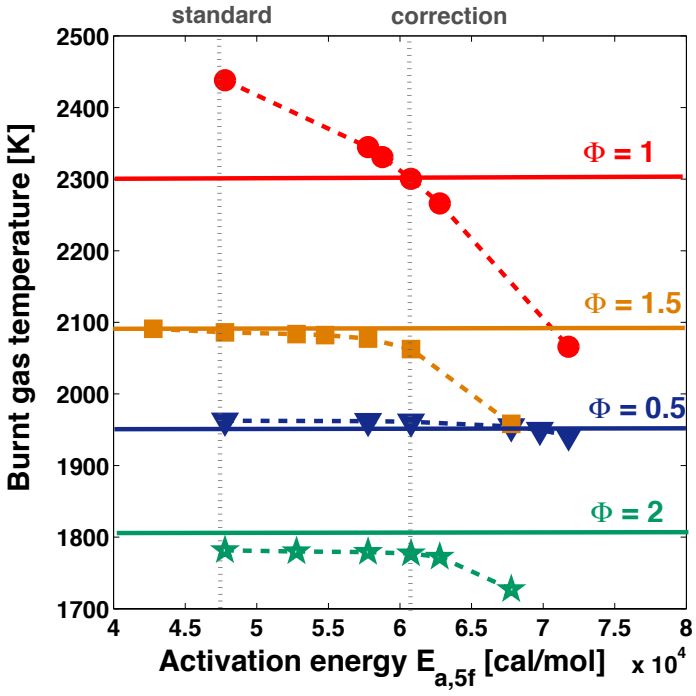
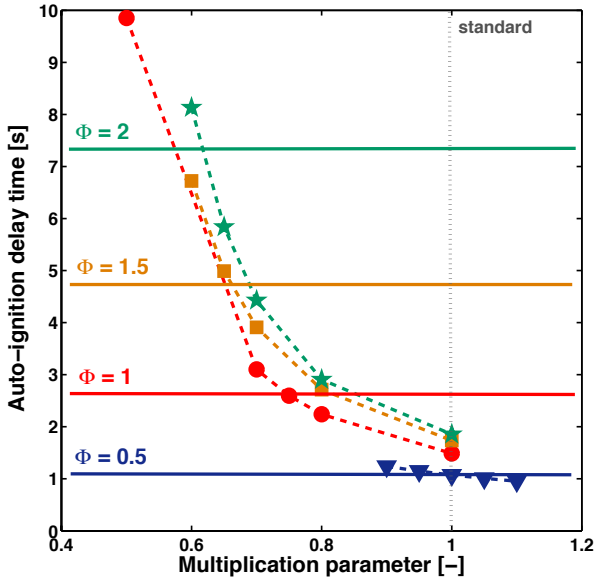


Figure D.1: *Burnt gas temperature with respect to $E_{a,5\mathrm{f}}$ for Φ varying between 0.5 and 2.0; the initial gas temperature is 1000 K. Solid lines correspond to GRI-Mech3.0 predictions; dashed lines correspond to 5S-GLO-pyr predictions (blue triangles for $\Phi = 0.5$, red dots for $\Phi = 1$, orange squares for $\Phi = 1.5$ and green stars for $\Phi = 2$).*

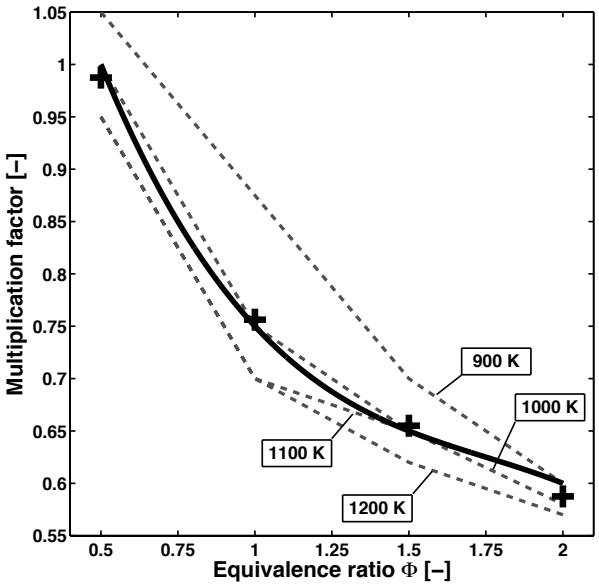
The optimal multiplication factor for the 1000 K initial gas temperature is determined from Fig. D.2(a) and corresponds to the value that minimizes the deviation from GRI-Mech3.0 predictions; this value changes with Φ. The corrected pre-exponential factor $k_{1,cor}$ shown in Fig. D.2(b) is obtained by taking the mean value of the optimal multiplication factor over the considered range of initial gas temperatures (from 900 K to 1200 K). This behavior is fitted over the whole range of equivalence ratios from 0.5 to 2 such that $k_{1,cor}$ reads:

$$k_{1,cor} = k_1 \, c_1(\Phi) = k_1 \, \left(0.20\,\Phi^2 - 0.76\,\Phi + 1.325\right). \tag{D.2}$$

Figure D.3 presents the resulting predictions of the calibrated global scheme (referred to as 5S-GLO-pyr*) for the 1000 K initial gas temperature. As expected, they show the overall significant improvement in adiabatic burnt gas temperatures and auto-ignition delay times of 5S-GLO-pyr* for different states of pyrolysis gases (in terms of the equivalence ratio and the initial gas temperature). Similar behavior is retrieved for the initial gas temperature varying between 900 K and 1200 K. Besides, the correction of the activation energy $E_{a,5\mathrm{f}}$ induces a change in the $\mathrm{CO}/\mathrm{CO}_2$ equilibrium; accordingly, the mass fractions of CO and CO_2 are modified compared to 5S-GLO-pyr, while still being consistent with GRI-Mech3.0 predictions. It is worth noting that both the global scheme 5S-GLO-pyr and its corrected counterpart 5S-GLO-pyr* tend to underestimate the mass fraction of CO_2 and consistently, overestimate the mass fraction of CO for high equivalence ratios Φ. Further improvement of 5S-GLO-pyr* would be required to quantify wildfire emissions.

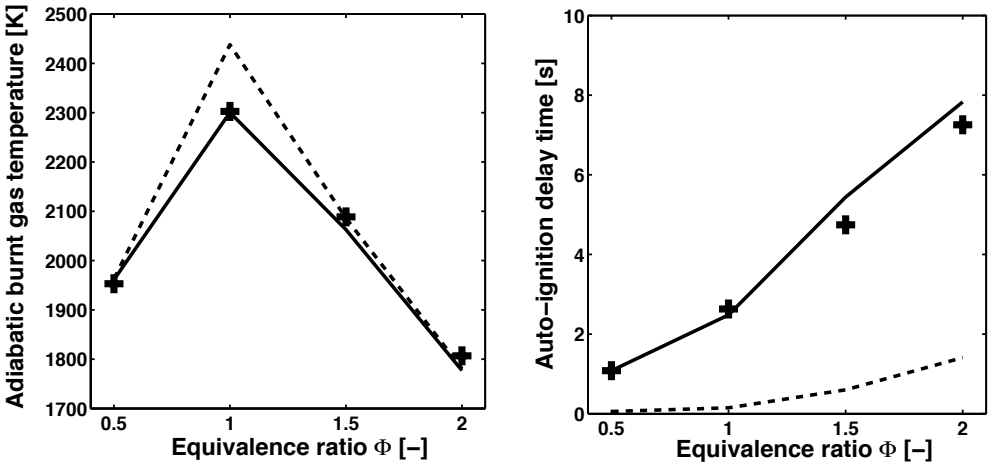


(a)

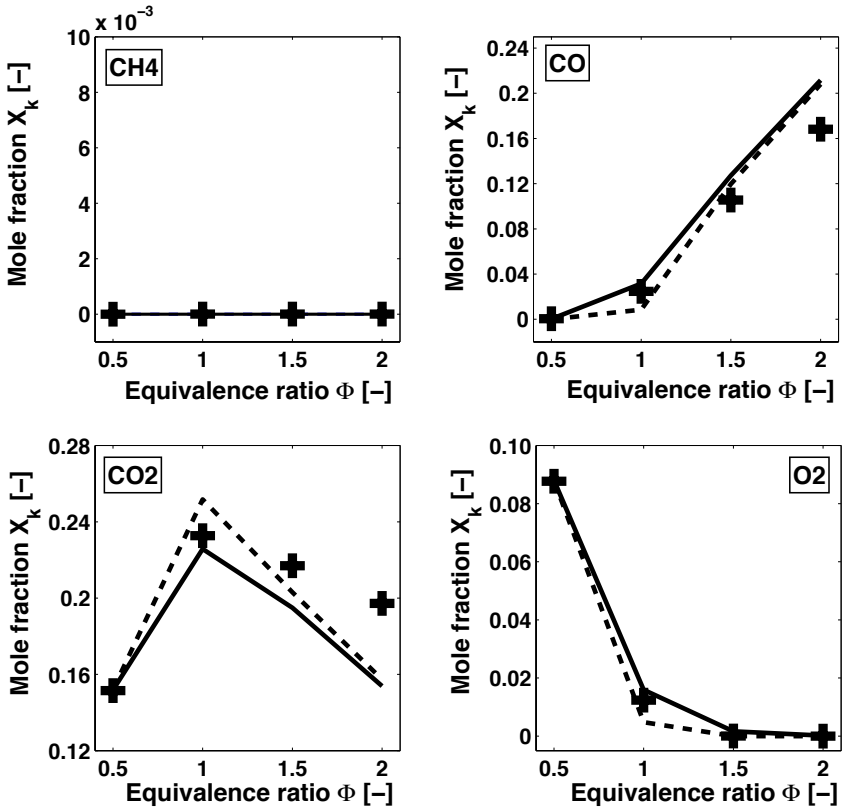


(b)

Figure D.2: *(a) Auto-ignition delay time with respect to the multiplication factor $c_1(\Phi)$ in the corrected formulation of the pre-exponential factor k_1 for Φ varying between 0.5 and 2.0; the initial gas temperature is 1000 K. Solid lines correspond to GRI-Mech3.0 predictions; dashed lines correspond to 5S-GLO-pyr predictions (blue triangles for $\Phi = 0.5$, red dots for $\Phi = 1$, orange squares for $\Phi = 1.5$ and green stars for $\Phi = 2$). (b) Multiplication factor $c_1(\Phi)$ with respect to Φ. Crosses correspond to the mean correction over the range of initial gas temperatures (i.e., 900 K to 1200 K); solid line corresponds to the analytical fit of the multiplication factor (see Eq. D.2); and dashed lines correspond to the optimal multiplication factor per initial gas temperature.*



(a) Burnt gas temperature (left) and auto-ignition delay time (right).



(b) Burnt gas composition.

Figure D.3: *Predictions of the 5S-GLO-pyr reduced kinetic scheme with (solid line) and without (dashed line) correction with respect to the equivalence ratio Φ for the 1000 K initial gas temperature.*

Appendix E

Basics and principles of particle image velocimetry

Particle image velocimetry (PIV) is an optical laser diagnostic that indirectly measures the gas flow velocity field through the reconstructed displacement of tracer particles in the flow. This appendix highlights the main aspects underlying the PIV optical diagnostic and our approach to characterize air entrainment induced by laboratory-scale buoyant flame propagation. The objective of this work (within the framework of the ANR-IDEA project) was to demonstrate the feasibility of PIV to quantify air entrainment effects induced by buoyancy in typical laboratory-scale fires.

Technical background

PIV measurements generally provide a two-dimensional flow velocity field and thereby, capture spatial, instantaneous features of the gas flow. The description of the different steps involved in these PIV measurements follows (Melling, 1997; Raffel et al., 1998; Adrian, 2005); the corresponding flowchart is illustrated in Fig. E.1.

- **Flow seeding with tracer particles.** The reconstruction of the velocity field is based on tracking tracer particles added to the flow. Typically, these particles are liquid oil droplets or solid material that can take different sizes (from 0.5 to 100 μm) depending on experimental requirements. The choice of these particles is a key aspect in the development of a PIV-based analysis since the tracer particles must be sufficiently small to track properly the flow motion (significant discrepancies in density between the gas flow and tracer particles could induce inertial forces). However, their size must not be too small in order to preserve their light-scattering efficiency properties. A balance between these two requirements must be found to ensure the good-quality of PIV measurements.

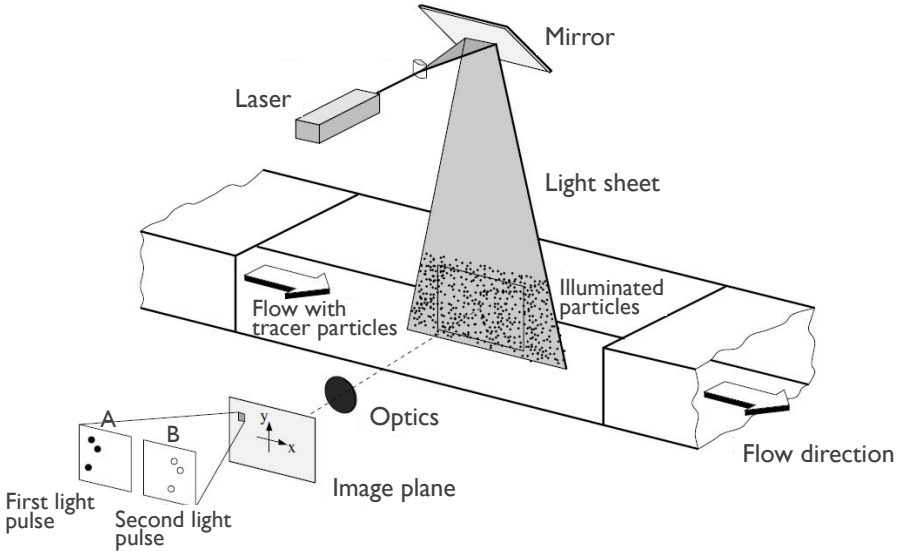


Figure E.1: *Principles of particle image velocimetry. Credit: Raffel et al. (1998).*

▷ **Illumination of particles using two successive laser pulses.** Tracer particles are illuminated using a laser in a two-dimensional section of the flow (i.e., the laser light-sheet) twice within a short time interval noted Δt (see Fig. E.2). In return, these particles scatter light, which is recorded on separate frames (named A and B) using a high-resolution digital camera.

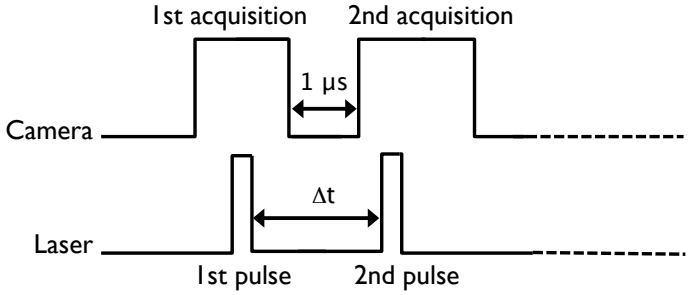


Figure E.2: *Digital timing diagram illustrating the sequential synchronization between the PIV camera and the two laser pulses. 1 μs corresponds to the minimal time delay between two camera acquisitions, and Δt corresponds to the time delay between two successive laser pulses to record frames A and B.*

▷ **Reconstruction of the tracer particle displacement field.** Between the two laser pulses, the tracer particles move according to the local gas flow velocity. Thus, changes in the spatial pattern of particle scattering within this time period are used to retrieve the displacement field of the particles based on sophisticated post-processing. For this purpose, the PIV frame (or visualization window) is divided into small areas referred to as *interrogation areas*

as shown in Fig. E.3 for the present fire spread application. Using statistical methods (e.g., auto- and cross-correlations), the displacement vector is locally estimated for each interrogation area assuming tracer particles move homogeneously within one interrogation area. Thus, the displacement field can be calculated for the whole PIV frame, and an instantaneous two-dimensional velocity field on the laser light-sheet can be derived (see Fig. E.4).

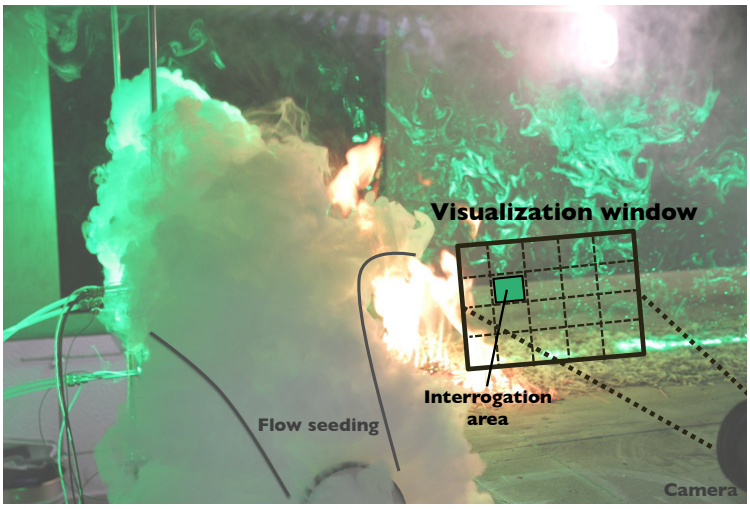


Figure E.3: *Schematic of the PIV frame or visualization window (115 mm × 59 mm, i.e., 13.6 pixel/mm) and its related corresponding areas for the PIV fire spread application (the visualization window is aligned with the top surface of the fuel bed).*

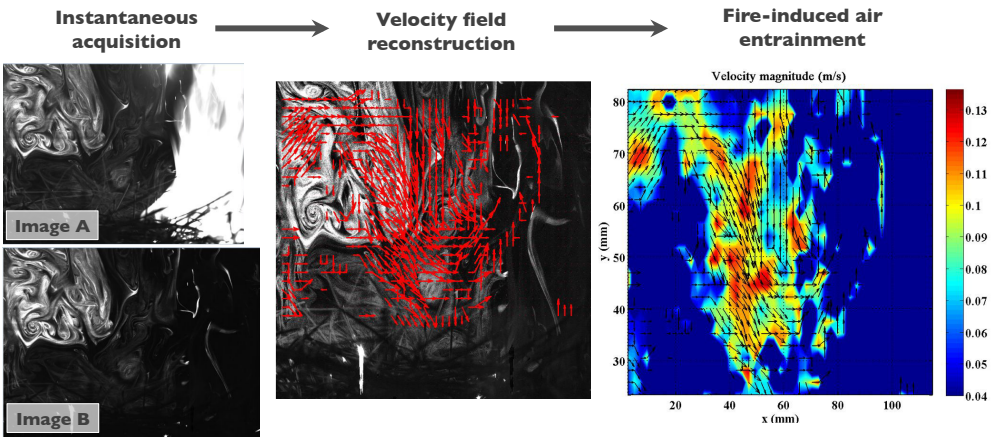


Figure E.4: *Example of PIV acquisition for the gas flow above the pre-heated biomass fuel ahead of the flame region in the unburnt zone.*

Note that the time delay between two successive laser pulses must be consistent with the mean flow velocity to be able to accurately reconstruct the gas flow velocity and its statistics. Note also that the spatial resolution of the PIV post-processing

and in particular the number of interrogation areas within the PIV frame must be chosen carefully, since each interrogation area must contain a minimum number of particles to provide reliable statistics of the tracer particle displacement.

Appendix F

Basics of probability and statistics

Estimation theory provides a broad and natural mathematical foundation for data assimilation science. Basic notations and definitions for probability and statistics are reviewed in this appendix in order to provide a comprehensive introduction to the data assimilation framework.

F.1 Probability density function & Statistical moments

F.1.1 Definitions

In probability and statistics, a random variable or stochastic variable $\mathbf{X}$ is a variable that can take a range of possible different values, each value with an associated probability P.

$\hookrightarrow$ Distribution and probability density function

Consider a multi-variate random variable $\mathbf{X} = [\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n]^{\mathrm{T}} \in \mathbb{R}^n$, each variable X_i $(i = 1, \cdots, n)$ being a scalar. $\mathbf{X}$ is related to a distribution function $F(\mathbf{x})$, $F : \mathbb{R}^n \longrightarrow [0, 1]$, describing the probability P that a realization of $\mathbf{X}$ takes a value less than or equal to $\mathbf{x}$. This distribution function is related to a probability density function (PDF), noted $\mathrm{p}_{\mathbf{X}}(\mathbf{x})$, which measures the probability that $\mathbf{X}$ takes a particular value $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n]^{\mathrm{T}}$ through the following relation:

$$F(\mathbf{x}) = P[\mathbf{X} < \mathbf{x}] = \int_{-\infty}^{\mathbf{x}_1} \cdots \int_{-\infty}^{\mathbf{x}_n} \mathrm{p}_{\mathbf{X}}(\xi)\, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_n. \tag{F.1}$$

We abbreviate this definition of $F(\mathbf{x})$ by the compact notation:

$$F(\mathbf{x}) = P[\mathbf{X} < \mathbf{x}] = \int_{-\infty}^{\mathbf{x}} \mathrm{p}_{\mathbf{X}}(\xi)\, \mathrm{d}\xi, \tag{F.2}$$

where the following conditions on the probability density function $\mathrm{p}_{\mathbf{X}}(\mathbf{x})$ are satisfied: (1) $\mathrm{p}_{\mathbf{X}}(\mathbf{x}) \geq 0$ for any realization $\mathbf{x}$; and (2) $\int_{\mathbb{R}^n} \mathrm{p}_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. Note that in the manuscript no notational distinction is made between a random variable and its realizations; the stand-alone notation $\mathbf{x}$ with its associated PDF $\mathrm{p}(\mathbf{x})$ is used.

↪ Joint and conditional probability density functions

A joint PDF $\mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ describes the probability of simultaneous occurrence of two events $\mathbf{x}$ and $\mathbf{y}$, and a conditional PDF $\mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y})$ describes the probability of occurrence of an event $\mathbf{x}$ given the occurrence of an event $\mathbf{y}$. The following expression holds:

$$\mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y})\,\mathrm{p}_{\mathbf{Y}}(\mathbf{y}) = \mathrm{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\,\mathrm{p}_{\mathbf{X}}(\mathbf{x}), \tag{F.3}$$

where the marginal densities $\mathrm{p}_{\mathbf{X}}(\mathbf{x})$ and $\mathrm{p}_{\mathbf{Y}}(\mathbf{y})$ are defined by:

$$\begin{cases} \mathrm{p}_{\mathbf{X}}(\mathbf{x}) = \displaystyle\int_{-\infty}^{\mathbf{y}} \mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\eta)\,\mathrm{d}\eta \\ \mathrm{p}_{\mathbf{Y}}(\mathbf{y}) = \displaystyle\int_{-\infty}^{\mathbf{x}} \mathrm{p}_{\mathbf{X},\mathbf{Y}}(\xi,\mathbf{y})\,\mathrm{d}\xi \end{cases}. \tag{F.4}$$

If $\mathbf{X}$ and $\mathbf{Y}$ are independent, the relation $\mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \mathrm{p}_{\mathbf{X}}(\mathbf{x})\,\mathrm{p}_{\mathbf{Y}}(\mathbf{y})$ leads to the intuitive result $\mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \mathrm{p}_{\mathbf{X}}(\mathbf{x})$, or alternatively $\mathrm{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \mathrm{p}_{\mathbf{Y}}(\mathbf{y})$.

F.1.2 Characteristics of the probability density function

↪ Scalar (univariate) random variable

Consider a scalar random variable $\mathbf{X}$. The PDF $\mathrm{p}_{\mathbf{X}}(\mathbf{x})$ contains a large amount of information that characterize the variability in the realizations of the random variable $\mathbf{X}$. Some important properties of the PDF are briefly defined here:

- ▷ **mode**: the value of the random variable $\mathbf{X}$ with highest probability (i.e., the location of the PDF peak);
- ▷ **head/tail**: the range of values of the random variable $\mathbf{X}$ over which the PDF is relatively high/low;
- ▷ **mean (or expected value)**: the weighted average of the possible values for $\mathbf{X}$ using their probabilities as their weights, noted $\mu_{\mathbf{X}}$ and expressed formally as:

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \int_{-\infty}^{+\infty} \xi\,\mathrm{p}_{\mathbf{X}}(\xi)\,\mathrm{d}\xi, \tag{F.5}$$

  with $\mathbb{E}[\,\cdot\,]$ referring to the mathematical expectation operator. Note that if the relation $\mathbb{E}[\mathbf{X}] = 0$ is satisfied, the random variable $\mathbf{X}$ is unbiased. By

extension, if $g(\mathbf{X})$ is a deterministic function of $\mathbf{X}$, then its mean value reads:

$$\mathbb{E}[g(\mathbf{X})] = \int_{-\infty}^{+\infty} g(\xi)\, \mathrm{p}_{\mathbf{X}}(\xi)\, \mathrm{d}\xi, \tag{F.6}$$

▷ **standard deviation/variance**: the standard deviation (STD) represents the average deviation from the mean (in the unit of the random variable) and thus, the square-root of the variance. This variance noted $(\sigma_{\mathbf{X}})^2$ is given by:

$$(\sigma_{\mathbf{X}})^2 = \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])^2\Big] = \int_{-\infty}^{+\infty} (\xi - \mathbb{E}[\mathbf{X}])^2\, \mathrm{p}_{\mathbf{X}}(\xi)\, \mathrm{d}\xi = \mathbb{E}[\mathbf{X}^2] - \mathbb{E}[\mathbf{X}]^2 \tag{F.7}$$

using the linearity property of the expectation operator $\mathbb{E}[\,\cdot\,]$. The variance is a convenient measure of the dispersion of the realizations around the mean value $\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}]$.

In the case of a joint PDF $\mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ or a conditional PDF $\mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y})$ associated with the scalar random variables $\mathbf{X}$ and $\mathbf{Y}$, we can define additional properties presented below:

▷ **conditional mean value** of $\mathbf{X}$ given $\mathbf{Y}$, noted $\mathbb{E}[\mathbf{X} \mid \mathbf{Y}]$ and expressed as a function of the random variable $\mathbf{Y}$ such that:

$$\mathbb{E}[\mathbf{X} \mid \mathbf{Y}] = \int_{-\infty}^{+\infty} \xi\, \mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\xi|\mathbf{y})\, \mathrm{d}\xi. \tag{F.8}$$

▷ **covariance** of $\mathbf{x}$ and $\mathbf{y}$, noted $\mathrm{cov}(\mathbf{X},\mathbf{Y})$ and given by:

$$\begin{aligned} \mathrm{cov}(\mathbf{X},\mathbf{Y}) &= \mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])\,(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])\Big] \\ &= \iint_{-\infty}^{+\infty} (\xi - \mathbb{E}[\mathbf{X}])(\eta - \mathbb{E}[\mathbf{Y}])\, \mathrm{p}_{\mathbf{X},\mathbf{Y}}(\xi,\eta)\, \mathrm{d}\xi\, \mathrm{d}\eta \\ &= \Big(\iint_{-\infty}^{+\infty} \xi\eta\, \mathrm{p}_{\mathbf{X},\mathbf{Y}}(\xi,\eta)\, \mathrm{d}\xi\, \mathrm{d}\eta\Big) - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}]. \end{aligned} \tag{F.9}$$

Note that the covariance of a random variable with itself is the variance, i.e., $\mathrm{cov}(\mathbf{X},\mathbf{X}) = (\sigma_{\mathbf{X}})^2$. Note also that if the random variables $\mathbf{X}$ and $\mathbf{Y}$ are independent, i.e., $\mathrm{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \mathrm{p}_{\mathbf{X}}(\mathbf{x})\mathrm{p}_{\mathbf{Y}}(\mathbf{y})$, the covariance $\mathrm{cov}(\mathbf{X},\mathbf{Y})$ becomes zero.

▷ **correlation** of $\mathbf{X}$ and $\mathbf{Y}$, noted $\rho(\mathbf{X},\mathbf{Y})$ and given by the Pearson's formula:

$$\rho(\mathbf{X},\mathbf{Y}) = \frac{\mathbb{E}\Big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])\,(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])\Big]}{\sigma_{\mathbf{X}}\,\sigma_{\mathbf{Y}}} = \frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\sigma_{\mathbf{X}}\,\sigma_{\mathbf{Y}}} \tag{F.10}$$

with $\sigma_{\mathbf{X}}$ and $\sigma_{\mathbf{Y}}$ the STD of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Equation (F.10) implies that the correlation between a pair of random variables is obtained by dividing the covariance of the two variables by the product of their STD, and consequently, the correlation of a random variable with itself is systematically 1. More generally, a correlation is a dimensionless quantity that indicates the degree of dependence between a set of random variables: the closer the correlation coefficient $\rho(\mathbf{X}, \mathbf{Y})$ is to either -1 or 1, the stronger the correlation between the random variables.

↪ Multi-variate random variable

In the context of a multi-variate random variable $\mathbf{X} = [\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n]^{\mathrm{T}} \in \mathbb{R}^n$, the definitions given above for a scalar random variable can be extended, in particular for the mean value $\mathbb{E}[\mathbf{X}]$ with:

$$\mu_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[\mathrm{X}_1] \\ \mathbb{E}[\mathrm{X}_2] \\ \vdots \\ \mathbb{E}[\mathrm{X}_n] \end{pmatrix} = \begin{pmatrix} \mu_{\mathrm{X}_1} \\ \mu_{\mathrm{X}_2} \\ \vdots \\ \mu_{\mathrm{X}_n} \end{pmatrix}, \tag{F.11}$$

and for the covariance $\mathrm{cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])\,(\mathbf{X} - \mathbb{E}[\mathbf{X}])]$ that becomes a $n \times n$ matrix (instead of a scalar for a univariate random variable) noted $\mathbf{C_{XX}}$ and defined as follows:

$$\mathbf{C_{XX}} = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{pmatrix} \tag{F.12}$$

$$\Rightarrow \mathbf{C_{XX}} = \begin{pmatrix} \sigma_{\mathrm{X}_1} & & \\ & \ddots & \\ & & \sigma_{\mathrm{X}_n} \end{pmatrix} \mathit{\Lambda}_{\mathbf{XX}} \begin{pmatrix} \sigma_{\mathrm{X}_1} & & \\ & \ddots & \\ & & \sigma_{\mathrm{X}_n} \end{pmatrix},$$

with

$$\mathit{\Lambda}_{\mathbf{XX}} = \begin{pmatrix} \rho(\mathrm{X}_1, \mathrm{X}_1) & \cdots & \rho(\mathrm{X}_1, \mathrm{X}_n) \\ \vdots & \ddots & \vdots \\ \rho(\mathrm{X}_n, \mathrm{X}_1) & \cdots & \rho(\mathrm{X}_n, \mathrm{X}_n) \end{pmatrix}. \tag{F.13}$$

The matrix element $C_{ij} = C_{ji} = \mathbb{E}[(\mathrm{X}_i - \mathbb{E}[\mathrm{X}_i])(\mathrm{X}_j - \mathbb{E}[\mathrm{X}_j])]$ corresponds to the covariance $\mathrm{cov}(\mathrm{X}_i, \mathrm{X}_j)$ of X_i and X_j, and where $\rho(\mathrm{X}_i, \mathrm{X}_j) = C_{ij}/(\sigma_{\mathrm{X}_i}\sigma_{\mathrm{X}_j})$ corresponds to the correlation of X_i and X_j with σ_{X_i} the STD of X_i, σ_{X_j} the STD of X_j and $\rho(\mathrm{X}_i, \mathrm{X}_i) = 1$.

