



HAL
open science

Approches RBF-FD pour la modélisation de la pollution atmosphérique urbaine et l'estimation de sources

Roman Lopez Ferber

► **To cite this version:**

Roman Lopez Ferber. Approches RBF-FD pour la modélisation de la pollution atmosphérique urbaine et l'estimation de sources. Automatique / Robotique. Université Grenoble Alpes [2020-..], 2024. Français. NNT : 2024GRALT006 . tel-04584408

HAL Id: tel-04584408

<https://theses.hal.science/tel-04584408>

Submitted on 23 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)

Spécialité : Automatique - Productique

Unité de recherche : Laboratoire d'Electronique et de Technologie de l'Information (LETI)

**Approches RBF-FD pour la modélisation de la pollution
atmosphérique urbaine et l'estimation de sources**

**RBF-FD approaches for modelling urban air pollution and source
estimation**

Présentée par :

Roman LOPEZ FERBER

Direction de thèse :

Didier GEORGES

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Sylvain LEIRENS

CEA Grenoble

Co-encadrant de thèse

Rapporteurs :

Isabelle CALMET

PROFESSEURE DES UNIVERSITES, ECOLE CENTRALE NANTES

Laurent AUTRIQUE

PROFESSEUR DES UNIVERSITES, Université d'Angers

Thèse soutenue publiquement le **31 janvier 2024**, devant le jury composé de :

Gildas BESANCON,

PROFESSEUR DES UNIVERSITES, GRENOBLE INP

Président

Didier GEORGES,

PROFESSEUR DES UNIVERSITES, GRENOBLE INP

Directeur de thèse

Isabelle CALMET,

PROFESSEURE DES UNIVERSITES, ECOLE CENTRALE NANTES

Rapporteuse

Laurent AUTRIQUE,

PROFESSEUR DES UNIVERSITES, Université d'Angers

Rapporteur

Patrick ARMAND,

DOCTEUR EN SCIENCES HDR, CEA CENTRE DE PARIS-SACLAY

Examineur

Invités :

Sylvain LEIRENS

Ingénieur de recherche, CEA-Leti



Remerciements

Les travaux présentés dans cette thèse n'auraient pas été possibles sans le soutien de nombreuses personnes auxquelles je tenais à exprimer ma reconnaissance.

Je tiens tout d'abord à sincèrement remercier Sylvain Leirens et Didier Georges. Leur soutien constant tout au long de cette thèse a été précieux. Travailler avec Sylvain Leirens m'a particulièrement fait mûrir humainement et professionnellement. Le présent travail doit beaucoup à Didier Georges, son expérience et son immense culture scientifique ont donné des directions déterminantes à la thèse. Leurs judicieux conseils ont également été essentiels pour faire progresser ce travail.

J'adresse ma profonde gratitude aux membres du jury, à savoir la Pr. Isabelle Calmet, le Pr. Laurent Autrique, le Pr. Gildas Besançon et le Dr. Patrick Armand. L'évaluation de ces travaux, à la lumière de leur expertise, est un véritable honneur.

Je tiens également à remercier chaleureusement Patrick Armand pour ses conseils avisés et les explications qu'il m'a apportés avec Christophe Duchenne sur le logiciel PMSS. Je remercie également Anne-Sophie Saffre et Maxime Nibart pour leur aide, leur rencontre a été très enrichissante. Merci également à Trever Hines pour nos échanges fructueux sur Github.

Je remercie les membres du laboratoire, pour toutes les discussions et moments partagés ensemble, je pense aux très bons moments d'amitié avec vous, à table, ou ailleurs, sur un vélo ou dans l'herbe. Chacun de vous m'a beaucoup apporté.

Comment pourrais-je oublier de remercier mon grand ami, Deepesh. Tu m'as toujours soutenu et su me motiver en toutes circonstances. Ta compagnie, tes biryanis et nos sorties à Ikéa, par des journées pluvieuses d'hiver, ont été un véritable baume au cœur. *Chala thanks for everything na mitrama.*

Merci à ma famille, mes parents et mes frères, pour tout leur soutien, et merci à ma mère et à mon père pour leur aide immense.

Enfin merci à Carole-Anne. Tu m'as toujours soutenu, conseillé et aidé, et bien plus encore. Vivre avec toi est un privilège.

Résumé

Depuis l'ère industrielle, les villes sont impactées par la pollution de l'air du fait de la densité de l'industrie, de la circulation de véhicules et de la densité d'appareils de chauffage à combustion. La pollution atmosphérique urbaine a des conséquences sur la santé auxquelles les pouvoirs publics et les citoyens s'intéressent toujours plus. Cette pollution peut entraîner l'aggravation de l'asthme ou encore de troubles cardiovasculaires. Le but de ce travail de thèse est de localiser et de quantifier des sources de pollution urbaine à l'aide d'un réseau dense de mesures bruitées. Nous avons fait le choix de développer des méthodes d'estimation de sources de pollution en se basant sur des modèles physiques de dispersion de polluants. Ainsi l'estimation de sources de pollution est contrainte par la connaissance de la physique du phénomène de dispersion. Ce travail de thèse porte ainsi sur la modélisation numérique de la dispersion de polluants en milieu urbain et sur l'estimation des termes sources.

À cause des nombreuses contraintes qu'impose le bâti urbain aux flux de polluants, la physique de dispersion est représentée par des modèles numériques coûteux en calcul. Nous avons développé un modèle numérique de dispersion basé sur la méthode des Différences Finies supportées par des Fonctions de Base Radiale (RBF-FD). Ces approches sont réputées frugales en calcul et adaptées au traitement de domaines de simulation à géométrie complexe. Notre modèle RBF-FD peut traiter des problèmes bidimensionnels (2D) et tridimensionnels (3D). Nous avons comparé cette modélisation face à un modèle analytique en 2D, et avons qualitativement comparé notre modèle en 3D à un modèle numérique de référence.

Des expériences d'estimations de source ont ensuite été réalisées. Elles considèrent de nombreuses mesures bruitées pour estimer un terme source quelconque sur tout le domaine de simulation. Les différentes études effectuées mettent en œuvre des expériences jumelles : nous générons nous-même des mesures simulées par un modèle numérique et évaluons les performances des estimations. Après avoir testé une approche d'apprentissage automatique sur un cas unidimensionnel en régime stationnaire, nous avons testé des méthodes d'estimation de terme source sur des cas tridimensionnels en régime permanent et transitoire, en considérant des géométries sans et avec présence d'obstacles. Nous avons testé des estimations en utilisant une méthode adjointe originale puis une méthode d'estimation inspirée de l'apprentissage automatique informé par la physique (PIML) et enfin un filtre de Kalman. L'approche inspirée du PIML testée en régime stationnaire conduit à une qualité d'estimation comparable au filtre de Kalman (où ce dernier considère un régime transitoire de dispersion à source stationnaire). L'approche inspirée du PIML exploite directement la frugalité du modèle direct de calcul RBF-FD, ce qui en fait une méthode prometteuse pour des estimations de source sur des domaines de calcul de grande taille.

Abstract

Since the industrial era, cities have been affected by air pollution due to the density of industry, vehicle traffic and the density of combustion heaters. Urban air pollution has health consequences that are of increasing concern to both public authorities and the general public. This pollution can aggravate asthma and cardiovascular problems. The aim of this thesis is to locate and quantify sources of urban pollution using a dense network of noisy measurements. We have chosen to develop methods for estimating pollution sources based on physical models of pollutant dispersion. The estimation of pollution sources is therefore constrained by knowledge of the physics of the dispersion phenomenon. This thesis therefore focuses on the numerical modelling of pollutant dispersion in an urban environment and on the estimation of source terms.

Because of the many constraints imposed on pollutant flows by urban buildings, the physics of dispersion is represented by computationally expensive numerical models. We have developed a numerical dispersion model based on the Finite Difference method supported by Radial Basis Functions (RBF-FD). These approaches are known to be computationally frugal and suitable for handling simulation domains with complex geometries. Our RBF-FD model can handle both two-dimensional (2D) and three-dimensional (3D) problems. We compared this model with a 2D analytical model, and qualitatively compared our 3D model with a reference numerical model.

Source estimation experiments were then carried out. They consider numerous noisy measurements in order to estimate any source term over the entire simulation domain. The various studies carried out involve twin experiments : we ourselves generate measurements simulated by a numerical model and evaluate the performance of the estimates. After testing a machine-learning approach on a one-dimensional steady-state case, we tested source term estimation methods on three-dimensional steady-state and transient cases, considering geometries without and with the presence of obstacles. We tested estimates using an original adjoint method, then an estimation method inspired by physics-informed machine learning (PIML) and finally a Kalman filter. The PIML-inspired approach, tested in a stationary regime, produces an estimation quality comparable to that of the Kalman filter (where the latter considers a transient dispersion regime with a stationary source). The PIML-inspired approach directly exploits the frugality of the RBF-FD direct computation model, which makes it a promising method for source estimates over large computational domains.

Table des matières

Remerciements	i
Résumé	ii
Abstract	iii
Table des matières	iv
1 Introduction	3
1.1 Contexte et objectif principal de la thèse	4
1.2 Principaux apports de la thèse	6
1.3 Organisation de la thèse	7
2 Etat de l'art	10
2.1 Introduction	12
2.2 Météorologie et dispersion de polluants	12
2.2.1 Modélisation de la météorologie	13
2.2.2 L'équation d'advection-diffusion, un modèle de dispersion	17
2.3 Méthodes de simulation de la dispersion atmosphérique de polluants	18
2.3.1 Modélisation de la météorologie	18
2.3.2 Un aperçu des méthodes de simulation de phénomènes de dispersion de polluants	20
2.4 Approximations RBF-FD et LS-RBF-FD des Equations aux Dérivées Partielles	25
2.4.1 Quelques applications des RBF-FD pour résoudre des EDP	26
2.4.2 Introduction aux RBFs	27
2.4.3 La méthode RBF globale	29
2.4.4 La méthode RBF-FD	30
2.4.5 Traitement des conditions aux limites	33
2.4.6 Approche RBF-FD moindres carrés (LS-RBF-FD)	33
2.4.7 Avantages et inconvénients des RBF-FD et LS-RBF-FD	36
2.4.8 Conclusion sur les méthodes RBF-FD et LS-RBF-FD	37
2.5 Assimilation de données	37
2.5.1 Introduction	38
2.5.2 Assimilation variationnelle	38
2.5.3 Assimilation Bayésienne	47
2.5.4 Méthodes hybrides d'Assimilation de Données	53
2.6 Exemples d'assimilation de données	54
2.7 Récapitulatif modèles météorologiques, techniques d'assimilation, étude de cas de la littérature	57
2.8 Conclusion sur l'état de l'art	58
3 Modélisation et simulation par RBF-FD	60
3.1 Introduction	62
3.2 Formalisme des méthodes RBF-FD appliqué à l'ADPDE	62
3.2.1 Mise en œuvre des opérateurs RBF-FD	63

3.2.2	Méthodes RBF-FD et LS-RBF-FD pour l'ADPDE stationnaire	64
3.3	Modélisation en 3D de domaines complexes	65
3.3.1	Le shapefile : des données d'entrée de bâti urbain.	65
3.3.2	Maillage (nuage de points) conforme et non conforme	66
3.3.3	Création d'un maillage grossier conforme aux bâtiments	67
3.3.4	Stratégie de placement de nœuds	69
3.3.5	Domaines d'études retenus	73
3.3.6	Définition formelle des sources et conversion de leurs unités	74
3.3.7	Perspectives	75
3.4	Analyse de performance des méthodes RBF-FD et LS-RBF-FD	75
3.4.1	Une solution explicite de référence en 2D	75
3.4.2	Résultats sur un cas simple en 2D	78
3.4.3	Simulations urbaines RBF-FD 2D à champs météorologiques uniformes	84
3.4.4	Résultats sur un cas simple en 3D, Résultats sur le quartier de Bonne à Grenoble en 3D	85
3.5	Une méthode LS-RBF-FD dans le cas instationnaire	89
3.5.1	Un schéma d'Euler implicite	89
3.5.2	Résultats de simulation	91
3.5.3	Conclusion	92
4	Estimation de sources	96
4.1	Introduction	98
4.2	Une étude comparative entre approches variationnelle et par apprentissage - cas 1D stationnaire	98
4.2.1	Un modèle 1D de transport-diffusion de polluants	99
4.2.2	Méthodes	101
4.2.3	Étude de cas : estimation de source dans une rue	103
4.2.4	Conclusions	107
4.3	Approche adjointe avec modèle LS-RBF-FD	108
4.3.1	Méthode de STE en utilisant un adjoint discret	110
4.3.2	Etude de cas STE dans sur une zone urbaine simulée	111
4.3.3	Conclusions	116
4.4	Approche de type PIML avec modèle LS-RBF-FD	117
4.4.1	Formalisme de la méthode STE-PIML, cas stationnaire	117
4.4.2	Résultats	118
4.4.3	Conclusions	126
4.5	Filtre de Kalman avec modèle transitoire LS-RBF-FD	127
4.5.1	Introduction	127
4.5.2	Mise en œuvre du filtre de Kalman	128
4.5.3	Résultats	129
4.5.4	Conclusions	134
4.6	Conclusion	134
5	Conclusion générale et perspectives	137
5.1	Conclusion générale	138
5.2	Perspectives en lien avec les méthodes abordées	139
5.3	Perspectives et contraintes d'application en temps réel et de prise en compte de capteurs mobiles	141
5.4	Placement optimal de capteurs et estimation avec capteurs mobiles	143

A	Annexe 1 : Suppléments sur la simulation numérique par RBF-FD	XIV
A.1	Expression numérique des opérateurs RBF-FD	XV
A.2	Structure des données de mailles triangulaires grossières	XVII

1

Introduction

Sommaire

1.1	Contexte et objectif principal de la thèse	4
1.2	Principaux apports de la thèse	6
1.3	Organisation de la thèse	7

1.1 Contexte et objectif principal de la thèse

La pollution urbaine a des conséquences sur la santé auxquelles les pouvoirs publics s'intéressent toujours plus. Cette préoccupation est également partagée par nombre d'habitants de grandes villes. Depuis l'ère industrielle, les villes sont impactées par la pollution de l'air du fait de la densité de l'industrie, de la circulation de véhicules et de la densité d'appareils de chauffage qui, dans le cas de ceux fonctionnant au fuel ou au bois sont notamment émetteurs de particules fines. La pollution urbaine a des conséquences sur la santé. Des autorités de l'union européenne comme l'Agence Européenne de l'Environnement accusent les particules fines de diamètres inférieurs à 2,5 μm (PM2.5) de causer près de 420 000 morts prématurées en 2015 [Guerreiro2018]. L'Institut national de l'environnement industriel et des risques (INERIS) a d'ailleurs été mandaté pour conseiller l'état dans l'établissement de norme pour le chauffage au bois et a conclu que le chauffage domestique au bois a été responsable de 44% des émissions de particules fines à l'échelle de la France ¹. Ailleurs également, la présence de particules fines, d'oxydes d'azote et d'oxydes de soufre, sont également source de préoccupation. Des auteurs ont notamment observé en Inde une corrélation entre ces concentrations de polluants et l'apparition de maladies [Liu2013]. Comme l'INERIS, la ville de Grenoble montre la responsabilité de l'humain dans l'émission de ces polluants. Elle communique également dans un rapport de 2022 ² que 67 % des particules fines observées à Grenoble sont générées par le chauffage au bois individuel tandis que le trafic routier est responsable de 63 % des émissions de dioxyde d'azote quand l'industrie en rejette bien moins (19 %). Lors d'un fort ensoleillement où le niveau de rayonnement ultraviolet est élevé, les oxydes d'azote participent à la formation d'ozone, un gaz polluant et agressif aux conséquences sensibles sur la santé telles que l'aggravation de l'asthme ou de troubles cardiovasculaires ³.