F.1.3 Normal/Gaussian probability density function

Important and commonly known probability distribution include the Gaussian PDF, defined solely by its mean $\mu_{\mathbf{X}}$ (first moment) and its variance $(\sigma_{\mathbf{X}})^2$ (second mo-

ment). If the random variable $\mathbf{X}$ follows a Gaussian PDF noted $\mathcal{N}(\mu_{\mathbf{X}}, (\sigma_{\mathbf{X}})^2)$, the PDF is of the following form:

$$\mathrm{p}_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sigma_{\mathbf{X}}\sqrt{2\pi}} \exp\left\{-\frac{(\mathbf{x}-\mu_{\mathbf{X}})^2}{2(\sigma_{\mathbf{X}})^2}\right\}, \tag{F.14}$$

Note that $\mu_{\mathbf{X}} \pm \sigma_{\mathbf{X}}$ spans 68% of the realizations of the random variable $\mathbf{X}$, while $\mu_{\mathbf{X}} \pm 2.57\sigma_{\mathbf{X}}$ spans 99% of its realizations. By applying the variable transformation $\mathbf{U} = (\mathbf{X}-\mu_{\mathbf{X}})/\sigma_{\mathbf{X}}$, the random variable $\mathbf{U}$ follows the standard normal distribution $\mathcal{N}(0,1)$ of zero mean and unit STD; this PDF is represented in Fig. F.1.

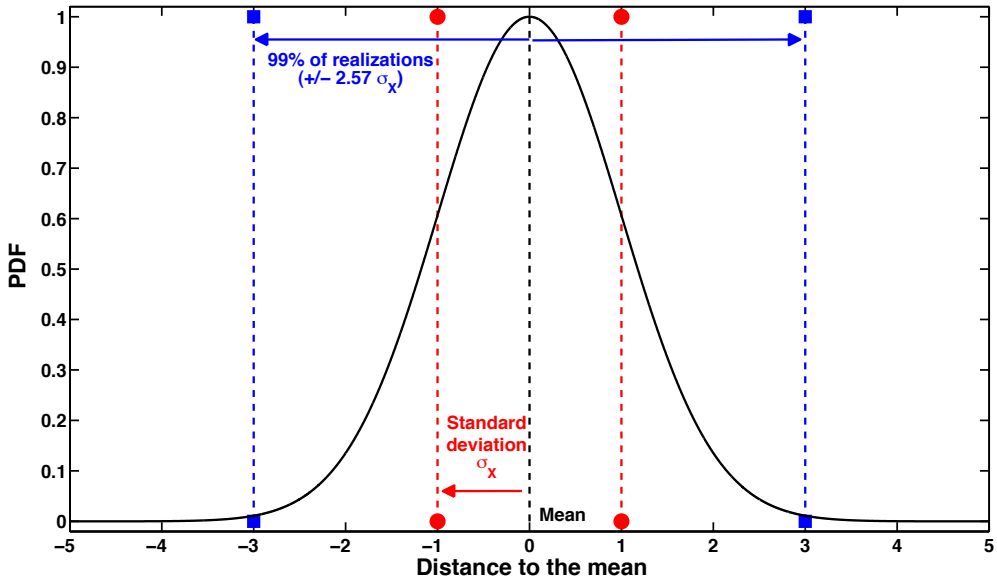


Figure F.1: *Standard normal probability density function $\mathcal{N}(0,1)$ (black solid line).*

F.2 Sample approximations

The exact evaluation of the PDF properties of a multi-variate random variable $\mathbf{X} \in \mathbb{R}^n$ becomes impractical when the dimension n increases. One alternative is to approximate these quantities using Monte Carlo (MC) techniques, considering a large number (sample) of independent realizations N_e from the PDF $\mathrm{p}_{\mathbf{X}}(\mathbf{x})$. In this context, the random variable $\mathbf{X}$ is approximated by a sample noted $[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N_e)}]$, where each realization $\mathbf{x}^{(k)}$ (the subscript k is the index of a particular realization of the random variable $\mathbf{X}$) is a vector of dimension n. From

this set of realizations, the MC-based sample mean $\mu_{\mathbf{X}}^{MC}$ is calculated as follows:

$$\mu_{\mathbf{X}}^{MC} = \frac{1}{N_e}\sum_{k=1}^{N_e}\mathbf{x}^{(k)} = \begin{pmatrix} \frac{1}{N_e}\sum_{k=1}^{N_e}\mathrm{x}_1^{(k)} \\ \vdots \\ \frac{1}{N_e}\sum_{k=1}^{N_e}\mathrm{x}_n^{(k)} \end{pmatrix} = \begin{pmatrix} \mu_{\mathrm{X}_1}^{MC} \\ \vdots \\ \mu_{\mathrm{X}_n}^{MC} \end{pmatrix}. \tag{F.15}$$

Similarly, the MC-based sample covariance matrix $(\mathbf{C_{XX}})^{MC}$ of dimension $n \times n$ reads:

$$(\mathbf{C_{XX}})^{MC} = \frac{1}{N_e - 1}\sum_{k=1}^{N_e}\left(\mathbf{x}^{(k)} - \mu_{\mathbf{X}}^{MC}\right)\left(\mathbf{x}^{(k)} - \mu_{\mathbf{X}}^{MC}\right)^{\mathrm{T}}. \tag{F.16}$$

The structure of $(\mathbf{C_{XX}})^{MC}$ is equivalent to the covariance matrix $\mathbf{C_{XX}}$ with:

$$(\mathbf{C_{XX}})^{MC} = \begin{pmatrix} (C_{11})^{MC} & \cdots & (C_{1n})^{MC} \\ \vdots & \ddots & \vdots \\ (C_{n1})^{MC} & \cdots & (C_{nn})^{MC} \end{pmatrix},$$

where the diagonal terms $(C_{ii})^{MC}$ corresponds to the sample variance associated with the variable X_i of the multi-variate random variable $\mathbf{X}$ such that:

$$(C_{ii})^{MC} = (\sigma_{\mathbf{X}_i}^{MC})^2 = \frac{1}{N_e - 1}\sum_{k=1}^{N_e}\left(\mathrm{x}_i^{(k)} - \mu_{\mathrm{X}_i}^{MC}\right)^2, \tag{F.17}$$

and where the off-diagonal terms $(C_{ij})^{MC} = (C_{ji})^{MC}$ $(i \neq j)$ corresponds to the sample covariance of the variables X_i and X_j such that:

$$(C_{ij})^{MC} = \mathrm{cov}^{MC}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{N_e - 1}\sum_{k=1}^{N_e}\left(\mathrm{x}_i^{(k)} - \mu_{\mathrm{X}_i}^{MC}\right)\left(\mathrm{x}_j^{(k)} - \mu_{\mathrm{X}_j}^{MC}\right). \tag{F.18}$$

Note that the statistical moments of the MC-based sample converge towards the exact properties ($\mu_{\mathbf{X}}$, $\mathbf{C_{XX}}$) with increasing sample size N_e; the decreasing trend of the sample error is proportional to $1/\sqrt{N_e}$ as shown in Fig. F.2 for the sample mean of the standard normal PDF $\mathcal{N}(0, 1)$. The errors in the sample mean are calculated in $\mathbb{L}_1$-norm as $\left|\mu_{\mathbf{X}} - \mu_{\mathbf{X}}^{MC}\right|$.

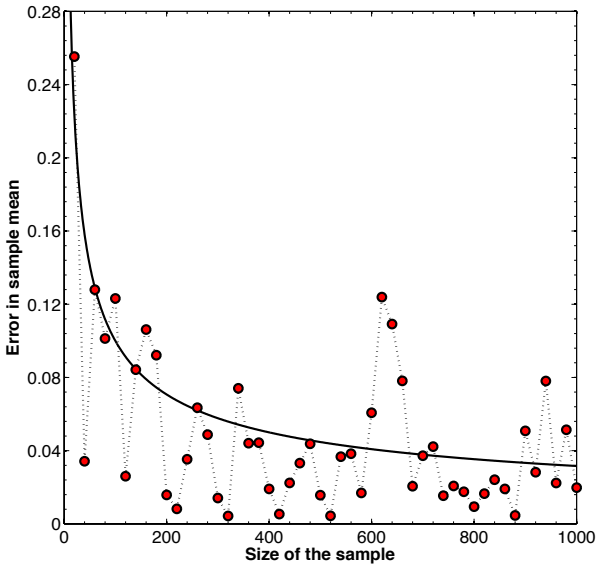


Figure F.2: *Errors in sample mean (dependent on the realizations and shown in red circles) of the standard normal PDF $\mathcal{N}(0,1)$ as a function of the sample size N_e, with the error function $1/\sqrt{N_e}$ represented in solid line.*

F.3 Bayes' theorem

Equation (F.3) leads to a fundamental result in probability theory called the Bayes' theorem and expressed as:

$$\mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{\mathrm{p}_{\mathbf{X}}(\mathbf{x})\, \mathrm{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})}{\mathrm{p}_{\mathbf{Y}}(\mathbf{y})}, \tag{F.19}$$

stating that the conditional PDF of a random variable $\mathbf{X}$ given $\mathbf{Y}$ (the posterior) is obtained by combining the conditional PDF of variable $\mathbf{Y}$ given $\mathbf{X}$ (the likelihood) and the marginal PDF of the random variable $\mathbf{X}$ (the prior). Note that the marginal PDF of the random variable $\mathbf{Y}$, noted $\mathrm{p}_{\mathbf{Y}}(\mathbf{y})$, in the denominator of Eq. (F.19) is a normalizing factor. Thus, the Bayes' theorem is often formulated as:

$$\mathrm{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) \propto \mathrm{p}_{\mathbf{X}}(\mathbf{x})\, \mathrm{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}), \tag{F.20}$$

where the symbol $\propto$ means *proportional to*. This theorem is the foundation of the Bayesian filtering technique, from which multiple existing data assimilation algorithms can be derived (e.g., the Kalman filter and variational approaches), see Chapter 5.

Appendix G

Numerical treatments in the FIREFLY wildfire spread simulator

Firefly simulates the time-evolving fire front location at regional scales (see Fig. G.1) using the following three components: (1) a submodel for the rate of spread (ROS) Γ parameterized with respect to local environmental conditions (e.g., vegetation, terrain topography and weather properties), (2) a level-set-based solver for the fire front propagation equation that simulates the propagating fire front along its normal direction according to Γ and (3) an iso-contour algorithm for the reconstruction of the fire front.

This appendix provides a detailed description of the numerical implementation of these components (briefly presented in Section 6.4) with a focus on the implementation of the level-set-based solver and on the treatment of complex terrain topography.

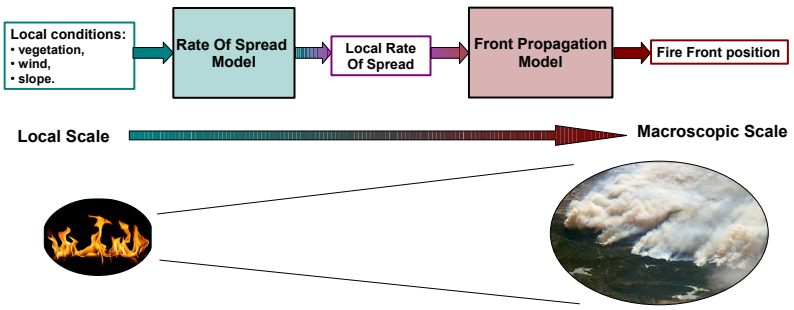


Figure G.1: *Regional-scale wildfire spread modeling. Credit: Delmotte et al. (2011).*

G.1 Implementation of the level-set-based solver

G.1.1 Lagrangian approach versus Eulerian approach

Propagating an interface (e.g., the fireline separating the burned and unburned regions in wildfire spread applications) with a numerical treatment that avoids numerical diffusion and ensures numerical stability requires the use of high-order numerical schemes. Historically, regional-scale simulators of wildfire spread have relied on a Lagrangian front-tracking approach (Fendell and Wolff, 2001) that describes the time-evolving fireline using a set of tracers, see FARSITE (Finney, 1998). However, in Lagrangian solvers the number of required tracers significantly increases with the fire size, and complex numerical treatments are necessary to handle fire crossovers and fire front merging as highlighted by Filippi et al. (2009). In contrast, Eulerian approaches naturally handle complex topology of fire fronts along with collisions and merging of fire fronts. These approaches inevitably exhibit a computational cost higher than Lagrangian approaches (Fendell and Wolff, 2001; Rehm and McDermott, 2009; Mallet et al., 2009; Mandel et al., 2011). An Eulerian level-set-based approach is adopted in FIREFLY, largely inspired from Rehm and McDermott (2009).

G.1.2 Numerical scheme underlying the FIREFLY level-set solver

↪ Definition of the propagating equation

This section describes in detail the numerical method used to solve for the two-dimensional scalar progress variable $c = c(x, y, t)$ in the computational domain Ω, $t \geq 0$: $c = 0$ in the unburnt vegetation, $c = 1$ in the burnt vegetation, and the flame is the region where c takes values between 0 and 1 (the flame front is identified as the progress variable iso-contour $c_{fr} = 0.5$). The propagation equation presented in Chapter 6 reads:

$$\frac{\partial c}{\partial t}(x, y, t) = \mathit{\Gamma}\, |\nabla c| = -\gamma \cdot \nabla c, \tag{G.1}$$

with $|\nabla c|$ the magnitude of the gradient of the progress variable c defined as:

$$|\nabla c| = \sqrt{\left(\frac{\partial c}{\partial x}\right)^2 + \left(\frac{\partial c}{\partial y}\right)^2}, \tag{G.2}$$

and with $\mathit{\Gamma} = \gamma \cdot \mathbf{n}_{fr}$ the normal component of the spread velocity vector γ, defined along the normal direction to the iso-contours of the progress variable c (also referred to as the *normal direction to the fire front*). This normal direction is represented using the normal vector $\mathbf{n}_{fr} = [n_{fr,x}, n_{fr,y}]^{\mathrm{T}}$ pointing towards the unburnt vegetation, i.e.,

$$\mathbf{n}_{fr} = \frac{-\nabla c}{|\nabla c|}. \tag{G.3}$$

Equation (G.1) can be alternatively formulated as follows:

$$\frac{\partial c}{\partial t}(x,y,t) + \gamma_x \left(\frac{\partial c}{\partial x}\right) + \gamma_y \left(\frac{\partial c}{\partial y}\right) = 0, \tag{G.4}$$

where γ_x and γ_y are the components of the spread velocity vector γ along the x- and y-directions with:

$$\begin{aligned} \gamma_x &= \Gamma n_{fr,x}, \\ \gamma_y &= \Gamma n_{fr,y}. \end{aligned} \tag{G.5}$$

Γ is the ROS of the fireline, conveniently identified a posteriori of the resolution of the progress variable $c = c(x,y,t)$ as the iso-contour $c_{fr} = 0.5$ in FIREFLY. One major difference with Rehm and McDermott (2009) is that the target variable c represents a two-dimensional bi-modal field (i.e., burned or non-burned states), instead of a fire front signed function that takes a given constant value (usually referred to as the *level curve*). In this sense, the FIREFLY simulator is not a proper level-set-based solver. Still, FIREFLY shares its Eulerian characteristics and tracks the location of the fireline, derived a posteriori from the solution $c = c(x,y,t)$ at time t.

↪ Basic steps in the numerical resolution

The numerical scheme used to solve Eq. (G.4) relies on a second-order Runge-Kutta scheme for time-integration and an advection algorithm for spatial discretization based on a second-order total variation diminishing (TVD) scheme combined with a Superbee slope limiter (Rehm and McDermott, 2009). This flux-limiting scheme based on the rate and direction of spread Γ is introduced to preserve monotonicity of the scalar field c and to avoid spurious oscillations (which could be induced by the convective terms in the absence of a flux-limiting scheme, when the solution exhibits discontinuities or sharp variations within the computational domain). The basic steps of the numerical scheme are listed below.

(1) Computation of the node-centered gradient using a centered finite difference scheme:

$$\begin{aligned} \left(\frac{\partial c}{\partial x}\right)^{\mathrm{t}}_{i,j} &= \frac{c^{\mathrm{t}}_{i+1,j} - c^{\mathrm{t}}_{i-1,j}}{2\,\Delta x}, \\ \left(\frac{\partial c}{\partial y}\right)^{\mathrm{t}}_{i,j} &= \frac{c^{\mathrm{t}}_{i,j+1} - c^{\mathrm{t}}_{i,j-1}}{2\,\Delta y}, \end{aligned} \tag{G.6}$$

with Δx a uniform mesh cell size along the x-direction and Δy its counterpart along the y-direction, i and j corresponding to the respective mesh index, and with t the time step index.

(2) Computation of the unit normal vector $(\mathbf{n}_{fr})^{\mathrm{t}}_{i,j}$, corresponding to the normal direction of the fireline at the grid node indexed by (i, j), using Eq. (G.2) and Eq. (G.3).

(3) Computation of the spread velocity vector $(\gamma)^{\mathrm{t}}_{i,j}$ using Eq. (G.5).

(4) Determination of the monotonicity preserving scalar gradient ∇c^{t} for the propagating equation with a Superbee slope limiter for the convective terms, at time t.

(5) Time-integration of the propagating Eq. (G.4), from time t to time $(t + 1)$, using a second-order Runge-Kutta scheme defined as a linear combination of two forward Euler steps.

This numerical scheme requires two different computations of the progress variable gradient ∇c, in step 1 for the determination of the normal vector $(\mathbf{n}_{fr})^{\mathrm{t}}_{i,j}$ as well as in step 4 for limiting the gradient of the progress variable c near discontinuities or sharp variations.

↪ Calculation of the monotonicity preserving scalar gradient (step 4)

The calculation of the progress variable gradient ∇c at time t is similar to a flux calculation in fluid mechanics (the time index t is not mentioned in this step for clarify purposes). The values of the progress variable at the cell boundaries (also referred to as *edges*) are calculated to determine the progress variable gradients (along x- and y-directions) required by Eq. (G.4). The x-gradient is defined in terms of the *East* and *West* values of the progress variable c, while the y-gradient is defined in terms of *North* and *South* values as illustrated in Fig. G.2. Using this formalism, the x- and y-gradients read:

$$\frac{\partial c}{\partial x} = \frac{c_{east} - c_{west}}{\Delta x} = \frac{c_{i+1/2,j} - c_{i-1/2,j}}{\Delta x}, \tag{G.7}$$

$$\frac{\partial c}{\partial y} = \frac{c_{north} - c_{south}}{\Delta y} = \frac{c_{i,j+1/2} - c_{i,j-1/2}}{\Delta y}, \tag{G.8}$$

where the values of the progress variable at each edge of the cell (i, j) are determined using the slope and local direction of the spread velocity vector $\gamma = [\gamma_x, \gamma_y]^{\mathrm{T}}$. The details for the calculation of c_{east}, c_{north}, c_{west} and c_{south} follow. The parameter r and the Superbee flux limiter $\mathcal{B}(r)$, based on Toro (1999) and Rehm and McDermott (2009), are defined as follows:

$$r = \frac{\delta c_{up}}{\delta c_{loc}}, \quad \mathcal{B}(r) = \max\Big(0, \min(2r, 1), \min(r, 2)\Big), \tag{G.9}$$

where the subscript up refers to *upwind* and loc refers to *local*. The parameter r represents the ratio of the upwind variations to the local variations; it is used as

an indicator of the discontinuities in the progress variable field c. Note that, in practice, $r = 0$ if $\delta c_{loc} = 0$.

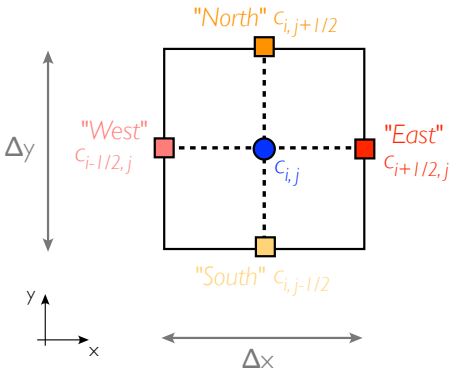


Figure G.2: *Schematic of the progress variable $c_{i,j}$ at the centered node (i,j) with its* East, North, West *and* South *counterparts corresponding to the cell boundaries.*

▷ **East edge**

$$\delta c_{loc} = c_{i+1,j} - c_{i,j}, \quad \delta c_{up} = \begin{cases} c_{i,j} - c_{i-1,j} & \text{for } (\gamma_x)_{i,j} > 0 \\ c_{i+2,j} - c_{i+1,j} & \text{for } (\gamma_x)_{i,j} < 0 \end{cases}$$

$$\Rightarrow c_{east} = \begin{cases} c_{i,j} + \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_x)_{i,j} > 0 \\ c_{i+1,j} - \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_x)_{i,j} < 0 \end{cases}. \tag{G.10}$$

▷ **West edge**

$$\delta c_{loc} = c_{i-1,j} - c_{i,j}, \quad \delta c_{up} = \begin{cases} c_{i-2,j} - c_{i-1,j} & \text{for } (\gamma_x)_{i,j} > 0 \\ c_{i,j} - c_{i+1,j} & \text{for } (\gamma_x)_{i,j} < 0 \end{cases}$$

$$\Rightarrow c_{west} = \begin{cases} c_{i-1,j} - \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_x)_{i,j} > 0 \\ c_{i,j} + \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_x)_{i,j} < 0 \end{cases}. \tag{G.11}$$

▷ **North edge**

$$\delta c_{loc} = c_{i,j+1} - c_{i,j}, \quad \delta c_{up} = \begin{cases} c_{i,j} - c_{i,j-1} & \text{for } (\gamma_y)_{i,j} > 0 \\ c_{i,j+2} - c_{i,j+1} & \text{for } (\gamma_y)_{i,j} < 0 \end{cases}$$

$$\Rightarrow c_{north} = \begin{cases} c_{i,j} + \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_y)_{i,j} > 0 \\ c_{i,j+1} - \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_y)_{i,j} < 0 \end{cases}. \tag{G.12}$$

▷ South edge

$$\delta c_{loc} = c_{i,j-1} - c_{i,j}, \quad \delta c_{up} = \begin{cases} c_{i,j-2} - c_{i,j-1} & \text{for } (\gamma_y)_{i,j} > 0 \\ c_{i,j} - c_{i,j+1} & \text{for } (\gamma_y)_{i,j} < 0 \end{cases}$$

$$\Rightarrow c_{south} = \begin{cases} c_{i,j-1} - \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_y)_{i,j} > 0 \\ c_{i,j} + \frac{1}{2}\,\mathcal{B}(r)\,\delta c_{loc} & \text{for } (\gamma_y)_{i,j} < 0 \end{cases} . \tag{G.13}$$

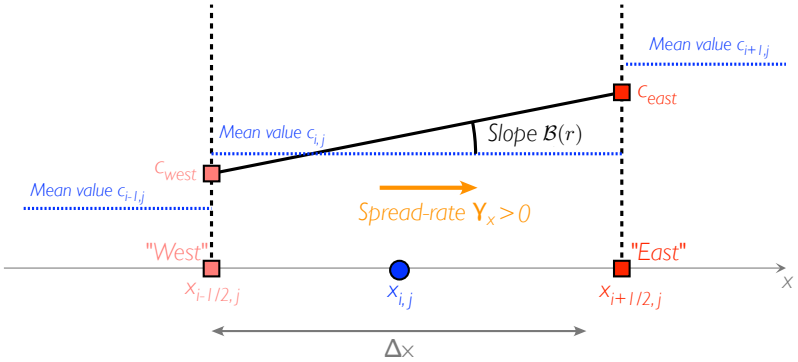


Figure G.3: *Schematic of the slope-limiting calculation of c_{west} and c_{east} along the x-direction. Credit: Delmotte et al. (2011).*

↪ Stability condition

The calculation of the monotonicity preserving scalar gradient leads to the following time-integration of Eq. (G.4) using a second-order Runge-Kutta method (step. 5):

$$c^*_{i,j} = c^{\mathrm{t}}_{i,j} - \Delta t \left\{ \gamma^{\mathrm{t}}_x \left(\frac{c^{\mathrm{t}}_{east} - c^{\mathrm{t}}_{west}}{\Delta x} \right) + \gamma^{\mathrm{t}}_y \left(\frac{c^{\mathrm{t}}_{north} - c^{\mathrm{t}}_{south}}{\Delta y} \right) \right\}$$

$$\Rightarrow c^{\mathrm{t}+1}_{i,j} = \frac{1}{2}\, c^{\mathrm{t}}_{i,j} + \frac{1}{2} \left\{ c^*_{i,j} - \Delta t \left\{ \gamma^*_x \left(\frac{c^*_{east} - c^*_{west}}{\Delta x} \right) + \gamma^*_y \left(\frac{c^*_{north} - c^*_{south}}{\Delta y} \right) \right\} \right\},$$

where the time step Δt is limited by the Courant-Friedrichs-Lewy (CFL) condition. Because of the accuracy of the numerical scheme in space, the CFL condition for Firefly is more restrictive:

$$\Delta t \left(\frac{\max(\gamma_x)}{\Delta x} + \frac{\max(\gamma_y)}{\Delta y} \right) \leq K, \tag{G.14}$$

where the value of the parameter K was empirically determined; $K = 0.3$ ensures a systematic numerical stability of the numerical scheme (Delmotte et al., 2011).

G.2 Iso-contour algorithms for front reconstruction

G.2.1 Selection scheme versus projection scheme

FIREFLY simulations indirectly provide the location of the iso-contour $c_{fr} = 0.5$, which is discretized using an iso-contour algorithm by N_{fr} markers, characterized by the two-dimensional coordinates $[(x_1, y_1), \ldots, (x_{N_{fr}}, y_{N_{fr}})]$. This discretization of the simulated fire front with N_{fr} markers corresponds to the first step of this iso-contour algorithm.

Within the framework of data assimilation, the innovation vector $\mathbf{d}^{\mathrm{f}}$ measures the distance between the simulated (forecast) estimates of the fire fronts and the observations. Note that the FIREFLY solver uses a high-resolution computational grid that allows for a detailed representation of the local conditions. In contrast, observations of the fire front position are likely to be provided with a much coarser resolution; in addition, observations may be incomplete and cover only a fraction of the fire front perimeter. Thus, we may expect the number of observed markers N^{o}_{fr} to be much lower than N_{fr}. Thus, the second step in the iso-contour algorithm (also referred to as *mapping*) consists in determining the equivalent of the N^{o}_{fr} observed markers onto the simulated iso-contour $c_{fr} = 0.5$. The observed markers are defined as

$$\mathbf{y}^{\mathrm{o}}_t = \left[(x^{\mathrm{o}}_1, y^{\mathrm{o}}_1), (x^{\mathrm{o}}_2, y^{\mathrm{o}}_2), \ldots, (x^{\mathrm{o}}_{N^{\mathrm{o}}_{fr}}, y^{\mathrm{o}}_{N^{\mathrm{o}}_{fr}})\right].$$

This mapping can be simply performed through a selection of the N^{o}_{fr} markers among the fine-grained discretization of the simulated fire front. In Section 6.4 (Chapter 6), the corresponding selection operator $\mathcal{H}$ is simply described as an operator that takes 1 out of every r markers, with $N^{\mathrm{o}}_{fr} = (N_{fr}/r)$ and r an integer taking values (much) larger than 1. Figure G.4 is an example of this simple selection procedure.

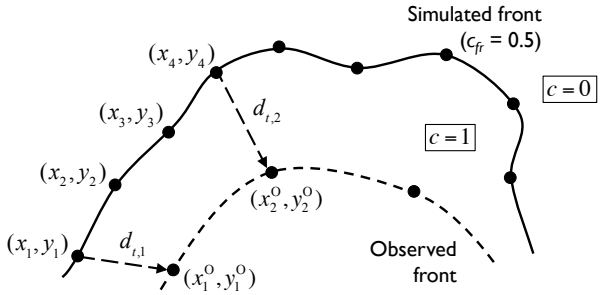


Figure G.4: *Calculation of the discrepancies between simulated and observed fire fronts. In this illustration,* $r = 4$.

However, this selection operator $\mathcal{H}$ may be defined in several ways, for instance using projection schemes (Rochoux et al., 2010; Delmotte et al., 2011). The objective of projection schemes is to determine the equivalent of the N^{o}_{fr} observed markers onto the simulated iso-contour $c_{fr} = 0.5$, *equivalent* meaning that the location

of the N_{fr}^{o} simulated markers is calculated by reconstructing the trajectory of the observed markers in space and by assuming that discrepancies between observed and simulated fire fronts are due to a temporal shift. Basically, these projection schemes are performed by advancing every observed marker i, originally located at (x_i^o, y_i^o), along a specified direction until it reaches the target simulated fire front $c_{fr} = 0.5$. The location at which this observed marker crosses $c_{fr} = 0.5$ is its equivalent simulated location (x_i, y_i). This projection requires two elements: (1) the choice of the projection direction, and (2) a convergence criterion to ensure that the target $c_{fr} = 0.5$ is reached.

G.2.2 Selected schemes for projection

Two different projection schemes could be considered: (1) a *progressive* scheme that updates the projection direction at each step of the projection and thereby, accounts for the variations of the observed fire front topology over the fire duration (see Fig. G.5); and (2) a *direct* scheme that projects the observed markers along a constant direction corresponding to the normal direction to the observed fire front $\mathbf{y}_t^o$ (see Fig. G.6). These two algorithms differ in their choice of the projection direction. A brief technical description follows.

▷ **Progressive projection scheme.** The markers of the observed fire front $\mathbf{y}_t^o$ are projected step-by-step towards the simulated fire front $c_{fr} = 0.5$. For each observed marker, each step can be divided into three stages:

(i) the calculation of the normal direction $\mathbf{n}_p^o$ at the location of the observed marker: the projection direction is updated at each step as the local normal direction $\mathbf{n}_p^o$ to the segment crossing the adjacent projected markers;

(ii) the translation of the marker along the projection direction $\mathbf{n}_p^o$ by a user-defined spatial shift Δr (see Fig. G.5);

(iii) the interpolation of the value of the simulated progress variable c at the location of the translated observed marker to check if the simulated iso-contour $c_{fr} = 0.5$ is found.

This iterative scheme ends when the reconstructed trajectory of the observed markers crosses the isoline $c_{fr} = 0.5$ (i.e., when c changes from 0.5^- to 0.5^+, or vice versa). Note that this type of projection guarantees the equivalence between (x_i^o, y_i^o) and its image (x_i, y_i) for each observed marker $i = 1, \cdots, N_{fr}^o$, since it reconstructs the time-history of the location of the marker i and indirectly, accounts for the local environmental conditions during the fire event. However, the definition of the successive normal directions $\mathbf{n}_p^o$ along the projection trajectory is a difficult task, especially for a small amount of observed markers (i.e., for a coarse resolution of the observations).