Pour mesurer la pollution atmosphérique, des organismes d'intérêt général agréés par l'État utilisent des réseaux de capteurs sophistiqués et coûteux qui fournissent alors des données fiables pour réaliser de l'estimation et de la prévision de la pollution atmosphérique. Nous pouvons citer par exemple l'observatoire Atmo Auvergne Rhône-Alpes pour la région homonyme ou encore AIRPARIF pour la région parisienne. L'observatoire AIRPARIF dresse des cartes précises de la pollution en réalisant de l'assimilation de données ⁴. Cette méthode consiste à utiliser un modèle numérique (ici de dispersion de polluants) dont les variables cibles (ici des concentrations de polluants) sont corrigées par la mesure permettant ainsi une interpolation optimale et réaliste du champ de mesures.

Avec le temps, certains types de capteurs, notamment de particules fines, se sont démocratisés [Alfano2020]. Plusieurs capteurs à bas coût d'une centaine d'euros ou moins sont à la vente tandis que des capteurs de particules fines équipent de nombreux purificateurs d'air intérieur abordables et disponibles au grand public. Il est alors envisageable d'utiliser ces capteurs comme un réseau dense de capteurs à bas coût. L'hypothèse de l'abondance de mesures à bas coût ouvre alors la voie d'une part à des assimilations de données précises spatialement car les mesures échantillonnent mieux l'espace qu'avec un réseau de capteurs de haute précision mais épars. Ce phénomène a été observé dans une étude qui a comparé les performances de l'estimation de concentration de polluants à l'aide d'un modèle physique, avec des 19 stations de mesures de haute qualité et à l'aide

1. Synthèse des études à l'émission réalisées par l'INERIS sur la combustion du bois en foyers domestiques, INERIS, consulté en Août 2023, <https://www.ineris.fr/fr/synthese-etudes-emission-realisees-ineris-combustion-bois-foyers-domestiques>

2. Pollution de l'air : Grenoble attaque l'État pour son inaction, Service presse de la ville de Grenoble, consulté en Août 2023, https://www.grenoble.fr/uploads/Externe/05/1619_002_DP-Recours-PPA-vweb.pdf

3. Qu'est-ce que l'ozone?, Atmo Auvergne-Rhône-Alpes, consulté en Août 2023, <https://www.atmo-auvergnerhonealpes.fr/article/quest-ce-que-lozone#:~:text=L'ozone%20c'est%20quoi%20%3F,-Contenu&text=L'Ozone%20est%20un%20polluant,d%C3%A9j%C3%A0%20pr%C3%A9sents%20dans%20l'air.>

4. La pollution en direct en Île-de-France, Airparif, consulté en Août 2023, <https://www.airparif.asso.fr/surveiller-la-pollution/la-pollution-en-direct-en-ile-de-france>

d'un réseau dense de 255 capteurs à bas coût [Lopez-Restrepo2021]. Le domaine de l'étude, la vallée d'Aburrá (Colombie) s'étend sur $111 \times 122 \text{ km}^2$ avec une résolution de $1,22 \text{ km}^2$ représentant 0,025 capteurs à bas coût par point de grille au sol. D'autre part, l'abondance de mesures ouvre la voie à la possibilité de résoudre des problèmes inverses, notamment, l'estimation de sources de pollution multiples ou de formes quelconques.

Par problème inverse, nous entendons au sens général, la détermination de la cause d'un phénomène à partir d'observations expérimentales de ses effets. Dans le cas présent, des sources de pollution causent une dispersion de polluants dans l'atmosphère. La résolution de problème inverse passe généralement par la maîtrise préalable d'une modélisation du phénomène dite problème direct qui permet de traduire les paramètres du modèle en effets expérimentalement observables. Dans le cadre de cette thèse, il s'agit de modéliser le phénomène de la dispersion atmosphérique de polluants.

Historiquement, les méthodes d'estimation de source de pollution ont servi à évaluer et à amoindrir l'impact de sites industriels rejetant des polluants par des cheminées. L'Estimation de Terme Source (STE) a également été employée dans le contrôle des accidents nucléaires ou encore la surveillance internationale des essais nucléaires [Pudykiewicz1998]. Aujourd'hui la STE connaît un intérêt croissant aux échelles urbaine [Septier2020, Mons2017, Xue2017] et régionale [Tang2013]. Généralement dans des exemples de STE, le problème inverse est résolu en appliquant plusieurs fois un modèle direct de dispersion atmosphérique de polluants.

Le présent travail s'intéresse au contexte de la pollution atmosphérique en milieu urbain. Le but est de localiser et de quantifier des sources de cette pollution. La thèse vise à développer des outils de modélisation de la pollution qui, utilisés avec des mesures fournies par un réseau de capteur à bas coût, permettent de remonter à des sources. Les capteurs à bas coût pourraient également être envisagés mobiles comme abordé récemment dans la littérature [Georges2013, Hasenfratz2015], tout comme les sources de pollutions qui pourraient être des véhicules particulièrement polluants. Cependant la mobilité des capteurs et des sources n'est pas traitée dans ce travail de thèse. L'INERIS a aussi émis des recommandations d'interdiction de cheminées et de poêles à bois polluants. Les méthodes développées pourraient également aider à identifier les chauffages domestiques les plus polluants en les considérant comme des sources ponctuelles fixes.

À cause des nombreuses contraintes qu'imposent les bâtiments aux flux de polluants, les modèles analytiques usuels comme le modèle gaussien des panaches ou les modèles gaussiens à bouffées ne peuvent pas être assez précis pour représenter correctement le phénomène de dispersion atmosphérique de polluants [Mendil2022]. Des modèles physiques à maille de calcul fine, précis et coûteux en calcul sont donc utilisés pour y représenter la physique.

Il est possible de se passer de l'utilisation d'un modèle physique pour réaliser une STE : les approches basées sur les données ou encore les approches basées sur les apprentissages profonds suscitent un intérêt certain pour l'estimation de source [Huang2021] ou pour la prévision de concentration de polluants [Zaini2022]. L'utilisation d'interpolation linéaire ou de krigeage combinée à des techniques de régression peut également permettre de résoudre des problèmes inverses [Bailly du Bois1995, Bailly du Bois2012]. Bailly du bois et al. ont par exemple utilisé une interpolation par krigeage du champ de concentration en traceurs radioactif dilués dans la mer et une méthode de régression pour estimer le terme source radioactif marin issu de l'accident de la centrale nucléaire de Fukushima en 2011 [Bailly du Bois2012].

Cependant, les méthodes basées sur des modèles ont été choisies dans ce travail. Les approches basées sur les modèles sont en effet réputées pour leur fiabilité par rapport à des méthodes statistiques ou encore d'interpolation ne se basant pas sur des modèles physiques. En effet, contrairement à un champ de polluants en milieu marin comme étudié dans [Bailly du Bois2012], un champ de concentration de polluants atmosphérique en milieu urbain est hautement perturbé par les obstacles ainsi que la météorologie, ce qui fausse toute interpolation du champ de concentration en polluants depuis quelques mesures. Ce travail de thèse a également abouti au développement d'un modèle urbain de dispersion atmosphérique de polluants basé sur les Différences Finies sup-

portées par des Fonctions de base Radiales (RBF-FD). Nous avons remarqué que ces méthodes sont prometteuses en temps de calcul et en précision. Nous faisons donc le choix d'utiliser des méthodes de reconstruction de source basées sur des modèles physiques capables de traiter des domaines d'étude hétérogènes et complexes. De plus, une méthode sans maillage telle que les RBF-FD sont très flexibles dans la création des domaines de calcul et présentent des caractéristiques prometteuses de raffinement local de nœuds de calcul et de traitement de géométries complexes [Fornberg2015a, Flyer2012]. Un modèle physique est utilisé pour simuler la dispersion de polluants dans l'air : un logiciel tiers modélise la météorologie (c. à. d. les circulations d'air), notre modèle RBF-FD modélise la dispersion atmosphérique de polluants consécutive de la météorologie. Ensuite, nous utilisons des méthodes qui permettent de localiser et quantifier la ou les source(s) candidate(s) sachant les mesures de concentration effectuées sur le terrain. Notre objectif principal dans cette thèse est de développer une méthodologie d'estimation de sources qui permette d'obtenir rapidement une estimation de sources de pollution en combinant des méthodes de réduction de modèles, et en s'appuyant sur un réseau de capteurs à bas coût, quitte à faire des compromis sur la précision. Ce travail contribue également au développement de jumeaux numériques en milieu fortement urbanisé.

1.2 Principaux apports de la thèse

Le but de ce travail est de localiser et quantifier des sources de pollution atmosphériques en milieu urbain avec l'utilisation d'un modèle physique de dispersion de polluants et la prise en compte de mesures de concentration de polluants prises avec un ou des capteurs géographiquement dispersés. Dans ce travail nous avons considéré deux scénarios d'Estimation de Terme Source (STE).

Premièrement un scénario de STE en contexte unidimensionnel et stationnaire où nous avons comparé une méthode de STE classique, la 3D-Var, à une approche originale d'apprentissage automatique. Ce travail a fait l'objet d'une publication dans la conférence IFAC CSC 2021.

Deuxièmement, nous avons considéré des expériences de STE en régime stationnaire et transitoire dans des environnements tridimensionnels (3D) urbains et non urbains où notre modèle de dispersion atmosphérique LS-RBF-FD a été utilisé. Dans ce contexte, la météorologie locale est calculée préalablement par un logiciel tiers. Notre modèle RBF-FD assure ensuite la dispersion de polluants dans un deuxième temps.

Une partie majeure du travail de thèse a été de développer un modèle de dispersion atmosphérique de polluants ainsi que de générer le domaine de calcul de ce modèle. Le modèle de dispersion prend en entrée des données météorologiques de champ de vitesse du vent et de champ de diffusion, ce dernier étant essentiellement induit par les circulations turbulentes de l'air à petite échelle. Le modèle choisi utilise les Différences Finies supportées par des Fonctions de Base Radiales résolues en moindres carrés (LS-RBF-FD). Le choix de cette méthode a été motivé par sa réputation de bonne précision malgré une faible densité de points de calcul [Flyer2012, Tominec2021c]. Ce modèle étant utilisé pour un problème inverse, la faible complexité (en termes de nombre de nœuds de calcul) du modèle de dispersion qui en résulte ouvre des perspectives encourageantes pour résoudre des problèmes d'estimation de terme source dans des domaines complexes.

Le développement d'un modèle de dispersion par RBF-FD et LS-RBF-FD pour résoudre un modèle 3D de dispersion atmosphérique de polluants est à notre connaissance unique dans l'état de l'art. La construction d'un domaine de simulation urbain sans maillage pour le calcul RBF-FD l'est également. De plus, la LS-RBF-FD, la variante de moindres carrés des RBF-FD est à notre connaissance testée pour la première fois sur des cas 3D. Elle apporte des améliorations notables par rapport à notre modèle de dispersion RBF-FD. Les modèles LS-RBF-FD et RBF-FD de dispersion

atmosphériques ont été testés dans divers cas 2D et 3D.

Pour faire avancer le modèle de dispersion LS-RBF-FD en temps un schéma numérique d'Euler implicite temporel a été développé. Le côté implicite du schéma assure la stabilité temporelle du schéma de calcul.

Le modèle LS-RBF-FD a servi à réaliser en milieu tridimensionnel les expériences d'Estimations de Terme(s) Source(s) (STE) suivantes :

- Nous avons développé une méthode adjointe unique et basée sur le modèle de dispersion LS-RBF-FD pour réaliser une STE dans un contexte de dispersion urbaine de polluants en régime stationnaire dans un domaine à trois obstacles. Ce travail a été soumis sous la forme d'article à la conférence ECC 2024 (European Control Conference).
- Nous avons développé une méthode de STE basée sur le modèle LS-RBF-FD inspirée des réseaux de neurones informés de la physique (PINN) [Raissi2019] et de l'apprentissage informé par la physique (PIML) dans un contexte stationnaire sur divers domaines et divers types de termes sources de pollution.
- Enfin une approche bayésienne basée sur le filtre de Kalman a été testée dans un cadre de dispersion de polluants en régime transitoire provenant de sources multiples stationnaires.

Nous observons que le modèle de dispersion atmosphérique de polluants LS-RBF-FD que nous avons développé pendant cette thèse montre de bonnes aptitudes pour traiter des problèmes inverses comme l'estimation de terme source.

Publications Ce travail a donné lieu aux publications suivantes :

1. Roman Lopez-Ferber, Sylvain Leirens et Didier Georges *Source term estimation : variational method versus machine learning applied to urban air pollution*, CSC 2022 - IFAC Workshop on Control for Smart Cities, Sozopol (virtual), Bulgaria, 2022
2. Roman Lopez-Ferber, Didier Georges et Sylvain Leirens, *A Fast Adjoint Based LS-RBF-FD Method for 3D Source Term Estimation of Air Pollution in Complex Urban Areas*, Soumis à European Control Conference (ECC), Stockholm, Sweden, 2024

1.3 Organisation de la thèse

Le reste de la thèse est organisé de la manière suivante :

Chapitre 2 : Ce chapitre rassemble l'état de l'art sur le phénomène de dispersion atmosphérique de polluants et présente quelques méthodes pour le modéliser. Les approximations d'Equations aux Dérivées Partielles (EDP) que réalisent la méthode des Différences Finies supportées par des Fonctions de Base Radiales (RBF-FD) et sa variante basée sur les moindres carrés (LS-RBF-FD) sont présentées ensuite. Le chapitre discute enfin de l'assimilation de données et présente différentes méthodes de l'état de l'art pour la réaliser. Une partie présente quelques études utilisant ces diverses techniques notamment dans le cadre de la pollution atmosphérique.

Chapitre 3 : Ce chapitre porte sur le flux de travail de la modélisation du transport de polluants RBF-FD et LS-RBF-FD. Le chapitre discute de l'implémentation des opérateurs d'approximation de l'Equation aux Dérivées Partielles de transport de polluants : l'équation d'Advection-Diffusion (ADPDE), de la génération des domaines de calculs, de la définition formelle des sources de pollution et présente le comportement des différentes méthodes LS-RBF-FD et RBF-FD sur des cas d'étude.

Chapitre 4 : Ce chapitre est consacré aux problèmes d'Estimation de Terme Source. Un premier travail d'étude porte sur un cas unidimensionnel où une méthode basée sur un apprentissage automatique a été comparée à une méthode classique d'assimilation de données 3D-Var. Un deuxième travail emploie une méthode adjointe dans un contexte de régime stationnaire de dispersion de polluants. Ensuite une approche originale inspirée de l'apprentissage automatique informé par la physique (Physics-Informed Machine Learning) et basé sur le modèle direct LS-RBF-FD est présentée. Enfin une méthode utilisant le filtrage de Kalman dans un contexte de régime transitoire est discutée avant de donner quelques conclusions.

Chapitre 5 : Ce dernier chapitre conclut le manuscrit et donne des perspectives pour une poursuite des travaux.

2

Etat de l'art

Le présent chapitre réalise un état de l'art d'abord sur la dispersion atmosphérique de polluants en milieu urbain. Puis nous présentons le principe des méthodes de résolution numérique/réduction de modèles sans maillage fondées sur les approches RBF-FD (à fonctions de base radiales et différence finies) et LS-RBF-FD (moindres carrés à fonctions de base radiales et différence finies) qui sont au cœur de cette thèse. Nous faisons enfin l'exposé des principales méthodes d'assimilation de données utilisant un modèle physique.

Sommaire

2.1	Introduction	12
2.2	Météorologie et dispersion de polluants	12
2.2.1	Modélisation de la météorologie	13
2.2.2	L'équation d'advection-diffusion, un modèle de dispersion	17
2.3	Méthodes de simulation de la dispersion atmosphérique de polluants	18
2.3.1	Modélisation de la météorologie	18
2.3.2	Un aperçu des méthodes de simulation de phénomènes de dispersion de polluants	20
2.4	Approximations RBF-FD et LS-RBF-FD des Equations aux Dérivées Partielles	25
2.4.1	Quelques applications des RBF-FD pour résoudre des EDP	26
2.4.2	Introduction aux RBFs	27
2.4.3	La méthode RBF globale	29
2.4.4	La méthode RBF-FD	30
2.4.5	Traitement des conditions aux limites	33
2.4.6	Approche RBF-FD moindres carrés (LS-RBF-FD)	33
2.4.7	Avantages et inconvénients des RBF-FD et LS-RBF-FD	36
2.4.8	Conclusion sur les méthodes RBF-FD et LS-RBF-FD	37
2.5	Assimilation de données	37
2.5.1	Introduction	38
2.5.2	Assimilation variationnelle	38

2.5.3	Assimilation Bayésienne	47
2.5.4	Méthodes hybrides d'Assimilation de Données	53
2.6	Exemples d'assimilation de données	54
2.7	Récapitulatif modèles météorologiques, techniques d'assimilation, étude de cas de la littérature	57
2.8	Conclusion sur l'état de l'art	58

2.1 Introduction

Nous présentons dans ce chapitre un état de l'art en 3 axes, nous discuterons de la modélisation physique de la dispersion atmosphérique de polluants en milieu urbain du point de vue théorique, ensuite, nous présenterons des aspects pratiques de modélisation numérique du phénomène, notamment grâce à la méthode de discrétisation par Différences Finies supportées par des Fonctions de Base Radiales (RBF-FD). Enfin nous présenterons des techniques d'assimilations de données dans le but de résoudre le problème inverse de l'Estimation de Terme Source (STE).

Les études présentées dans cette thèse s'intéressent principalement à la modélisation de la dispersion atmosphérique de polluants, notamment par la méthode RBF-FD, et à l'estimation de terme source. La méthode RBF-FD est une méthode de discrétisation et de résolution déterministe d'Equation aux Dérivées Partielles (EDP). Nous n'avons donc pas choisi de développer un *modèle direct* produit par des méthodes d'apprentissage dans cette thèse. Entendons par "*modèle direct*" le modèle qui traduit l'EDP ou les EDP impliquées dans un problème d'Assimilation de Données.

Cependant, dans la perspective de réaliser des STE en temps réel ou encore sur du matériel embarqué, disposer de modèles numériques de dispersion approchés par apprentissage est d'intérêt. En effet, les méthodes d'Assimilation de Données ont pour point commun de réaliser de nombreux appels de *modèle direct*. Dans le cadre de cette thèse, ces modèles ont un coût de calcul élevé.

Pour répondre au besoin de réaliser de l'Assimilation de Données en temps réel ou encore pour utiliser des calculs de *modèle direct* sur des appareils embarqués, une approche pertinente est d'approximer le *modèle direct* de dispersion par des méthodes d'apprentissage (Machine Learning en anglais).

Khodayi-Mehr et al. ont par exemple utilisé la Décomposition de Modes Propres (POD ¹) dans le contexte de la pollution atmosphérique contrainte par des obstacles. Les auteurs ont ainsi disposé d'un modèle de dispersion approché permettant de réaliser de l'estimation en temps réel [Khodayi-mehr2019]. Ding et al. ont aussi utilisé la POD cette fois pour approximer dans champs de vélocité de vent dans une agglomération urbaine complexe [Ding2021]. Par ailleurs, les modèles auto-associatifs, extensions non-linéaires des méthodes POD, peuvent produire des modèles réduits de meilleure qualité que ces dernières [Girard2020].

Concernant les approches utilisant des réseaux de neurones pour approximer un modèle physique, les réseaux de neurones informés de la physique sont une référence [Raissi2019, Lu2021, Gayray2023].

L'état de l'art qui suit présente la théorie et les méthodes utilisées par les études menées dans cette thèse.

2.2 Météorologie et dispersion de polluants

Les études d'Estimations de Terme Source (STE) de pollution atmosphérique couplent leurs méthodes de recherches optimales de variables de contrôle (telles que des conditions initiales, conditions limites, terme de source etc.) avec des modèles physiques de transports de polluants déterministes (par exemple : modèle de convection-diffusion). Ces modèles sont couplés à la mécanique des fluides [Defforge2021, Krysta2007a, Mons2017] pour prendre en compte le rôle du vent qui transporte et mélange les polluants.

Le séjour d'un polluant dans l'atmosphère peut s'identifier à trois processus [Zannetti1990] :

- Le transport sous l'influence des circulations d'air dans l'atmosphère (vent, diffusion). Les

1. La méthode POD peut aussi être vue comme un Analyse en Composantes Principales

turbulences de l’atmosphère ont une grande influence dans le transport.

- Les transformations physico-chimiques, condensation, évaporation, réaction chimique, filiation radioactive. En général, les événements autres que la radioactivité ne sont pas pris ou peu pris en compte dans des modèles à fenêtre de temps réduite.
- Les dépôts secs ou obtenus par lessivage.

La modélisation des champs de vitesses du vent et de sa turbulence a donc une influence prépondérante sur le transport de polluants. A l’échelle micro-météorologique (0 à 1 km de hauteur par rapport au sol), les champs de vents ainsi que leur turbulence sont en priorité physiquement régis par la conservation de la masse et de la quantité du mouvement et de l’énergie que modélisent les équations de Navier-Stokes [Arya1988, p. 117].

Dans quelques cas particuliers, quelques hypothèses simplificatrices sur le champ de vent sont permises et dispensent alors de résoudre la physique de la mécanique des fluides qui régit le vent. Ces hypothèses en modélisation de la pollution atmosphérique conduisent à l’utilisation de modèles gaussiens de panaches qui calculent la dispersion des polluants grâce à des formules semi-analytiques. Cependant, dans des environnements contraints par de nombreux obstacles comme une ville, la simplification du champ de vent, et donc le modèle gaussien des panaches, devient peu pertinente. Ce problème est illustré en Figure 2.1 où la simulation qui tient compte des obstacles et donc calcule les champs météorologiques représente un panache de pollution qui vient s’étaler dans les rues d’un quartier urbain (Figure 2.1, droite) tandis que le panache ne tenant pas compte des obstacles est beaucoup moins étalé (Figure 2.1, gauche).

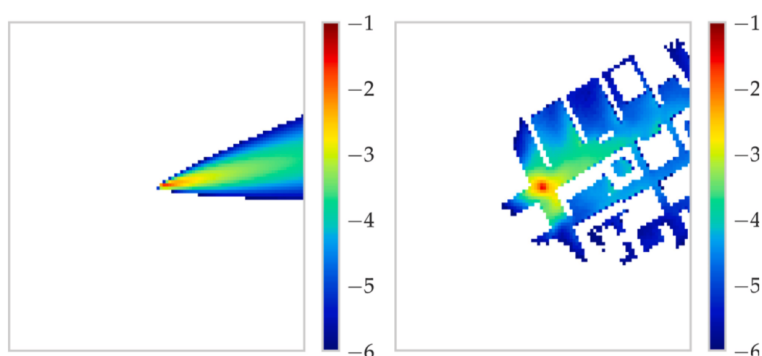


FIGURE 2.1 – A gauche, panache de pollution urbaine calculé avec un modèle analytique des panaches. A droite, panache de pollution calculé en résolvant la physique discrétisée du transport. Figure construite à partir de [Mendil2022] (Figure 8).

2.2.1 Modélisation de la météorologie

Les échelles de météorologie

A l’échelle urbaine, le type de météorologie pertinente à prendre en compte est la micro-météorologie. Par ce terme, on entend la modélisation de la météorologie locale à l’échelle du km^2 ou de la centaine de km^2 . A cette échelle, le déplacement de grandes masses d’air dans les hautes couches de l’atmosphère n’est pas prépondérant tandis que la topographie du terrain et des obstacles (villes, montagnes, rivières) est importante. La modélisation de la météorologie à l’échelle urbaine a ainsi lieu dans la zone basse de la troposphère. Les différentes couches atmosphériques discutées sont visibles en Fig. 2.2. Retenons que la hauteur du phénomène modélisé ne dépasse pas le kilomètre.

2. Source : https://www.umr-cnrm.fr/ecole_lidar/IMG/pdf/16._vm_structure_atmosphere.pdf, p. 4, consulté le 25 Août 2023

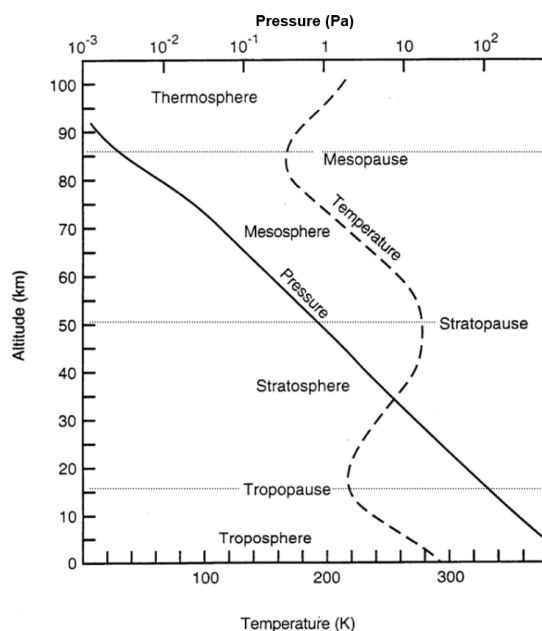


FIGURE 2.2 – Schéma des couches atmosphériques. Source : Centre National de Recherches Météorologiques ².

Notre but est alors de modéliser les circulations d'air dans la couche limite atmosphérique illustrée en Fig. 2.3. A cette basse altitude au sein de la troposphère (moins d'1 km de hauteur), les circulations d'air sont impactées par la rugosité du terrain (présence de bâtiments), et la topographie ou la constitution du terrain (présence de végétation, d'eau, ou de béton).

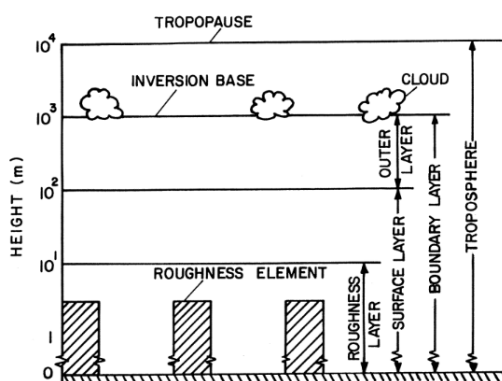


FIGURE 2.3 – Schéma de la couche limite atmosphérique comme partie de la troposphère. Source : [Arya1988]. Roughness layer désigne la couche de rugosité, Surface layer désigne la couche de surface. Boundary layer désigne la couche limite atmosphérique.

Les gradients de température et de pression en fonction de la hauteur au sol jouent également un rôle dans les circulations d'air. En milieu urbain ces gradients ont une influence sur les conditions aux limites des champs de vitesse du domaine de modélisation.

La stabilité atmosphérique

Il existe différents types de gradients de température et de pression dans la couche limite atmosphérique et les connaître permet d'estimer un profil de vent servant de condition limite aux modèles de circulation d'air à l'échelle urbaine. On appelle ces types de gradients la stabilité

atmosphérique. Le motif des dispersions de polluants est très sensible à ces conditions comme illustré en Fig. 2.4 de par la sensibilité de la dispersion face aux circulations d'air. Une atmosphère est qualifiée d'instable lorsque le gradient thermique est inférieur à -1°C par 100 mètres de hauteur, neutre à -1°C par 100 mètres, et stable pour des valeurs au delà ($+0^\circ\text{C}/100\text{m}$, $+1^\circ\text{C}/100\text{m}$...). Les classes de Pasquill catégorisent six classes de stabilités atmosphériques et sont utilisées comme un standard pour la modélisation analytique de la dispersion de polluants. Dans le cadre de la modélisation de la dispersion atmosphérique de polluants, il est en pratique difficile de mesurer les gradients de température. Plusieurs méthodes existent pour l'approximer, par exemple lier la stabilité à la nébulosité (couverture nuageuse) observée.

Le calcul du profil de vélocité de vent incident à des obstacles renseigne une condition limite pour le calcul des circulations d'air en milieu urbain. Ce profil de vélocité de vent peut être donné par une simulation météorologique tierce réalisée à plus grande échelle. Lorsque des valeurs de vent incident sont connues à une certaine hauteur par rapport au sol, ce profil de vélocité de vent incident peut être extrapolé à partir de ces quelques valeurs connues. Pour ce faire, une méthode de référence est d'utiliser la théorie de la similarité de Monin-Obukhov dont la stabilité atmosphérique est un paramètre d'entrée. Ce point est détaillé dans [Businger1971] et des mises à jour récentes de cette théorie sont visibles en [Soulhac2011]. En résumé, la stabilité atmosphérique est un paramètre important qui : (1) change les conditions aux limites de calcul de champ météorologiques à l'échelle urbaine, (2) est utilisée pour des modèles analytiques qui ainsi dispensent du calcul de météorologie à petite échelle.

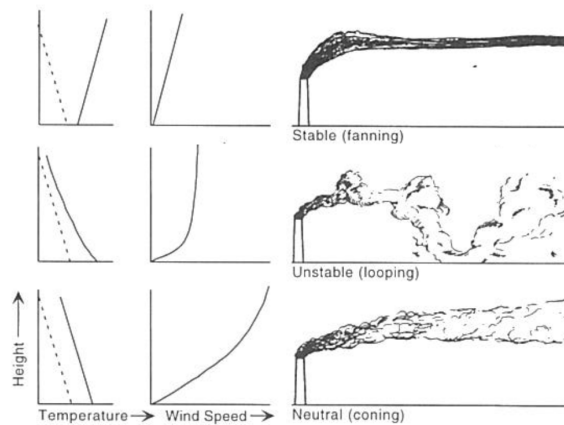


FIGURE 2.4 – Description schématique de panaches de pollution selon la stabilité atmosphérique (de haut en bas : stable, instable et neutre) et profils de vents et de températures correspondant à ces situations. Source : [Defforge2019a] p. 34.