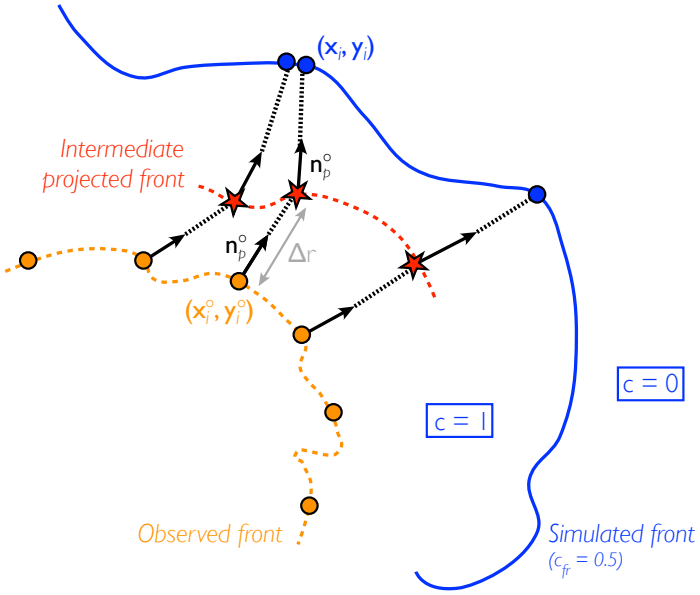


Figure G.5: *Schematic of the progressive projection scheme. Illustration of the projection of the observed marker i located at $(x_i^{\circ}, y_i^{\circ})$ onto the iso-contour $c_{fr} = 0.5$ to determine the location of its equivalent simulated marker (x_i, y_i) through the update of the local normal direction $\mathbf{n}_p^{\circ}$.*

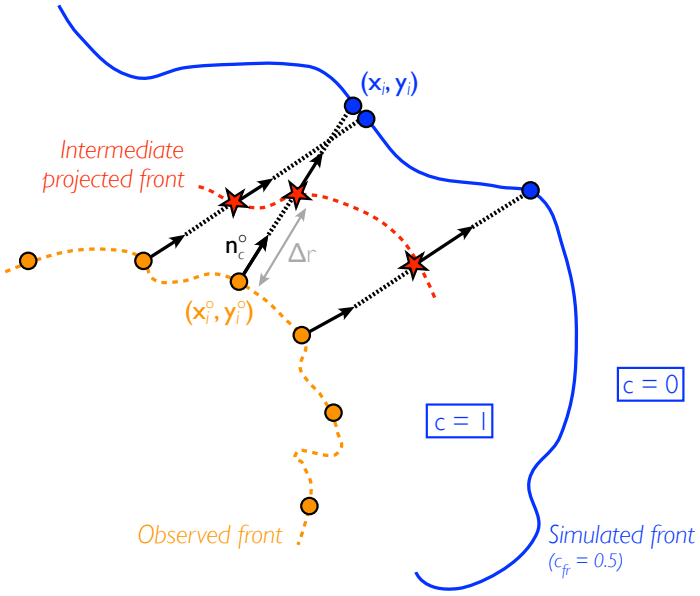


Figure G.6: *Schematic of the direct projection scheme. Illustration of the projection of the observed marker i located at $(x_i^{\circ}, y_i^{\circ})$ onto the iso-contour $c_{fr} = 0.5$ to determine the location of its equivalent simulated marker (x_i, y_i) through the translation along the constant normal direction $\mathbf{n}_c^{\circ}$.*

▷ **Direct projection scheme.** The markers of the observed fire front $\mathbf{y}_t^{\mathrm{o}}$ are directly projected onto the simulated fire front $c_{fr} = 0.5$, following the normal direction to the observed fire front noted $\mathbf{n}_c^{\mathrm{o}}$. This implies that there is no update of the projection direction, while the observed markers are projected towards $c_{fr} = 0.5$. This projection direction is maintained constant and therefore does not depend on the time-history of the topology of the observed fire front. Thus, this projection scheme can be regarded as a translation of the observed markers along the constant direction $\mathbf{n}_c^{\mathrm{o}}$ as illustrated in Fig. G.6.

In summary, these projection schemes perform the projection of the observed markers onto the simulated iso-contour $c_{fr} = 0.5$. The simplifications underlying the direct projection scheme may introduce some errors in the evaluation of the distance between simulated and observed fire fronts, especially when they are far from each other. However, it is less computationally expensive than the progressive projection scheme (since it avoids the update of the projection direction at each step and for each observed marker). Besides, within the framework of data assimilation, simulated fire fronts are expected to remain nearby the observed fire front due to the frequently-renewed data assimilation.

These two projection schemes were compared for the estimation of the proportionality coefficient P in the Rothermel's ROS formulation ($n = 1$), using the extended Kalman filter without outer-loop presented in Section 5.6 (Chapter 6). Firefly simulates a fire spread during 800 s (with a constant time step $\Delta t = 0.5$ s), over a horizontal fuel layer of dimension 200 m $\times 200$ m (with a uniform grid cell size $\Delta x = \Delta y = 1$ m) and characterized by a random distribution $\delta_v(x, y)$. There is no external flow ($u_w = 0$). An anisotropic propagation is obtained, even though P is uniform and constant over the fire duration. The true trajectory of the fire front associated with $\mathbf{x}^{\mathrm{t}} = P^{\mathrm{t}} = 0.1$ s^{-1} is presented in Fig. G.7.

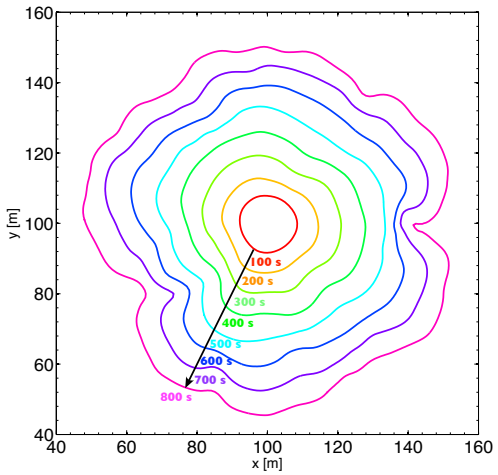


Figure G.7: *Time-evolving location of the true fire front, each* 100 *s from* $t = 100$ *s to* $t = 800$ *s, associated with the true control parameter* $P^{\mathrm{t}} = 0.1$ *s*$^{-1}$.

The analysis P^{a} is performed for different prior values P^{f} ranging from 0.02 to 0.18 m/s; the associated error STD $\sigma^{\mathrm{f}} = P^{\mathrm{f}} - P^{\mathrm{t}}$ varies therefore from -80 % to $+80$ % of the true value P^{t}. The observed fronts are discretized with 50 markers at 100 s intervals; observation errors are small. EKF-based results are presented in Table G.1 for a reaction-diffusion solver[1] and the current FIREFLY level-set solver.

Table G.1: *Comparison of EKF-based results for progressive and constant projection schemes as well as for the simple selection scheme. These results are obtained for a wildfire spread model based on a reaction-diffusion equation (rd) or the level-set propagating equation (ls). The control parameter is the proportionality coefficient P and its true value is $P^{\mathrm{t}} = 0.1$ s^{-1}.*

Prior P^{f}	Analysis P^{a}			
	Progressive (rd)	Direct (rd)	Direct (ls)	Selection (ls)
0.02 (-80 %)	0.1106	0.1166	0.0970	-
0.07 (-30 %)	0.0995	0.1001	0.1010	0.1000
0.13 ($+30$ %)	0.1000	0.0993	0.1002	0.0999
0.18 ($+80$ %)	0.1033	0.0999	0.1010	0.0999

Results show that the different projection schemes provide consistent analyses P^{a} that retrieve the true ROS value for a large range of perturbations. Thus, projection schemes appear as a promising approach for properly capturing the topology of the fire front along with the heterogeneities of wildfire spread and thereby, for applying data assimilation to wildfire spread. Still, the data assimilation results presented in this manuscript rely on the selection scheme (presented in Fig. G.4), since this scheme is computationally efficient and exhibits a sufficient accuracy for exploring which data assimilation algorithm is the most adequate for wildfire spread forecast. Besides, further investigations (out of the scope of this work) are required to extend the projection schemes and to optimize their algorithms for tracking more complex fire front topology.

[1] Preliminary developments of FIREFLY presented in previous works, Rochoux et al. (2010) and Delmotte et al. (2011), were based on a reaction-diffusion equation for tracking wildfire spread, instead of a level-set-based propagation equation.

G.3 Treatment of wind- and slope-induced wildfire spread

This section briefly explains how a complex terrain topography is accounted for in the FIREFLY wildfire spread simulator. A two-dimensional description of the time-evolving location of the fire front is maintained by projecting the three-dimensional propagation onto the horizontal plane (x, y). This formalism induces modifications in the Rothermel's ROS model, since the original formulation involving the wind and slope correction coefficients ϕ_w^* and ϕ_s^* (see Appendix A) describes the ROS Γ in the upslope direction only. More details on these modifications are provided in Emery et al. (2013), largely inspired from references due to Sharples (2008) and Lautenberger (2013).

G.3.1 Mathematical variables describing the terrain topography

A rectangular Cartesian coordinate system $\mathcal{R}(x_0, y_0, z_0)$ is introduced. The x_0-direction is the horizontal direction pointing towards the East and the y_0-direction towards the North; the z_0-direction is the vertical direction. This reference frame is illustrated in Fig. G.8. The downslope direction is described by the topographical aspect angle α_a, defined in a clockwise representation, where $0°$ indicates the North direction (i.e., the y_0-direction). The reference frame around the axis z_0 by the angle $(\alpha_a + 180°)$ defines the aspect frame noted $\mathcal{R}_a(x_a, y_a, z_0)$, where the x_a-direction indicates the normal to the slope direction in the horizontal plane and the y_a-direction indicates the uphill direction. The slope frame $\mathcal{R}_{sl}(x_a, y_{sl}, z_{sl})$ is defined as the rotation of the aspect frame $\mathcal{R}_a(x_a, y_a, z_0)$ around the axis x_a, with the slope angle α_{sl}; α_{sl} takes values between $0°$ (flat terrain) and $90°$ (vertical wall). In this slope frame, y_{sl} indicates the upslope direction, while z_{sl} indicates the normal direction to the slope plane. Thus, any terrain topography can be locally characterized by the pair of aspect and slope angles noted (α_a, α_{sl}).

G.3.2 Adaptation of the Rothermel's spread-rate model to complex terrain topography

A two-dimensional modification of the slope contribution to the ROS due to Rothermel was proposed by Lautenberger (2013) to account for wildfire spread in other directions than the uphill direction. The direction of fire spread is noted α_{fr} and the new slope correction coefficient in the modified Rothermel's formulation (in contrast to the original slope correction coefficient ϕ_{sl}^*) is noted ϕ_{sl}. This modification relies on the assumptions listed below.

- ▷ When the wildfire propagates in the upslope direction (i.e., $\alpha_{fr} = \alpha_a + 180°$), the slope contribution to the ROS is maximal and therefore, $\phi_{sl} = \phi_{sl}^*$.
- ▷ If this wildfire propagation occurs in the normal direction to upslope or downslope (i.e., $\alpha_{fr} = \alpha_a \pm 90°$), the slope does not contribute to the propagation,

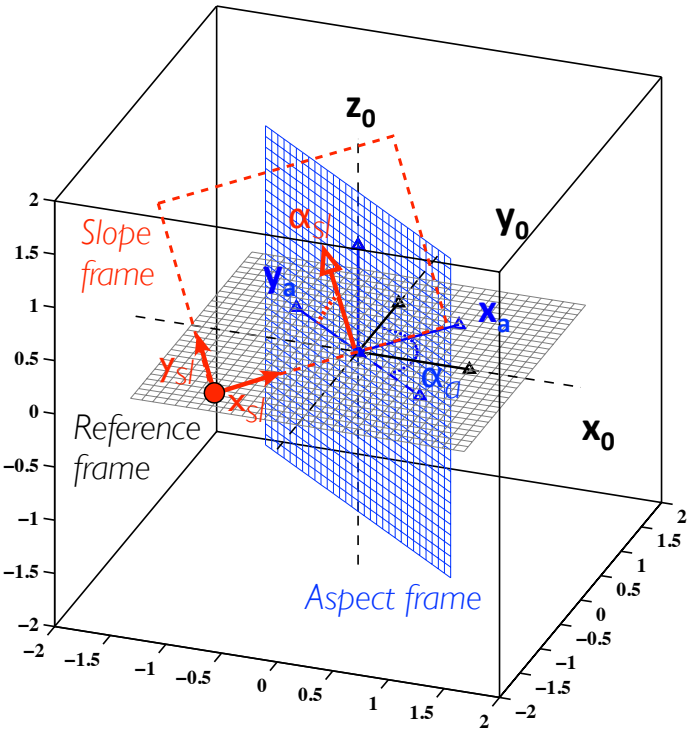


Figure G.8: *Three-dimensional reference frames to describe terrain topography.*

implying $\phi_{sl} = 0$.

- ▷ If the wildfire propagates in the downslope direction implying that the direction of the fire spread satisfies $\alpha_{fr} \in [\alpha_a - 90°; \alpha_a + 90°]$, an arbitrary treatment is required for the slope correction coefficient ϕ_{sl}. Since it is commonly assumed that a fire cannot spread at a lower ROS than the no-slope no-wind ROS Γ_0, the ROS Γ is forced to the value Γ_0 for a downslope configuration.

These assumptions for the no-wind ROS $\Gamma_{\mathcal{R}_{sl}}$ are reformulated as follows:

$$\Gamma_{\mathcal{R}_{sl}} = \Gamma_0 \max\left(1,\, 1 + \underbrace{\cos\left(\alpha_{fr} - \alpha_a - \pi\right) \phi_{sl}^*}_{\phi_{sl}}\right). \tag{G.15}$$

$\Gamma_{\mathcal{R}_{sl}}$ corresponds here to the evaluation of the ROS in the slope frame and this value must be projected onto the horizontal reference frame $\mathcal{R}(x_0, y_0, z_0)$ to obtain $\Gamma = \Gamma_{\mathcal{R}}$, the main physical quantity in FIREFLY. This projection is explained in detail in the following.

Figure G.9 shows the variations of the slope-aided ROS Γ evaluated using Eq. (G.15), with respect to the slope angle α_{sl} for different aspect angles α_a varying between $0°$ (dark-blue-plain line) and $90°$ (brown-plain line). For $\alpha_a = 0°$, the fire spreads in the upslope direction with a ROS reaching up to 1 m/s for a slope angle α_{sl}

above $65°$; the effect of the slope is considerable since the ROS Γ can be multiplied by a factor up to 25 compared to the no-slope no-wind ROS $\Gamma_0 = 0.048$ m/s. To the contrary, for $\alpha_a = 90°$, the fire propagates in the transverse direction to the slope, implying that the slope does not modify the ROS and $\Gamma = \Gamma_0 = 0.048$ m/s.

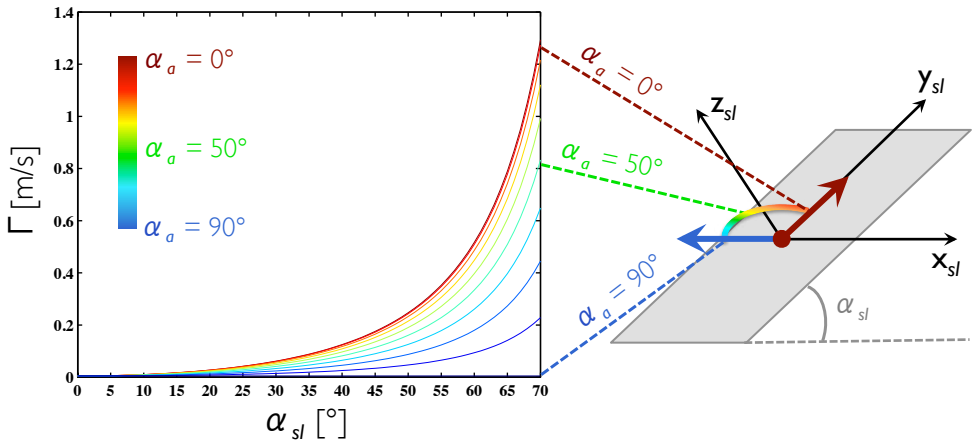


Figure G.9: *Slope-aided ROS Γ with respect to the slope angle α_{sl} for different values of the aspect angle α_a (represented by the colormap) for a plane configuration and $\Gamma_0 = 0.048$ m/s. Credit: Emery et al. (2013).*

G.3.3 Combined wind and slope effects

In addition to the slope correction coefficient ϕ_{sl}, Lautenberger (2013) proposed a new formulation of the wind correction coefficient noted ϕ_w in the modified Rothermel's formulation (in constrast to the original wind correction coefficient ϕ_w^*). The wind angle is noted α_w in the reference frame $\mathcal{R}(x_0, y_0, z_0)$ and is defined in the horizontal plane, starting from the North direction and turning clockwise around the axis z_0. Attention must be paid here since the wind angle α_w represents the direction from which the wind blows, implying that the wind velocity vector is provided in the direction $(\alpha_w + \pi)$ in this new formalism. Figure G.10 provides a representation of the topographical aspect angle α_a and wind angle α_w in the horizontal plane to clarify these notations. Besides, Fig. G.11 represents the wind blowing direction in both the horizontal plane and projected onto the slope frame. The wind velocity vector is assumed to be provided in the slope frame in Firefly, since available meteorological data account for the effects of the terrain topography onto the surface wind conditions.

The Rothermel's formulation of the wind- and slope-aided ROS $\Gamma_{\mathcal{R}_{sl}}$ becomes:

$$\Gamma_{\mathcal{R}_{sl}} = \Gamma_0 \max\left(1,\, 1 + \underbrace{\cos(\alpha_{fr} - \alpha_w - \pi)\phi_w^*}_{\phi_w} + \underbrace{\cos(\alpha_{fr} - \alpha_a - \pi)\phi_s^*}_{\phi_{sl}}\right). \quad \text{(G.16)}$$

As stated previously, $\Gamma_{\mathcal{R}_{sl}}$ is the ROS defined with respect to the environmental

conditions in the slope frame $\mathcal{R}_{sl}(x_a, y_{sl}, z_{sl})$. This ROS must be now projected onto the two-dimensional reference frame to be combined with the level-set solver in FIREFLY. This is the purpose of the next section.

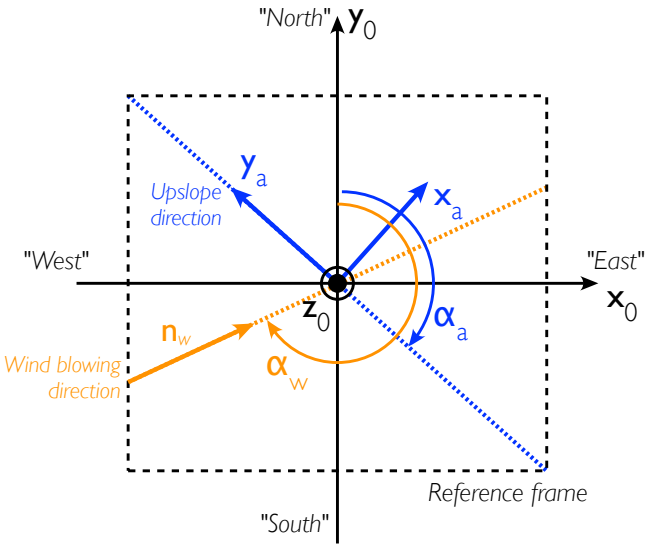


Figure G.10: *Representation of the topographical aspect angle α_a and the wind angle α_w in the horizontal reference frame $\mathcal{R}(x_0, y_0, z_0)$.*

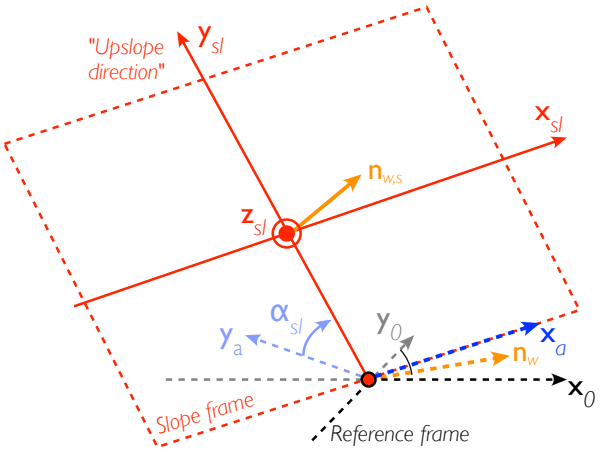


Figure G.11: *Representation of the slope angle α_{sl} and the wind blowing direction, in both horizontal reference frame $\mathcal{R}(x_0, y_0, z_0)$ (noted $\mathbf{n}_w$) and projected onto the slope frame $\mathcal{R}_{sl}(x_a, y_{sl}, z_{sl})$ (noted $\mathbf{n}_{w,s}$).*

G.3.4 Projection onto the horizontal plane

The wind correction coefficient ϕ_w in Eq. (G.16) differs from the one implemented in FIREFLY as described in Chapter 6. The wind correction coefficient ϕ_w^* is calculated

with two different approaches: the wind velocity vector is projected onto the normal direction to the front $\mathbf{n}_{fr}$ in Chapter 6 and the resulting wind magnitude u_w (i.e., the input of the Rothermel's ROS) is used to evaluate ϕ_w^*. This approach is referred to as the FIREFLY *classical approach*. On the other hand, the input parameter u_w of the Rothermel's model is taken directly as the wind velocity vector to calculate ϕ_w^* and thereby, the modified coefficient ϕ_w in Eq. (G.16) in Lautenberger (2013). This approach is referred to as the *Lautenberger approach*. Figure G.12 compares these two approaches in terms of simulated fire fronts for a wind blowing from the South-West direction at 0.75 m/s. These results show that these approaches have a significant impact on the topology of the front at the head of the fire. However, further investigations (out of the scope of this work) are required to assess which formulation of the wind correction coefficient ϕ_w^* is the most consistent with the dynamics of wildfires. The results presented in this work are obtained using the classical FIREFLY approach, unless mentioned otherwise.

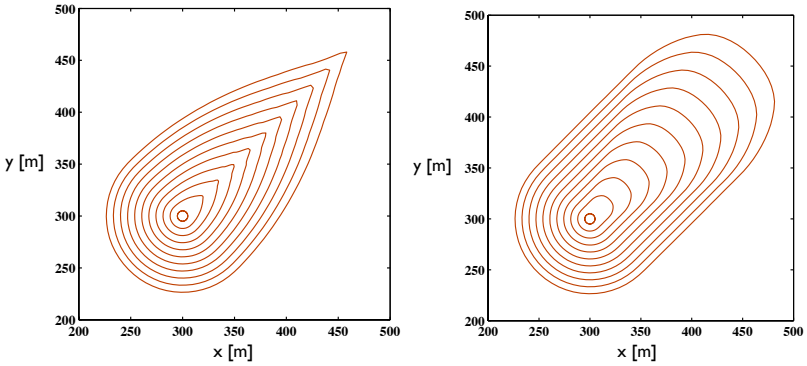


Figure G.12: *Time-evolving location of the simulated fire fronts each* 100 *s on a flat terrain with a South-West blowing wind* ($\alpha_w = 225°$). *Left:* FIREFLY *classical approach. Right: Lautenberger approach.*

The ROS in Eq. (G.16) is defined in the slope frame $\mathcal{R}_{sl}$. Representing the simulated fire fronts in a three-dimensional viewpoint would imply adding the z-coordinate in the description of the fire fronts and thereby, increase the size of the state and control vectors within the data assimilation framework. This would translate into an additional computational cost (due partly to the increased size of the error covariance matrices), which is not affordable for real-time wildfire spread forecast. For the purpose of data assimilation, we therefore adopt a strategy consisting in projecting the ROS $\Gamma_{\mathcal{R}_{sl}}$ onto the reference frame $\mathcal{R}(x_0, y_0, z_0)$, in order to simulate the propagation of the fire fronts onto the horizontal plane. Thus, the definition of data assimilation variables remain limited to the two-dimensional x- and y-coordinates as explained in Chapter 6. Based on geometrical considerations (Emery et al., 2013), the projected ROS Γ reads:

$$\Gamma = \Gamma_{\mathcal{R}_{sl}} \left(\sqrt{1 + \tan^2(\alpha_{sl}) \cos^2(\alpha_a - \alpha_{fr})} \right)^{-1}, \tag{G.17}$$

where α_{fr} indicates the direction of wildfire spread, consistently represented by the normal direction to the fire front $\mathbf{n}_{fr}$.

↪ Altimetric data

The terrain topography is specified in FIREFLY as altimetric data $h(x,y)$, i.e., as the topographic elevation h at specified grid nodes (x_h, y_h) that are interpolated on the FIREFLY computational grid. This interpolation step is required since altimetric data are mostly provided with a much coarser resolution than the spatial resolution required by FIREFLY. Vico and Porporato (2009) proposed to reconstruct the slope aspect $\alpha_a(x,y)$ and angle $\alpha_s(x,y)$ from altimetric data $h(x,y)$ as follows:

$$\tan \alpha_{sl} = \sqrt{\left(\frac{\partial h(x,y)}{\partial x}\right)^2 + \left(\frac{\partial h(x,y)}{\partial y}\right)^2} \tag{G.18}$$

$$\begin{pmatrix} \sin \alpha_a \\ \cos \alpha_a \end{pmatrix} = -\frac{1}{\tan \alpha_{sl}} \begin{pmatrix} \dfrac{\partial h(x,y)}{\partial x} \\ \dfrac{\partial h(x,y)}{\partial y} \end{pmatrix}. \tag{G.19}$$

The calculation of the gradient of the topographic elevation $h(x,y)$ is performed through a classical centered finite difference scheme.

G.3.5 Validation of the modified Rothermel's rate of spread

A series of three test cases is presented to illustrate the effects of a non-uniform terrain topography on the ROS and on the shape of the fire front as simulated by FIREFLY using the Lautenberger approach. This terrain topography corresponds to (1) a slope plane, (2) a canyon, and (3) a complex non-uniform elevation $h(x,y)$ in a mountainous region.

↪ Fire propagation on a slope plane

The slope-plane fire spread is simulated within a uniform vegetal fuel layer characterized by a fuel depth $\delta_v = 1$ m, a fuel moisture content $M_v = 15$ % and a fuel particle surface-area-to-volume ratio $\Sigma_v = 11500$ m^{-1}. The terrain is a simple slope plane, tilted by $\alpha_{sl} = 15°$ with respect to the horizontal plane, whose aspect angle is $\alpha_a = 225°$ and whose dimensions are 600 m $\times 600$ m (with a mesh stepsize $\Delta x = \Delta y = 1$ m). The initial condition is described by a circular front centered in ($x_0 = 300$ m, $y_0 = 300$ m) and of radius $r_0 = 5$ m; there is no external flow ($u_w = 0$). FIREFLY is integrated during 1000 s (with a time step $\Delta t = 0.5$ s). Results presented in Fig. G.13 show that the slope induces a constant upslope propagation, while the spread of the back of the fire remains very limited. The effective simulated ROS of the head of the fire is diagnosed a posteriori and is equal to

0.264 m/s, which is consistent with the theoretical value 0.261 m/s provided by Eq. (G.16)/(G.17). The no-slope ROS I_0 is equal to 0.068 m/s; the slope induces a propagation that is four times faster than in a no-slope configuration.

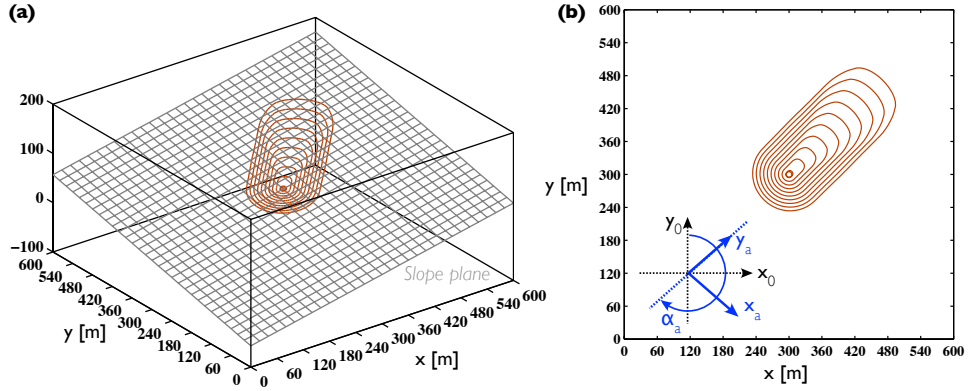


Figure G.13: *Time-evolving location of the simulated fire fronts each* 100 *s on a slope plane with the topographical aspect angle* $\alpha_a = 225°$. *(a) Three-dimensional representation. (b) Projected representation (onto the horizontal plane). Credit: Emery et al. (2013).*

↪ Fire propagation in a canyon

A canyon (illustrated in Fig. G.14) is the combination of two planes (P_1) and (P_2) of respective aspect angle $\alpha_{a,1}$ and $\alpha_{a,2}$ and of respective slope angle $\alpha_{sl,1}$ and $\alpha_{sl,2}$. The intersection between these two planes is a line of aspect angle α_c in the reference frame $\mathcal{R}(x_0, y_0, z_0)$, which can also be tilted with respect to the horizontal plane with a slope angle $\alpha_{sl,c}$. More details on the mathematical description of a canyon is provided in Emery et al. (2013). To guarantee a realistic description in Firefly, the user must specify the values of the angles α_c, $\alpha_{sl,c}$, $\alpha_{sl,1}$ and $\alpha_{sl,2}$ with the following constraints:

$$\begin{cases} \alpha_c \in]-90°;\ 90°[\,, \\ \alpha_{sl,c} \in [0°, \min(\alpha_{sl,1}, \alpha_{sl,2})]\,, \\ \alpha_{sl,1}, \alpha_{sl,2} \in [0°, 90°[\,. \end{cases} \tag{G.20}$$

These constraints determine a unique pair of aspect angles ($\alpha_{a,1}$,$\alpha_{a,2}$) for the two slope planes, which satisfies the following conditions:

$$\alpha_c - \pi \leq \alpha_{a,1} \leq \alpha_c, \quad \alpha_c \leq \alpha_{a,2} \leq \alpha_c + \pi. \tag{G.21}$$

The canyon configuration tends asymptotically to a plane configuration when the slope angles $\alpha_{sl,1}$ and $\alpha_{sl,2}$ converge towards $\alpha_{sl,c}$, meaning that the resulting

slope plane is defined by the slope angle $\alpha_{sl} = \alpha_{sl,c}$ and by the aspect angle $\alpha_a = \alpha_c \pm 180°$.

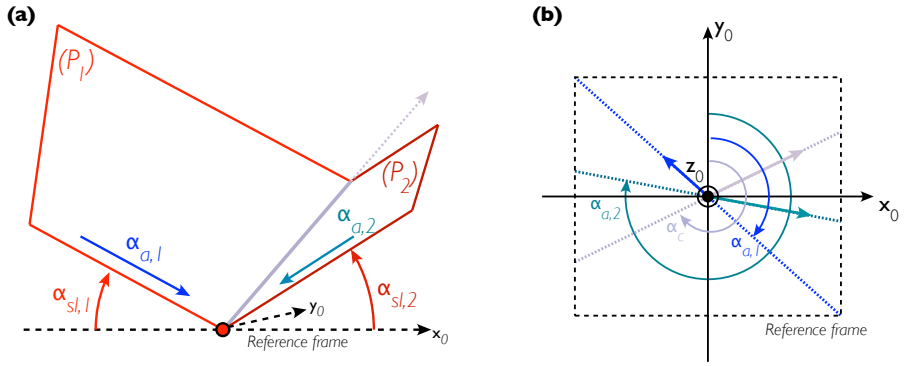


Figure G.14: *Definition of a canyon terrain. (a) Three-dimensional viewpoint. (b) Representation of the aspect angles in the horizontal reference frame* $\mathcal{R}(x_0, y_0, z_0)$.