Physique du vent dans la couche limite atmosphérique

A basse altitude, c'est-à-dire dans la Couche Limite Atmosphérique (0 à 1 km de hauteur par rapport au sol), la circulation de l'air est considérée comme régie par les équations de Navier-Stokes comme discutée en introduction de la section 2.2. Tout d'abord, la physique impose la conservation de la masse (équation de continuité),

$$\partial_t \rho + \sum_{i=1}^3 \partial_i (\rho u_i) = 0 \quad (2.1)$$

avec, t, ρ : le temps et la masse volumique du fluide définie dans $(x \in \Omega \subset \mathbb{R}^3, t)$. Les coordonnées cartésiennes (x, y, z) sont désignées par l'utilisation des indices $i = \{1, 2, 3\}$. ∂_t indique une dérivée partielle temporelle tandis que ∂_i indique une dérivée spatiale sur la composante i . De

même ∂_i indiquera une dérivée de deuxième ordre sur la composante i pour la suite. u_i est la i -ème composante de la vélocité du flux. Pour des fluides compressibles, la masse volumique du fluide peut varier dans le temps en un point x de l'espace. Cette variation temporelle de masse s'oppose aux variations spatiales de la masse dans un volume élémentaire de fluide. Pour des petites échelles, par exemple urbaine ou de l'ordre de quelques kilomètres, la variation temporelle de masse volumique $\partial_t \rho$ est généralement négligeable et l'air peut généralement être considérée comme incompressible [Arya1988, p. 118-119]. Sous hypothèse d'incompressibilité, ρ est une constante et la conservation de la masse pour les liquides incompressibles traduit une divergence nulle du champ de vitesse $\mathbf{u} = (u_1, u_2, u_3)$,

$$\sum_{i=1}^3 \partial_i u_i = 0 \quad (2.2)$$

pour tout $x, t \in \Omega \times T$. De manière générale, la conservation de la quantité de mouvement s'écrit pour chaque composante i de l'espace,

$$\partial_t u_i + \sum_{j=1}^3 u_j \partial_j u_i = -\frac{1}{\rho} \partial_i p + g_i + \frac{1}{\rho} \sum_{j=1}^3 \partial_j \tau_{ij} + C_i, \quad (2.3)$$

avec p la pression, g_i la i -ème composante de la force de gravité, τ_{ij} le tenseur des contraintes visqueuses et C_i la i -ème composante de la force de Coriolis qui a petite échelle est souvent négligée. Les équations de la conservation de la masse (2.2) et de la quantité de mouvement (2.3) constituent les équations de Navier Stokes.

Le tenseur des contraintes visqueuses τ peut être explicité formellement pour le cas où le fluide est dit newtonien, c'est à dire que les contraintes visqueuses du fluide sont proportionnelles à son gradient de vitesse. Ainsi $\tau = \mu(\nabla(\mathbf{u}) + \nabla(\mathbf{u}^T))$ où \cdot^T désigne une transposition, avec μ désignant la viscosité dynamique du fluide considérée homogène dans le fluide. En ce qui concerne la micro-météorologie, l'hypothèse d'incompressibilité et la prise en compte de l'air comme un fluide newtonien sont souvent considérées dans la littérature [Winiarek2014a, p. 35], [Oldrini2013, Defforge2021, Mortensen2011]. Ainsi les équations de Navier-Stokes utilisées pour modéliser la circulation de l'air dans la couche limite atmosphérique sont,

$$\partial_t u_i + \sum_{j=1}^3 u_j \partial_j u_i = -\frac{1}{\rho} \partial_i p + g_i + \sum_{j=1}^3 \partial_j \nu (\partial_i u_j + \partial_j u_i) \quad (2.4)$$

$$\sum_{i=1}^3 \partial_i u_i = 0 \quad (2.5)$$

avec, $\nu = \mu/\rho$, la viscosité cinématique du fluide. Il n'existe pas de solution analytiques à ces équations. Elles peuvent se résoudre de manière discrète de plusieurs façons. Cependant, la résolution de (2.4) et (2.5) est rendue numériquement difficile voire impossible pour des flux turbulents agissant sur des grands domaines. En effet, la turbulence implique des fluctuations du champ de vitesse rapides en temps et à petites échelles en espace. Cela exige des résolutions spatiale et temporelle très fines pour résoudre numériquement le modèle (2.4)-(2.5). Une technique classique pour palier cette difficulté dans les problèmes de propagation de polluants consiste à considérer la turbulence comme un phénomène aléatoire de variation de la vélocité d'un fluide qui s'additionne à une valeur moyenne de vélocité. On appelle ce principe la décomposition de Reynolds. Elle consiste à considérer une variable Φ liée à un écoulement comme la somme de sa valeur moyenne et des fluctuations causées par les turbulences :

$$\Phi = \langle \Phi \rangle + \Phi', \quad (2.6)$$

avec $\langle \Phi \rangle$ et Φ' la valeur moyenne et la valeur fluctuante de Φ . L'injection de cette relation en choisissant le champ de vélocité \mathbf{u} comme variable cible dans les équations couplées (2.4)-(2.5) aboutit aux équations de Navier-Stokes moyennées appelées RANS (Reynolds Averaged Navier Stokes). Différents modèles peuvent ensuite être utilisés pour modéliser les termes turbulents résultants.

2.2.2 L'équation d'advection-diffusion, un modèle de dispersion

Dans nombre de travaux [Defforge2021, Krysta2007b, Mons2017, Singh2015, Kumar2015], un modèle d'advection-diffusion (ou convection-diffusion) pour modéliser la dispersion atmosphérique de polluants est utilisé. Ce modèle est obtenu à partir d'un bilan de masse sur des espèces en atmosphère. Cette équation de conservation de la masse, « précisément représentative » de la physique à toute les échelles spatiales et temporelles est tirée d'un bilan de masse des espèces polluantes présentes dans un fluide (ici l'air) et peut être formulée de la manière suivante [Farágó2008, Quélo2004],

$$\frac{\partial c}{\partial t} + \text{div}(\mathbf{u}c) = \text{div}(D_m \nabla c) + s + \chi(c) + P + R(t, c) \quad \text{dans } \Omega \quad (2.7)$$

avec pour conditions initiales et limites,

$$\begin{aligned} c(\mathbf{x}, t=0) &= c_0 & \mathbf{x} \in \Omega \\ c(\mathbf{x}, t) &= d(\mathbf{x}, t) & \mathbf{x} \in \Gamma_D \\ \nabla c \cdot \mathbf{n} &= g(\mathbf{x}, t) & \mathbf{x} \in \Gamma_N, \end{aligned} \quad (2.8)$$

où \mathbf{u} est le champ de vent, c , la concentration d'une espèce, D_m , sa diffusion moléculaire, et s , le terme de source de pollution liée à une espèce i . Les conditions aux limites peuvent être de Dirichlet sur Γ_D , et de type Neumann sur Γ_N , avec c_d une valeur de concentration sur le domaine limite et g une fonction liée à une condition de Neumann. Les autres termes, χ , P , R représentent la filiation radioactive, les pertes par lessivage/déposition et la réaction chimique du polluant. La considération de χ , P , R dépend du polluant. Certains par exemple ne sont pas radioactifs. Cependant les calculs sur une telle équation sont rendus difficiles par les turbulences. C'est pourquoi, la décomposition de Reynolds est conseillée pour simplifier le modèle et prendre en compte les turbulences [Winiarek2014a, Zannetti1990]. Dans [Zannetti1990], cette décomposition en termes moyens et turbulents sur l'équation d'advection-diffusion aboutit à une équation portant des termes croisés de turbulence. Nous écrivons cette équation décomposée en ignorant les termes χ , P , R pour plus de lisibilité,

$$\partial_t \langle c \rangle = -\langle \mathbf{u} \rangle \cdot \nabla \langle c \rangle - \nabla \langle c' \mathbf{u}' \rangle + D_m \nabla^2 \langle c \rangle + \langle s \rangle \quad (2.9)$$

Le terme croisé des grandeurs turbulentes, $\nabla \langle c' \mathbf{u}' \rangle$, est difficile à calculer. Il est approximé par la théorie gradient-transport comme dans [Winiarek2014a, p. 26]. Cette théorie assimile la moyenne des termes turbulents croisés à un phénomène convectif tel que,

$$\langle c' \mathbf{u}' \rangle = -\mathbf{K} \nabla \langle c \rangle \quad (2.10)$$

où \mathbf{K} représente le tenseur de diffusion turbulente induite par le phénomène de la turbulence. En se séparant des notations de Reynolds pour s'intéresser aux valeurs moyennes du champ de vent et de concentration, nous obtenons l'équation d'advection-diffusion simplifiée,

$$\partial_t c + \mathbf{u} \cdot \nabla c - \nabla \cdot (\mathbf{K} \nabla c) + D_m \nabla^2 c = s \quad (2.11)$$

De plus, la diffusion moléculaire est bien souvent négligeable, ainsi l'équation aux dérivées partielles d'Advection-Diffusion (ADPDE) simplifiée s'écrit,

$$\partial_t c + \mathbf{u} \cdot \nabla c - \nabla \cdot (\mathbf{K} \nabla c) = s \quad (2.12)$$

où le champ vectoriel de vélocité de l'air \mathbf{u} et le champ de concentration c sont considérés en valeurs moyennes au sens de Reynolds. En pratique, le tenseur de diffusion \mathbf{K} est construit de manière telle que seuls les champs de diffusion horizontaux et verticaux sont non nuls. Ainsi, le tenseur de diffusion \mathbf{K} est diagonal [Winiarek2014a, p.74] ce qui amène à considérer l'équation d'ADPDE suivante,

$$\partial_t c + \sum_{i=1}^3 u_i \partial_i c - \sum_{i=1}^3 \partial_i (K_i \partial_i) c = s \quad (2.13)$$

où u_i et K_i représentent les composantes du champ moyen de vélocité de l'air et de diffusion turbulente selon la composante x_i .

Conditions aux limites dans le cadre de la pollution atmosphérique : Généralement, pour la pollution urbaine, deux types de conditions aux limites sont à considérer : des conditions de flux nul ($\nabla(c(\mathbf{x})) \cdot \mathbf{n} = 0$) sur le sol et les obstacles, et l'imposition de concentrations nulles ($c(\mathbf{x}) = 0$) sur les côtés du domaine de simulation. D'après nos observations et à la manière du logiciel PMSS, une condition de flux nul est de préférence à considérer sur le plafond de simulation. Nous n'avons pas considéré dans cette thèse de dépôt de polluants au sol par lessivage ou dépôt sec autrement la condition de flux au sol ne serait plus nulle et serait remplacée par ces conditions. Des détails sur les conditions de lessivage et de dépôt sec sont précisés dans [Bruse2007]. Quant à la condition initiale, elle peut être fixée comme nulle si le niveau de concentration initiale n'est pas connu ou alors initialisée à des valeurs connues calculées à un temps précédent, estimée grâce à des mesures ou encore fournie par un modèle tiers à grande maille.

Nous avons présenté le phénomène physique d'intérêt à modéliser afin de simuler la dispersion de polluants. Nous introduisons dans la suite des méthodes de simulation de la circulation de l'air et du transport de polluants.

2.3 Méthodes de simulation de la dispersion atmosphérique de polluants

2.3.1 Modélisation de la météorologie

Un vaste champ d'étude existe pour simuler la circulation de l'air, notamment à l'échelle urbaine. Une modélisation correcte des champs de circulation d'air est nécessaire dans les études de pollution urbaine car ces champs sont très sensibles aux obstacles dans l'espace. Plusieurs approches sont considérées pour modéliser la circulation de l'air. Les méthodes visant à résoudre numériquement les équations agissant sur un fluide (dans ce travail, il s'agit de l'air) peuvent être catégorisées dans la mécanique des fluides numériques (MFN ou CFD en anglais). Nous pouvons citer les méthodes DNS (simulation numérique directe), LES (simulation des grandes structures de la turbulence) ou encore RANS (équations de Navier-Stokes moyennées à la manière de Reynolds). Des méthodes de calcul existent en dehors des CFD comme l'utilisation de modèles analytiques comme la méthode de Röckle utilisée dans [Moussafir2004, Brown2015, Bernard2023] qui met en oeuvre des lois empiriques pour simuler des champs de vent. Présentons quelques méthodes de mécanique des fluides numériques, en particulier les RANS qui sont utilisées dans cette thèse.

DNS

La simulation numérique directe (DNS en anglais) consiste à résoudre les équations de Navier-Stokes sans simplifications. Par exemple, la turbulence dans l'équation du mouvement en (2.4)

est totalement prise en compte. Les DNS sont précisément représentatives du phénomène mais sont très coûteuses en calcul [Mortensen2011]. Bien souvent, l'utilisation de DNS est limitée à la recherche, ce qui ne l'empêche pas d'être utilisée dans des travaux d'assimilation de données comme dans [Chandramouli2020] ou [Mons2016] pour des applications sur des flux gazeux.

LES

Les Simulations des Grandes Structures de la turbulence (SGS), nommées LES en anglais sont une simplification des équations de Navier Stokes. Les LES ne traitent que les turbulences de grande échelle spatiale. Ce filtrage spatial selon l'intensité de la turbulence permet de simuler des flux réalistes tout en économisant beaucoup de temps de calcul [Zhiyin2015]. Cependant les LES restent coûteuses dans des applications pratiques. Par exemple dans [Mons2017], 8 h de calcul sur 96 coeurs AMD Opteron 8431 sont nécessaires pour réaliser une simulation LES sur un quartier de $2 \times 2 \text{ km}^2 \times 0,6 \text{ km}$.