The simulation of a fire spread in a canyon is performed with FIREFLY, considering the same uniform vegetation as in the slope plane case without external flow ($u_w = 0$). The initial circular front is centered at ($x_0 = 300$ m, $y_0 = 300$ m) and its radius is $r_0 = 5$ m. FIREFLY is integrated during 200 s (with a time step $\Delta t = 0.5$ s) and provides the propagation of the fire fronts in the canyon corresponding to $\alpha_c = 0°$, $\alpha_{sl,c} = 15°$, $\alpha_{sl,1} = \alpha_{sl,2} = 25°$ and $\alpha_{a,1} = -\alpha_{a,2} = -125.1°$. These fronts are represented every 20 s in Fig. G.15 and show that the highest values for the ROS are, consistently, in the directions of steepest ascent (i.e., upslope the planes (P_1) and (P_2)). No significant fire propagation is found in the downslope direction, i.e., at the rear of the fire.

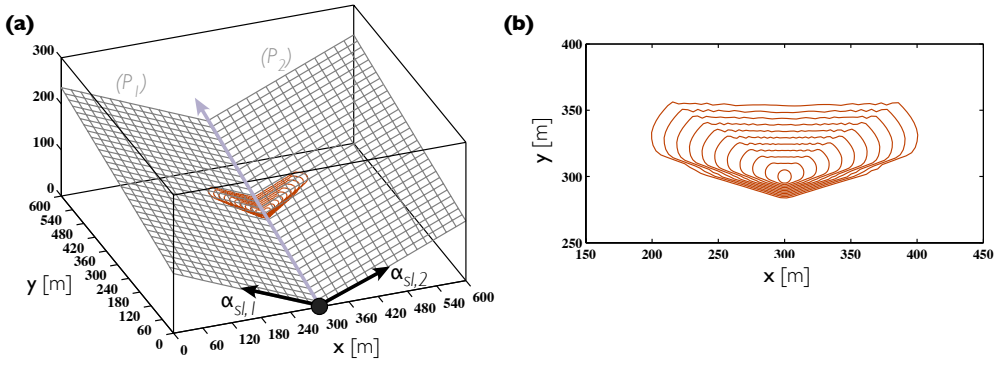


Figure G.15: *Time-evolving location of the simulated fire fronts each* 20 *s on a canyon topography. (a) Three-dimensional representation. (b) Projected representation (onto the horizontal plane). Credit: Emery et al. (2013)*

↪ Fire propagation on a complex terrain topography

A simulation of wildfire spread induced by a complex terrain topography illustrated in Fig. G.16 is performed for moderate wind conditions characterized by $u_w = 0.75$ m/s and $\alpha_w = 315°$. The vegetation is uniformly distributed over the 200 m $\times$ 200 m computational domain, in particular the vegetation layer depth is $\delta_v = 1$ m, the fuel moisture content is $M_v = 20$ % and the fuel particle surface-area-to-volume ratio is $\Sigma_v = 10000$ m^{-1}.

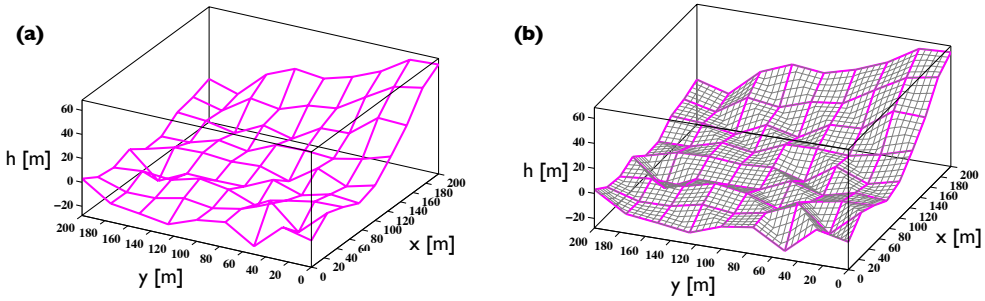


Figure G.16: *Altimetric data. (a) Reference data (database). (b) Interpolation of reference data $h(x, y)$ on* FIREFLY *mesh grid.*

The initial circular front is centered at ($x_0 = 100$ m, $y_0 = 100$ m) and its radius is $r_0 = 5$ m. FIREFLY is integrated during 1500 s, providing, every 75 s, the simulated fire front discretized with $N_{fr} = 2000$ markers. Figure G.17 illustrates the growth of the burnt area over time. Even though validating the physical consistency of the FIREFLY simulator for the treatment of wind and slope effects is difficult, this simulation corresponds to the main features of wildfire spread commonly used in the fire research field. The fastest ROS is reached upslope (the slope effect is high compared to the moderate wind effect, partly due to the high topographical elevation in one corner of the computational domain).

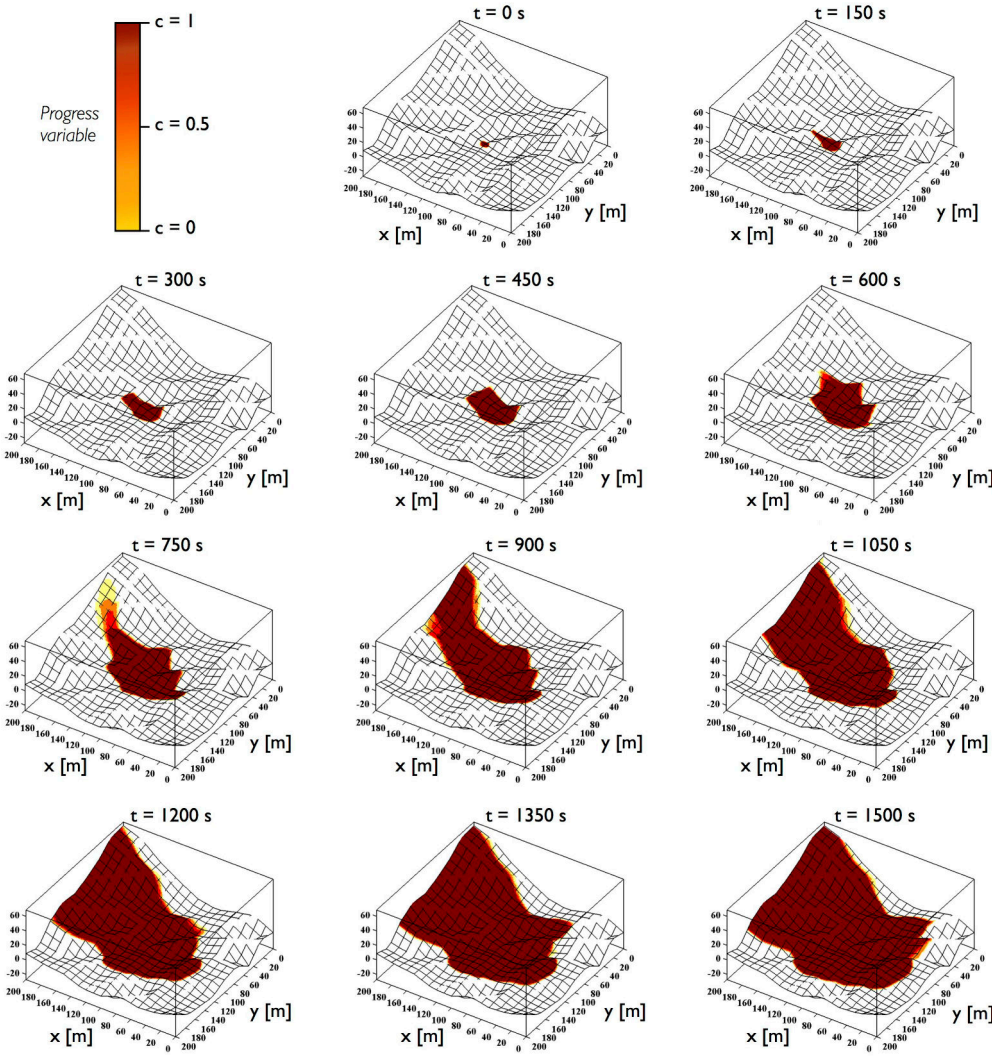


Figure G.17: *Growth of the burnt area from $t = 0$ s to $t = 1500$ s simulated with* FIREFLY *over a complex terrain topography; horizontal wind conditions with $u_w = 0.75$ m/s and $\alpha_w = 315°$. Credit: Emery et al. (2013).*

Appendix H

Application of particle filters to regional-scale wildfire spread

Particle filters, also known as sequential Monte Carlo filters, are considered as a more general solution to the Bayesian filtering problem than the Kalman filter and its extensions. From a theoretical viewpoint, it is not limited by linear and Gaussian error statistics assumptions as Kalman filtering and thereby, provides a complete description of the PDF through a point-mass representation (i.e., particles). In contrast, error statistics in Kalman filtering are represented through an error covariance model and are therefore limited to the second-order moments of the PDF.

The potential of particle filters for wildfire spread forecast was explored in this thesis within the framework of a collaboration between CERFACS (France), Mines Albi (France) and the Federal University of Rio de Janeiro (Brazil), in particular through a common research project with Wellington da Silva. This work led to the following publications da Silva et al. (2014, HTHP) and da Silva et al. (2013, IPDO).

In this appendix, the objective is to summarize the main ideas developed in this work on particle filters for wildfire spread and to provide the most recent study presented in da Silva et al. (2013, IPDO) at the 4th International Symposium on Inverse Problems, Design and Optimization (IPDO).

Summary

The purpose of this work is to show the capability of particle filters (see Section 5.7, Chapter 5) for improving wildfire spread simulation and forecast accuracy. This improvement is obtained through the sequential correction of input parameters in

the rate of fire spread (ROS) model as new observations become available. The performance of the Sequential Importance Resampling (SIR) and Auxiliary Sequential Importance Resampling (ASIR) algorithms is evaluated for a real-world case study corresponding to the controlled grassland fire experiment (see Section 6.3, Chapter 6).

da Silva et al. (2014, HTHP) proposed a two-parameter estimation, in which the estimation targets (the control vector) includes the fuel moisture content M_v and the fuel particle surface-area-to-volume ratio Σ_v with $\mathbf{x} = [M_v, \Sigma_v]$. This estimation problem therefore assumed that the main uncertainties are related to vegetation properties; the wind properties were assumed constant over the fire duration, and measurements were assimilated at 14 s time intervals (as for the comparison between the EnKF-based parameter and state estimation approaches presented in Section 7.3, Chapter 7). Both SIR and ASIR algorithms are found to sequentially track the propagation of the observed fire fronts. As the Kalman filter and its extensions (see Chapter 7), particle filters suffered in this study from the equifinality issue (also referred to as *sample variability*), since they did not manage to converge towards a unique solution for the set of control parameters (in the parameter space). Several sets of control parameters lead indeed to the same simulated fire front close to the observations (the error statistics in the observation space are converged). Still, it was shown that the ASIR algorithm retrieves more certain values of the parameters (with a narrower 99 %-confidence interval) than the SIR algorithm at a lower computational cost.

Since the sample variability may be reduced by including more sources of uncertainties in the Bayesian filtering procedure, da Silva et al. (2013, IPDO) added the wind magnitude u_w to the fuel moisture content M_v and the fuel surface-area-to-volume ratio Σ_v in the control vector for the same controlled grassland burning experiment (with assimilation at 28 s time intervals), i.e., $\mathbf{x} = [M_v, \Sigma_v, u_w]$. This study showed that controlling the wind magnitude allowed to track the observed fire front further in time. The estimation of u_w was validated against independent in-situ wind measurements and features a reduced scatter of the analysis estimates in comparison to the 2-parameter estimation study. Still, the estimation of M_v and Σ_v remained consistent with the results presented in da Silva et al. (2014, HTHP), showing that the analysis estimates provide physical values for the uncertain environmental conditions. Note that the polynomial chaos (PC) strategy presented in Section 7.2.3 (Chapter 7) is also considered as a promising approach to limit the computational cost of particle filters (based on Monte Carlo sampling as the ensemble Kalman filter).

4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)
Albi, France, June 26-28, 2013

APPLICATION OF PARTICLE FILTERS IN MOVING FRONTIER PROBLEMS

W. B. da Silva[a,d], M. C. Rochoux[b,c], H. R. B. Orlande[a], M. J. Colaço[a], O. Fudym[d], M. El Hafi[d], B. Cuenot[b], S. Ricci[b]

[a] *Department of Department of Mechanical Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, wellingtonuff@yahoo.com.br, helcio@mecanica.ufrj.br, colaco@ufrj.br;*

[b] *CERFACS; CNRS URA-1875, 42 Ave. G. Coriolis, F-31057 Toulouse cedex 01, France, benedicte.cuenot@cerfacs.fr, sophie.ricci@cerfacs.fr*;

[c] *Ecole Centrale Paris; CNRS UPR-288, Grande voie des vignes, F-92295 Châtenay-Malabry cedex, France, melanie.rochoux@ecp.fr*;

[d] *Université de Toulouse; Mines Albi; CNRS; Centre RAPSODEE, Albi, France, olivier.fudym@mines-albi.fr , elhafi.meh@gmail.com;*

Abstract

This paper demonstrates the capability of particle filters to combine measurements to model simulations along with their uncertainties, in order to formulate some feedback information on the uncertain model variables and thereby, improve the simulation forecast in moving frontier problems such as wildfire spread. Sequential Importance Resampling (SIR) and Auxiliary Sequential Importance Resampling (ASIR) filters were built on top of a level-set based front-tracking simulator in order to assimilate the time-evolving location of the fire front. This work focuses primarily on the uncertainty in the input parameters of the fire spread-rate model (characterizing the vegetation properties and the wind conditions) and considers those as the main source of errors in the simulated front positions, neglecting simplifications in the fire spread model structure. The good performance of the SIR and ASIR filters for the sequential estimation of the input model parameters is illustrated for a controlled grassland burning experiment; results indicate that the ASIR filter is able to track the observed fire fronts at the expense of a reasonable computational cost in comparison to the SIR filter.

Introduction

Because wildfire spread is a complex multi-physical multi-scale problem, our ability to predict the behavior of wildfires at large regional scales (i.e., at scales ranging from a few tens of meters up to several kilometers) remains limited [1]. The propagation speed of wildfires, also called the Rate Of Spread (ROS), is modeled in current wildfire spread simulators as a semi-empirical function of a reduced number of parameters that locally characterize the vegetation properties, the weather conditions and the terrain topography [2,3]. In such simulators, the wildfire spread is described as a front propagating towards the unburned vegetation (fuel) at the ROS that is relevant to the local conditions, using a standard level-set or Lagrangian front-tracking technique. The input model parameters are not easily measurable and are therefore embedded with significant levels of uncertainties. For the wildfire spread simulation to be predictive and compatible with operational applications, these uncertainties need to be quantified and reduced. For this purpose, an inverse modeling approach, based on particle filters for the solution of a state estimation problem, is proposed in this study.

State estimation problems consist in using the available measurements together with prior knowledge about the physical phenomena and the associated uncertainties, in order to sequentially produce more

accurate estimates of the dynamic variables of interest. Such problems can be solved using the Bayesian filtering approach to statistics [4-8].

The Kalman filter and its extensions [4,7] are widely used in geosciences in fields like hydrology or oil reservoir modeling. However, this filter is limited to linear models and Gaussian assumptions regarding the statistical description of errors; the quality of its feedback correction can be indeed significantly degraded in cases involving high non-linearities and non-Gaussian error statistics. Since particle filters do not make assumptions on the linearity of the model or the form of the statistical errors in their general formulation, they appear as a promising alternative in those cases [8,9]. The idea behind the Sequential Importance Sampling (SIS) technique was to describe the Probability Density Function (PDF) of the control variables as a set of random particles (prior); each particle was then associated with a weight that was calculated using the measurements along with their uncertainties; the values of the particles and their associated weights allowed a more accurate PDF (posterior) to be retrieved. To avoid the degeneracy problem (i.e., to avoid that only a few particles participate effectively in the filtering process), Gordon *et al.* [10] added a resampling approach into the SIS filter. Resampling can be either applied if the number of effective particles falls below a specified threshold number, or at every step in a technique known as the Sequential Importance Resampling (SIR) filter. A large number of recent studies have highlighted the performance of the SIR filter over a wide range of applications [11]. Despite these applications, the SIR filter remains computationally intensive, as a large number of particles are required to obtain a complete and accurate statistical description of the control variables. In order to overcome these difficulties, Pitt and Shephard [12] introduced the auxiliary particle filters, whose main idea is to improve the prior information by using an additional set of particles (called auxiliary particles), so as to reduce the computational cost without degrading the accuracy of the result. In this perspective, Silva *et al.* [13] applied the Auxiliary Sequential Importance Resampling (ASIR) filter to solve a non-linear solidification problem, where simulated temperature measurements were used to estimate a transient line heat sink as well as the solidification front. Colaço *et al.* [14] compared the performance of the SIR and ASIR filters in the estimation of the heat flux applied to a square cavity in a natural convection problem; this study showed excellent estimates for the time variation of the unknown quantity. Also, the sequential propagation of modeling errors was studied to improve the choice of the particles at the next observation time (i.e., at the next assimilation cycle), in particular in the case of combined parameter/state estimation [8,15].

The application of inverse methods in the context of fire modeling has been considered only recently [16-18]. Gu [17] applied the SIR algorithm to synthetic cases of wildfire spread, in order to estimate average wind magnitude or wind direction of a semi-empirical model in the fire area using ground-based temperature sensor data.

The objective of this paper is to address the challenges specific to the development of a robust inverse modeling approach for realistic wildfire spread. While the preliminary approach adopted in Rochoux *et al.* [18] did provide good results, since data assimilation is to be applied to more realistic cases (i.e., to large regional-scale fires that are strongly coupled to atmospheric dynamics), it should be able to deal with heterogeneous vegetation properties as well as non-constant wind direction and magnitude. To better take into account the underlying model non-linearities and thus to provide a more accurate posterior distribution of the control parameters, we propose here a particle filter strategy based on the assimilation of the time-evolving fire front locations and the front-tracking fire spread simulator FIREFLY as in Ref. [18].

1. Physical problem and mathematical formulation

The propagation of wildfires results from complex interactions between pyrolysis, combustion, heat transfer and flow dynamics, and atmospheric dynamics. These interactions occur over a wide range of scales: vegetation scales that characterize the biomass fuel; topographical scales that characterize the terrain and vegetation boundary layer; and meteorological micro-/meso-scales that characterize atmospheric conditions. As in current operational wildfire spread models [3], we adopt in this study a regional-scale perspective and simulate a wildfire as a thin flame zone (i.e., as a front) that self-propagates normal to itself towards unburned vegetation. In this representation, the main quantity of interest is the ROS, which is the local propagation speed of the front.

In this approach based on Rothermel's model [2], the ROS is formulated as an empirical function of a reduced number of parameters that locally characterize the vegetation (fuel) properties, the weather conditions and the terrain topography. The local ROS, denoted as Γ [m/s], can be written as [2]

$$\Gamma = \Gamma(x,y,t) = P\left(M_v, \Sigma_v, \mathbf{u}_w(x,y,t)\right)\delta_v(x,y) \tag{1}$$

where δ_v [m] is the fuel depth (e.g., the vegetation layer thickness) and P [1/s] is a function of the fuel moisture content M_v (mass of water divided by mass of dry fuel), the fuel particle surface-area-to-volume ratio Σ_v [1/m], and the wind velocity (at mid-flame height) $\mathbf{u}_w$ [m/s], among others. In this paper, M_v, Σ_v, and δ_v are treated as spatially-uniform parameters. Note that $\mathbf{u}_w$ corresponds to the wind velocity vector (defined by the wind velocity magnitude m_w and direction d_w) projected along the normal direction to the front $\mathbf{n} = \mathbf{n}(x,y,t)$, meaning that $\mathbf{u}_w$ is a time-varying two-dimensional field.

In the FIREFLY simulation capability, the propagation of the fire front at the ROS given by Eq. (1) is simulated using a standard level-set front-tracking technique [18]. A progress variable noted c is introduced as a flame marker, so that: $c = 0$ in the unburned vegetation, $c = 1$ in the burnt vegetation; and the flame front is identified by the two-dimensional isocontour $c = 0.5$, as shown in Fig. 1.

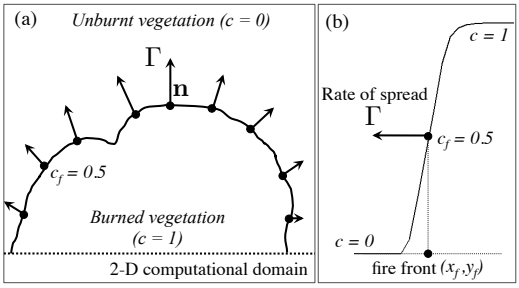


Figure 1: Schematic of the fire propagation model: (a) 2-D surface fire spread at the ROS Γ along the normal direction $\mathbf{n}$ to the front (b) Profile of the progress variable c throughout $c_f = 0.5$.

The flame front locations are reconstructed using the two following steps:

i) The spatio-temporal evolution of the progress variable $c = c(x,y,t)$ is calculated as a solution of the following propagation equation using the ROS model (1):

$$\frac{\partial c}{\partial t} = \Gamma \, | \nabla c |, \tag{2}$$

with Γ the ROS (m/s) along the normal direction **n** to the isocontours of the progress variable *c*. Equation (2) is solved using a second-order Runge-Kutta scheme for time-integration and a second-order total variation diminishing scheme with a Superbee slope limiter [18] for spatial discretization.

ii) The instantaneous position of the fire front (x_f, y_f) is extracted using a simple isocontour algorithm, verifying $c(x_f, y_f, t) = c_f$ with $c_f = 0.5$.

2. State estimation problem

State estimation problems, also designated as nonstationary inverse problems [5], are of great interest in innumerable practical applications. In such problems the available measured data is used together with prior knowledge about the physical phenomena and the measuring devices, in order to sequentially produce estimates of the desired dynamic variables. This is accomplished in such a manner that the error is minimized statistically.

Consider a model for the evolution of the state variables **x** in the form:

$$\mathbf{x}_k = \mathbf{f}_k\left(\mathbf{x}_{k-1}, \mathbf{v}_k\right) \tag{3}$$

where **f** is, in the general case, a non-linear function of **x** and of the state noise or uncertainty vector given by $\boldsymbol{v}_k \in \mathbf{R}^{\mathbf{n}}$. The vector $\mathbf{x}_k \in \mathbf{R}^{\mathbf{n}}$ is called the state vector and contains the *n* variables to be dynamically estimated. This vector advances in time in accordance with the *state evolution model* (3). The subscript $k = 1, 2, 3, \ldots$, denotes a time instant t_k.

The observation model describes the dependence between the state variables **x** to be estimated and the measurements $\boldsymbol{z}^{\boldsymbol{obs}}$ through the general, possibly non-linear, function $\boldsymbol{h}$. This can be represented by:

$$\mathbf{z}_k^{obs} = \mathbf{h}_k\left(\mathbf{x}_k, \mathbf{n}_k\right) \tag{4}$$

where $\boldsymbol{z}_k^{obs} \in \mathbf{R}^{\mathbf{nz}}$ are available at times t_k. Equation (4) is referred to as the *observation/measurement model*. The vector $\boldsymbol{n}_k \in \mathbf{R}^{\mathbf{nz}}$ represents the measurement noise or uncertainty. The evolution and observation models given by Eqs. (3) and (4) are based on the following assumptions [5,8,9]:

(a) The sequence $\mathbf{x}_k$ for $k = 1, 2, 3, \ldots$, is a Markovian process, that is,

$$\pi\left(\mathbf{x}_k \middle| \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{k-1}\right) = \pi\left(\mathbf{x}_k \middle| \mathbf{x}_{k-1}\right) \tag{5.a}$$

(b) The sequence $\boldsymbol{z}_k^{obs}$ for $k=1, 2, 3, \ldots$, is a Markovian process with respect to the history of $\mathbf{x}_k$,

$$\pi\left(\mathbf{z}_k^{obs} \middle| \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k\right) = \pi\left(\mathbf{z}_k^{obs} \middle| \mathbf{x}_k\right) \tag{5.b}$$

(c) The sequence $\mathbf{x}_k$ depends on the past observations only through its own history, that is,

$$\pi\left(\mathbf{x}_k \middle| \mathbf{x}_{k-1}, \mathbf{z}_1^{obs}, \mathbf{z}_2^{obs}, \ldots, \mathbf{z}_{k-1}^{obs}\right) = \pi\left(\mathbf{x}_k \middle| \mathbf{x}_{k-1}\right) \tag{5.c}$$

where $\pi(\mathbf{a}|\mathbf{b})$ denotes the conditional probability of **a** when **b** is given. For the state and observation noises, the following assumptions are made [5,8,9]:

(a) For $i \neq j$, the noise vectors $\mathbf{v}_i$ and $\mathbf{v}_j$, as well as $\mathbf{n}_i$ and $\mathbf{n}_i$, are mutually independent and also mutually independent of the initial state $\mathbf{x}_0$.

(b) The noise vectors $\mathbf{v}_i$ and $\mathbf{n}_j$ are mutually independent for all *i* and *j*.

Different problems can be considered for the evolution-observation models described above, such as [5,8,9]:

(i) The *prediction problem*, in which the objective is to obtain $\pi(\mathbf{x}_k|\, \mathbf{z}_{1:k-1}^{obs})$;

(ii) The *filtering problem*, in which the objective is to obtain $\pi(\mathbf{x}_k|\, \mathbf{z}_{1:k}^{obs})$;

(iii) The *fixed-lag smoothing problem*, in which the objective is to obtain $\pi(\mathbf{x}_k|\, \mathbf{z}_{1:k+p}^{obs})$, where $p \geq 1$ is the fixed lag.

(iv) The *whole-domain smoothing problem*, in which the objective is to obtain $\pi(\mathbf{x}_k|\, \mathbf{z}_{1:K}^{obs})$, where $\mathbf{z}_{1:K}^{obs} = \{\mathbf{z}_i^{obs}, i=1,...,K\}$ is the complete set of measurements.

We consider here the filtering problem. By assuming that $\pi(\mathbf{x}_0|\, \mathbf{z}_0^{obs}) = \pi(\mathbf{x}_0)$ is available, the posterior probability density $\pi(\mathbf{x}_k|\, \mathbf{z}_{1:K}^{obs})$ is then obtained with Bayesian filters in two steps [5,8,9]: *prediction and update*. The most widely known Bayesian filter method is the Kalman filter [5,6,8,9]. However, the application of the Kalman filter is limited to linear models with additive Gaussian noises. Extensions of the Kalman filter were developed in the past for less restrictive cases by using linearization techniques. Similarly, Monte Carlo methods have been developed in order to represent the posterior density in terms of random samples and associated weights. Such Monte Carlo methods, usually referred to as particle filters among other designations found in the literature, do not require the restrictive assumptions of the Kalman filter. Hence, particle filters can be applied to non-linear models with non-Gaussian errors [5,6,8,9].

The idea in particle filters is to represent the required posterior density function by a set of random samples with associated weights and to compute the estimates based on these samples and weights [5,8,9]. Let $\{\mathbf{x}^i_k, i=0,...,N\}$ be the particles with associated weights $\{\mathbf{w}^i_k, i=0,...,N\}$ and $\mathbf{x}_{0:k}=\{\mathbf{x}_j, j=0,...,k\}$ be the set of all states up to t_k, where N is the number of particles. The weights are normalized, so that $\Sigma_i\, w^i_k=1$. Then, the posterior density at t_k can be discretely approximated by:

$$\pi\left(x_{0k}\left|z_{1:k-1}\right.\right) \approx \sum_{i=1}^{I} w_k^i \delta\left(x_{0k} - x_{sk}^i\right), \tag{6}$$

where $\delta(.)$ is the Dirac delta function. Using assumptions (5.a-c), the posterior density in Eq. (6) can be written as $\pi(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \approx \Sigma_i\, w^i_k\, \delta(\mathbf{x}_k - \mathbf{x}^i_k)$.

A common problem with the particle filter method is the degeneracy phenomenon; after a few estimations all but one particle may have negligible weight. The degeneracy implies that a large computational effort is devoted to update particles, whose contribution to the posterior density function is almost zero. This problem can be overcome by increasing the number of particles, or more efficiently by appropriately selecting the importance density as the prior density $\pi(\mathbf{x}_k|\mathbf{x}^i_{k-1})$. In addition, the use of the resampling technique is recommended to avoid the degeneracy of the particles [5,8,9].

Resampling generally involves a mapping of the random measure $\{\mathbf{x}^i_k, w^i_k\}$ into a random measure $\{\mathbf{x}^{i^*}_k, 1/N\}$ with uniform weights. It can be performed if the number of effective particles with large weights falls below a certain threshold number. Alternatively, resampling can also be applied indistinctively at every instant t_k, as in the *Sequential Importance Resampling* (SIR) algorithm [8,9].

Although the resampling step reduces the effects of the degeneracy problem, it may lead to a loss of diversity and the resulting sample can contain many repeated particles. This problem, known as sample impoverishment, can be severe in the case of small evolution model noise. In this case, all particles collapse to a single particle within a few instants t_k. Another drawback of the particle filters is related to

the large computational cost due to the Monte Carlo sampling, which may limit its application only to fast computing problems.

Different algorithms for the implementation of the particle filters can be found in [9], including those that allow for the simultaneous estimation of constant parameters appearing in the model and the transient states. One of such algorithms is the *Auxiliary Sequential Importance Resampling* (ASIR) method.

In the ASIR filter algorithm, the indexes $i^{(j)}$ (j = 1,…,N) are obtained by resampling (i.e., by resampling the particles with higher weights). According to reference [9], the advantage of the ASIR algorithm over the SIR algorithm is that it naturally generates points from the sample at (k-1), which, conditioned on the current measurement, are most likely to be close to the true state. Still, as described in [9], the ASIR filter can be viewed as a resampling at the previous time step, based on some point estimates μ^i_k that characterize $\pi(\mathbf{x}_k|\mathbf{x}^i_{k-1})$. Since a single point μ^i_k is not able to accurately characterize $\pi(\mathbf{x}_k|\mathbf{x}^i_{k-1})$ for a large process noise, the use of the ASIR filter is limited to small process noises.

3. Results and discussions

In this paper, the SIR and ASIR particle filter algorithms are applied to a natural fire propagation in order to calibrate several physical parameters involved in the formulation of the Rothermel-based ROS within the FIREFLY simulator. Data were taken from an experimental database corresponding to a small-scale (4 m x 4 m) open-field grassland fire occurring under moderate fluctuating wind conditions [18]. The time-varying wind magnitude m_w was measured during the controlled burning experiment. The fire spread was recorded during 350 s using a thermal-infrared camera; the resulting observations are the time-evolving positions of the fire front (see Fig. 2) identified as the zones where the temperature reaches the value 600 K (generally considered as the temperature of vegetation ignition).