RANS

Les simulations RANS sont des méthodes populaires mises en avant dans les années 1970 grâce aux travaux de Launder et Spalding [Launder1974]. Elle s'appuient sur la décomposition de Reynolds des équations de Navier-Stokes pour contourner le calcul de la turbulence locale du flux d'intérêt. La plupart du temps, les modèles RANS appliqués à la micro-météorologie considèrent des fluides newtoniens incompressibles. Dans ce cas, les équations de Navier-Stokes moyennées deviennent pour chaque composante i ($i = \{1, 2, 3\}$),

$$\partial_t u_i + \sum_{j=1}^3 u_j \partial_{x_j} u_i = -\frac{1}{\rho} \partial_i p + g_i + \sum_{j=1}^3 \partial_j \nu (\partial_j u_i + \partial_j u_i) + \sum_{j=1}^3 \partial_j R_{ij} \quad (2.14)$$

$$\sum_{i=1}^3 \partial_i u_i = 0$$

où u_i est la composante du champ de vent moyen associé à une coordonnée cartésienne x_i et où R est le tenseur de contraintes de Reynolds. R prend en compte les contraintes exercées par les fluctuations turbulentes au sein du fluide. L'expression de chaque élément de R est, $R_{ij} = -\langle u'_i u'_j \rangle$. Ainsi R_{ij} est la moyenne du produit de termes fluctuants issus de la décomposition de Reynolds. Plusieurs modèles existent pour expliciter ce terme de contraintes en fonction des valeurs moyennes de pression et de vitesse. Le modèle le plus populaire en RANS reste le modèle $k - \epsilon$ pour les applications liées à la météorologie et à la pollution. Ce modèle est utilisé par divers auteurs [Defforge2019b, Hammond2019, Mons2017]. D'autres modèles de fermeture de la turbulence existent comme celui de Spalart-Allmaras [Spalart1992]. Dans le cadre de ce travail nous utilisons MICRO-SWIFT-M de la suite PMSS qui implémente un modèle de fermeture de la turbulence développé par [Oldrini2013]. Ce modèle permet d'obtenir des champs météorologiques à l'état stationnaire avec un temps de calcul contenu. Le tenseur de contraintes de turbulence de Reynolds est obtenu par un modèle de fermeture turbulente à l'ordre 0 (sans ajout d'équation pour exprimer le tenseur de Reynolds) en utilisant la théorie de la longueur de mélange. La longueur de mélange est obtenue par le produit de la constante de Von Karman κ avec la distance aux frontières solides d_b ,

$$l_{mix} = \kappa d_b \quad (2.15)$$

La viscosité cinématique turbulente est modélisée par,

$$\nu_t = l_{mix}^2 \sqrt{\sum_{i=1}^3 \sum_{j=1}^3 S_{ij}^2} \quad (2.16)$$

avec S désignant le tenseur des déformations défini comme $S_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$. L'équation du mouvement devient alors,

$$\partial_t u_i + \sum_{j=1}^3 u_j \partial_j u_i = -\frac{1}{\rho} \partial_i p + g_i + \sum_{j=1}^3 \partial_j (\nu + \nu_t) (\partial_i u_j + \partial_j u_i) \quad (2.17)$$

L'équation de continuité est modifiée par l'ajout d'une compressibilité artificielle [Oldrini2013]. La vitesse artificielle β modélisant les vagues de pression est alors introduite dans l'équation de continuité,

$$\frac{1}{\beta} \partial_\tau P = - \sum_{i=1}^3 \partial_i U_i \quad (2.18)$$

La solution se calcule itérativement grâce à une méthode à pas de temps fractionnaire. Le champ de vitesse est d'abord calculé en excluant la pression, puis la pression est mise à jour en se servant de l'équation de continuité. La pression mise à jour est injectée dans l'équation du mouvement. Les itérations se succèdent jusqu'à convergence d'un champ de vitesse stationnaire. Le détail du schéma numérique est explicité dans [Oldrini2013].

Les RANS ne sont pas valides aux conditions limites. C'est pourquoi des fonctions semi analytiques sont utilisées aux bords appelées « wall functions » en anglais. Un cours de l'Université de Stanford détaille les points à considérer pour utiliser les bonnes approximations selon les situations ³.

Conclusion sur les modèles de vent

Un modèle de vent de type RANS peut convenir en réalisme et en temps de calcul pour des études de pollution urbaine. Le développement de tels solveurs est complexe, la plupart des études mentionnées [Defforge2019a, Hammond2019, Mons2017] utilisent des solveurs établis tels que Code_Saturne, COMSOL Multiphysics ou encore PowerFLOW. Le logiciel Micro-SWIFT-M [Oldrini2013] faisant partie de la suite PMSS [Moussafir2004] est utilisé dans cette étude. Ce logiciel validé en soufflerie offre une bonne précision pour calculer les champs météorologiques de vitesse et de diffusion donnés en entrée du modèle de transport de polluants (2.13) et est économe en temps de calcul.

2.3.2 Un aperçu des méthodes de simulation de phénomènes de dispersion de polluants

Il existe différentes catégories de modèles capables de simuler un transport de polluants réaliste qui approximent les solutions de l'équation d'ADPDE (2.13).

Il existe divers types de modèles de dispersion de polluants : les modèles gaussien de dispersion, les modèles à boîtes et les modèles à physique discrétisée. Les modèles gaussiens et modèles à bouffées gaussiennes ne nécessitent pas le calcul de la micro-météorologie sur le domaine de simulation. Le vent y est considéré comme uniforme et commandé par sa direction ainsi que son intensité ou alors, la météorologie est fourni par une grille de vent à maille grossière obtenu par interpolation de mesures météorologiques ou encore par les données fournies par un modèle météorologique à méso-échelle comme WRF (Weather Research and Forecasting [Skamarock2008]). Les modèles à boîtes (ou canyon) considèrent des modèles de transports gaussiens en espaces ouverts (plaines et montagnes) tandis que la météorologie et la dispersion en milieu urbain dense sont calculées par bilan de masse au sein de "rues-canyon". Enfin, une troisième catégorie de modèles sont les modèles à physique discrétisée où les équations différentielles qui régissent la circulation

³. Cours de RANS par l'université de Stanford : <http://web.stanford.edu/class/me469b/handouts/turbulence.pdf>

de l'air et le transport des polluants sont discrétisées et résolues sur des cellules remplissant un espace de simulation.

Modèles gaussiens

Modèle gaussien des panaches Dans l'hypothèse d'un vent homogène décrit par sa direction et son intensité, connaissant la stabilité atmosphérique et à source polluante constante et ponctuelle, il existe une solution analytique au problème d'advection diffusion (2.13) telle que la concentration en polluant en tout point de l'espace devient, en un axe x placé dans l'axe du vent,

$$c(x, y, z) = \frac{Q}{U} \frac{1}{2\pi\sigma_y\sigma_z} \exp\left(\frac{-y^2}{2\sigma_y^2}\right) \exp\left(\frac{-z^2}{2\sigma_z^2}\right) \quad (2.19)$$

où c désigne la concentration en $\mu\text{g m}^{-3}$, Q , le débit de la source en $\mu\text{g s}^{-1}$ et U , la vitesse du vent en m s^{-1} . Les écarts-types σ_y et σ_z permettent de décrire la forme du panache en abscisse x . σ_y et σ_z dépendent de la stabilité atmosphérique et leur valeur décroît lorsque x croît. Plusieurs modèles différents peuvent modéliser la décroissance des écarts-types. Des variations de ce modèle existent pour prendre compte de la hauteur de rejet ou encore de la topographie du terrain. Le modèle gaussien des panaches est par exemple utilisé dans un travail d'Estimation de Terme Source [Singh2015]. Compte tenu de sa faible capacité à prendre en compte des domaines complexes, nous n'avons pas retenu cette approche.

Modèle à bouffées gaussiennes Ce modèle est une variante du modèle gaussien des panaches capable de traiter l'instationnarité du débit de source ainsi que les variations spatiales et temporelles des conditions météorologiques telles que la direction et l'intensité du vent. Ces modèles sont applicables et représentatifs des transports de polluants lorsque le domaine de simulation n'est pas contraint par des obstacles physiques au sol. Ainsi il est adapté aux modélisations qui considèrent des plaines ou lorsque les polluants voyageant assez haut dans l'atmosphère pour considérer que le domaine de simulation n'est pas contraint à hauteur de rejet. Les modèles de panaches à bouffées peuvent par ailleurs être adaptés à la topographie et rendent par exemple possible la modélisation de déplacement de panache le long de vallées de manière satisfaisante. Une telle modélisation est visible en figure 2.5 où le modèle à bouffées gaussiennes (figure 2.5, droite) est capable de simuler de manière réaliste le déplacement d'un panache de polluants dans une vallée tandis que le modèle gaussien échoue (figure 2.5, gauche).

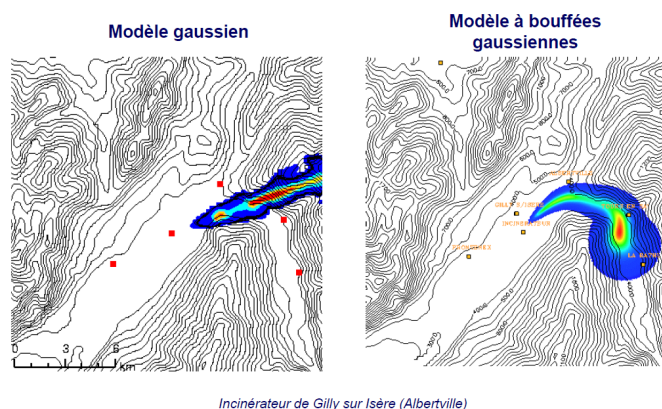


FIGURE 2.5 – Exemple d'application où le modèle de panache à bouffées conduit à un résultat réaliste et où l'approche du panache gaussien échoue. Illustration reproduite avec l'aimable autorisation d'ARIA Technologies.

De plus, les modèles gaussiens de panaches étant formulés à partir d'une source ponctuelle, une paramétrisation d'une source non ponctuelle comme combinaison de sources ponctuelles est nécessaire. Soulhac et al. utilisent par exemple des modèles à bouffées gaussiennes pour modéliser le transport de polluants au dessus d'une canopée urbaine [Soulhac2017].

Les modèles gaussiens ont pour avantage leur facilité de mise en oeuvre et leur temps de calcul restreint. Cependant ils ne peuvent considérer des domaines hétérogènes comme les quartiers urbains que de manière grossière. C'est la raison pour laquelle nous n'avons pas retenu cette approche dans cette thèse. Ils restent aujourd'hui une référence dans la modélisation de la qualité de l'air.

Modèles à boîtes

Les modèles à boîtes considèrent des bilans de masses pour modéliser un transport d'espèces chimiques. De par la géométrie des quartiers urbains denses tels que les centres des grandes villes, il est possible de considérer de tels quartiers comme des réseaux de canyons propices à raisonner en bilans de masse à leurs extrémités (en amont et en aval du canyon et au-dessus). Une illustration de ce raisonnement est visible sur la figure ci-dessous,

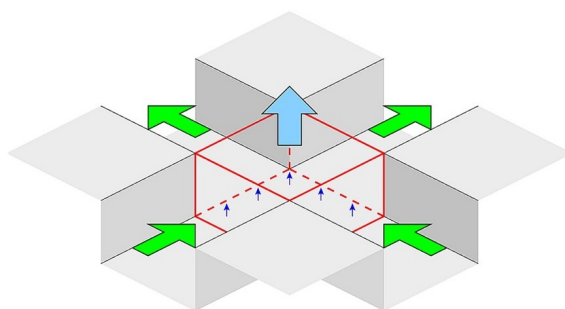


FIGURE 2.6 – Modélisation par un réseau de rues, [Soulhac2011]

SIRANE est un modèle à boîtes bien connu capable de modéliser le déplacement de polluants en milieu urbain. Il retourne des concentrations moyennes sur des rues. Le modèle des panaches à bouffées [Soulhac2016] permet de simuler le transport de polluants qui sortent du haut des canyons ainsi que le transport dans des zones non obstruées par le bâti telles que les parcs et les rivières. Toutefois cette approche est peu adaptée à la résolution de problèmes inverses du fait que les concentrations soient moyennées dans les rues-canyons. C'est la raison pour laquelle nous ne l'avons pas retenue.

Modèles lagrangiens

Les modèles de dispersion peuvent aussi être résolus de manière lagrangienne. Dans le principe, la méthode suit le mouvement de particules de polluants soumises aux contraintes physiques de modèle. La simulation du déplacement d'un grand nombre de particules donne alors une moyenne représentative du phénomène réellement en jeu. Les approches lagrangiennes deviennent populaires pour modéliser précisément la dispersion atmosphérique de polluants et offrent des performances comparables voire supérieures aux approches eulériennes [Bahlali2018]. Le modèle PSPRAY de la suite PMSS [Moussafir2004] est également capable de modéliser un transport de polluant de manière lagrangienne. L'approche lagrangienne mise en oeuvre dans PMSS servira de référence de comparaison dans la suite de cette thèse.

Méthode de discrétisation pour les modèles eulériens

Le transport de polluant peut être vu de manière eulérienne. Dans ce cas, on cherche à résoudre une équation aux dérivées partielles en procédant à une discrétisation de l'espace où on essaie d'approcher la mathématique du phénomène, par exemple en écrivant des différences finies, ou encore employant une méthode de Galerkin par éléments finis.

Les modèles eulériens pour le calcul du transport de polluants cherchent à résoudre l'équation (2.13) en se plaçant dans un domaine spatial fixe (référentiel eulérien). Plusieurs méthodes de résolution de modèles eulériens existent telles que, la méthode des différences finies, des volumes finis, des éléments finis ou encore les différences finies supportées par des fonctions de base radiales (RBF-FD).

Différences finies : La première des méthodes est bien souvent jugée comme la plus facile à programmer. Par conséquent elle est couramment abordée dans les universités et dans nombre d'études. Elle consiste à discrétiser des équations différentielles par des approximations de Taylor par exemple à l'ordre 1, pour un schéma centré,

$$\left(\frac{\partial u}{\partial x}\right)_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x} + o(\Delta x)^2 \quad (2.20)$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_i = \frac{u_{i+1} + u_{i-1} - 2u_i}{\Delta x^2} + o(\Delta x)^2 \quad (2.21)$$

Ces approximations conduisent à une discrétisation des équations différentielles à résoudre.

Les différences finies ont quelques avantages mais aussi des inconvénients. Parmi les avantages : la formulation est simple dans le principe rendant la programmation relativement aisée ; la méthode est efficace pour les domaines simples (rectangulaires avec conditions limites aux bords). Comme inconvénients on peut citer le manque d'efficacité et la difficulté à programmer des différences finies en présence d'un domaine géométriquement contraint (ex : non rectangulaire, obstacles au milieu etc). La méthode présente également des instabilités pour modéliser des phénomènes discontinus comme les problèmes de rupture de barrage ou encore dans la modélisation de phénomènes très advectifs [Cellier2018].

Exemple de Méthode des Différences Finies : Schématiquement pour modéliser un phénomène d'Advection-Diffusion en 2D, l'espace est discrétisé en une grille généralement carrée contenant n_x lignes et n_y colonnes contenant alors $N = n_x \times n_y$ points de discrétisation. Une variable c peut alors être approximée sur un ensemble de points $\{x^k\}_{k=1}^N$. Un point d'approximation positionné en coordonnées de grille (i, j) est illustré en Fig. 2.7. La discrétisation en différences finies pour l'ADPDE en régime stationnaire et à champs de vents et diffusions uniformes s'écrit aux nœuds intérieurs du domaine de simulation (hors conditions de bord),

$$u_x \frac{c_{i+1,j} - c_{i,j}}{\Delta x} + u_y \frac{c_{i,j+1} - c_{i,j}}{\Delta y} - k_x \frac{c_{i+1,j} + c_{i-1,j} - 2c_{i,j}}{\Delta x^2} - k_y \frac{c_{i,j+1} + c_{i,j-1} - 2c_{i,j}}{\Delta y^2} = s_{i,j} \quad (2.22)$$

Le but des différences finies est ensuite d'assembler un système linéaire du type,

$$\mathbf{A}\mathbf{c} = \mathbf{s} \quad (2.23)$$

où \mathbf{A} est une matrice de différentiation de taille $N \times N$, avec les composantes du vecteur \mathbf{c} de taille $n_x \times n_y$ contenant les composantes $c_{i,j}$ $(i, j) \in \{1, \dots, n_x\} \times \{1, \dots, n_y\}$ ordonnées de la manière suivante :

$$\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_y})^T \quad (2.24)$$

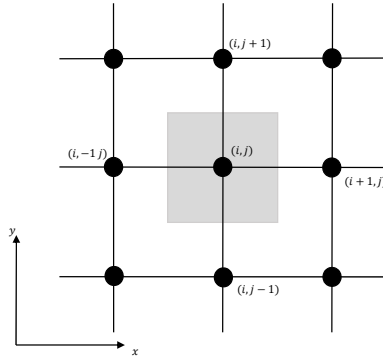


FIGURE 2.7 – Exemple de discrétisation 2D pour les différences finies. Espace grisé : cellule de calcul centré au point de coordonné (i, j) . Les points voisins servent à calculer les dérivées avec le développement de Taylor.

chaque vecteur \mathbf{c}_i de dimension n_x représente les valeurs de la solution recherchée sur une ligne de grille. Similairement, le terme source-conditions-limites \mathbf{s} se construit de la même manière. La discrétisation des opérateurs différentiels montrés en (2.22) étant exprimée dans le système assemblé (2.23). Dans cet exemple de différences finies pour l'ADPDE à champs diffusifs et advectifs uniformes, nous pouvons poser un opérateur différentiel qui traduit la physique points intérieurs de Ω (Ω),

$$\mathcal{L} = \sum_{i=1}^2 \mathbf{U}_i \partial_i + \mathbf{K}_i \partial_{ii}, \quad (2.25)$$

avec \mathbf{U}_i et \mathbf{K}_i désignant les vecteurs discrets de composantes des champs vectoriels de vitesse de l'air et de diffusion turbulente. Similairement, un opérateur \mathcal{B} tenant compte des conditions limites sur les bords de domaine $\partial\Omega$ peut être considéré tel que le problème à discrétiser se ramène à

$$\mathcal{L}c(x) = s(x), \quad x \in \Omega, \quad \mathcal{B}c(x) = 0, \quad x \in \partial\Omega. \quad (2.26)$$

Les différences finies permettent alors d'exprimer l'opérateur différentiel \mathcal{L} (et similairement l'opérateur aux bords \mathcal{B}) appliqué à c comme une somme pondérée de n valeurs de c voisines,

$$\begin{aligned} \mathcal{L}c(\mathbf{x}_k) &= \sum_{i=1}^n w_i(\mathbf{x}_k) c(\mathbf{x}_i) \\ &= \mathbf{w}^T \mathbf{c} \end{aligned} \quad (2.27)$$

et la matrice de différentiation \mathbf{A} impose cette pondération pour tout $\mathbf{x}_k \in \Omega$.