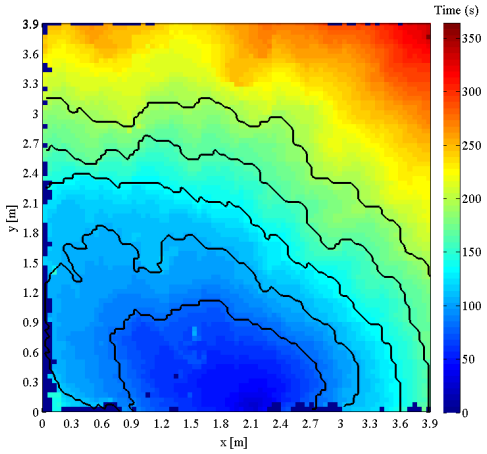


Figure 2: Arrival times of the fire front (colormap), and observed fire fronts (black solid lines) separated by 28 s (at t = 78 s, 106 s, 134 s, 162 s and 190 s).

Details of the measurement technique to retrieve the temperature field from thermal imaging are given in Wooster *et al.* [20]. In the following state estimation process, we assimilate measurements of fire front locations every 28 s from t = 64 s to t = 190 s (the associated fronts are represented in black solid lines in Fig. 2). This means that, in the particle filters, the update step is performed successively at

t = 78 s, 106 s, 134 s, 162 s and 190 s, and also that the prediction step allows the PDF of the control parameters to be integrated during 28 s between two consecutive observation times. Each observed front is represented with 200 markers of coordinates (x_f, y_f), whose error standard deviation is estimated to 0.047 m (based on the spatial resolution of the thermal-infrared camera). This error standard deviation is used to describe the measurement uncertainty vector $\boldsymbol{\varepsilon}$ and the observation error covariance matrix $\mathbf{W}$.

The fire spread simulator assumes uniform properties of the (fuel) short grass with a fuel layer thickness equal to δ_v = 8 cm (field measure), a moisture content equal to M_v = 22 % (field measure), a particle fuel surface-area-to-volume ratio Σ_v = 11480 1/m (values taken from Rothermel's database [2]) and the wind velocity magnitude and direction are initialized to the mean wind conditions recorded before ignition, m_w = 1 m/s and d_w = 307° (in a clockwise representation, where 0° indicates the North direction), respectively. The objective of the SIR and ASIR particle filters is to search for the effective posterior PDF of the fuel moisture content M_v, the fuel particle surface-area-to-volume ratio Σ_v and the wind velocity magnitude m_w, which are subject to significant uncertainties. The error standard deviations associated with the vegetal fuel parameters are taken to be 30 % of their initial mean values, that is σ_M = 6.6 % for M_v and σ_Σ = 3444 1/m for Σ_v. As to the wind velocity magnitude, its error standard deviation is set to σ_w = 0.1 m/s for m_w.

The three control parameters are assumed to be spatially-uniform, meaning that particle filters aim at retrieving the time-profile of these parameters as the fire front propagates through the 4 m x 4 m domain. It was found in Rochoux *et al.* [18] that these prior values of the control parameters significantly underestimate the position of the fire fronts (the associated simulation is called the free run) and that a state estimation procedure is required to produce fire spread simulations that are more consistent with observations. The 4 m x 4 m domain is discretized with a regular mesh ($\Delta x = \Delta y$ = 0.047 m), and the time step for integration of the progress variable equation is fixed to Δt = 0.02 s. For each triplet of control parameters taken in the associated Gaussian PDF, the fire spread simulation is initialized using the observed fire position at time t = 50 s, and is then integrated by time period of 28 s to update the posterior PDF of M_v, Σ_v and m_w at the five different observation times (t = 78 s, 106 s, 134 s, 162 s, 190s). As there is no explicit formulation of the control vector evolution between two successive observation times, a random walk model is applied so that the error standard deviation introduced in the parameters from time t_{k-1} to t_k is equal to σ_M for M_v, σ_Σ for Σ_v and σ_w for m_w, respectively. It reads

$$M_v(t_k) = M_v(t_{k-1}) + \sigma_M R_M \tag{7}$$

$$\Sigma_v(t_k) = \Sigma_v(t_{k-1}) + \sigma_\Sigma R_\Sigma \tag{8}$$

$$m_w(t_k) = m_w(t_{k-1}) + \sigma_w R_w \tag{9}$$

with R_M, R_Σ and R_w random numbers following a Normal distribution, with zero mean and unitary standard deviation. Since M_v , Σ_v and m_w directly influence the progress variable c (see Eq. 2), such quantity is also included in the inverse problem formulation as a state variable. The state evolution model for the vector containing the values of the progress variable at each of the grid points, $\mathbf{c}(t_k)$, is obtained from the discrete integration of Eq. (2). Uncertainties for $\mathbf{c}(t_k)$ are assumed to be additive, Gaussian, with zero mean and a constant standard deviation of 0.01.

The performance of the particle filters is analyzed in the observation space, in terms of the Root Mean Square (RMS) error between the simulated and observed fire front positions, at each observation time. At time t_k, the RMS is calculated as follows:

$$RMS_k = \sqrt{\frac{1}{p}\sum_{j=1}^{p}\left(\mathbf{z}_{k,j}^{obs} - \mathbf{z}_{k,j}\right)^2}, \tag{10}$$

where $\mathbf{z}_k$ contains the p simulated fire front positions given by Eq. (4), and $\mathbf{z}_k^{obs}$ represents the corresponding observations. The 99%-credible interval, denoted $I_{99\%}$ and defined in the parameter space, is used as an additional diagnostic of the performance of particle filters. It reads:

$$I_{99\%} = \hat{\mathbf{x}}_k + 2.576\sigma_x, \tag{11}$$

where $\hat{\mathbf{x}}_k$ represents the estimated mean value of the control parameter and σ_x represents its associated error standard deviation. The performance of both SIR and ASIR particle filters is presented in Table 1 in terms of RMS error at each observation/assimilation time and of the required computational time for the whole sequential Bayesian process for different numbers of particles. The different solutions of the particle filters are also compared to the free run configuration (using standard Rothermel's database).

Figure 3 represents, along with the observations, the simulated time-evolving fire fronts (from t = 78 s to 190 s) using the mean of the posterior PDF of the control parameters obtained through the SIR and ASIR filters, respectively. These results show that both the SIR and ASIR filters are able to significantly reduce the distance between predicted and observed fire fronts and thereby, to closely track the observed fire fronts along time. The free run presents indeed the highest RMS errors for all observation times; the RMS errors for the SIR and ASIR filters are reduced by a factor of at least 2 for all observation times. Furthermore, these results indicate that the distance to the observations remains significant at t = 190 s due to the particular shape of the front as shown in Fig. 3. Note that there is no spatial correction of the fire front position per observation time since we assumed that the control parameters were spatially uniform. Tracking all the variations of the fire front topology at a given time was out of the scope of this study. Still, this representation is able to efficiently describe the propagation of the front in the wind direction and to accurately track the head of the fire, which is the main quantity of interest within an operational fire spread framework.

Table 1 shows for the SIR algorithm with 200 particles is similar to the ASIR algorithm with 50 particles, while the solution provided by the ASIR algorithm is fairly more accurate. The ASIR filter presents indeed the smallest RMS total error, 8.831 m, of all test cases with only 50 particles (this total error is obtained by summing the RMS errors for all observation times). Table 1 also shows that even though the number of particles is increased to 200, the SIR algorithm does not succeed in converging towards a solution closer to the observations than for 50 particles, whereas the computational cost is multiplied by 25. This might be due to the existence of multiple solutions to the problem, meaning that several triplets of control parameters can result in a similar simulated front close to the observation.

Table 1: RMS errors and computation time for SIR and ASIR particle filters.

Filter	Particle nb. (N)	RMS error (78 s)	RMS error (106 s)	RMS error (134 s)	RMS error (162 s)	RMS error (190 s)
SIR	25	0.1037 m	0.2976 m	2.5683 m	2.8106 m	3.2251 m
SIR	50	0.1002 m	0.2680 m	2.5669 m	2.8011 m	3.0948 m
SIR	100	0.1067 m	0.2980 m	2.5664 m	2.8055 m	3.0966 m
SIR	200	0.0956 m	0.2642 m	2.5710 m	2.8089 m	3.0989 m
ASIR	25	0.1675 m	0.2842 m	2.5778 m	2.8121 m	3.0995 m
ASIR	50	0.1033 m	0.2647 m	2.5681 m	2.8078 m	3.0969 m
ASIR	100	0.1718 m	0.2883 m	2.5715 m	2.8090 m	3.0988 m
ASIR	200	0.1139 m	0.2640 m	2.5650 m	2.8036 m	3.0965 m

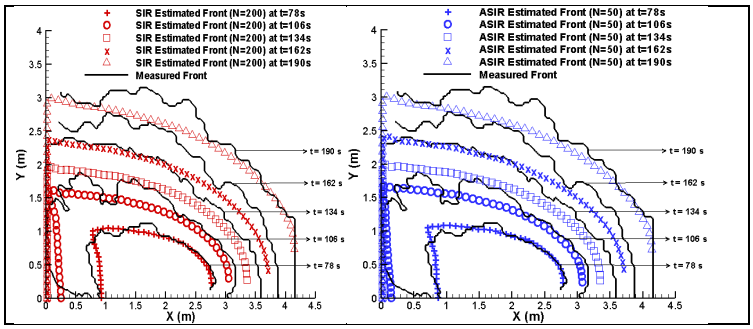


Figure 3: Comparison between simulated and measured fire front positions from t = 78 s to t = 190 s using the SIR filter (left) and the ASIR filter (right).

These results indicate that the ASIR filter converges better than the SIR filter for the fire spread problem. However, it is difficult to assess which of the particle filters provides the most realistic fire spread. It is therefore important to perform an additional analysis in the parameter space to further analyze the performance of SIR and ASIR filters. Figures 4, 5 and 6 show the mean value of the posterior sample along with the 99%-credible interval $I_{99\%}$, associated with the fuel moisture content M_v, the fuel particle surface-area-to-volume ratio Σ_v and the wind speed m_w, respectively. The SIR results are shown for $N = 200$ particles, while the ASIR results are shown for $N = 50$ particles. The posterior mean value found in Rochoux *et al.* [18] for the control of the vegetal fuel parameters only (M_v and Σ_v) with the EKF algorithm is also represented. It is found that the EKF solution is within the confidence interval and relatively close to the mean solution of particle filters. Both data assimilation approaches provide consistent results, meaning that the EKF algorithm behaves reasonably well in this case, despite of its linearity assumption on the observation model. On the other hand, while the SIR filter with 200 particles is found to provide the mean of the posterior PDF that is the closest to the EKF result, the ASIR filter with 50 particles provides a solution that reduces more effectively the size of the credible interval for both control parameters. The ASIR filter with 50 particles provides a more reliable solution and thereby, features a better approximation to the real fire spread than SIR. Besides, Fig. 6 shows that the estimation of the wind magnitude is in excellent agreement with the in-situ measurement made during the controlled burning experiment (the mean value of the wind velocity magnitude follows the same evolution as measurements). This result can be viewed as an additional validation of the state estimation problem.

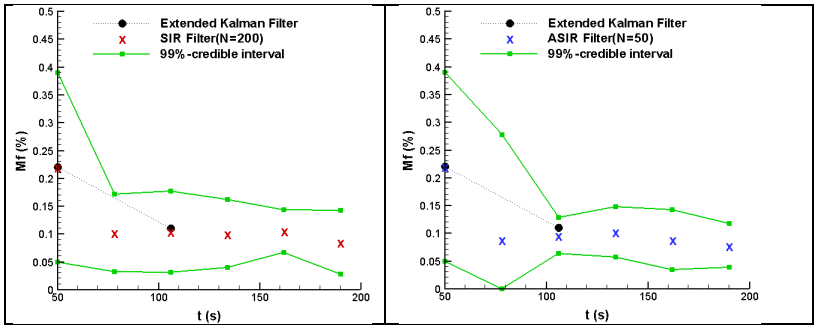


Figure 4: Sequential comparison of the fuel moisture content M_v provided by the SIR (left) / ASIR (right) algorithm with the EKF [18].

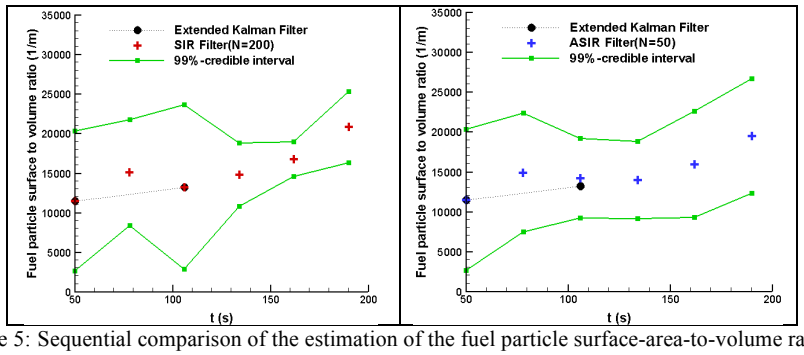


Figure 5: Sequential comparison of the estimation of the fuel particle surface-area-to-volume ratio Σ_v provided by the SIR (left) / ASIR (right) algorithm with the EKF [18].

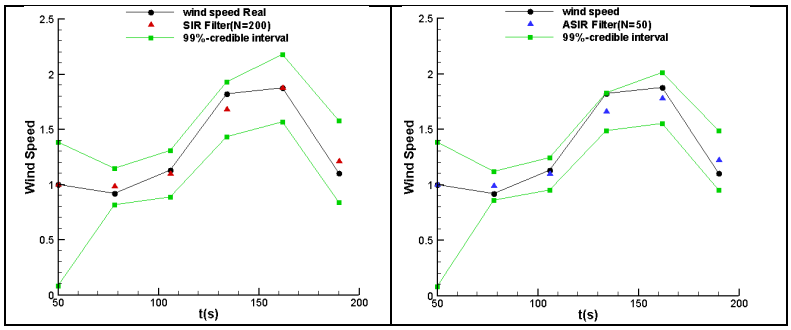


Figure 6: Sequential comparison of the estimation of the wind velocity magnitude provided by the SIR (left) / ASIR (right) algorithm.

4. Conclusion

This paper has explored the capability of particle filters (or sequential Monte Carlo approach) to improve the predictions of wildfire spread simulations using measurements of a reduced-scale controlled grassland burning experiment. The proposed inverse modeling technique relied on the estimation of a triplet of parameters characterizing the properties of the short grass vegetation on the one hand, the fluctuating wind conditions on the other hand. While both Sequential Importance Resampling (SIR) and Auxiliary Sequential Importance Resampling (ASIR) filters were able to sequentially track the displacement of the observed fire fronts, the ASIR filter was more efficient at retrieving accurate values of the control parameters at the expense of a lower computational cost than the SIR filter. Further analysis showed that the two-parameter Extended Kalman filter (EKF) estimation [18] provided similar results at the first assimilation time t = 78 s as the three-parameter SIR and ASIR estimations. Thus, the EKF algorithm was already able to retrieve physical values of the biomass fuel parameters. Results provided by the SIR and ASIR particle filters also showed that controlling the wind velocity magnitude in addition to these two biomass fuel properties allowed to keep tracking the fire front further in time (until time t = 190 s), even though the wind conditions were subject to significant fluctuations. These SIR and ASIR estimations were validated against independent observations, i.e., the in-situ measurement of the wind velocity magnitude. Ongoing research aims at further improving the Bayesian filtering strategy in order to better account for modeling uncertainties and to limit the computational cost of Bayesian

filtering problems using approaches such as Polynomial Chaos approximation [19]. Still, particle filters have already shown potential to relate comprehensively computational fire modeling and fire sensor technology as it is highly needed in the fire research area.

Acknowledgements

The financial support provided by FAPERJ, ANP (PRH37), CAPES and CNPq, Brazilian agencies for the fostering of science, as well as the *Centre National pour la Recherche Scientifique* (CNRS) and the French ministry of foreign affairs (MAEE), is greatly appreciated. The authors also acknowledge the program STIC-AmSud for supporting the project «11STIC06-I3PE-Inverse Problems in Physical Property Estimation».

References

1. Viegas, D.X., *Overview of Forest Fire Propagation Research*, Proc. Intl. Assoc. Fire Safety Science, 10 (2011) 95-108.
2. Rothermel, R.C., *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*, Research Paper INT-115, US Department of Agriculture Forest Service (1972).
3. Sullivan, A.L., *Wildland surface fire spread modeling, 1990-2007: 2. Empirical and quasi-empirical models*, International Journal of Wildland Fire, 18:4 (2009) 369-386.
4. Gelb, A., *Applied Optimal Estimation*, Cambridge Massachusetts MIT Press (1974).
5. Kaipio, J., Somersalo, E., *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences 160, Springer-Verlag (2004).
6. Kalman, R., *A New Approach to Linear Filtering and Prediction Problems*, ASME J. Basic Engineering, 82 (1960) 35-45.
7. Talagrand, O., *Assimilation of Observations, an Introduction*, J. Meteor. Soc. Japan, 75:1B (1997), 191-209.
8. Doucet, A., Freitas, N., Gordon, N., *Sequential Monte Carlo Methods in Practice*, Springer, New York (2001).
9. Ristic, B., Arulampalam, S., Gordon, N., *Beyond the Kalman Filter*, Artech House, Boston (2004).
10. Gordon, N.J., Salmond, D.J., Smith, A.F.M., *Novel approach to non-linear/non-Gaussian Bayesian state estimation*, Proceedings of IEEE, 140:2 (1993), 107-113.
11. Orlande, H.R.B., Colaço, M.J., Dulikravich, G., Vianna,F., Silva, W.B., Fonseca, H., Fudym, O., *State Estimation Problems In Heat Transfer*, International Journal for Uncertainty Quantification, 2:3 (2012), 239–258.
12. Pitt, M., Shephard, N., *Filtering via simulation: Auxiliary particle filters*, J. Amer. Statist. Assoc., 94:446 (1999), 590–599.
13. Silva, W.B., Orlande, H.R.B., Colaço, M. J., Fudym, O., *Application Of Bayesian Filters To A One-Dimensional Solidification Problem*, In: 21st Brazilian Congress of Mechanical Engineering, Natal, RN, Brazil, (2011).
14. Colaço, M.J., Orlande, H.R.B., Silva, W.B., Dulikravich, G., *Application Of A Bayesian Filter To Estimate Unknown Heat Fluxes In A Natural Convection Problem*, Journal of Heat Transfer, 134:9 (2012), 092501.
15. West, M., *Approximating posterior distributions by mixture*, *J.R.Statist. Soc. B*, 55 (1993), 409–422.

16. Mandel, J., Bennethum, L.S., Beezley, J.D., Coen, J.L., Douglas, C.D., Kim, M., Vodacek, A., *A Wildland Fire Model With Data Assimilation*, Mathematics and Computers in Simulation, 79 (2008), 584-606.
17. Gu, F., *Dynamic Data Driven Application System for Wildfire Spread Simulation*, PhD thesis, Georgia State University (2010).
18. Rochoux, M.C., Delmotte, B., Cuenot, B., Ricci, S., Trouvé, A., *Regional-Scale Simulations of Wildland Fire Spread informed by Real-time Flame Front Observations*, Proc. Combust. Inst., 34:2 (2012), 2641-2647.
19. Rochoux, M.C., Ricci, S., Lucor, D., Cuenot, B., Trouvé, A., Bart, J-M., *Towards predictive simulations of wildfire spread using a reduced-cost Ensemble Kalman Filter based on Polynomial Chaos approximations*, In Proceedings of the Summer Program, Center for turbulence Research, NASA AMES, Stanford University, USA, July (2012)
20. Wooster, M.J., Roberts, G., Perry, G., Kaufman, Y.J., *Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release*, Journal of Geophysical Research, 110, D2431 (2005).

Bibliography

Adrian, R. (2005). Twenty years of particle image velocimetry. *Experiments in Fluids 39*, 159–169. *(p. 150, 467)*

Albini, F. (1985). A model for fire spread in wildland fuels by radiation. *Combustion Science and Technology 42*, 229–258. *(p. 15, 23)*

Albini, F. (1986). Wildland fire spread by radiation - A model including fuel cooling by natural convection. *Combustion Science and Technology 45*(1-2), 101–113. *(p. 23)*

Alexander, G., J. Weinman, and J. Schols (1998). The use of digital warping of microwave integrated water vapor imagery to improve forecasts of marine extratropical cyclones. *Monthly Weather Review 126*(6), 1469–1496. *(p. 271)*

Alexander, M. (1982). Calculating and interpreting forest fire intensities. *Canadian Journal of Botany 60*(4), 349–357. *(p. 34)*

Amaya, J., E. Collado, B. Cuenot, and T. Poinsot (2010). Coupling LES, radiation and structure in gas turbine simulations. In *Proceedings of the 2010 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*. *(p. 93, 94, 136, 140, 441)*

Ambrosia, V. and S. Wegener (2009). Unmanned airborne platforms for disaster remote sensing support. In Pei-Gee and P. Ho (Eds.), *Geoscience and Remote Sensing*, Chapter 5, pp. 91–114. Vukovar, Croatia: InTech. *(p. 33)*

Anand, M. and S. Pope (1987). Calculations of premixed turbulent flames by PDF methods. *Combustion and Flame 67*(2), 127–142. *(p. 92)*

Anderson, J. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review 129*, 2884–2903. *(p. 252)*

Anderson, J. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A 59*(2), 210–224. *(p. 257)*

Anderson, J. and S. Anderson (2003). A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review 131*, 634–642. *(p. 252, 256)*

Andreadis, K., E. Clark, D. Lettenmaier, and D. Alsdorf (2007, May). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters 34*(10), L10403. *(p. 276)*

Andrews, P. (1986). BEHAVE: Fire behavior prediction and modeling system, BURN Subsystem Part 1. Technical report, U.S. Department of Agriculture, Forest Service. *(p. 24, 31, 417)*

Andrews, P., C. Blevins, and R. Seli (2008). BehavePlus fire modeling system, version 4.0: User's guide. Technical report, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. *(p. 24)*

Angelberger, C., F. Egolfopoulos, and D. Veynante (2000). Large eddy simulations of chemical and acoustic effects on combustion instabilities. *Flow, Turbulence and Combustion 65*(2), 205–222. *(p. 91)*

Arias, L. and E. Pezoa (2012). Automated algorithm for removing continuous flame spectrum based on sampled linear bases. *World Academy of Science, Engineering and Technology 6*(3), 365–372. *(p. 156)*

Arino, O. and J. Melinotte (1995). Fire index atlas. *Earth Observation, European Space Agency Quarterly 50*, 11–16. *(p. 32)*

Arino, O. and J. Mellinotte (1998). The 1993 africa fire map. *International Journal of Remote Sensing 19*, 2019–2023. *(p. 38)*

Arulampalam, S., S. Maskell, N. Gordon, and T. Clapp (2001). A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing 50*, 174–188. *(p. 258, 259, 260)*

Ashton, C., G. Rein, J. Rivera, J. Torero, C. Leff, M. Davies, and A. Gray (2007). Experiments and observations of peat smouldering fires. In *International Meeting of Fire Effects on Soil Properties*, Barcelona. *(p. 11)*

Aster, R., B. Borchers, and C. Thurber (2012). *Parameter estimation and inverse problems* (Second ed.). Elsevier. *(p. 198)*

Auzillon, P., B. Fiorina, R. Vicquelin, N. Darabiha, O. Gicquel, and D. Veynante (2011). Modeling chemical flame structure and combustion dynamics in les. *Proceedings of the Combustion Institute 33*(1), 1331–1338. *(p. 90, 91, 92, 93)*

Auzillon, P., O. Gicquel, N. Darabiha, D. Veynante, and B. Fiorina (2012). A filtered tabulated chemistry model for LES of stratified flames. *Combustion and Flame 159*(8), 2704–2717. *(p. 91)*

Balbi, J.-H., F. Morandini, X. Silvani, J.-B. Filippi, and F. Rinieri (2009, December). A physical model for wildland fires. *Combustion and Flame 156*(12), 2217–2230. *(p. 23, 27, 97, 417, 428, 429, 433)*

Bartoli, P. (2011). *Forest fires: Improvement of the knowledge of the coupling between flames and vegetative fuels*. Ph. D. thesis, Université de Corse, France - University of Edinburgh, UK. *(p. 115)*

Battin-Leclerc, F., J. Simmie, and E. Blurock (2013). *Cleaner combustion. Developing detailed chemical kinetic models*. Green Energy and Technology. Springer. *(p. 63)*

Beezley, J. (2009). *High-dimensional data assimilation and morphing ensemble Kalman filters with applications in wildfire modeling*. Ph. D. thesis, University of Colorado, USA. *(p. 269, 272, 275, 305, 378)*

Beezley, J. and J. Mandel (2008). Morphing ensemble Kalman filters. *Tellus 60A*, 131–140. *(p. 252, 269, 271, 272, 276, 305, 346, 378)*

Beji, T., S. Verstockt, R. Van de Walle, and B. Merci (2011). Data assimilation based numerical simulations to assist real-time smoke control management in large spaces. In *7th Proceedings of the Mediterranean Combustion Symposium, Cagliari, Sardinia, Italy*. *(p. 269)*

Beji, T., S. Verstockt, R. Van der Walle, and B. Merci (2012). Prediction of smoke filling in large volumes by means of data assimilation-based numerical simulations. *Journal of Fire Sciences 30*(4), 300–317. *(p. 269)*

Beltrán, C. (2012). *Problèmes directs et inverses en interaction fluide-structure. Application à l'hémodynamique.* Ph. D. thesis, Université Pierre et Marie Curie - Paris VI. *(p. 205)*

Bergman, N. (1999). *Recursive Bayesian estimation: Navigation and tracking applications*. Ph. D. thesis, Linköping University, Sweden. *(p. 258, 260)*

Bergthorsson, P. and B. Döös (1955). Numerical weather map analysis. *Tellus 7*(3), 329–340. *(p. 202)*

Beven, K. and J. Freer (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology 249*, 11–29. *(p. 328)*

Bilger, R. (1989). Turbulent diffusion flames. *Annual Review of Fluid Mechanics 21*, 101–135. *(p. 65, 166)*

Bilger, R. (1993). Conditional moment closure for turbulent reacting flow. *Physics of Fluids A 5*(2), 436–444. *(p. 92)*

Bishop, C., B. Etherton, and S. Majumdar (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review 129*(3), 420–436. *(p. 252)*

Blanchard, E., A. Sandu, and C. Sandu (2010). PSM: A polynomial chaos-based Kalman filter approach for parameter estimation of mechanical systems. *Journal of Dynamic Systems, Measurement and Control 132*(6), 061404–061404–18. *(p. 360)*

Bockhorn, H., P. Hbisreuther, and M. Hettel (2009). *Numerical modeling of technical combustion*, pp. 325–340. 100 volumes of "Notes on numerical fluid mechanics". Berlin, Heidelberg: Springer. *(p. 79)*

Bocquet, M. (2011). Introduction aux principes et méthodes de l'assimilation de données en géophysique. Technical report, CEREA, Ecole des Ponts ParisTech. *(p. 198, 206, 210, 242)*

Boé, J., J. Terray, E. Martin, and F. Habets (2009). Projected changes in components of the hydrological cycle in French river basins during the 21st century. *Water Resources Research 45*, W08426. *(p. 5)*

Boger, M., D. Veynante, H. Boughanem, and A. Trouvé (1998). Direct numerical simulation analysis of flame surface density concept for large eddy simulation of turbulent premixed combustion. *Proceedings of the Combustion Institute 27*, 917–925. *(p. 90, 92)*

Bohren, C. and D. Huffman (1983). *Absorption and scattering of light by small particles*. New York: Wiley. *(p. 197)*

Boileau, M., G. Staffelbach, B. Cuenot, T. Poinsot, and C. Bérat (2008). LES of an ignition sequence in a gas turbine engine. *Combustion and Flame 154*(1-2), 2–22. *(p. 78, 91)*

Boivin, P., C. Jiménez, A. Sanchez, and F. Williams (2011). An explicit reduced mechanism for H2-air combustion. *Proceedings of the Combustion Institute 33*, 517–523. *(p. 82)*

Boles, S. and D. Verbyla (1999). Effect of scan angle on AVHRR fire detection accuracy in interior Alaska. *International Journal of Remote Sensing 20*, 3437–3443. *(p. 34)*

Bonan, B. (2013). *Assimilation de données pour l'initialisation et l'estimation de paramètres d'un modèle d'évolution de calotte polaire*. Ph. D. thesis, Université de Grenoble. *(p. 252)*

Boonmee, N. (2004). *Theoretical and experimental study of auto-ignition of wood*. Ph. D. thesis, University of Maryland, USA. *(p. 111)*

Boonmee, N. and J. Quintiere (2005). Glowing-ignition of wood: The onset of surface combustion. *Proceedings of the Combustion Institute 30*, 2303–2310. *(p. 20, 111)*

Bouttier, F. and P. Courtier (1999). Data assimilation concepts and methods. Technical report, ECMWF. *(p. 198, 202, 203, 204, 226, 242, 244)*

Boyaval, S. (2012). A fast monte-carlo method with a reduced basis of control variables applied to uncertainty propagation and bayesian estimation. *Computer Methods in Applied Mechanics and Engineering 241*, 190–205. *(p. 360)*

Bradley, D., L. Kwa, A. Lau, and M. Missaghi (1988). Laminar flamelet modelling of recirculation premixed methane and propane-air combustion. *Combustion and Flame 71*, 109–122. *(p. 84)*

Bradshaw, L., J. Deeming, R. Burgan, and J. Cohen (1984). The 1978 national fire-danger rating system: Technical documentation. Technical Report General Technical Report INT-169, U.S. Department of Agriculture, Forest Service, Intermountain forest and range experiment station, Ogden, Utah, USA. *(p. 7)*

Brankart, J.-M., E. Cosme, C.-E. Testut, P. Brasseur, and J. Verron (2010). Efficient adaptive error parameterizations for square root or ensemble kalman filters: Application to the control of ocean mesoscale signals. *Monthly Weather Review 138*(3), 932–950. *(p. 257)*

Brankart, J.-M., C. Ubelmann, C.-E. Testut, E. Cosme, P. Basseur, and J. Verron (2009). Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: Application to ocean altimetry. *Monthly Weather Review 137*, 1908–1927. *(p. 218, 315)*

Briess, K., H. Jahn, E. Lorenz, D. Oertel, W. Skrbek, and B. Zhukov (2003). Fire recognition potential of the bi-spectral infrared detection (bird) satellite. *International Journal of Remote Sensing 24*(4), 865–872. *(p. 39)*

Broido, A. (1969). A simple, sensitive graphical method of treating thermogravimetric analysis data. *Journal of Polymer Science. Part A-2: Polymer Physics 7*(10), 1761–1773. *(p. 106)*

Broido, A. (1991). Chemistry of pyrolysis and combustion of cellulose. In *Proceedings from Nordic Workshop on Combustion of Biomass*. *(p. 114)*

Bronli, M. (1996). *A theoretical and experimental study of the thermal degradation of wood*. Ph. D. thesis, Norwegian University of Science and Technology, Faculty of Mechanical Engineering, Division of Thermal Energy and Hydro Power, Trondheim, Norway. *(p. 103, 107, 114)*