Volumes finis : Les volumes finis reposent sur une discrétisation des cellules définies en volumes. L'équation différentielle à résoudre est intégrée en temps et en espace. Les intégrations en espace sont simplifiées par le théorème de Gauss et les intégrales résultantes sont discrétisées sur les éléments de volume. Cette méthode est appréciée des physiciens pour sa programmation plutôt représentative de la physique. Elle est en vogue depuis les années 2000 lorsque Randall J. LeVeque publie un ouvrage de référence sur la méthode [LeVeque2002]. Plusieurs solveurs réputés utilisent cette méthode tels que *OpenFOAM* et *Code_Saturne* [Archambeau2004, Guerrero2021]. Une bibliothèque développée pour la science des matériaux, FiPy, est disponible sur Python. Elle ne présente cependant pas d'exemple dans le domaine de la CFD [Guyer2009].

Éléments finis : La méthode des éléments finis est une méthode de discrétisation reposant sur des maillages en général polygonaux. Elle utilise une fonction test qui multiplie le système d'équations différentielles à résoudre. Grâce à des intégrations par parties, le problème se ramène à résoudre un système variationnel. Les polynômes d'interpolation, souvent de Lagrange, sont utilisés pour discrétiser la forme variationnelle là où les coordonnées des fonctions de test correspondent aux nœuds du maillage. Plus de détails sur la méthode sont décrits dans [Logg2012, Chapitre 2].

RBF-FD : Les méthodes sans maillage, en particulier les différences finies supportées par des fonctions de base radiales (RBF-FD) sont sujettes à un intérêt croissant depuis la fin des années 2000. Ces méthodes sans maillage, sont efficaces et tolèrent les obstacles. La bibliothèque la plus aboutie à ce sujet est *Medusa* [Slak2021]. Trevor Hines a aussi programmé une bibliothèque prometteuse en Python [Hines2015]. Cette méthode d'approximation d'EDP a en particulier la réputation d'être précise malgré un faible nombre de points de calcul [Flyer2012, Tominec2021c] comparée à des Méthodes des Éléments Finis. Dans cette thèse, nous avons choisi d'utiliser des approches à base de RBF-FD que nous allons présenter de manière détaillée en 2.4.

Conclusion sur les méthodes de résolution d'équations différentielles

Pour conclure la méthode des différences finies est «instinctive» mais peut montrer des limites notamment pour la modélisation de phénomènes où la diffusion est faible par rapport à l'advection. Les méthodes des volumes finis et éléments finis sont adaptées à toutes les géométries mais sont plus difficiles à programmer. Les éléments finis demandent cependant un niveau d'abstraction supplémentaire, en revanche la méthode est plus précise que les volumes finis [Guyer2009]. Dans cette thèse nous faisons le choix de modéliser l'équation de dispersion représentée par l'ADPDE par des approches basées sur les RBF-FD qui ont des propriétés intéressantes de précision avec un faible nombre de points, d'opérateurs différentiels creux ainsi qu'un traitement intéressant du domaine de simulation sans maillage qui apporte de la flexibilité et de la simplicité à ces méthodes. Détaillons cette approche dans la prochaine section.

2.4 Approximations RBF-FD et LS-RBF-FD des Equations aux Dérivées Partielles

Les Equations aux Dérivées Partielles constituent des modèles en dimension infinie. Un grand nombre de problèmes physiques comme l'équation d'Advection-Diffusion (ADPDE) ne comportent pas de solution analytique pour des domaines où la géométrie est contrainte par la topologie et la présence d'obstacles. Le modèle analytique des panaches gaussiens ou à bouffées ne convient notamment pas pour modéliser le transport de polluants en milieu urbain. Il faut donc se ramener à des problèmes discrets où on cherche à construire un vecteur de solution discret dont la dimension dépend de la résolution et de la taille d'un domaine discrétisé.

Les modèles eulériens d'advection diffusion ont prouvé leur capacité à modéliser le transport scalaire [Archambeau2004, Bruse2007] et peuvent se concevoir formellement assez facilement ce qui facilite ensuite leur programmation numérique. De plus les approches eulériennes ont l'avantage de discrétiser les équations différentielles sous la forme de systèmes linéaires de type $\frac{d}{dt} \mathbf{c} = \mathbf{A} \mathbf{c} + \mathbf{b}$ où \mathbf{A} est une matrice système représentative de la physique, \mathbf{c} le vecteur de solution discrétisé à retrouver et \mathbf{b} , le vecteur de contraintes comprenant le terme source et les conditions aux limites. L'intégration en temps de ce genre de système peut alors se faire de manière relativement aisée en utilisant des techniques d'intégration numérique comme les schémas de Runge-Kutta ou d'autres

schémas divers. L'utilisation de modèles eulériens permet donc d'exprimer assez explicitement la physique et d'utiliser des procédures d'inversion comme le filtre de Kalman pour les problèmes inverses. Cependant, les modèles eulériens produisent de la diffusion numérique qui est source d'erreur de calcul.

Les approches lagrangiennes ont prouvé leur capacité à modéliser le transport scalaire [Mousafir2004, Bahlali2019] mais nécessitent la formulation de modèles effectuant des lancers de nombreuses particules puis d'appliquer à chaque pas de temps une loi de trajectoire sur chaque particule, une loi de trajectoire dépendante du vent et des obstacles. Il faut également simuler un grand nombre de particules par cellule de grille pour obtenir un nombre moyen de particules par grille cohérent ce qui peut être exigeant en temps de calcul surtout avec des mailles fines et des grands domaines de simulation. Les modèles lagrangiens ont l'avantage de ne pas présenter de problème de diffusion numérique, et peuvent fournir des précisions comparables aux modèles eulériens [Bahlali2019, Armand2023]. En résumé les modèles lagrangiens comme eulériens peuvent donner des résultats proches comme présenté dans [Nielinger2004].

Le logiciel *Envi-met* [Bruse2007] montre que la modélisation tridimensionnelle de dispersion atmosphérique en différences finies est possible. Tout en partageant des similitudes avec les différences finies classiques, les différences finies supportées par des fonctions de base radiales (RBF-FD) ont démontré sur un certain nombre de cas de tests des capacités à résoudre avec grande précision et sur des nœuds de calcul très épars des problèmes complexes [Tominec2021c]. Les auteurs cités ont par exemple fait converger des solutions LS-RBF-FD (Least Squares Radial Basis Functions Finite Differences) vers des solutions de références produites par des modèles à éléments finis exécutés sur des domaines de calcul plus résolus. Ce type de comportement a été également observé par Flyer et al. sur une étude où les RBF-FD offraient une précision supérieure aux méthodes de Galerkin Discontinues à résolution de domaine de calcul égale [Flyer2012]. Cette méthode sans maillage offre également une flexibilité sur le placement des points de calculs et permet de produire des domaines structurés ou non structurés de simulation d'une manière relativement simple. Pour ces raisons, nous avons choisi d'explorer ce type d'approche dans le cadre de cette thèse.

2.4.1 Quelques applications des RBF-FD pour résoudre des EDP

Bien qu'encore peu connues, les RBF-FD ont été appliquées dans de nombreux exemples pour la science et l'ingénierie. Elles ont été appliquées aux géosciences [Bayona2015, Flyer2012, Matthews2022, Martin2017], à l'ingénierie de la combustion [Bayona2021], à des problèmes biomédicaux [Tominec2021c], sismiques [Martin2017], pour la modélisation des interfaces, la mécanique des fluides appliquée aux barrages [Dehghan2017, Raoult2019], le calcul à haute performance [Bollig2012, Depolli2022, Jančić2020], ou encore l'épidémiologie [Oliver2022] pour citer quelques exemples. Des bibliothèques informatiques ont également été développées comme Medusa, une bibliothèque en C++ avec un interfaçage de haut niveau pour construire et résoudre des EDP de tout type y compris fluidiques. Sur Python, la bibliothèque RBF connaît également un certain succès [Slak2021, Hines2015]. Pour donner du contexte aux travaux sus-mentionnés, Bayona et Flyer ont modélisé un circuit électrique global à l'échelle de la terre et en prenant compte de la topologie de la croûte terrestre pour estimer la densité de courant électrique dans l'atmosphère [Bayona2015]. Flyer et al. ont aussi proposé des modélisations de équations de Saint-Venant (SWE, Shallow Water Equations) pour des domaines sphériques et ont montré que les RBF-FD sont plus rapides en calcul à précision équivalente que la méthode de Galerkin Discontinue faisant référence dans ce domaine [Flyer2012]. Tominec et al. ont modélisé avec succès la déformation d'un diaphragme humain [Tominec2021c] avec une méthode proche des RBF-FD, les RBF-FD en moindres carrés et ont obtenu des solutions de précision proches à des calculs en éléments finis le tout avec moins de points de calculs. Les modèles physiques à méso-échelles sont aussi étudiés : Dehghan et al. ont utilisé une modélisation RBF-FD pour modéliser les équations de Saint-Venant

pour la rupture de barrages, et ont utilisé les solutions pour construire un modèle réduit par Analyse en Composantes Principales [Dehghan2017]. Les équations d'advection-diffusion ont aussi été à l'étude. Liu et al. ont par exemple modélisé un transport de polluants en régime transitoire sur un domaine urbain 2D [Liu2019].

2.4.2 Introduction aux RBFs

Attardons-nous à présent sur quelques propriétés des RBF. Les RBF présentent quelques caractéristiques remarquables pour travailler avec des équations aux dérivées partielles :

- Les RBF ont la propriété d'interpoler de manière satisfaisante toute fonction assez lisse définie dans un espace de dimension quelconque 2D, 3D ou plus.
- Les RBF peuvent servir à décrire des domaine spatiaux en partitionnant l'espace par des centres de bases. Les EDP peuvent ainsi être résolues sans maillage.
- Les RBF-FD peuvent être utilisées pour construire des opérateurs différentiels spatiaux comme des gradients ou des laplaciens directement à partir de la forme forte ou faible (LS-RBF-FD) des EDP.

Interpolation RBF Un grand nombre de fonctions de base radiales (RBF) existent pour résoudre des EDP. Les RBF montrent d'excellentes propriétés d'interpolation et cela pour des espaces de dimensions quelconques (2D, 3D, 3D etc.). Une fonction quelconque $f : y \in \Omega \mapsto f(y) \in f(\Omega) \subset \mathbb{R}^d$ (d étant la dimension de l'espace) peut être approximée par un nombre fini de fonctions de bases (RBF) ϕ centrées sur des nœuds $\{x_i\}_{i=1}^{N_x} = X \subset \Omega$. Notons \hat{f} la fonction interpolante de f alors,

$$\hat{f}(y) = \sum_{i=1}^{N_x} a_i \phi(\|y - x_i\|), \quad \forall y \in \Omega. \quad (2.28)$$

où $\|\cdot\|$ désigne la norme euclidienne appliquée à un vecteur de dimension d . La distance $\|y - x_i\|$ est qualifiée de rayon de la RBF. Ce rayon est noté $r = \|y - x_i\|$ avec $y \in \Omega$ un point quelconque de l'espace et $x_i \in X \subset \Omega$ le centre d'une base RBF. En considérant une fonction interpolante \hat{f} évaluée sur l'ensemble X des points de centre de bases $\{x_i\}_{i=1}^{N_x}$ telle que $\hat{\mathbf{f}} = (\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_{N_x}))$, nous pouvons exprimer le problème d'interpolation évalué sur les points $\{x_i\}_{i=1}^{N_x}$, sous la forme du système linéaire

$$\hat{\mathbf{f}} = \mathbf{\Phi} \mathbf{a}, \quad (2.29)$$

avec $\mathbf{a} = (a_1, \dots, a_{N_x})^T$ le vecteur de coefficient d'interpolation et $\mathbf{\Phi}$ une matrice de collocation de Fonctions de Base Radiales telle que,

$$\mathbf{\Phi} = \begin{pmatrix} \phi(\|x_1 - x_1\|) & \dots & \phi(\|x_{N_x} - x_1\|) \\ \vdots & \ddots & \vdots \\ \phi(\|x_{N_x} - x_1\|) & \dots & \phi(\|x_{N_x} - x_{N_x}\|) \end{pmatrix}. \quad (2.30)$$

Ainsi si $\mathbf{\Phi}$ est inversible les coefficients d'interpolation se trouvent en inversant le système avec,

$$\mathbf{a} = \mathbf{\Phi}^{-1} \hat{\mathbf{f}}. \quad (2.31)$$

Mathématiquement, il a été prouvé que \hat{f} est une bonne approximation de f sous condition d'espace régulier des centres de bases x_i . Le problème d'interpolation (2.29) admet alors une unique solution de vecteur de coefficients $\mathbf{a} = (a_1, \dots, a_{N_x})^T$ pour différentes bases [Fornberg2015c, Chapitre 3 p. 39-47] en particulier pour les bases "PHS" (splines polyharmoniques) [Barnett2015a, p. 13 théorème 2].