Buehner, M., P. Houtekamer, C. Charette, H. Mitchell, and B. He (2010a). Intercomparison of variational data assimilation and the ensemble Kalman filter for

global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review 138*, 1567–1586. *(p. 210)*

Buehner, M., P. L. Houtekamer, C. Charette, H. Mitchell, and B. He (2010b). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review 138*, 1550–1566. *(p. 210)*

Buehner, M. and P. Malanotte-Rizzoli (2003). Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation. *Journal of Geophysical Research 108*(C6). *(p. 204)*

Buis, S., A. Piacentini, D. Déclat, and t. P. Group (2006). PALM: A computational framework for assembling high performance computing applications. *Concurrency and computation: Practice and Experience 18*(2), 231–245. *(p. 440)*

Burgan, R. (1988). 1988 revisions to the 1978 national fire-danger rating system. Technical Report Research Paper SE-273, US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC, USA. *(p. 7)*

Burgan, R., R. Hartford, and J. Eidenshink (1996). Using ndvi to assess departure from average greenness and its relation to fire business. Technical Report General Technical Report INT-GTR-333, US Department of Agriculture, Forest Service, Intermountain forest and range experiment station, Ogden, Utah, USA. *(p. 7)*

Burgers, G., P. Van Leeuwen, and G. Evensen (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review 126*, 1719–1724. *(p. 250, 251, 252, 310)*

Butterfield, B. (1980). *Handbook of three-dimensional structure of wood* (2nd ed.)., pp. 99–115. London, UK: Chapman and Hall. *(p. 102)*

Byram, G. (1959). Combustion of forest fuels. In K. Brown (Ed.), *Forest Fire: Control and Use*, pp. 65–89. New York: McGraw-Hill. *(p. 29, 34)*

Byram, G. and W. Fons (1952). Thermal properties of forest fuels. Technical report, U.S. Department of Agriculture, Forest Service. *(p. 417)*

Cabrit, O. (2009). *Modélisation des flux pariétaux sur les tuyères des moteurs à propergol solide*. Ph. D. thesis, University Montpellier II, France. *(p. 136)*

Cabrit, O. and F. Nicoud (2010). Direct numerical simulation of a reacting turbulent channel flow with thermochemical ablation. *Journal of Turbulence 11*(44), 1–33. *(p. 136)*

Calle, A. and J. Casanova (2012). *Forest fires and remote sensing*. Earth and Planetary Sciences. *(p. 29, 34, 40)*

Calle, A., J. Casanova, and A. Romo (2006). Fire detection and monitoring using msg spinning enhanced visible and infrared imager (seviri) data. *Journal of Geophysical Research 111*(G4). *(p. 40)*

Camia, A. and G. Bovio (2000). European forest fire information system (effis): Meteorological indices. Technical report, European Commission. *(p. 8)*

Candel, S., T. Schmitt, and N. Darabiha (2011). Progress in transcritical combustion: Experimentation, modeling and simulation. In *23rd International Colloquium on the Dynamics of Explosions and Reactive Systems*, Irvine, California, USA. *(p. 120)*

Caraeni, D., C. Bergström, and L. Fuchs (2000). Modeling of liquid fuel injection, evaporation and mixing in a gas turbine burner using large eddy simulations. *Flow, Turbulence and Combustion 65*, 223–244. *(p. 78)*

Cardinali, C., S. Pezzuli, and E. Andersson (2004). Influence matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society 130*, 2767–2786. *(p. 242)*

Caudal, J. (2013). *Simulation numérique du reformage autothermique du méthane*. Ph. D. thesis, Ecole Centrale Paris, France. *(p. 120)*

Cavaliere, A. and M. de Joannon (2004). Mild combustion. *Progress in Energy and Combustion Science 30*(4), 329–366. *(p. 84)*

Cetegen, B. (1998). A phenomenological model of near-field fire entrainment. *Fire Safety Journal 31*, 299–312. *(p. 133, 134)*

Chandler, C., N. Cheney, P. Thomas, L. Trabaud, and D. Williams (1983). *Fire in forestry. Volume I: Forest Fire behavior and effects*. New York: John Wiley & Sons. *(p. 7)*

Chapnik, B., G. Desroziers, and F. Rabier (2004). Properties and first application of an error-statistics tuning method in variational assimilation. *Quarterly Journal of the Royal Meteorological Society 130*, 2253–2275. *(p. 208, 220)*

Chapnik, B., G. Desroziers, and F. Rabier (2006). Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quarterly Journal of the Royal Meteorological Society 132*, 543–565. *(p. 208, 220, 242)*

Charlette, F., C. Meneveau, and D. Veynante (2002). A powerful flame wrinkling model for LES of premixed turbulent combustion. Part i: Non-dynamic formulation. *Combustion and Flame 131*(1/2), 159–180. *(p. 92)*

Chaudhury, K. and K. Ramakrishnan (2007). Stability and convergence of the level-set method in computer vision. *Pattern Recognition Letters 28*, 884–893. *(p. 293)*

Chen, J. and R. Dibble (1991). Applications of reduced chemical mechanisms for the prediction of turbulent nonpremixed methane jet flames. In M. Smooke (Ed.), *Reduced Chemical Mechanisms and Asymptotic Approximations for Methane-Air Flames*, Volume 384 of *Lecture Notes in Physics*, pp. 193–226. Springer Berlin Heidelberg. *(p. 82)*

Chen, J., T. Kaiser, and W. Kollman (1993). Transient behavior of simplified reaction mechanisms for methane non-premixed combustion. *Combustion Science and Technology 92*, 313–347. *(p. 82)*

Chen, Y. and C. Snyder (2006). Assimilating vortex position with an ensemble Kalman filter. *Monthly Weather Review 135*, 1828–1845. *(p. 271)*

Cheney, N., J. Gould, and W. Catchpole (1993). The influence of fuel, weather and fire shape variables on fire-spread in grasslands. *International Journal of Wildland Fire 3*(1), 31–44. *(p. 6, 21, 24)*

Cheney, N., J. Gould, and W. Catchpole (1998). Prediction of fire spread in grasslands. *International Journal of Wildland Fire 8*(1), 1–13. *(p. 23, 24, 25, 417)*

Chomiak, J. and J. Nisbet (1985). Modeling variable density effects in turbulent flows - some basic considerations. *Combustion and Flame 102*, 371–386. *(p. 135)*

Clandillon, S. and H. Yesou (2011). GMES Emergency Response - Fire mapping experience within EO-based Rapid Mapping. In *EGU General Assembly 2011*, Volume 13. *(p. 40)*

Clark, T., J. Coen, and D. Latham (2004). Description of a coupled atmosphere-fire model. *International Journal of Wildland Fire 13*(1), 49–63. *(p. 26)*

Clements, C. (2007). Fireflux - Observing wildland grass fire dynamics. *Bulletin of the American Meteorological Society 88*(9), 1369–1382. *(p. 27)*

Coelho, P. (2007). Numerical simulation of the interaction between turbulence and radiation in reactive flows. *Progress in Energy and Combustion Science 33*, 311–383. *(p. 93)*

Coheur, P., L. Clarisse, S. Turquety, D. Hurtmans, and C. Clerbaux (2009). Iasi measurements of reactive trace species in biomass burning plumes. *Atmospheric Chemistry and Physics 9*, 5655–5667. *(p. 40)*

Cohn, S. (1997). An introduction to estimation theory. Technical report, Office Note Series on Global Modeling and Data Assimilation, DAO Office Note 97-01, Goddard Space Flight Center, Greenbelt, Maryland. *(p. 215, 223, 242)*

Colaço, M., H. Orlande, W. Da Silva, and G. Dulikravich (2011). Application of a Bayesian filter to estimate unknown heat fluxes in a natural convection problem. In B. Dennis and J. Michopoulos (Eds.), *ASME International Design Engineering*

Technical Conferences and Computers and Information in Engineering Conference, Number No. DETC2011-47652, Washington, USA. *(p. 262)*

Colin, O., F. Ducros, D. Veynante, and T. Poinsot (2000). A thickened flame model for large eddy simulations of turbulent premixed combustion. *Physics of Fluids 12*(7), 1843–1863. *(p. 91, 92, 93)*

Colucci, P., F. Jaberi, P. Givi, and S. Pope (1998). Filtered density function for large eddy simulation of turbulent reacting flows. *Physics of Fluids 10*(2), 499–515. *(p. 92)*

Consalvi, J., F. Nmira, A. Fuentes, P. Mindykowski, and B. Porterie (2011). Numerical study of piloted ignition of forest fuel layer. *Proceedings of the Combustion Institute 33*(2), 2641–2648. *(p. 19, 106, 109, 113)*

Cook, A. and J. Riley (1998). Subgrid-scale modeling for turbulent reacting flows. *Combustion and Flame 112*(4), 593–606. *(p. 92)*

Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994). A strategy for operational implementation of 4D-VAR using an incremental approach. *Quarterly Journal of the Royal Meteorological Society 120*, 1367–1387. *(p. 205, 228, 230, 235)*

Cowlard, A., W. Jahn, C. Abecassis-Empis, G. Rein, and J. Torero (2010). Sensor assisted fire fighting. *Fire Technology 46*, 719–741. *(p. 269, 274)*

Cressman, G. (1959). An operational objective analysis system. *Monthly Weather Review 87*, 367–374. *(p. 202)*

Crombette, P. (2010). Optimisation et poursuite des développements du système Livefire de géolocalisation automatisée et temps-réel de prises de vues aéroportées. Technical report, Université de Toulouse - Service Départmental d'Incendie et de Secours des Pyrénées-Orientales (SDIS-66), France. *(p. 41, 280)*

Cruz, M., A. Sullivan, J. Gould, N. Sims, A. Bannister, J. Hollis, and R. Hurley (2012). Anatomy of a catastrophic wildfire: The Black Saturday Kilmore East fire in Victoria, Australia. *Forest Ecology and Management 284*, 269–285. *(p. 5)*

Cuenot, B. and T. Poinsot (1995). Effects of curvature and unsteadiness in diffusion flames. implications for turbulent diffusion flames. *Proceedings of the Combustion Institute 25*, 1383–1390. *(p. 130, 132)*

Curry, J. and W. Fons (1938). Rate of spread of surface fires in the ponderosa pine type of California. *Journal of Agriculture Research 57*(4), 239–267. *(p. 15)*

da Silva, W., H. Orlande, M. Colaço, and O. Fudym (2011). Application of Bayesian filters to a one-dimensional solidification problem. In *Proceedings of 21st Brazilian Congress of Mechanical Engineering, October 24-28*, Natal, Brazil. *(p. 205, 262)*

da Silva, W., M. Rochoux, H. Orlande, M. Colaço, O. Fudym, M. El Hafi, B. Cuenot, and S. Ricci (2013a). Application of particle filters in moving frontier problems. In *4th Inverse Problems, Design and Optimization Symposium, June 26-28, Albi, France*. *(p. 53, 358, 404, 503, 504)*

da Silva, W., M. Rochoux, H. Orlande, M. Colaço, O. Fudym, M. El Hafi, B. Cuenot, and S. Ricci (2013b). Application of particle filters to regional-scale wildfire spread. *High Temperatures - High Pressures, International Journal of Thermophysical Properties Research*. (under revision). *(p. 52, 358, 404, 503, 504)*

Dagaut, P. and F. Lecomte (2003). Experimental and kinetic modeling study of the reduction of no by hydrocarbons and interactions with so2 in a jsr at 1 atm. *Fuel 82*(9), 1033–1040. *(p. 63)*

Daget, N. (2008). *Estimation d'ensemble des paramètres des covariances d'erreur d'ébauche dans un système d'assimilation variationnelle de données océaniques*. Ph. D. thesis, Université de Toulouse, France. *(p. 205)*

Daley, R. (1991). *Atmospheric data analysis*. Cambridge University Press. *(p. 198, 202, 216, 383)*

Damköhler, G. (1940). Zeitschrift für elektrochemie und angewandte physikalische chemie. *English Translation: NASA TM No. 1112 (1947) 46*, 601–626. *(p. 70)*

Davies, D., S. Kumar, and J. Descloitres (2004). Global fire monitoring using MODIS near-real-time satellite data. *GIM International 18*(4), 41–43. *(p. 2)*

Davis, C., B. Brown, and R. Bullock (2006). Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review 134*(7), 1772–1784. *(p. 271)*

Davis, W. and G. Forney (2001). Sensor-driven fire model version 1.1. Technical Report Report NISTIR 6705, National Institute of Standards and Technology. *(p. 268)*

de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen (2013). A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society 139*(674), 1199–1213. *(p. 247)*

DeCecchis, D., L. Drummond, and J. Castillo (2011). Design of a distributed coupling toolkit for high performance computing environment. *Mathematical and Computer Modelling 57*(9-10), 2267–2278. *(p. 442)*

Deckmyn, A. and L. Berre (2005). A wavelet approach to representing background error covariances in a limited area model. *Monthly Weather Review 133*, 1279–1294. *(p. 216, 247)*

Deeming, J., R. Burgan, and J. Cohen (1978). The national fire-danger rating system. Technical Report General Technical Report INT-39, U.S. Department of Agriculture, Forest Service, Intermountain forest and range experiment station, Ogden, Utah, USA. *(p. 7)*

Delmotte, B., M. Rochoux, S. Ricci, and A. Trouvé (2011). Parameter calibration using data assimilation for simulation of forest fire spread. Technical Report TR-CMGC-11-73, CERFACS - University of Maryland. *(p. 192, 288, 295, 296, 298, 481, 486, 487, 491)*

Denham, M., K. Wendt, G. Bianchini, A. Cortés, and T. Margalef (2012). Dynamic data-driven genetic algorithm for forest fire spread prediction. *Journal of Computational Science 3*, 398–404. *(p. 273)*

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society 131*, 3385–3396. *(p. 208, 220, 257)*

Desroziers, G., P. Brousseau, and B. Chapnik (2005). Use of randomization to diagnose the impact of observations on analyses and forecasts. *Quarterly Journal of the Royal Meteorological Society 131*, 2821–2837. *(p. 208)*

Desroziers, G. and S. Ivanov (2001). Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society 127*, 1433–1452. *(p. 242)*

Di Blasi, C. (1993). Modeling and simulation of combustion processes of charring and non-charring solid fuels. *Progress in Energy and Combustion Science 19*, 71–104. *(p. 100, 448, 450)*

Di Blasi, C. (2008). Modeling chemical and physical processes of wood and biomass pyrolysis. *Progress in Energy and Combustion Science 34*, 47–90. *(p. 100, 107, 108, 113)*

Doche, C., S. Ricci, and M. Rochoux (2012). Vers l'assimilation de données pour la simulation des feux de forêts à l'échelle régionale. Technical report, Météo-France - CERFACS, France, TR-CMGC-13-5. *(p. 38)*

Dopazo, C., L. Valino, and F. Fuego (1997). Statistical description of the turbulent mixing of scalar fields. *International Journal of Modern Physics B 11*(25), 2975–3014. *(p. 92)*

Doucet, A., N. Freitas, and N. Gordon (2001). *Sequential Monte Carlo methods in practice*. New York: Springer. *(p. 257, 262, 278)*

Doucet, A., S. Godsill, and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing 10*, 197–208. *(p. 258, 259, 260)*

Dowdy, A., G. Mills, K. Finkele, and W. de Groot (2009). Australian fire weather as represented by the mcarthur forest fire danger index and the canadian forest fire weather index. Technical Report CAWCR Technical Report No. 10, Centre for Australian Weather and Climate Research. *(p. 7, 8)*

Dozier, J. (1981). A method for satellite identification of surface temperature fields of subpixel resolution. *Remote Sensing of Environment 11*, 221–229. *(p. 35)*

Driscoll, J. (2008). Turbulent premixed combustion: Flamelet surface and its effect on turbulent burning velocities. *Progress in Energy and Combustion Science 34*(1), 91–134. *(p. 70)*

Duchaine, F., A. Corpron, L. Pons, V. Moureau, F. Nicoud, and T. Poinsot (2009). Development and assessment of a coupled strategy for conjugate heat transfer with large eddy simulation: Application to a cooled turbine blade. *International Journal of Heat Fluid Flow 30*(6), 1129–1141. *(p. 136, 441)*

Ducros, F., F. Nicoud, and T. Poinsot (1998). Wall-adaptating local eddy-viscosity models for simulations in complex geometries. In M. Baines (Ed.), *ICFD*, pp. 293–300. *(p. 80)*

Dupuy, J. and D. Morvan (2005). Numerical study of a crown fire spreading toward a fuel break using a multiphase physical model. *International Journal of Wildland Fire 14*, 141–151. *(p. 18, 19)*

Dupuy, J. and J.-C. Valette (1997). An analysis of semi-empirical and physical models for fire spread in wildland fuels. Technical report, Forest Fire Prevention Research Team, Institut National de la Recherche Agronomique. *(p. 289, 419)*

Durand, M., K. Andreadis, D. Alsdorf, D. Lettenmaier, D. Moller, and M. Wilson (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters 35*, L20401. *(p. 205, 276, 307, 333)*

Duwig, C. (2007). Study of a filtered flamelet formulation for large eddy simulation of premixed turbulent flames. *Flow, Turbulence and Combustion 79*(4), 433–454. *(p. 90)*

Emery, C., M. Rochoux, S. Ricci, and A. Trouvé (2013). State estimation using data assimilation for simulation of regional-scale wildfire spread with complex topography. Technical report, CERFACS - University of Maryland. *(p. 192, 288, 291, 314, 380, 492, 494, 496, 498, 499, 501)*

Emmons, H. (1964). Fire in the forest. *Fire Research Abstract and Reviews 5*, 163–178. *(p. 15, 417)*

Engel, C., T. Lane, M. Reeder, and M. Rezny (2013). The meteorology of Black Saturday. *Quarterly Journal of the Royal Meteorological Society 139*(672), 585–599. *(p. 5)*

Ern, A. and V. Giovangigli (1995). Fast and accurate multicomponent transport property evaluation. *Journal of Computational Physics 120*(1), 105–116. *(p. 75)*

European Forest Institute (2009). *Living with wildfires: What science can tell us. A contribution to the Science-Policy dialogue*. Yves Birod Editions. *(p. 2, 3)*

Evensen, G. (1992). Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model. *Journal of Geophysical Research 97*(C11), 17905–17924. *(p. 248)*

Evensen, G. (1993). Open boundary conditions for the extended Kalman filter with a quasi-geostrophic model. *Journal of Geophysical Research 98*(C9), 16529–16543. *(p. 248)*

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research 99*(C5), 10143–10162. *(p. 204, 244, 247, 248, 250, 278, 305, 333)*

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics 53*, 343–367. *(p. 251)*

Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics 54*, 539–560. *(p. 252)*

Evensen, G. (2007). *Data assimilation: the ensemble Kalman filter*. Berlin: Springer. *(p. 204, 244)*

Fendell, F. and M. Wolff (2001). Wind-aided fire spread. In E. Johnson and K. Miyanishi (Eds.), *Forest fires - Behavior and ecological effects*, Chapter 6, pp. 171–223. San Diego, USA: Academic Press. *(p. 22, 293, 482)*

Fernandez-Tarrazo, E., A. Sanchez, A. Linan, and F. Williams (2006). A simple one-step chemistry model for partially premixed hydrocarbon combustion. *Combustion and Flame 147*, 32–38. *(p. 83)*

Ferrya, N., E. Rémya, P. Brasseurb, and C. Maesc (2007). The MERCATOR global ocean operational analysis system: Assessment and validation of an 11-year reanalysis. *Journal of Marine Systems 65*, 540–560. *(p. 440)*

Fertig, E., S.-J. Baek, B. Hunt, E. Ott, I. Szunyogh, J. Aravéquia, E. Kalnay, H. Li, and J. Liu (2009). Observation bias correction with an ensemble Kalman filter. *Tellus 61*(2), 210–226. *(p. 208)*

Filippi, J.-B., F. Bosseur, C. Mari, C. Lac, P. Le Moigne, B. Cuenot, D. Veynante, D. Cariolle, and J.-H. Balbi (2009). Coupled atmosphere-wildland fire modelling. *Journal of Advances in Modeling Earth Systems 1*(4), Quarter 4. *(p. 27, 482)*

Filippi, J.-B., F. Bosseur, X. Pialat, P.-A. Santoni, S. Strada, and C. Mari (2011). Simulation of coupled fire/atmosphere interaction with the MesoNH-ForeFire models. *Journal of Combustion 2011*(ID540390), 1–13. *(p. 27, 31, 414)*

Filippi, J.-B., V. Mallet, and B. Nader (2013). Representation and evaluation of wildfire propagation simulations. (in press). *(p. 47)*

Filippi, J.-B., F. Morandini, J.-H. Balbi, and D. Hill (2011). Discrete event front-tracking simulation of a physical fire-spread model. *Simulation, Transactions of the Society for Modeling and Simulation International 87*, 555–580. *(p. 27)*

Filippi, J.-B., X. Pialat, and C. Clements (2013). Assessment of ForeFire/Meso-NH for wildland fire/atmosphere coupled simulation of the FireFlux experiment. *Proceedings of the Combustion Institute 34*, 2633–2640. *(p. 27, 28, 31, 414)*

Finney, M. (1998). FARSITE: Fire Area Simulator - Model Development and Evaluation. Technical Report February, US Department of Agriculture, Forest Service, Rocky Mountain Research Station. *(p. 24, 31, 104, 273, 417, 482)*

Finney, M., I. Grenfell, C. Mchugh, R. Seli, D. Trethewey, R. Stratton, and S. Brittain (2011). A method for ensemble wildland fire simulation. *Environmental Modeling and Assessment 16*, 153–167. *(p. 276, 277)*

Finney, M. and S. Mcallister (2011). A review of fire interactions and mass fires. *Journal of Combustion 2011*, ID548328. *(p. 4, 15, 68)*

Fiorina, B. (2004). *Modélisation de la combustion turbulente pour l'amélioration de la prédiction des pollutants dans un brûleur industriel*. Ph. D. thesis, Ecole Centrale Paris, France. *(p. 84)*

Fiorina, B., R. Baron, O. Gicquel, D. Thévenin, S. Carpentier, and N. Darabiha (2003). Modelling non-adiabatic partially-premixed flames using flame prolongation of ILDM. *Combustion Theory and Modelling 7*(3), 449–470. *(p. 84)*

Fiorina, B., D. Veynante, and S. Candel (2013). Modeling combustion chemistry in large eddy simulation of turbulent flames. In *International Symposium on Turbulence and Shear Flow Phenomena (TSFP-8), August 28-30, Poitiers, France*. *(p. 89)*

Fisher, M. (1998). Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*, Reading, UK, pp. 364–385. ECMWF. *(p. 204, 227)*

Fisher, M. (2003). Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. Technical report, ECMWF, Reading, UK. *(p. 242)*

Fishman, G. (1996). *Monte Carlo: Concepts, algorithms and application*. Springer-Verlag. *(p. 359)*

Fons, W. (1946). Analysis of fire spread in light forest fuels. *Journal of Agricultural Research 72*, 93–121. *(p. 10, 15, 417, 418)*

Forkel, H. and J. Janicka (2000). Large eddy simulation of a turbulent hydrogen diffusion flame. *Flow, Turbulence and Combustion 65*, 163–175. *(p. 78)*

Forthofer, J. (2007). Modeling wind in complex terrain for use in fire spread prediction. Technical report, Master thesis, Colorado State University, USA. *(p. 24, 25)*

Fouilloux, A. and A. Piacentini (1999). The PALM project: MPMD paradigm for an oceanic data assimilation software. *Lecture Notes in Computer Science 1685*, 1423–1430. *(p. 440)*

Fournier, Y., J. Bonelle, C. Moulinec, Z. Shang, A. Sunderland, and J. Uribe (2011). Optimizing Code_Saturne computations on Petascale systems. *Computers and Fluids 45*(1), 103–108. *(p. 442)*

Frandsen, W. (1971). Fire spread through porous fuels from the conservation of energy. *Combustion and Flame 16*, 9–16. *(p. 289, 417, 418, 419, 420)*

Franzelli, B. (2011). *Impact of the chemical description on direct numerical simulations and large eddy simulations of turbulent combustion in industrial aeroengines*. Ph. D. thesis, Université de Toulouse - MeGeP Dynamique des Fluides, France. *(p. 83, 85, 86, 87, 88, 93, 126, 163)*

Franzelli, B., E. Riber, M. Sanjosé, and T. Poinsot (2010). A two-step chemical scheme for kerosene-air premixed flames. *Combustion and Flame 157*(7), 1364–1373. *(p. 82, 88)*

Freeborn, P., M. Wooster, W. Hao, C. Ryan, B. Nordgren, S. Baker, and C. Ichoku (2008). Relationships between energy release, fuel mass loss, and trace gas and aerosol emissions during laboratory biomass fires. *Journal of Geophysical Research 113*(D01301). *(p. 36)*

Freitag, M., C. Lacor, A. Sadiki, and J. Janicka (2007). *Investigation of subgrid scale wrinkling models and their impact on the artificially thickened flame model in large eddy simulations*, Volume 56 of *Lecture Notes in Computational Science and Engineering*, Chapter Complex Effects in Large Eddy Simulations, pp. 353–369. Springer. *(p. 91)*

Friedman, H. (1964). Kinetics of thermal degradation of char-forming plastics from thermogravimetry. application to a phenolic plastic. *Journal of Polymer Science. Part C: Polymer Symposia 6*(1), 183–195. *(p. 115)*

Gandin, L. (1963). *Objective analysis of meteorological fields*. Leningrad, USSR: Gidrometeorologicheskoe Izdate'stvo. *(p. 202)*

Gao, F. and E. O'Brien (1993). A large-eddy simulation scheme for turbulent reacting flows. *Physics of Fluids A 5*(6), 1282–1284. *(p. 92)*

Garmory, A. and E. Mastorakos (2011). Capturing localised extinction in sandia flame F with LES-CMC. *Proceedings of the Combustion Institute 33*(1), 1673–1680. *(p. 92)*

Gauthier, P., M. Buehner, and L. Fillion (1999). Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses. In *Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems*, Reading, UK, pp. 131–145. *(p. 204, 216)*

Gauthier, P., M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau (2007). Extension of 3D-Var to 4D-Var: Implementation of 4D-Var at the meteorological service of Canada. *Monthly Weather Review 135*, 2339–2354. *(p. 205)*

Gelb, A. (1974). *Applied optimal estimation*. Cambridge MIT Press, Massachusetts, USA. *(p. 198, 221, 244, 247)*

Germano, M., U. Piomelli, P. Moin, and W. Cabot (1991). A dynamic subgrid-scale eddy viscosity model. *Physics of Fluids 3*(7), 1760–1765. *(p. 80, 92)*

Ghanem, R. and P. Spanos (1991). *Stochastic finite elements, a spectral approach*. Dover. *(p. 360, 361)*

Ghil, M., S. Cohn, J. Tavanzis, K. Bube, and E. Isaacson (1981). *Application of estimation theory to numerical weather prediction. Dynamic meteorology: Data assimilation methods*, Volume 75. Springer Verlag. *(p. 203)*

Gicquel, L., N. Gourdain, J.-F. Boussuge, H. Deniau, G. Staffelbach, P. Wolf, and T. Poinsot (2011). High performance parallel computing of flows in complex geometries. *Comptes Rendus Mecanique 339*(2-3), 104–124. *(p. 441)*

Gicquel, L., G. Staffelbach, and T. Poinsot (2012). Large eddy simulations of gaseous flames in gas turbine combustion chambers. *Progress in Energy and Combustion Science 38*, 782–817. *(p. 79)*

Gicquel, O., N. Darabiha, and D. Thévenin (2000). Laminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ILDM with differential diffusion. *Proceedings of the Combustion Institute 28*(2), 1901–1908. *(p. 84)*

Giglio, L., J. Descloitres, C. Justice, and Y. Kaufman (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment 87*, 273–282. *(p. 2, 34, 35, 38)*

Giglio, L. and J. Kendall (2001). Application of the Dozier retrieval to wildfire characterization: A sensitivity analysis. *Remote Sensing of Environment 77*, 34–49. *(p. 35)*

Giglio, L., J. Kendall, and C. Justice (1999). Evaluation of global fire detection algorithms using simulated AVHRR infrared data. *International Journal of Remote Sensing 20*, 1947–1985. *(p. 32, 34)*

Glimm, J. and D. Sharp (1999). Prediction and the quantification of uncertainty. *Physica D 133*, 152–170. *(p. 360)*

GODAE (2009). Special issue on the revolution in global ocean forecasting. godae: 10 years of achievement, oceanography society. *(p. 191, 205)*

Goody, R. and Y. Yung (1952). *Atmospheric radiation*. Oxford University Press. *(p. 93)*

Gordon, N., D. Salmond, and A. Smith (1993). Novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEEE Proceedings F, Radar and Signal Processing 140*(2), 107–113. *(p. 257, 260)*

Gorin, V. and M. Tsyrulnikov (2011). Estimation of multivariate observation-error statistics for AMSU-A data. *Monthly Weather Review 139*, 3765–3780. *(p. 218, 315)*

Goutiere, V., A. Charette, and L. Kiss (2002). Comparative performance of non-gray gas modeling techniques. *Numerical Heat Transfer: Part B: Fundamentals 41*, 361–381. *(p. 95)*

Goutiere, V., F. Liu, and A. Charette (2000). An assessment of real-gas modelling in 2d enclosures. *Journal of Quantitative Spectroscopy and Radiative Transfer 64*, 299–326. *(p. 95)*

Grishin, A. (1997). Mathematical modeling of forest fires and new methods of fighting them. Technical report, Publishing House of Tomsk State University, English translation edition, Tomsk, Russia. *(p. 17, 18, 106, 110, 113, 118, 447)*

Gu, F. (2010). *Dynamic data driven application system for wildfire spread simulation*. Ph. D. thesis, Georgia State University, USA. *(p. 274)*

Gu, X., M. Haq, M. Lawes, and R. Woolley (2000). Laminar burning velocity and markstein lengths of methane-air mixtures. *Combustion and Flame 121*(1-2), 41–58. *(p. 85, 86)*

Gürol, S. (2013). *Solving regularized nonlinear least-squares problem in dual space*. Ph. D. thesis, Université de Toulouse, France. *(p. 205)*

Gustafsson, N. (2007). Discussion on "4D-Var or EnKF?". *Tellus 59A*, 774–777. *(p. 210)*

Habert, J., S. Ricci, A. Piacentini, G. Jonville, T. Morel, F. Duchaine, A. Thévenin, E. Le Pape, O. Thual, N. Goutal, F. Zaoui, and R. Ata (2012). Estimation of

lateral inflows using data assimilation in the context of real-time flood forecasting for the Marne catchment in France. In *XIXth TELEMAC-MASCARET User Conference in Oxford, UK, October 18-19*. *(p. 440)*

Hadamard, J. (1902). *Sur les problèmes aux dérivées partielles et leur signification physique*. Princeton University Bulletin. *(p. 197)*

Hadden, R., G. Rein, and C. Belcher (2013). Study of the competing chemical reactions in the initiation and spread of smouldering combustion in peat. *Proceedings of the Combustion Institute 34*, 2547–2553. *(p. 11)*

Hamill, T., J. Whitaker, and C. Snyder (2001). Distance-dependent filtering of background error covariance in an ensemble kalman filter. *Monthly Weather Review 129*(11), 2776–2790. *(p. 256)*

Hammersley, J. and D. Hanscomb (1964). *Monte Carlo methods*. London: Chapman and Hall. *(p. 257)*

Hanson, H., M. Bradley, J. Bossert, R. Linn, and L. Younker (2000). The potential and promise of physics-based wildfire simulation. *Environmental Science and Policy 3*, 161–172. *(p. 17, 18)*

Harader, E., V. Borrell-Estupina, S. Ricci, M. Coustau, O. Thual, A. Piacentini, and C. Bouvier (2012). Correcting the radar rainfall forcing of a hydrological model with data assimilation: application to flood forecasting in the Lez catchment in Southern France. *Hydrology and Earth System Sciences 16*, 4247–4264. *(p. 205, 245, 247)*

Hawkes, E. and S. Cant (2000). A flame surface density approach to large eddy simulation of premixed turbulent combustion. *Proceedings of the Combustion Institute 28*, 51–58. *(p. 92)*

Hekestad, G. (1998). Dynamics of the fire plume. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences 356*(1748), 2815–2833. *(p. 134)*

Hirsch, K. (1996). Canadian forest fire behavior prediction (FBP) system: User's guide. Technical report, Northern Forest Centre, Special Report No. 7. *(p. 7)*

Hirschfelder, J., C. Curtis, and B. Bird (1954). *Molecular theory of gases and liquids*. New York: John Wiley & Sons. *(p. 75)*

Hoffman, R., Z. Liu, J.-F. Louis, and C. Grassoti (1995). Distortion representation of forecast errors. *Monthly Weather Review 123*(9), 2758–2770. *(p. 271)*

Hopkins, D. and J. Quintiere (1996). Material fire properties and predictions of thermoplastics. *Fire Safety Journal 26*, 241–268. *(p. 105)*

Houtekamer, P. and H. Mitchell (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review 126*, 796–811. *(p. 244, 247, 250, 305)*

Huggett, C. (1980). Estimation of rate of heat release by means of oxygen consumption measurements. *Fire and Materials 4*(2), 61–65. *(p. 106, 146)*

Hunt, B., E. Kostelich, and I. Szunyogh (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D 230*, 112–126. *(p. 256)*

Ichoku, C. and Y. Kaufman (2005). A method to derive smoke emission rates from MODIS fire radiative energy measurements. *IEEE Transactions on Geoscience and Remote Sensing 43*(11), 2636–2649. *(p. 35)*

Ichoku, C., Y. Kaufman, L. Giglio, Z. Li, R. Fraser, J. Jin, and M. Park (2003). Comparative analysis of daytime fire detection algorithms using avhrr data for the 1995 fire season in canada: Perspective for modis. *International Journal of Remote Sensing 24*, 1669–1690. *(p. 38)*

Ide, K., P. Courtier, M. Ghil, and A. Lorenc (1997). Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan 75*(1B), 181–189. *(p. 198, 221)*

Ihme, M. and H. Pitsch (2008). Modeling of radiation and nitric oxide formation in turbulent nonpremixed flames using a flamelet/progress variable formulation. *Physics of Fluids 20*(5), 055110. *(p. 84)*

Ince, N. and B. Launder (1989). On the computation of buoyancy-driven turbulent flows in rectangular enclosures. *International Journal of Heat and Fluid Flow 10*(2), 110–117. *(p. 135)*

Incropera, F. and D. DeWitt (1996). *Fundamentals of heat and mass transfer* (fourth ed.). Chichester, UK: Wiley. *(p. 112)*

Jahn, W. (2010). *Inverse modelling to forecast enclosure fire dynamics*. Ph. D. thesis, University of Edinburgh, UK. *(p. 269)*

Jahn, W., G. Rein, and J. Torero (2012). Forecasting fire dynamics using inverse computational fluid dynamics and tangent linearisation. *Advances in Engineering Software 47*(1), 114–126. *(p. 269, 274)*

Jaluria, Y. (1980). *Natural convection heat and mass transfer*. New York: Pergamon Press. *(p. 69)*

Janicka, J. and A. Sadiki (2005). Large eddy simulation of turbulent combustion systems. *Proceedings of the Combustion Institute 30*(1), 537–547. *(p. 79)*

Janjić, T. and S. Cohn (2006). Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review 134*, 2900–2915. *(p. 214)*

Jauré, S., F. Duchaine, and L. Gicquel (2011). Comparisons of coupling strategies for massively parallel conjugate heat transfer with large eddy simulation. In *IV International Conference on Computational Methods for Coupled Problems in Science and Engineering*. *(p. 136, 441, 442)*

Jazwinski, A. (1970). *Stochastic processes and filtering theory*. San Diego, California, USA: Academic Press. *(p. 204, 223, 242)*

Jean-Baptiste, N., P.-O. Malaterre, C. Dorée, and J. Sau (2011). Data assimilation for real-time estimation of hydraulic states and unmeasured perturbations in a 1D hydrodynamic model. *Mathematics and Computers in Simulation 81*, 2201–2214. *(p. 260)*

Jensen, K., J. Ripoll, A. Wray, D. Joseph, and M. Hafi (2007). On various modeling approaches to radiative heat transfer in pool fires. *Combustion and Flame 148*(4), 263–279. *(p. 93)*

Jones, W. and R. Lindstedt (1988). Global reaction schemes for hydrocarbon combustion. *Combustion and Flame 73*(3), 222–233. *(p. 82, 83, 126)*

Joppich, W. and M. Kürschner (2006). MPCCI - A tool for the simulation of coupled applications. *Concurrency and computation: Practice and Experience 18*(2), 183–192. *(p. 442)*

Joulain, P. (1996). Convective and radiative transport in pool and wall fires: 20 years of research in poitiers. *Fire Safety Journal 26*, 99–149. *(p. 68, 133)*

Julier, S. (1998). A skewed approach to filtering. *Signal and Data Processing of Small Targets SPIE Volum*, 271–282. *(p. 254, 255)*

Justice, C., L. Giglio, S. Korontzi, J. Owens, J. Morisette, D. Roy, J. Descloitres, S. Alleaume, F. Petitcolin, and Y. Kaufman (2002). The MODIS fire products. *Remote Sensing of Environment 83*, 244–262. *(p. 32, 35)*

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering 82*, 35–45. *(p. 201)*

Kalman, R. and R. Bucy (1961). New results in linear filtering and prediction theory. *Transactions of the ASME - Journal of Basic Engineering 83*(1), 95–108. *(p. 201)*

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press. *(p. 198, 206, 221)*

Kaufman, Y., C. Justice, L. Flynn, J. Kendall, E. Prins, L. Giglio, D. Ward, W. Menzel, and A. Setzer (1998). Potential global fire monitoring from eos-modis. *Journal of Geophysical Research 103*(D24), 32215–32238. *(p. 34, 35)*

Kaufman, Y., R. Remer, D. Ottmar, D. Ward, L. Rong, R. Kleidman, R. Fraser, L. Flynn, D. McDougal, and G. Shelton (1996). Relationship between remotely sensed fire intensity and rate of emission of smoke: SCAR-C experiment. In J. Levine (Ed.), *Global biomass burning*, pp. 685–696. Cambridge, MA: MIT Press. *(p. 35)*

Kee, R., M. Coltrin, and P. Glarborg (2003). *Chemically reacting flow: Theory and Practice*. Hoboken, New Jersey, USA. John Wiley & Sons. *(p. 119)*

Kee, R., F. Rupley, and J. Miller (1993). Chemkin-ii: A fortran chemical kinetics package for the analysis of gas phase chemical kinetics. Technical report, SANDIA National Laboratories. *(p. 120)*

Keetch, J. and G. Byram (1968). A drought index for forest fire control. Technical Report Research Paper SE-38, US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC, USA. *(p. 7)*

Kennedy, P., A. Belward, and J. Gregoire (1994). An improved approach to fire monitoring in West Africa using AVHRR data. *International Journal of Remote Sensing 15*, 2235–2255. *(p. 32)*

Kerstein, A., W. Ashurst, and F. Williams (1988). Field equation for interface propagation in an unsteady homogeneous flow field. *Physical Review A 37*(7), 2728–2731. *(p. 90, 293)*

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian non-linear state space models. *Journal of Computational and Graphical Statistics 5*(1), 1–25. *(p. 260)*

Klimenko, A. and R. Bilger (1999). Conditional moment closure for turbulent combustion. *Progress in Energy and Combustion Science 25*(6), 595–687. *(p. 92)*

Knikker, R., D. Veynante, and C. Meneveau (2002). A priori testing of a similarity model for large eddy simulations of turbulent premixed combustion. *Proceedings of the Combustion Institute 29*, 2105–2111. *(p. 92)*

Knudsen, E. and H. Pitsch (2010). A dynamic model for the turbulent burning velocity for large eddy simulation of turbulent combustion. *Physics of Fluids 22*(11), 740–760. *(p. 92)*

Kochanski, A., M. Jenkins, J. Mandel, J. Beezley, and S. Krueger (2013). Real time simulation of 2007 Santa Ana fires. *Forest Ecology and Management 294*, 136–149. *(p. 3, 26, 31)*

Koo, S.-H., J. Fraser-Mitchell, and S. Welch (2010). Sensor-steered fire simulation. *Fire Safety Journal 45*(3), 193–205. *(p. 268, 269)*

Kundu, P., I. Cohen, and D. Dowling (2011). *Fluid Mechanics* (5th Edition ed.). Academic Press. *(p. 68)*

Kunii, O., S. Kanagawa, I. Yajima, Y. Hisamatsu, S. Yamamura, T. Amagai, and I. Ismail (2002). The 1997 haze disaster in indonesia: Its air quality and health effects. *Archives of Environmental Health 57*(1), 16–22. *(p. 4)*

Lafore, J., J. Stein, N. Asencio, P. Bougeault, V. Ducrocq, J. Duron, C. Fisher, P. Hereil, P. Mascart, V. Masson, P. Pinty, J. Redelsperger, E. Richard, and J. Vilà-Guerau de Arellano (1998). The Meso-NH atmospheric simulation system. Part 1: Adiabatic formulation and control simulations. *Annales Geophysicae 16*, 90–109. *(p. 27)*

Lamouroux, J., M. Ihme, B. Fiorina, and O. Gicquel (2013). Flamelet-progress variable approach for mild combustion. (submitted). *(p. 84, 91)*

Larini, M., F. Giroud, B. Porterie, and J. Loraud (1998). A multiphase formulation for fire propagation in heterogeneous combustible media. *International Journal of Heat and Mass Transfer 41*(6-7), 881–897. *(p. 17, 18, 106, 109, 447)*

Laroche, S., P. Gauthier, M. Tanguay, S. Pellerin, and J. Morneau (2007). Impact of the different components of 4D-Var in the global forecast system of the meteorological service of Canada. *Monthly Weather Review 135*, 2355–2364. *(p. 205)*

Launder, B. and D. Spalding (1974). The numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering 3*(2), 269–289. *(p. 76)*

Lautenberger, C. (2013). Wildland fire modeling with an Eulerian level set method and automated calibration. (in press). *(p. 273, 274, 275, 279, 293, 492, 494, 496)*

Lautenberger, C., G. Rein, and A. Fernandez-Pello (2006). Application of a genetic algorithm to estimate material properties for fire modeling from bench-scale fire test data. *Fire Safety Journal 41*, 204–214. *(p. 273)*

Law, C. and C. Sung (2000). Structure, aerodynamics and geometry of premixed flamelets. *Progress in Energy and Combustion Science 26*, 459–505. *(p. 70)*

Lawson, W. and J. Hansen (2005). Alignment error models and ensemble-based data assimilation. *Monthly Weather Review 133*(6), 1687–1709. *(p. 271)*

Le Dimet, F. and O. Talagrand (1986). Variational algorithm for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus 38A*, 97–110. *(p. 205)*

Le Maître, O. and O. Knio (2010). *Spectral methods for uncertainty quantification with applications to computational fluid dynamics*. Scientific Computation. Springer. *(p. 360)*

Leblanc, M. and A. Trouvé (2009). Inverse zone modeling of enclosure fire dynamics. In *6th US Combustion Meeting*, Ann Arbor, MI, USA. *(p. 268, 269)*

Lecoustre, V., P. Narayanan, H. Baum, and A. Trouvé (2011). Local extinction of diffusion flames in fires. *Fire Safety Science – Proc. Tenth International Symposium, International Association for Fire Safety Science 10*, 583–595. *(p. 130)*

Lee, W. and S.-K. Lee (2005). The estimation of fire location and heat release rate by using sequential inverse method. *Journal of the Chinese Society of Mechanical Engineers 26*, 201–207. *(p. 268)*

Légier, J., T. Poinsot, and D. Veynante (2000). Dynamically thickened flame LES model for premixed and non-premixed turbulent combustion. In *Proceedings of the 2000 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*, pp. 157–168. *(p. 91)*

Lentile, L., Z. Holden, A. Smith, M. Falkowski, A. Hudak, P. Morgan, S. Lewis, P. Gessler, and N. Benson (2006). Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire 15*(3), 319–345. *(p. 37)*

Leroy, V. (2007). *Contribution à la modélisation des feux de forêt: Cinétique de dégradation thermique et Cinétique de combustion des végétaux*. Ph. D. thesis, Université Pascal Paoli, Corsica, France. *(p. 103, 114, 460)*

Leroy, V., E. Leoni, and P.-A. Santoni (2008). Reduced mechanism for the combustion of evolved gases in forest fires. *Combustion and Flame 154*, 410–433. *(p. 460)*

Li, J. and D. Xiu (2008). On numerical properties of the ensemble kalman filter for data assimilation. *Computer Methods in Applied Mechanics and Engineering 197*, 3574–3583. *(p. 360)*

Li, J. and D. Xiu (2009). A generalized polynomial chaos based ensemble Kalman filter. *Journal of Computational Physics 228*, 5454–5469. *(p. 252, 360, 415)*

Li, Z., Y. Kaufman, C. Ichoku, R. Fraser, A. Trishchenkp, L. Giglio, J. Jin, and X. Yu (2002). *A review of AVHRR-based active fire detection algorithms: Principles, limitations and recommandations in global and regional vegetation fire monitoring from space: Planning a coordinated international effort*, pp. 199–225. The Hague, Netherlands: SPB Academic Publishing. *(p. 34, 38)*

Li, Z., S. Nadon, J. Chilar, and B. Stocks (2000). Satellite mapping of canadian boreal forest fires: Evaluation and comparison of algorithms. *International Journal of Remote Sensing 21*, 3071–3082. *(p. 38)*

Linn, R. (1997). *A transport model for prediction of wildfire behavior*. Ph. D. thesis, New Mexico State University. *(p. 20)*

Linn, R., J. Reisner, J. Colman, and J. Winterkamp (2002). Studying wildfire behavior using FIRETEC. *International Journal of Wildland Fire 11*, 233–246. *(p. 17, 20, 31, 105, 447)*

Linn, R., J. Winterkamp, J. Colman, and C. Edminster (2005). Modeling interactions between fire and atmosphere in discrete element fuel beds. *International Journal of Wildland Fire 14*, 37–48. *(p. 20, 21, 45)*

Liu, C., Q. Xiao, and B. Wang (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review 136*, 3363–3373. *(p. 210)*

Liu, J. and R. Chen (1998). Sequential Monte Carlo methods for dynamical systems. *Journal of American Statistical Association 93*, 1032–1044. *(p. 260)*

Liu, Y., J. Stanturf, and S. Goodrick (2010). Trends in global wildfire potential in a changing climate. *Forest Ecology and Management 259*(4), 685–697. *(p. 5)*

Lopez, S., J. San-Miguel-Ayanz, and R. Burgan (2002). Integration of satellite sensor data, fuel type maps, and meteorological observations for the evaluation of forest fire risk at the pan-European scale. *International Journal of Remote Sensing 23*(13), 2713–2719. *(p. 24)*

Lorenc, A. (1981). A global three-dimensional multivariate statistical interpolation scheme. *Quarterly Journal of the Royal Meteorological Society 109*, 701–721. *(p. 203)*

Lorenc, A. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society 112*(474), 1177–1194. *(p. 202, 206, 214, 220, 223, 242)*

Lucas, C., K. Hennessy, G. Mills, and J. Bathols (2007). Bushfire weather in Southeast Australia: Recent trends and projected climate change impacts. Technical report, Bushfire CRC Consultancy Report to the Climate Institute. *(p. 5)*

Lucor, D., J. Meyers, and P. Sagaut (2007). Sensitivity analysis of les to subgrid-scale model parametric uncertainty using polynomial chaos. *Journal of Fluid Mechanics 585*, 255–279. *(p. 199, 360)*

Maas, U. and S. Pope (1992). Simplying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combustion and Flame 88*, 239–264. *(p. 83)*

Magnussen, B. (2005). The eddy dissipation concept - A bridge between science and technology. In *ECCOMAS Thematic Conference on Computational Combustion, Lisbon, June 21-24*. *(p. 18, 90)*

Maheu, N., V. Moureau, P. Domingo, P. Duchaine, and G. Balarac (2012). Large-eddy simulations of flow and heat transfer around a low-mach number turbine blade. In *Proceedings of the 2012 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*. *(p. 136, 441)*

Mallet, V., D. Keyes, and F. Fendell (2009). Modeling wildland fire propagation with level-set methods. *Computers and Mathematics with Applications 57*(7), 1089–1101. *(p. 22, 293, 482)*

Mandel, J., J. Beezley, and A. Kochanski (2011). Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geoscientific Model Development 4*, 591–610. *(p. 26, 31, 269, 293, 482)*

Mandel, J., J. Beezley, A. Kochanski, V. Kondratenko, and M. Kim (2012). Assimilation of perimeter data and coupling with fuel moisture in a wildland fire–atmosphere DDDAS. *Procedia Computer Science 9*, 1100–1109. *(p. 269)*

Mandel, J., L. Bennethum, J. Beezley, J. Coen, C. Douglas, M. Kim, and A. Vodacek (2008, December). A wildland fire model with data assimilation. *Mathematics and Computers in Simulation 79*, 584–606. *(p. 269, 270, 275, 346)*

Mandin, P. and J.-M. Most (2000). Characterization of the puffing phenomenon on a pool fire. *Fire Safety Science 6*, 1137–1148. *(p. 134)*

Marle, C. (1982). On macroscopic quantities governing multiphase flow with diffusion and chemical reactions in porous media. *International Journal of Engineering Science 20*(5), 643–662. *(p. 448)*

Massart, S. (2003). *Méthodologies de l'assimilation de données. Application à la chimie atmosphérique et à la mécanique des structures*. Ph. D. thesis, Institut National Polytechnique de Toulouse, France. *(p. 203, 205, 237)*

Massart, S., C. Clerbaux, D. Cariolle, A. Piacentini, S. Turquety, and J. Hadji-Lazaro (2009). First steps towards the assimilation of IASI ozone data into the MOCAGE-PALM sytem. *Atmospheric Chemistry and Physics 9*, 5073–5091. *(p. 440)*

Maybeck, P. (1979). *Stochastic models, estimation and control*. New York: Academic Press. *(p. 223, 237)*

Mcallister, S. and M. Finney (2014). Convection ignition of live forest fuels. In *Proceedings of the Intl. Association of Fire Safety Science*. *(p. 105)*

McArthur, A. (1966). Weather and grassland fire behaviour. Technical report, Forest Research Institute, Department of National Development, Forest and Timber Bureau, CSIRO, Canberra. *(p. 7)*

Mell, W., J. Charney, M. Jenkins, N. Cheney, and J. Gould (2005). Numerical simulations of grassland fire behavior from the LANL-FIRETEC and NIST-WFDS models. In *Proceedings of the EastFIRE Conference, May 11-13, Fairfax, VA*. *(p. 20, 21, 22, 25)*

Mell, W., M. Jenkins, J. Gould, and P. Cheney (2007). A physics-based approach to modelling grassland fires. *International Journal of Wildland Fire 16*, 1–22. *(p. 20, 31)*

Melling, A. (1997). Tracer particles and seeding for particle image velocimetry. *Measurement Science and Technology 8*, 1406–1416. *(p. 150, 467)*

Menon, S. and W. Jou (1991). Large eddy simulations of combustion instability in an axisymmetric ramjet combustor. *Combustion Science and Technology 75*, 53–72. *(p. 77)*

Mercier, R., P. Auzillon, V. Moureau, N. Darabiha, O. Gicquel, D. Veynante, and B. Fiorina (2013). Modeling flame stabilization by heat losses using filtered tabulated chemistry for LES. In *International Symposium on Turbulence and Shear Flow Phenomena (TSFP-8), August 28-30th, Poitiers, France*. *(p. 84, 91)*

Merlet, N. (2008). Evaluation des possibilités de géolocalisation automatisée et temps-réel de prises de vue aéroportées dans le cadre de la lutte contre les feux de forêt. Technical report, Université de Toulouse - Service Départmental d'Incendie et de Secours des Pyrénées-Orientales (SDIS 66), France. *(p. 41, 280)*

Michel, Y. (2013a). Estimating deformations of random processes for correlation modelling in a limited area model. *Quarterly Journal of the Royal Meteorological Society 139*(671), 534–547. *(p. 247)*

Michel, Y. (2013b). Estimating deformations of random processes for correlation modelling: Methodology and the one-dimensional case. *Quarterly Journal of the Royal Meteorological Society 139*(672), 771–783. *(p. 247)*

Miller, R., E. Carter, and S. Blue (1999). Data assimilation into nonlinear stochastic models. *Tellus 51A*, 167–194. *(p. 248)*

Milly, P., R. Wetherald, K. Dunne, and T. Delwort (2002). Increasing risk of great floods in a changing climate. *Nature 415*, 514–517. *(p. 5)*

Miranda, A., J. Ferreira, J. Valente, P. Santos, J. Amorim, and C. Borrego (2005). Smoke measurements during Gestosa-2002 experimental field fires. *International Journal of Wildland Fire 14*, 107–116. *(p. 36, 46)*

Miranda, A., A. Monteiro, V. Martins, A. Carvalho, M. Schaap, P. Builtjes, and C. Borrego (2008). Forest fire impact on air quality over Portugal. In C. Borrego and A. Miranda (Eds.), *Air Pollution Modeling and Its Application*, pp. 190–198.

Netherlands: NATO Science for Peace and Security Series C: Environmental Security, Springer. *(p. 3, 36)*

Mirouze, I. (2010). *Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques*. Ph. D. thesis, Université de Toulouse, France. *(p. 205, 223)*

Mitchell, H. and P. Houtekamer (2000). An adaptive ensemble kalman filter. *Monthly Weather Review 128*, 416–433. *(p. 257)*

Modest, M. (2003). *Radiative heat transfer*. Academic Press. *(p. 93)*

Moireau, P. (2008). *Assimilation de données par filtrage pour les systèmes hyperboliques du second ordre. Applications à la mécanique cardiaque.* Ph. D. thesis, Ecole Polytechnique, France. *(p. 205)*

Moireau, P. and D. Chapelle (2011, March). Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems. *Control, Optimisation and Calculus of Variations 17*(2), 380–405. *(p. 255)*

Montmerle, T. and L. Berre (2010). Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quarterly Journal of the Royal Meteorological Society 136*, 1408–1420. *(p. 247)*

Moradkhani, H. and K.-L. Hsu (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research 41*, W05012. *(p. 258)*

Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources 28*, 135–147. *(p. 205, 276, 307, 309, 333, 413)*

Morvan, D. (2007). A numerical study of flame geometry and potential for crown fire initiation for a wildfire propagating through shrub fuel. *International Journal of Wildland Fire 16*, 511–518. *(p. 18, 19)*

Morvan, D. (2011). Physical phenomena and length-scales governing the behaviour of wildfires: A case for physical modelling. *Fire Technology 47*, 437–460. *(p. 46)*

Morvan, D. and J. Dupuy (2001). Modeling of fire spread through a forest fuel bed using a multiphase formulation. *Combustion and Flame 127*, 1981–1994. *(p. 18)*

Morvan, D. and J. Dupuy (2004). Modeling the propagation of a wildfire through a Mediterranean shrub using a multiphase formulation. *Combustion and Flame 138*, 199–210. *(p. 17, 18, 19, 20, 31, 105, 106, 110, 116, 118, 447, 454)*

Morvan, D., J. Dupuy, B. Porterie, and M. Larini (2000). Multiphase formulation applied to the modeling of fire spread through a forest fuel bed. *Proceedings of the Combustion Institute 28*, 2803–2809. *(p. 110)*

Morvan, D. and M. Larini (2001). Modeling of one-dimensional fire spread in pine needles with opposing air flow. *Combustion Science and Technology 164*, 37–64. *(p. 19, 113)*

Morvan, D., S. Méradji, and G. Accary (2009). Physical modelling of fire spread in grasslands. *Fire Safety Journal 44*, 50–61. *(p. 19)*

Moureau, V., B. Fiorina, and H. Pitsch (2009). A level set formulation for premixed combustion LES considering the turbulent flame structure. *Combustion and Flame 156*, 801–812. *(p. 90)*

Moureau, V., G. Lartigue, Y. Sommerer, C. Angelberger, O. Colin, and T. Poinsot (2005). Numerical methods for unsteady compressible multi-component reacting flows on fixed and moving grids. *Journal of Computational Physics 202*, 710–736. *(p. 78, 93)*

Neviackas, A. and A. Trouvé (2007). Sensor-driven inverse zone modeling of enclosure fire dynamics. In *SFPE Professional Development Conference and Expo*, Las Vegas, NV, USA. *(p. 268, 269)*

Nicolette, V., S. Tieszen, A. Black, S. Domino, and T. O'Hern (2005). A turbulence model for buoyant flows based on vorticity generation. Technical Report SAND2005-6273, SANDIA National Laboratories. *(p. 135)*

Nicoud, F., H. Toda, O. Cabrit, S. Bose, and J. Lee (2011). Using singular values to build a subgrid-scale model for large eddy simulations. *Physics of Fluids 23*, 085106. *(p. 80)*

Nijhuis, M. (2012). Forest fires: Burnt out. *Nature 489*, 352–354. *(p. 4, 5)*

Noble, I., G. Bary, and A. Gill (1980). McArthur's fire danger meters expressed as equations. *Australian Journal of Ecology 5*, 201–203. *(p. 7)*

Noonan-Wright, E., T. Opperman, M. Finney, G. Zimmerman, R. Seli, L. Elenz, D. Calkin, and J. Fiedler (2011). Developing the US wildland fire decision support system. *Journal of Combustion 2011*, ID168473. *(p. 1, 6)*

Novozhilov, V., B. Moghtaderi, D. Fletcher, and J. Kent (1996). Computational fluid dynamics modeling of wood combustion. *Fire Safety Journal 27*, 69–84. *(p. 105)*

Oliver, D. and Y. Chen (2011). Recent progress on reservoir history matching: A review. *Computers and Geosciences 15*, 185–221. *(p. 205)*

Orfao, J., F. Antunes, and J. Figueiredo (1999). Pyrolysis kinetics of lignocellulosic materials - three independent reactions model. *Fuel 78*, 349–358. *(p. 104)*

Orlande, H., M. Colaço, G. Dulikravich, F. Vianna, W. Da Silva, H. Fonseca, and O. Fudym (2012). State estimation problems in heat transfer. *International Journal for Uncertainty Quantification 2*, 239–258. *(p. 205, 260, 262)*

Orszag, S. (1970). Analytical theories of turbulence. *Journal of Fluid Mechanics 41*, 363–386. *(p. 76)*

Osher, S. and J. Sethian (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics 79*, 12–49. *(p. 293)*

Page, S., F. Siegert, J. Rieley, H. Boehm, A. Jaya, and S. Limin (2002). The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature 420*, 61–63. *(p. 12)*

Pagni, P. and T. Peterson (1973). Flame spread through porous fuels. *Proceedings of the Combustion Institute 14*(1), 1099–1107. *(p. 18)*

Palmer, T. and J. Räisänen (2002). Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature 415*, 512–514. *(p. 5)*

Pannekoucke, O., L. Berre, and G. Desroziers (2008). Background error correlation length-scale estimates and their sampling statistics. *Quarterly Journal of the Royal Meteorological Society 134*, 497–511. *(p. 216, 247, 383)*

Parrish, D. and J. Derber (1992). The national meteorological center's spectral statistical interpolation analysis system. *Monthly Weather Review 120*, 1747–1763. *(p. 204)*

Pastor, E., L. Zarate, E. Planas, and J. Arnaldos (2003). Mathematical models and calculation systems for the study of wildland fire behaviour. *Progress in Energy and Combustion Science 29*(2), 139–153. *(p. 17)*

Paugam, R., M. Wooster, and G. Roberts (2013). Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread. *Geoscience and Remote Sensing 51*, 3385–3399. *(p. 34, 37, 283)*

Pecquery, F. (2013). *Développement d'un modèle numérique de prédiction des émissions d'oxydes d'azote pour la simulation aux grandes échelles de chambres de combustion aéronautiques*. Ph. D. thesis, Institut National des Sciences Appliquées, Rouen, France. *(p. 84)*

Pérez-Ramirez, Y., P.-A. Santoni, N. Darabiha, V. Leroy-Cancellieri, and E. Leoni (2012). A global kinetic model for the combustion of the evolved gases in wildland fires. *Combustion Science and Technology 184*(9), 1380–1394. *(p. 46, 63, 121, 459, 460, 461)*

Pérez-Ramirez, Y., V. Tihay, and P.-A. Santoni (2010). Etude des gaz de pyrolyse, de leur composition et des émissions atmosphériques. Technical report, CNRS-SPE, Université de Corte, France. *(p. 118, 119)*

Perry, G. (1998). Current approaches to modelling the spread of wildland fire: A review. *Progress in Physical Geography 22*(2), 222–245. *(p. 17)*

Peters, B. and C. Bruch (2003). Drying and pyrolysis of wood particles: Experiments and simulation. *Journal of Analytical and Applied Pyrolysis 70*, 233–250. *(p. 100)*

Peters, N. (1984). Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in Energy and Combustion Science 10*, 319–339. *(p. 65, 84)*

Peters, N. (1985). Numerical and asymptotic analysis of systematically reduced reaction schemes for hydrocarbon flames. In R. Glowinsky, B. Larrouturou, and R. Temam (Eds.), *Numerical simulation of combustion phenomena*, pp. 90–109. Berlin: Springer-Verlag. *(p. 82, 83)*

Peyrounette, M. (2013). Participation au développement d'un outil numérique de la prévision de la pression artérielle. Technical report, Université de Pau et des Pays de l'Adour - Institut de Mathématiques et de Modélisation de Montpellier. *(p. 410)*

Pham, D., J. Verron, and C. Roubaud (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems 16*, 323–340. *(p. 204)*

Piacentini, A., T. Morel, A. Thévenin, and F. Duchaine (2011). Open-palm: an open source dynamic parallel coupler. In *IV International Conference on Computational Methods for Coupled Problems in Science and Engineering*. *(p. 440, 441)*

Pitsch, H. (2005). A consistent level set formulation for large eddy simulation of premixed turbulent combustion. *Combustion and Flame 143*(4), 587–698. *(p. 90)*