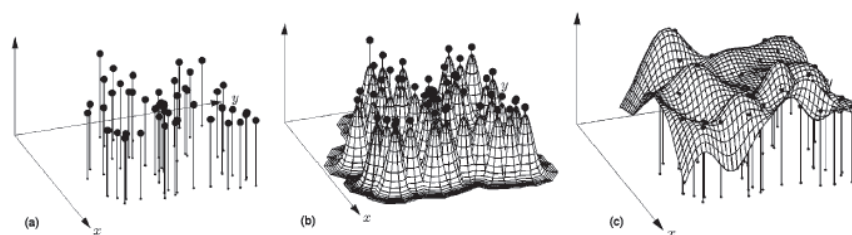


FIGURE 2.8 – (a) Exemple de données dispersées sur un plan, (b) Ensemble de RBF gaussiennes centrées sur chaque point de donnée, (c) Fonction interpolante comme unique combinaison de fonctions RBF gaussiennes passant par les points de données. Source : [Fornberg2015a, p. 40].

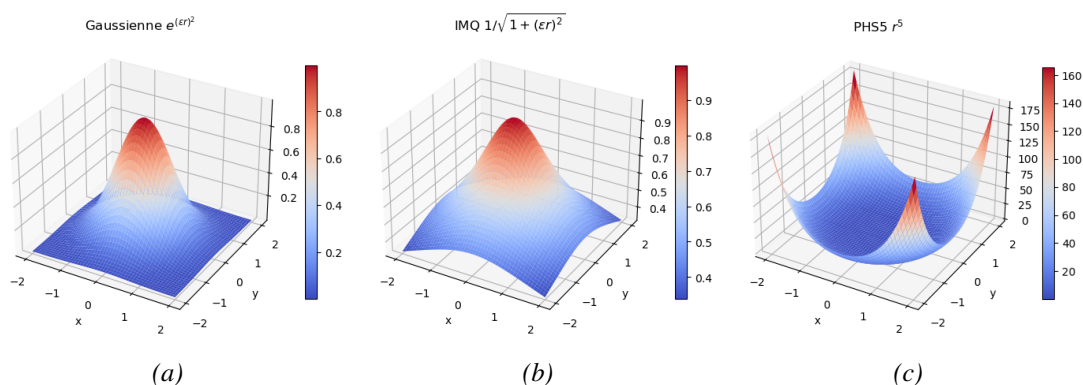


FIGURE 2.9 – Visualisation de RBF centrées en $x_i = (0, 0)$ et affichées sur $(x, y) \in [-2, 2]^2$. (a) RBF gaussienne (Gaussienne), (b) RBF Inverse multiquadrique (IMQ), (c) RBF spline polyharmonique d'ordre 5 (PHS5).

La reconstruction d'une fonction définie dans l'espace par interpolation basée sur des RBF est illustrée en Fig. 2.8.

Le choix des RBF pour résoudre des PDE est en soi un champ d'étude. Aujourd'hui, les plus populaires sont les splines polyharmoniques (PHS) [Raoult2019, Mathews2022, Tominec2021c, Jančić2020, Depolli2022] mais les RBF gaussiennes [Martin2017, Dehghan2017] et multiquadriques [Kansa1990b, Liu2019, Bayona2015] restent très utilisées car elles conduisent parfois à une réduction du nombre de fonctions nécessaires pour obtenir une approximation de précision donnée. Quelques fonctions de bases sont présentées en Tableau 2.1. Exceptées les PHS, les RBF requièrent un facteur de forme ϵ qui commandent leur étalement dans l'espace (propriété de localité, similaire au choix de l'échelle d'une ondelette). Le choix judicieux de ce facteur de forme doit permettre le meilleur conditionnement numérique de l'interpolation RBF présenté en (2.28). Les PHS gagnent en intérêt grâce à leur absence de facteur de forme. Leur robustesse est aussi accrue par l'utilisation de polynômes de régularisation dans le sens où l'interpolation est moins sujette à des erreurs d'oscillations [Bayona2017, Flyer2016b]. On qualifie d'erreur d'oscillations un type d'erreur où la fonction interpolante oscille autour de la valeur véritable à interpoler. Ce phénomène est particulièrement visible quand la fonction à interpoler présente des forts gradients, par exemple des valeurs élevées voisines avec des valeurs proches de 0.

Type de RBF	Fonction radiale $\phi(r)$
Spline Polyharmonique (PHS)	$r^{2m-1}, m \in \mathbb{N}, m \geq 1$
Gaussienne (GA)	$e^{-(\epsilon r)^2}$

Type de RBF	Fonction radiale $\phi(r)$
Multiquadrique (MQ)	$\sqrt{1 + (\epsilon r)^2}$
Inverse Multiquadrique (IMQ)	$1/\sqrt{1 + (\epsilon r)^2}$
Inverse Quadratique (IQ)	$1/(1 + (\epsilon r)^2)$

TABLEAU 2.1 – Fonctions de Base Radiales (RBF) usuelles

Les RBF ont des formes variées qu'il est par exemple possible de visualiser sur des points de coordonnées 2D comme visibles en Figure 2.9.

Le rayon d'une RBF centrée en un point x_i , s'écrit,

$$r = \|y - x_i\|, \quad (2.32)$$

avec $\|\cdot\|$, la norme euclidienne (L_2). Ainsi, une base RBF peut être évaluée quelque soit $y \in \Omega$ par exemple, la base PHS5 suivante,

$$\phi(y) = r^5. \quad (2.33)$$

2.4.3 La méthode RBF globale

Un des premiers auteurs à utiliser les RBF pour résoudre des EDP fut Kansa dès les années 1990 [Kansa1990b, Kansa1990a]. De nos jours, ce concept est appelé "méthode de Kansa" ou "méthode RBF globale". L'idée réside dans l'approximation d'une solution u régi par une équation différentielle et des conditions aux limites constituant un problème \mathcal{P} par une somme pondérée de RBF telle que,

$$u(x) = \sum_{i=1}^N \lambda_i \phi(\|x - x_i\|), \quad x \in \Omega, \quad x_i \in X \subset \Omega \quad (2.34)$$

est une solution approchée du problème \mathcal{P} . Avec $X = \{x_i\}_{i=1}^N \subset \Omega$ un ensemble de points centraux de fonctions de bases radiale servant à approximer la solution. L'enjeu est alors de trouver la pondération des RBF $\lambda \in \mathbb{R}^N$.

Considérons par exemple un problème \mathcal{P} stationnaire régi par un opérateur différentiel \mathcal{L} sur le domaine intérieur Ω , et par une condition de Dirichlet au bord sur $\partial\Omega$ tel que,

$$\mathcal{L}u(x) = f(x) \quad \text{sur } \Omega, \quad u(x) = g(x) \quad \text{sur } \partial\Omega. \quad (2.35)$$

Considérons un domaine de calcul sans maillage discrétisé $X = \{x_i\}_{i=1}^{N_x}$ comportant N_x centres de bases RBF divisées en N_Ω points d'intérieur du domaine et $N_{\partial\Omega}$ points de bords de Ω . La solution \mathbf{u} issue de l'approximation (2.34) est obtenue en calculant le vecteur de poids λ solution du système linéaire, (plus de détails dans [Fornberg2015c, p. 94]),

$$\begin{bmatrix} \Phi_{\partial\Omega} \\ \text{-----} \\ \mathcal{L}\Phi_\Omega \end{bmatrix} [\lambda] = \begin{bmatrix} \mathbf{g} \\ \mathbf{f} \end{bmatrix}, \quad (2.36)$$

avec $\Phi_{\partial\Omega}$, une matrice $N_{\partial\Omega} \times N_x$ d'évaluation des RBF aux points de bords du domaine X du domaine et $\mathcal{L}\Phi_\Omega$, une matrice $N_\Omega \times N_x$ qui évaluent les RBF dérivées par l'opérateur \mathcal{L} aux points intérieurs au domaine X . Cette méthode utilise donc N_x centres de base x_i pour approximer un vecteur de taille N_x (ici \mathbf{u}) en imposant les conditions de la physique en évaluant

les opérateurs spatiaux et les contraintes (conditions limites et termes sources) en ces N_x points. Le fait d'approximer une solution en imposant qu'elle satisfait exactement l'EDP aux points d'approximation de la solution est appelée *collocation*. Un exemple de points RBF en 2D disposés en collocation est présenté à la Figure 2.10.

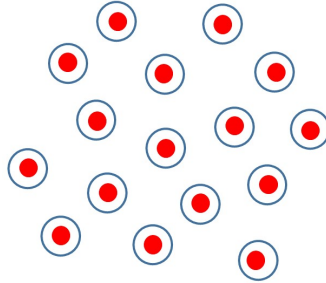


FIGURE 2.10 – Exemple 2D de placement de points de collocation : en rouge, localisation de l'expression du modèle physique, en bleu, localisation des centres de base. Ici les coordonnées où est évalué le modèle physique sont partagées avec les centres de base.

Cependant, la construction d'un système RBF-global impose que la matrice de collocation $\Phi_{\partial\Omega}$ ainsi que ses opérateurs de dérivées ici $\mathcal{L}\Phi_{\Omega}$ soient évalués sur la totalité du domaine. Cela implique de résoudre un système non-creux de taille $N_x \times N_x$. En pratique, pour des grands domaines, par exemple 50 000 points, cela implique la résolution d'un système à 50000×50000 donc 2,5 milliards d'éléments. Il faut ainsi construire une matrice système de 18,5 Go (gigaoctets) ce qui est impossible pour une station de travail équipée de 16 Go de mémoire vive. Par ailleurs quand bien même il serait possible de mettre une telle matrice en mémoire, la résolution du système serait rendue très difficile. Cette limitation a soulevé le besoin de développer des méthodes basées sur les RBF qui permettent de construire des systèmes creux.

2.4.4 La méthode RBF-FD

L'approche RBF-FD, telle que présentée dans [Flyer2016b, Fornberg2015c] ou encore [Barnett2015b], peut être vue comme une évolution de la méthode RBF globale. Cette méthode utilise l'interpolation RBF de manière locale dans le domaine physique de simulation Ω pour augmenter le caractère creux du système algébrique à résoudre et ainsi accélérer sa résolution. Les opérateurs différentiels représentatifs de la physique tels que les opérateurs intérieur et aux bords de Ω , \mathcal{L} et \mathcal{B} présentés en (2.35) sont approximés par des différences finies grâce à la méthode RBF-FD. Les RBF sus-mentionnées ont des propriétés intéressantes d'interpolations spatiales. Considérons de nouveau un ensemble de nœuds de calcul X , avec $N_x = \text{card}(X)$ définis dans l'intérieur du domaine Ω et sur son bord $\partial\Omega$. Ces nœuds servent de centres de bases aux RBF. X est muni de deux sous ensembles X_{Ω} et $X_{\partial\Omega}$ localisés respectivement à l'intérieur et sur les bords du domaine Ω . Posons $N_{x_{\Omega}} = \text{card}(X_{\Omega})$ et $N_{x_{\partial\Omega}} = \text{card}(X_{\partial\Omega})$, avec $N_x = \text{card}(X) = N_{x_{\Omega}} + N_{x_{\partial\Omega}}$.

Nous nommerons de manière similaire des sous domaines de $\partial\Omega$ comme Γ_D et $\Gamma_N \subset \partial\Omega$ pour distinguer différentes frontières et conditions limites. Les sous-domaines pourront être discrétisés sur X , avec X_{Γ_D} et X_{Γ_N} de cardinalités $N_{x_{\Gamma_D}}$ et $N_{x_{\Gamma_N}}$.

Approximation des opérateurs spatiaux par RBF-FD

L'idée dans cette méthode est d'approximer les contraintes imposées sur une variable u régie par une EDP en un point de l'espace $x = y_k \in X$. Des centres de bases/points d'EDP discrétisés (X) formant n voisins du point d'évaluation $y_k \in X$ servent à exprimer ces contraintes. La notion de voisinage est clé ici : le voisinage représente un petit sous-ensemble local du domaine ($n \ll N$), d'où une formulation creuse après construction d'une matrice de différentiation pour tout point

$y_k \in X$ avec $k = \{1, \dots, N_x\}$. Appelons $S_{y_k} \subset X$ l'ensemble de points voisins au nœud central y_k et nommons cet ensemble de voisins un "stencil". Un exemple d'un tel stencil est visible en Figure 2.11.

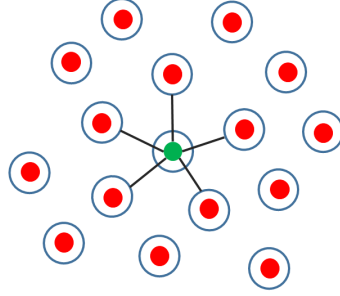


FIGURE 2.11 – Exemple 2D de placement de stencil à 5 points-voisins, les points-voisins étant des centres de RBF (bleu) de X voisins d'un point d'évaluation $y_k \in X$ (vert). Le point d'évaluation est une coordonnée où est exprimé le modèle physique et partage ici sa localisation avec les centres de RBF. Les autres points d'évaluation $y_j \in X$, $j \neq k$ (rouge) ne sont pas utilisés dans le stencil centré en y_k .

Soit $x \in S_{y_k}$, il est possible d'interpoler localement une variable $u \in \Omega$ telle que $\forall x \in S_{y_k} \subset X$,

$$u_a(y_k) = \sum_{i=1}^n w_i \phi(\|y_k - x_i\|) + \sum_{i=1}^m \lambda_i p_i(y_k). \quad (2.37)$$

Avec w_i et λ_i les coefficients de pondération de l'approximation et les p_i des monômes servant à régulariser l'approximation pour imposer $\sum_{i=1}^n w_i p_j(x_i) = 0 \forall j \in \{1, \dots, m\}$. On note u_a la fonction interpolante locale de u . Notons $\mathbf{u}_a(y_k)$ le vecteur où la fonction interpolante locale est évaluée aux n points voisins du stencil centré en y_k avec les $u_a(x_i)$ pour tout $x_i \in S_{y_k}$. L'interpolation locale (2.37) peut être vectorisée telle que,

$$\underbrace{\begin{pmatrix} \Phi(y_k) & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{pmatrix}}_{\mathbf{A}(y_k)} \begin{pmatrix} \mathbf{w} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{u}_a(y_k) \\ \mathbf{0} \end{pmatrix} \quad (2.38)$$

où Φ est la matrice de collocation RBF du stencil S_{y_k} telle que $\forall x_j \in \{x_j\}_{j=1}^n = S_{y_k}$,

$$\Phi(y_k) = \begin{pmatrix} \phi(\|x_1 - x_1\|) & \dots & \phi(\|x_n - x_1\|) \\ \vdots & \ddots & \vdots \\ \phi(\|x_n - x_n\|) & \dots & \phi(\|x_n - x_n\|) \end{pmatrix}. \quad (2.39)$$

Les vecteurs \mathbf{w} et λ sont obtenus en inversant la matrice $\mathbf{A}(y_k)$,

$$\begin{pmatrix} \mathbf{w} \\ \lambda \end{pmatrix} = \mathbf{A}(y_k)^{-1} \begin{pmatrix} \mathbf{u}_a(y_k) \\ \mathbf{0} \end{pmatrix}. \quad (2.40)$$

D'une manière similaire, on peut interpoler localement un opérateur différentiel \mathcal{L} sur u en un point $y_k \in X$ de la manière suivante,

$$\mathcal{L}u_a(y) = \sum_{i=1}^n w_i \mathcal{L}\phi(\|y - x_i\|) + \sum_{i=1}^m \lambda_i \mathcal{L}p_i(y). \quad (2.41)$$

La relation (2.41) peut être reformulée à la manière de (2.38) comme suit,

$$\mathcal{L}\mathbf{u}_a(y_k) = (\mathcal{L}\Phi)^T(y_k) \mathbf{w} + (\mathcal{L}\mathbf{P})^T(y_k) \lambda. \quad (2.42)$$