Pitsch, H. (2006). Large-eddy simulation of turbulent combustion. *Annual Review of Fluid Mechanics 38*, 453–482. *(p. 79, 90, 92)*

Pitt, M. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599. *(p. 262)*

Poinsot, T. (1996). Using direct numerical simulation to understand turbulent premixed combustion. *Proceedings of the Combustion Institute 26*, 219–232. *(p. 76)*

Poinsot, T. and S. Lele (1992). Boundary-conditions for direct simulations of compressible viscous flows. *Journal of Computational Physics 101*(1), 104–129. *(p. 93, 158)*

Poinsot, T. and D. Veynante (2005). *Theoretical and numerical combustion* (2nd ed.). R.T. Edwards. *(p. 69, 71, 77, 79, 126, 295)*

Poitou, D. (2009). *Modélisation du rayonnement dans la simulation aux grandes échelles de la combustion turbulente*. Ph. D. thesis, Université de Toulouse - Institut National Polytechnique de Toulouse, France. *(p. 93, 94)*

Poitou, D., J. Amaya, M. El Hafi, and B. Cuenot (2012). Analysis of the interaction between turbulent combustion and thermal radiation using unsteady coupled LES/DOM simulations. *Combustion and Flame 159*, 1605–1618. *(p. 95, 140, 441)*

Poitou, D., M. El Hafi, and B. Cuenot (2011). Analysis of radiation modeling for turbulent combustion: development of a methodology to couple turbulent combustion and radiative heat transfer in LES. *Journal of Heat Transfer 133*(6), 062701. *(p. 94, 441)*

Poludnenko, A. and E. Oran (2010). The interaction of high-speed turbulence with flames: Global properties and internal flame structure. *Combustion and Flame 157*, 995–1011. *(p. 76)*

Poludnenko, A. and E. Oran (2011). The interaction of high-speed turbulence with flames: Turbulent flame speed. *Combustion and Flame 158*, 301–326. *(p. 76)*

Pope, S. (2000). *Turbulent flows*. Cambridge University Press. *(p. 80, 89)*

Porterie, B., J. Consalvi, A. Kaiss, and J. Loraud (2005). Predicting wildland fire behavior and emissions using a fine-scale physical model. *Numerical Heat Transfer: Part A: Applications 47*(6), 571–591. *(p. 17, 18, 19)*

Prins, E., Y. Govaerts, and C. Justice (2004). Report on the joint gofc/gold fire and ceos lpv working group workshop on global geostationary fire monitoring applications. Technical Report Report No. 19, GOFC/GOLD, EUMETSAT, Darmstadt, Germany. *(p. 40)*

Prins, E. and W. Menzel (1994). Trends in south american burning detected with the goes vas from 1983-1991. *Journal of Geophysical Research 99*(D8), 16719–16735. *(p. 40)*

Pyne, S. (2001). *Year of the fires: The story of the Great Fires of 1910*. New York: Viking. *(p. 15)*

Rabier, F. (2005). Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society 131*, 3215–3233. *(p. 191, 205, 242)*

Raffel, M., C. Willert, and J. Kompenhans (1998). *Particle image velocimetry*. Berlin, Heidelberg: Springer. *(p. 150, 467, 468)*

Ravela, S., K. Emmanuel, and D. McLaughlin (2007). Data assimilation by field alignment. *Physica D: Nonlinear Phenomena 230*(1-2), 127–145. *(p. 271)*

Refloch, A., B. Courbet, A. Murrone, P. Villedieu, C. Laurent, P. Gilbank, J. Troyes, L. Tessé, G. Chaineray, J. Dargaud, E. Quémerais, and F. Vuillot (2011). CFD platforms and coupling - CEDRE software. Technical report, The ONERA Journal Aerospace Lab. *(p. 442)*

Rehm, R. and R. McDermott (2009). Fire front propagation using the level set method. Technical Report March, National Institute of Standards and Technology. *(p. 22, 293, 294, 482, 483, 484)*

Reichle, R. (2008). Data assimilation methods in the Earth sciences. *Advances in Water Resources 31*, 1411–1418. *(p. 207, 237, 243)*

Reichle, R., J. Walker, R. Koster, and P. Houser (2002). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology 3*, 728–740. *(p. 249)*

Rein, G., S. Cohen, and A. Simeoni (2008). Carbon emissions from smouldering peat in shallow and strong fronts. *Proceedings of the Combustion Institute 32*, 2489–2496. *(p. 12)*

Repellin, V. (2006). *Optimisation des paramètres durée et température d'un traitement thermique du bois. Modifications des propriétés d'usage du bois en relation avec les modifications physico-chimiques et ultrastructurales occasionnées par le traitement thermique*. Ph. D. thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, France. *(p. 107)*

Reuss, D., R. Adrian, and C. Landreth (1989). Two-dimensional velocity measurements in a laminar flame using particle image velocimetry. *Combustion Science and Technology 67*, 73–77. *(p. 150)*

Revel, J., J. Boettner, M. Cathonnet, and J. Bachman (1994). Derivation of a global chemical kinetic mechanism for methane ignition and combustion. *Journal de Chimie Physique 91*, 365–382. *(p. 460)*

Riber, E., V. Moureau, M. Garcia, T. Poinsot, and O. Simonin (2009). Evaluation of numerical strategies for LES of particulate to-phase recirculating flows. *Journal of Computational Physics 228*(2), 539–564. *(p. 79, 93)*

Ricci, S. (2004). *Assimilation variationnelle océanique: Modélisation multivariée de la matrice de covariance d'erreur d'ébauche*. Ph. D. thesis, Université de Toulouse, France. *(p. 203, 205, 216)*

Ricci, S., A. Piacentini, O. Thual, E. Le Pape, and G. Jonville (2011). Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting. *Hydrology and Earth System Sciences 15*, 3555–3575. *(p. 205)*

Richards, G. (1995). A general mathematical framework for modelling two-dimensional wildland fire spread. *International Journal of Wildland Fire 5*, 63–72. *(p. 24)*

Richards, R., B. Munk, and O. Plumb (1997). Fire detection, location and heat release rate through inverse problem solution. Part 1. Theory. *Fire Safety Journal 28*(4), 323–350. *(p. 268)*

Riggan, P. and J. Hoffman (2003). FireMapper: A thermal-imaging radiometer for wildfire research and operations. In *Proceedings of the IEEE Aerospace Conference, Paper no. 1522*, Big Sky, Montana. *(p. 41)*

Riggan, P., R. Lockwood, R. Tissell, J. Brass, J. Pereira, H. Miranda, A. Miranda, T. Campos, and R. Higgins (2004). Remote measurement of wildfire energy and carbon flux from wildfires in Brazil. *Ecological Applications 14*(3), 855–872. *(p. 35, 41)*

Riggan, P. and G. Robert (2009). Airborne remote sensing of wildland fires. In A. Bytnerowicz, M. Arbaugh, C. Andersen, and A. Riebau (Eds.), *Wildland Fires and Air Pollution. Developments in Environmental Science 8*, pp. 139–168. Amsterdam, The Netherlands: Elsevier. *(p. 32, 33, 41, 42)*

Rios, O. (2013). Forecasting wind-driven wildfires using an inverse modelling approach. Technical report, University of Edinburgh, UK. *(p. 274)*

Ristic, B., S. Arulampalam, and N. Gordon (2004). *Beyond the Kalman filter*. Boston: Artech House. *(p. 257, 263)*

Roberts, G. and M. Wooster (2008, April). Fire detection and fire characterization over Africa using Meteosat SEVIRI. *IEEE Transactions on Geoscience and Remote Sensing 46*(4), 1200–1218. *(p. 35, 37)*

Robinson, J. (1991). Fire from space: Global evaluation using infrared remote sensing. *International Journal of Remote Sensing 12*, 3–24. *(p. 32, 35)*

Rochoux, M., B. Cuenot, S. Ricci, A. Trouvé, B. Delmotte, S. Massart, R. Paoli, and R. Paugam (2013). Data assimilation applied to combustion. *Comptes Rendus Mécanique 341*, 266–276. *(p. 52, 320, 332, 403)*

Rochoux, M., B. Delmotte, B. Cuenot, S. Ricci, and A. Trouvé (2013). Regional-scale simulations of wildland fire spread informed by real-time flame front observations. *Proceedings of the Combustion Institute 34*, 2641–2647. *(p. 52, 320, 332, 403)*

Rochoux, M., C. Emery, S. Ricci, B. Cuenot, and A. Trouvé (2014a). Ensemble-based data assimilation for regional-scale simulations of wildland fire spread. *Proceedings of the Combustion Institute*. (submitted for publication). *(p. 52, 378, 404, 405)*

Rochoux, M., C. Emery, S. Ricci, B. Cuenot, and A. Trouvé (2014b). Towards predictive data-driven simulations of wildfire spread. Part II: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread. *Natural Hazards and Earth System Sciences, Special Issue: numerical wildland combustion, from the flame to the atmosphere*. (submitted for publication). *(p. 52, 378, 405)*

Rochoux, M., C. Emery, S. Ricci, B. Cuenot, and A. Trouvé (2014c). Towards predictive simulation of wildfire spread at regional-scale using ensemble-based data assimilation to correct the fire front position. In *Proceedings of the Intl. Association of Fire Safety Science*. (accepted for publication). *(p. 52, 378, 404)*

Rochoux, M., S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé (2012). Towards predictive simulation of wildfire spread using a reduced-cost ensemble Kalman filter based on polynomial chaos approximation. In *Proceedings of the 2012 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*, pp. 199–208. *(p. 52, 333, 360, 365, 404, 405)*

Rochoux, M., S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé (2014). Towards predictive data-driven simulations of wildfire spread. Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation. *Natural Hazards and Earth System Sciences, Special Issue: numerical wildland combustion, from the flame to the atmosphere*. (accepted for discussion). *(p. 51, 333, 405)*

Rochoux, M., S. Ricci, S. Massart, R. Paoli, B. Cuenot, and A. Trouvé (2010). Preliminary investigation of data assimilation methodologies for forest fire propagation. Technical Report TR-CMGC-10-99, CERFACS - University of Maryland. *(p. 49, 288, 295, 487, 491)*

Rodgers, C. (2000). *Inverse methods for atmospheres: theories and practice*. Singapore: World Scientific Publ. *(p. 241)*

Ros, D. and M. Borga (1997). Adaptive use of a conceptual model for real time flood forecasting. *Nordic Hydrology 28*, 169–188. *(p. 245)*

Rosic, B., A. Kucerov, J. Sykora, O. Pajonk, A. Litvinenko, and H. Matthies (2013). Parameter identification in a probabilistic setting. *Engineering Structures 50*, 179–196. *(p. 360, 415)*

Rothermel, R. (1972). A mathematical model for predicting fire spread in wildland fuels. Technical report, U.S. Department of Agriculture, Forest Service, Inter-

mountain forest and range experiment station, Ogden, Utah, USA. *(p. 15, 23, 24, 34, 97, 103, 104, 275, 289, 290, 320, 417, 419, 420, 421, 432, 433)*

Rothermel, R. (1983). How to predict the spread and intensity of forest and range fires. Technical Report General Technical Report, INT-143, U.S. Department of Agriculture, Forest Service. *(p. 23)*

Rubinstein, R. and D. Kroese (2008). *Simulation and the Monte Carlo method* (2nd ed.). Wiley Series in Probability and Statistics. Wiley. *(p. 359)*

Saad, G. (2007). *Stochastic data assimilation with application to multi-phase flow and health monitoring problems*. Ph. D. thesis, University of Southern California, USA. *(p. 360)*

Said, N., H. Mhiri, H. Bournot, and G. Le Palec (2008). Experimental and numerical modelling of the three-dimensional incompressible flow behavior in the near wake of circular cylinders. *Journal of Wind Engineering and Industrial Aerodynamics 96*(5), 471–502. *(p. 150)*

San-Miguel-Ayanz, J., N. Ravail, V. Kelha, and A. Ollero (2005). Active fire detection for fire emergency management: Potential and limitations for the operational use of remote sensing. *Natural Hazards 35*, 361–376. *(p. 29, 38)*

Santoni, P.-A., J.-B. Filippi, J.-H. Balbi, and F. Bosseur (2011). Wildland fire behaviour case studies and fuel models for landscape-scale fire modeling. *Journal of Combustion 2011*, ID613424. *(p. 27, 28)*

Santoni, P.-A., F. Morandini, and T. Barboni (2010). Steady and unsteady fireline intensity of spreading fires at laboratory scale. *The Open Thermodynamics Journal 4*, 212–219. *(p. 34, 146, 147, 149)*

Sarti, F., H. Yesou, R. Andreoli, and S. Clandillon (2005). Exploitation of Proba Chris/HRC data in the context of the International Charter Space and Major Disasters. In *3rd ESA CHRIS/Proba Workshop, 21-23 March, ESRIN*, Frascati, Italy. ESA SP-593. *(p. 40)*

Sasaki, Y. (1958). An objective analysis based on the variational method. *Journal of the Meteorological Society of Japan II-36*, 77–88. *(p. 204)*

Sasaki, Y. (1970). Some basic formalisms in numerical variational analysis. *Monthly Weather Review 98*, 875–883. *(p. 204)*

Schmitt, P., T. Poinsot, B. Schuermans, and K. Geigle (2007). Large-eddy simulation and experimental study of heat transfer, nitric oxide emissions and combustion instability in a swirled turbulent high pressure burner. *Journal of Fluid Mechanics 570*, 17–46. *(p. 79, 91, 93)*

Schönfeld, T. and M. Rudgyard (1999). Steady and unsteady flows simulations using the hybrid flow solver AVBP. *AIAA Journal 37*(11), 1378–1385. *(p. 93)*

Sedano, F., P. Kempeneers, J. San Miguel, P. Strobl, and P. Vogt (2013). Towards a pan-european burnt scar mapping methodology based on single data medium resolution optical remote sensing data. *20*, 52–59. International Journal of Applied Earth Observation and Geoinformation (in press). *(p. 39)*

Seiler, W. and P. Crutzen (1980). Estimates of gross and net fluxes of carbon between the biosphere and atmosphere from biomass burning. *Climatic Change 2*, 207–247. *(p. 36)*

Selle, L., L. Benoit, T. Poinsot, F. Nicoud, and W. Krebs (2006). Joint use of compressible large-eddy simulation and helmholtz solvers for the analysis of rotating modes in an industrial swirled burner. *Combustion and Flame 145*(1-2), 194–205. *(p. 79)*

Selle, L., G. Lartigue, T. Poinsot, R. Koch, U. Schildmacher, W. Krebs, B. Prade, P. Kaufmann, and D. Veynante (2004). Compressible large eddy simulation of turbulent combustion in complex geometry on unstructured meshes. *Combustion and Flame 137*(4), 489–505. *(p. 78, 82, 91, 93)*

Séro-Guillaume, O. and J. Margerit (2002). Modeling forest fires. Part I: A complete set of equations derived by extended irreversible thermodynamics. *International Journal of Heat and Mass Transfer 45*, 1705–1722. *(p. 18, 105, 110, 447, 448, 450, 451, 454)*

Sethian, J. (1999). *Level set methods and fast marching methods*. Cambridge University Press. *(p. 293)*

Shafizadeh, F. (1982). Introduction to pyrolysis of biomass. *Journal of Analytical and Applied Pyrolysis 3*(4), 283–305. *(p. 100, 450)*

Shafizadeh, F. and P. Chin (1977). *Thermal deterioration of wood*, Volume 23 of *ACS Symposium Series*, Chapter Wood Technology: Chemical aspects, pp. 57–81. ASC Washington. *(p. 100)*

Shafizadeh, F. and G. McGinnis (1971). Chemical composition and thermal analysis of cottonwood. *Carbohydrate Research 16*, 273–277. *(p. 114)*

Sharples, J. (2008). Review of formal methodologies for wind-slope correction of wildfire rate of spread. *International Journal of Wildland Fire 17*, 179–193. *(p. 492)*

Siau, J. (1984). *Transport processes in wood*, Volume 2. Springer Berlin Heidelberg. *(p. 102)*

Simmie, J. (2003). Detailed chemical kinetic models for the combustion of hydrocarbon fuels. *Progress in Energy and Combustion Science 29*(6), 599–634. *(p. 81)*

Smagorinsky, J. (1963). General circulation experiments with the primitive equations. I. The basic experiment. *Monthly Weather Review 91*, 99–164. *(p. 80)*

Spalding, D. (1971). Mixing and chemical reaction in steady confined turbulent flame. *Proceedings of the Combustion Institute 12*(1), 649–657. *(p. 89)*

Staffelbach, G., L. Gicquel, G. Boudier, and T. Poinsot (2009). Large eddy simulation of self excited azimutahl modes in annular combustors. *Proceedings of the Combustion Institute 32*, 2909–2916. *(p. 91)*

Stern, F., R. Wilson, H. Coleman, and E. Paterson (2001). Comprehensive approach to verification and validation of cfd simulations - part 1: Methodologies and procedures. *Journal of Fluids Engineering 123*(803-810). *(p. 360)*

Strada, S., C. Mari, J.-B. Filippi, and F. Bosseur (2012, May). Wildfire and the atmosphere: Modelling the chemical and dynamic interactions at the regional scale. *Atmospheric Environment 51*, 234–249. *(p. 4, 27, 31, 36, 47, 414)*

Sullivan, A. (2009a). Wildland surface fire spread modeling, 1990-2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire 18*, 349–368. *(p. 18, 20)*

Sullivan, A. (2009b). Wildland surface fire spread modeling, 1990-2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire 18*, 369–386. *(p. 23)*

Sun, L., X. Zhou, S. Mahalingam, and D. Weise (2005). Experimental investigation of the velocity field in buoyant diffusion flames using piv and tpiv algorithm. *Fire Safety Science 8*, 939–950. *(p. 150)*

Suzuki, T. (2012). Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and diect numerical simulation. *Journal of Fluid Mechanics 709*, 249–288. *(p. 205)*

Szunyogh, I., E. Kostelich, G. Gyarmati, E. Kalnay, B. Hunt, E. Ott, E. Satterfield, and J. Yorke (2008). A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A 60*, 113–130. *(p. 360)*

Taine, J. and J.-P. Petit (1993). *Heat transfer*. Prentice Hall. *(p. 93)*

Takeno, T. and M. Nishioka (1993). Species conservations and emission indices for flames described by similarity solutions. *Combustion and Flame 92*, 465–468. *(p. 166)*

Talagrand, O. (1997). Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan 75*, 191–209. *(p. 198, 208, 237)*

Talagrand, O. and P. Courtier (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation, part1: Theory. *Quarterly Journal of the Royal Meteorological Society 113*, 1311–1328. *(p. 230, 242)*

Tang, W. (1967). Effect of inorganic salts on pyrolysis of wood, alpha-cellullose and lignin. Technical Report FPL Research Paper no. 71, US Department of Agriculture, Forest Products Laboratory. *(p. 114)*

Tansey, K., J. Grégoire, P. Defourny, R. Leigh, J. Peckel, E. Bogaert, and J. Bartholome (2008). A new global multi-annual (2000-2007) burnt area product at 1 km resolution. *Geophysical Research Letters 35*, L01401. *(p. 2)*

Tarantola, A. (1987). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics. *(p. 198, 221, 276)*

Teague, B., R. McLeod, and S. Pascoe (2010). 2009 victorian bushfires royal commission final report. Technical report, Parliament of Victoria. *(p. 4)*

Thirel, G., E. Martin, J.-F. Mahfouf, S. Massart, S. Ricci, and F. Habets (2010). A past discharges assimilation system for ensemble streamflow forecasts over France. *Hydrology and Earth System Sciences 14*, 1623–1637. *(p. 245)*

Tieszen, S. (2001). On the fluid mechanics of fires. *Annual Review of Fluid Mechanics 33*, 67–92. *(p. 68, 133)*

Tieszen, S., H. Pistch, G. Blanquart, and S. Abarzhi (2004). Toward the development of a les-sgs closure model for buoyant plumes. In *Proceedings of the 2004 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*. *(p. 135)*

Tihay, V. and P. Gillard (2010). Pyrolysis gases released during the thermal decomposition of three mediterranean species. *Journal of Analytical and Applied Pyrolysis 88*, 168–174. *(p. 118)*

Tihay, V., P.-A. Santoni, A. Simeoni, J.-P. Garo, and J.-P. Vantelon (2009). Skeletal and global mechanisms for the combustion of gases released by crushed forest fuels. *Combustion and Flame 156*, 1565–1575. *(p. 104, 106, 109)*

Tikhonov, A. and V. Arsenin (1977). *Solution of ill-posed problems*. Washington: Winston and Sons. *(p. 198)*

Tikhonov, A. and A. Leonov (1998). *Nonlinear ill-posed problems*. Chapman and Hall. *(p. 198)*

Tippett, M., J. Anderson, C. Bishop, T. Hamill, and J. Whitaker (2003). Ensemble square root filters. *Monthly Weather Review 131*, 1485–1490. *(p. 252)*

Todling, R. and S. Cohn (1994). Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Monthly Weather Review 122*(11), 2530–2557. *(p. 221, 243)*

Toro, E. (1999). *Reimann solvers and numerical methods for fluid dynamics: A practical introduction* (Second ed.). Springer. *(p. 484)*

Trémolet, Y. (2007a). Incremental 4D-Var convergence study. *Tellus 59A*, 706–718. *(p. 205)*

Trémolet, Y. (2007b). Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society 133*, 1267–1280. *(p. 226)*

Trick, K., T. Saliba, and S. Sandhu (1997). A kinetic model of the pyrolysis of phenolic resin in a carbon/phenolic composite. *Carbon 35*(3), 393–401. *(p. 115, 455)*

Trouvé, A., D. Veynante, K. Bray, and T. Mantel (1994). The coupling between flame surface dynamics and species mass conservation in premixed turbulent combustion. In *Proceedings of the 1994 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*, pp. 95–124. *(p. 76)*

Turner, J. and B. Lawson (1978). Weather in the canadian forest fire danger rating system. a user guide to national standards and practices. Technical Report BC-X-177, Environment Canada, Pacific Forest Research Centre, Victoria, BC, Canada. *(p. 7)*

Turns, S. (2000). *An introduction to combustion. Concepts and Applications.* (Second ed.). McGraw-Hill. *(p. 71)*

Urbanski, S. (2013, July). Combustion efficiency and emission factors for wildfire-season fires in mixed conifer forests of the northern Rocky Mountains, US. *Atmospheric Chemistry and Physics 13*(14), 7241–7262. *(p. 36)*

Vagelopoulos, C. and F. Egolfopoulos (1998). Direct experimental determination of laminar flame speeds. In T. C. Institute (Ed.), *27th International Symposium on Combustion*, Volume No. 1, Pittsburgh, pp. 513–519. *(p. 86)*

Valcke, S. (2012). The oasis3 coupler: A european climate modelling community software. *Geoscientific Model Development 5*, 2139–2178. *(p. 442)*

van der Werf, G., J. Randerson, L. Giglio, G. Collatz, M. Mu, P. Kasibhatla, D. Morton, R. DeFries, Y. Jin, and T. van Leeuwen (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009). *Atmospheric Chemistry and Physics 10*, 11707–11735. *(p. 3)*

Van Leeuwen, P. and G. Evensen (1996). Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review 124*(12), 2898–2913. *(p. 252)*

van Oijen, J., F. Lammers, and L. de Goey (2001). Modeling of complex premixed burner systems by using flamelet-generayed manifolds. *Combustion and Flame 127*(3), 2124–2134. *(p. 83)*

Van Wagner, C. (1987). Development and structure of the canadian forest fire weather index system. Technical Report Technical Report 35, Canadian Forestry Service, Ottawa, ON, Canada. *(p. 7, 8)*

Vervisch, L. and T. Poinsot (1998). Direct numerical simulation of non premixed turbulent flames. *Annual Review of Fluid Mechanics 30*, 665–692. *(p. 76)*

Veynante, D., T. Schmitt, M. Boileau, and V. Moureau (2012). Analysis of dynamic models for turbulent premixed combustion. In *Proceedings of the 2012 Summer Program, Center for Turbulence Research, NASA AMES, Stanford University, USA*. *(p. 92)*

Veynante, D. and L. Vervisch (2002). Turbulent combustion modeling. *Progress in Energy and Combustion Science 28*, 193–266. *(p. 69)*

Vico, G. and A. Porporato (2009). Probabilistic description of topographic slope and aspect. *Journal of Geophysical Research 114*, 1–13. *(p. 497)*

Vicquelin, R. (2010). *Tabulation de la cinétique chimique pour la modélisation et la simulation de la combustion turbulente*. Ph. D. thesis, Ecole Centrale Paris, France. *(p. 91)*

Vicquelin, R., B. Fiorina, S. Payet, N. Darabiha, and O. Gicquel (2011). Coupling tabulated chemistry with compressible CFD solvers. *Proceedings of the Combustion Institute 33*(1), 1481–1488. *(p. 91, 93)*

Viegas, D. (1998). Forest Fire Propagation. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences 356*, 2907–2928. *(p. 10, 15)*

Viegas, D. (2004). On the existence of a steady-state regime for slope and wind driven fire. *International Journal of Wildland Fire 13*, 101–117. *(p. 25)*

Viegas, D. (2011). Overview of forest fire propagation research. *Proceedings of the International Association of Fire Safety Science 10*, 95–108. *(p. 10, 24)*

Viegas, D., G. Bovio, A. Ferreira, A. Nosenzo, and B. Sol (1999). Comparative study of various methods of fire danger evaluation in southern europe. *International Journal of Wildland Fire 9*(4), 235–246. *(p. 9)*

Viegas, D. and L. Pita (2004). Fire spread in canyons. *International Journal of Wildland Fire 13*(274), 1–22. *(p. 26)*

Viegas, D. and A. Simeoni (2010). Eruptive behaviour of forest fires. *Fire Technology 47*(2), 303–320. *(p. 15)*

Viskanta, R. ad Mengüç, M. (1987). Radiation heat transfer in combustion systems. *Progress in Energy and Combustion Science 13*(2), 97–160. *(p. 93)*

Wan, E. and R. Van der Merwe (2000). The unscented Kalman filter for non-linear estimation. In *Proceedings of Symposium 2001 on Adaptive Systems for Signal Processing, Communications and Control*, Lake Louise, Canada. *(p. 254, 256)*

Wang, G., M. Boileau, and D. Veynante (2011). Implementation of a dynamic thickened flame model for large eddy simulations of turbulent premixed combustion. *Combustion and Flame 158*(11), 2199–2213. *(p. 92)*

Wang, G., M. Boileau, D. Veynante, and K. Truffin (2012). Large eddy simulation of a growing turbulent premixed flame kernel using a dynamic flame surface density model. *Combustion and Flame. Special Issue on Turbulent Combustion 159*(8), 2742–2754. *(p. 92)*

Weaver, A. and P. Courtier (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society 127*, 1815–1846. *(p. 247)*

Weaver, A., C. Deltel, E. Machu, S. Ricci, and N. Daget (2005). A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society 131*, 3605–3625. *(p. 205, 207)*

Weaver, A. and I. Mirouze (2012). On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Quarterly Journal of the Royal Meteorological Society 139*(670), 242–260. *(p. 383)*

Weber, R. (1991). Toward a comprehensive wildfire spread model. *International Journal of Wildland Fire 4*, 245–248. *(p. 17)*

Weckman, E. and A. Sobiesiak (1988). The oscillatory behavior of medium scale pool fires. *Proceedings of the Combustion Institute 22*(1), 1299–1310. *(p. 134)*

Weise, D. and G. Biging (1997). A qualitative comparison of fire spread models incorporating wind and slope effects. *Forest Science 43*(2), 170–180. *(p. 23, 433, 434)*

West, M. (1993). Approximating posterior distributions by mixture. *Journal of the Royal Statistical Society 55*(2), 409–422. *(p. 262, 309)*

Westbrook, C. and F. Dyer (1981). Simplified reaction mechanism for the oxidation of hydrocarbon fuels in flames. *Combustion Science and Technology 27*(1-2), 31–43. *(p. 82)*

Westbrook, C., Y. Mizobuchi, T. Poinsot, P. Smith, and J. Warnatz (2005). Computational combustion. *Proceedings of the Combustion Institute 30*, 125–157. *(p. 79)*

Whelan, R. (1995). *The ecology of fire*. New York: Cambridge University Press. *(p. 34)*

Whitaker, J. and T. Hamill (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review 130*, 1913–1924. *(p. 252, 256)*

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics 60*, 897–936. *(p. 360)*

Williams, F. (1982). Urban and wildland fire phenomenology. *Progress in Energy and Combustion Science 8*, 317–354. *(p. 12)*

Wolfrum, J. (1998). Laser in combustion: From basic theory to practical devices. In *27th International Symposium on Combustion*. *(p. 150)*

Wooster, M., T. Richards, and K. Kidwell (1995). Noaa 11 avhrr/2 - thermal channel calibration update. *International Journal of Remote Sensing 16*(2), 359–363. *(p. 34)*

Wooster, M., G. Robert, A. Smith, J. Johnston, P. Freeborn, S. Amici, and A. Hudak (2013). *Thermal remote sensing of active vegetation fires and biomass burning events*, Volume Remote Sensing and Digital Image Processing 17, Chapter Thermal infrared remote sensing, pp. 347–390. Springer. *(p. 29, 38)*

Wooster, M., G. Roberts, G. Perry, and Y. Kaufman (2005). Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research 110*, D24311. *(p. 32, 34, 35, 36, 283, 285)*

Wooster, M., B. Zhukov, and D. Oertel (2003). Fire radiative energy for quantitative study of biomass burning: Derivation from the BIRD experimental satellite and comparison to MODIS fire products. *Remote Sensing of Environment 86*, 83–107. *(p. 34, 35, 39)*

Wu, C.-C., G.-Y. Lien, J.-H. Chen, and F. Zhang (2010). Assimilation of tropical cyclone track and structure based on the ensemble Kalman filter (EnKF). *Journal of the Atmospheric Sciences 67*(12), 3806–3822. *(p. 271)*

Xiu, D. (2010). *Numerical methods for stochastic computations. A spectral method approach*. Princeton University Press. *(p. 360)*

Xiu, D. and G. Karniadakis (2002). The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing 24*(2), 619–644. *(p. 361)*

Zhou, X. and J. Gore (1996). Study of entrainment and flow patterns in pool fires using particle imaging velocimetry. Technical Report Report GCR 97-706, National Institute of Standards and Technology. *(p. 150)*

Zhou, X. and S. Mahalingam (2002). A flame surface density based model for large eddy simulation of turbulent nonpremixed combustion. *Physics of Fluids 14*(11), L77. *(p. 66)*

Zhou, X., S. Mahalingam, and D. Weise (2005). Modeling of marginal burning state of fire spread in live chaparral shrub fuel bed. *Combustion and Flame 143*, 183–198. *(p. 45, 145)*

Zhou, X., S. Mahalingam, and D. Weise (2007). Experimental study and large eddy simulation of effect of terrain slope on marginal burning in shrub fuel beds. *Proceedings of the Combustion Institute 31*, 2547–2555. *(p. 45, 145)*

Zhou, X., L. Sun, D. Weise, and S. Mahalingam (2003). Thermal particle image velocity estimation of fire plume flow. *Combustion Science and Technology 175*, 1293–1316. *(p. 150)*