En introduisant (2.40) dans (2.41) on obtient,

$$\mathcal{L}\mathbf{u}_a(y_k) = \left[\left((\mathcal{L}\Phi)^T(y_k), (\mathcal{L}\mathbf{P})^T(y_k) \right) \mathbf{A}^{-1} \right] \begin{pmatrix} \mathbf{u}_a(y_k) \\ \mathbf{0} \end{pmatrix}. \quad (2.43)$$

Et en posant,

$$\left(\mathbf{w}^{\mathcal{L}T}, \lambda^{\mathcal{L}T} \right) = \left((\mathcal{L}\Phi)^T(y_k), (\mathcal{L}\mathbf{P})^T(y_k) \right) \mathbf{A}^{-1}, \quad (2.44)$$

on obtient après avoir enlevé les termes polynomiaux qui s'annulent,

$$\mathcal{L}\mathbf{u}_a(y_k) = \mathbf{w}^{\mathcal{L}T} \mathbf{u}_a(y_k). \quad (2.45)$$

Ce qui est équivalent à,

$$\mathcal{L}\mathbf{u}_a(y_k) = \mathbf{w}^{\mathcal{L}T} (u_a(x_0), \dots, u_a(x_n))^T, \quad x_i \in S_{y_k} \subset X \quad (2.46)$$

Pour exprimer les différences finies appliquées sur l'opérateur $\mathcal{L}u(y)|_{y=y_k}$, $y_k \in X = \{x_i\}_{i=1}^{N_x}$ les vecteurs de poids $w^{\mathcal{L}}$ et $\lambda^{\mathcal{L}}$ sont trouvés en résolvant le système,

$$\begin{bmatrix} \Phi & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}^{\mathcal{L}} \\ \lambda^{\mathcal{L}} \end{bmatrix} = \begin{bmatrix} (\mathcal{L}\Phi)(y = y_k) \\ (\mathcal{L}\mathbf{P})(y = y_k) \end{bmatrix}. \quad (2.47)$$

Il est ainsi possible d'approximer chaque opérateur différentiel $\mathcal{L}u(y_k) \forall y_k \in X$ comme somme pondérée de n valeurs voisines de u définies dans X . Il est possible de réaliser cette approximation sur d'autres opérateurs différentiels tels que celui de conditions limites \mathcal{B} . Un tel opérateur \mathcal{B} pourrait par exemple intervenir dans le problème \mathcal{P}_1 suivant,

$$\begin{aligned} \mathcal{L}u(x) &= f(x) && \text{sur } \Omega \\ \mathcal{B}(u)(x) &= h(x) && \text{sur } \partial\Omega, \end{aligned} \quad (2.48)$$

où h serait une fonction imposant une des valeurs de conditions aux limites et \mathcal{B} l'opérateur de conditions aux limites qui impose à la variable de champ u des conditions aux limites (p. ex. Neumann ou Dirichlet) sur le bord du domaine $\partial\Omega$. Il est alors possible de créer une matrice de différentiation discrète \mathbf{D} à dimensions $N_x \times N_x$ qui pour tout $x_k = y_k \in X$ exprime les poids de différentiation qui s'y appliquent. Les poids sont stockés lignes à lignes tels que $\forall x_k \in X$,

$$\mathbf{D}_{k, \cdot} = \begin{cases} [\mathbf{0}, \dots, w_1^{\mathcal{L}}, \dots, w_n^{\mathcal{L}}, \dots, \mathbf{0}] & \text{si } x_k \in \Omega^o \\ [\mathbf{0}, \dots, w_1^{\mathcal{B}}, \dots, w_n^{\mathcal{B}}, \dots, \mathbf{0}] & \text{si } x_k \in \partial\Omega \end{cases} \quad (2.49)$$

Les $\{w_i^{\mathcal{L}}\}_{i=1}^n$ et $\{w_i^{\mathcal{B}}\}_{i=1}^n$ étant respectivement les poids de différentiation qui discrétisent l'opérateur \mathcal{L} et l'opérateur de conditions aux limites \mathcal{B} . Les $\mathbf{0}$ étant ajoutés dans le vecteur de poids de D pour les nœuds de X ne font pas partie du stencil S_{y_k} .

Nous avons ainsi détaillé la construction d'une matrice de différentiation carrée de dimension $N_x \times N_x$ obtenue par RBF-FD. Ainsi il est possible de discrétiser puis de résoudre par exemple le problème stationnaire \mathcal{P}_1 suivant,

$$\begin{aligned} \mathcal{L}u(x) &= s(x) && x \in \Omega \\ \mathcal{B}_N u(x) &= g(x) && x \in \Gamma_N \subset \partial\Omega \\ \mathcal{B}_D u(x) &= d(x) && x \in \Gamma_D \subset \partial\Omega \end{aligned} \quad (2.50)$$

où \mathcal{L} , \mathcal{B}_N , \mathcal{B}_D représentent les opérateurs différentiels respectivement aux nœuds intérieurs de domaine, au bord où des conditions de Neuman s'appliquent, et aux bords où les conditions de Dirichlet s'appliquent. d et g sont des fonctions de l'espace commandant les termes de "droite" de Dirichlet et de Neuman et s un terme source. La méthode sus-mentionnée permet de construire les matrices de différentiation, $\mathbf{L} \in \mathbb{R}^{N_x, N_i}$, $\mathbf{B}_N \in \mathbb{R}^{N_x, N_n}$ et $\mathbf{B}_D \in \mathbb{R}^{N_x, N_d}$ qui contiennent pour tout point de domaine les pondérations similaires à celle de \mathcal{L} déjà montrée en (2.46). Le problème \mathcal{P}_1 revient alors à résoudre le système linéaire suivant,

$$\mathbf{D}\mathbf{u}_a = \mathbf{f} \quad (2.51)$$

où \mathbf{u}_a est le vecteur de dimension N_x qui approxime la solution évaluée à chaque point de collocation de X . \mathbf{f} est le vecteur contenant \mathbf{s} qui sont les valeurs de terme source évaluées aux nœuds X_Ω et contient les valeurs numériques des conditions de bords évaluées aux points de collocation $\partial\Omega$. \mathbf{D} est la matrice de différentiation de dimension $N_x \times N_x$. Ainsi les points de la solution approximée \mathbf{u}_a peuvent empiler ces différentes conditions de la physique de sorte que la matrice de différentiation \mathbf{D} et le terme source aggloméré aux conditions de bords \mathbf{f} peuvent s'écrire,

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{\Omega^\circ} \\ \mathbf{D}_{\Gamma_D} \\ \mathbf{D}_{\Gamma_N} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{s}(X_{\Omega^\circ}) \\ \mathbf{d}(X_{\Gamma_D}) \\ \mathbf{g}(X_{\Gamma_N}) \end{pmatrix}, \quad (2.52)$$

Le système algébrique (2.51) peut être résolu en inversant la matrice de différentiation \mathbf{D} tel que,

$$\mathbf{u}_a = \mathbf{D}^{-1}\mathbf{f}. \quad (2.53)$$

En pratique l'inversion de lourdes matrices peut limiter la rapidité des calculs. C'est pourquoi nous conseillons de déclarer \mathbf{D} comme matrice creuse ainsi que d'utiliser des solveurs adaptés au système (2.51) tel que le solveur GMRES muni d'un préconditionnement adéquat. Il est aussi possible d'ajouter au problème \mathcal{P}_1 en (2.50) un terme de variation temporelle. Dans ce cas un schéma temporel numérique tel qu'un schéma d'Euler ou Runge-Kutta ou encore une méthode des lignes utilisant ce type d'approche peut être utilisé.

2.4.5 Traitement des conditions aux limites

Les approches basées sur RBF-FD ont l'avantage d'être assez robustes faces aux conditions aux limites. De manière générale, il n'y a donc pas de nécessité particulière à imposer une forte densité de points de calcul près des frontières. Les conditions aux limites de Dirichlet ne nécessitent pas de mesures particulières dans la disposition des nœuds comparées aux nœuds intérieurs. En revanche les conditions de Neumann nécessitent soit une densification locale de points voisins des obstacles et placés régulièrement ou alors l'usage de points hors domaines dits "points fantômes". Ces points vont aider à une meilleure pondération des opérateurs discrets RBF-FD au voisinage des frontières. La figure empruntée à [Chinchapatnam2009] en Figure 2.12 illustre les choix de placement de points s'imposant pour traiter des conditions de Neumann. Les nœuds fantômes sont moins complexes à créer que des nœuds de bords localement orthogonaux et forment ainsi une approche populaire. Cependant, la gestion des nœuds fantômes peut être complexe, notamment lorsque des variables de champ sont hétérogènes comme dans le cas de champs de vents. Malgré les nœuds fantômes, la présence de nombreuses conditions de Neumann peut diminuer la précision des RBF-FD et compromettre sa stabilité numérique [Tominec2021b].

2.4.6 Approche RBF-FD moindres carrés (LS-RBF-FD)

Pour les problèmes physiques très advectifs et peu diffusifs tels que la modélisation de phénomènes météorologiques, C.A.J. Fletcher a observé que des méthodes à collocation peuvent être

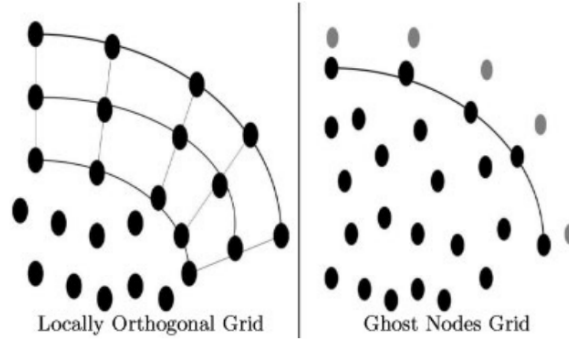


FIGURE 2.12 – Exemple de placement de nœuds frontières. Gauche placement "orthogonal". Droite nœuds fantômes. Emprunté à [Chinchapatnam2009].

inopérantes au contraire de méthodes de Galerkin [Fletcher1998] p.154. En s'inspirant directement de la méthode de Galerkin [Fletcher1998, p.98], Tominec et al. ont eu l'idée de supprimer la collocation dans les RBF-FD c'est-à-dire séparer un ensemble de points d'évaluation Y où sont évalués les résidus d'EDP d'un ensemble X où la solution est approximée à l'endroit de centres de RBF [Tominec2021b]. Ainsi le vecteur de solution \mathbf{u}_a satisfait un problème des moindres carrés où la forme forte du résidu d'EDP est minimisée.

Détaillons la manière de considérer deux ensembles de calculs :

- Un ensemble d'approximation $X = \{x_i\}_{i=1}^{N_x}$ de nœuds couvrant Ω , avec $N_x = \text{card}(X)$. Ces nœuds sont utilisés comme centre de RBF comme support pour réaliser des approximations locales à la manière des x_i présents en (2.28). La distribution des nœuds fait appel à l'algorithme de "minimisation d'énergie" qui utilise par exemple la suite pseudo aléatoire de Halton pour bien couvrir le domaine spatial.

- Un ensemble d'évaluation $Y = \{y_k\}_{k=1}^{N_y}$ de nœuds d'évaluation muni de deux sous-ensembles de nœuds Y_Ω et $Y_{\partial\Omega}$ correspondant respectivement aux nœuds intérieurs au domaine c.à.d. dans Y_Ω , et ceux présents en bord, $Y_{\partial\Omega}$. Cet ensemble joue le rôle des coordonnées y servant à évaluer une fonction interpolante telle que montrée en (2.28).

- Y est généré d'une manière similaire à X . Posons $N_{y_\Omega} = \text{card}(Y_\Omega)$ et $N_{y_{\partial\Omega}} = \text{card}(Y_{\partial\Omega})$, avec $N_y = \text{card}(Y) = N_{y_\Omega} + N_{y_{\partial\Omega}}$.

Nous nommerons de manière similaire des sous-domaines de $\partial\Omega$ comme Γ_D et $\Gamma_N \subset \partial\Omega$ pour distinguer différentes frontières et conditions limites. Les sous-domaines pourront être discrétisés sur Y , avec Y_{Γ_D} et Γ_N de cardinalités $N_{y_{\Gamma_D}}$ et $N_{y_{\Gamma_N}}$. Un exemple de domaine discrétisé en partitions X et Y est montré en Figure 2.13.

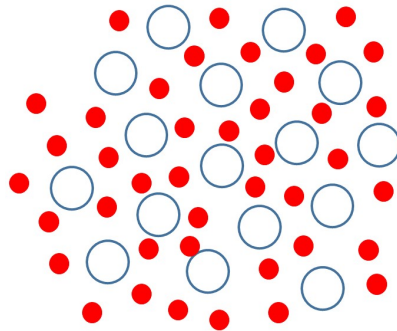


FIGURE 2.13 – Exemple de sous-domaine LS-RBF-FD. Nœuds d'"évaluations" où sont exprimées les contraintes du modèle physique $\{y_k\}_{k=1}^{N_y} \in Y$ rouge. Cercles Bleus : centre de bases $\{x_i\}_{i=1}^{N_x} \in X$

Similairement aux RBF-FD, la méthode LS-RBF-FD vise à créer des matrices de différentiation. Lors de cette étape, la différence par rapport à la méthode RBF-FD est que le stencil utilisé pour la construction de la matrice de différentiation est modifié puisque les points d'évaluation y_k où s'expriment les contraintes de modèle physique diffèrent des centres RBF-FD qui supportent l'interpolation RBF-FD c.à.d. $Y \neq X$. Un exemple de stencil adapté à la LS-RBF-FD est visible en Figure 2.14.

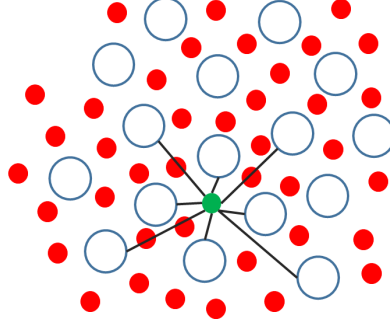


FIGURE 2.14 – Stencil à $n = 8$ voisins centré autour d'un noeud d'évaluation $y_k \in Y$ vert. Cercles bleus reliés : centre de base $\{x_i\}_{i=1}^n \in S_{y_k}$ servant à l'approximation des opérateurs spatiaux sur y_k . Cercles bleus : autres points de RBF dans X . Points rouges, autres points d'évaluation dans Y .

Ainsi la construction des matrices de différentiation exprimées par (2.37)-(2.47) restent valides avec la méthode LS-RBF-FD à la différence près que les nœuds d'évaluation $\{y_k\}_{k=1}^{N_y}$ appartiennent à un ensemble Y distinct des centres RBF X .

On peut interpréter l'approche comme une méthode de Galerkin des moindres carrés sur une version discrétisée de la norme L_2 du résidu $r(v)$, $v \in V \subset$ espace de Sobolev $W^{2,2}(\Omega)$ construit à partir du problème \mathcal{P}_1 exposé en (2.50),

$$\begin{aligned} u_a &= \arg \min_{v \in V} \|r(v)\|_{L_2(\Omega)}^2 \\ &= \arg \min_{v \in V} \int_{\Omega} (\mathcal{L}v(y) - s(y))^2 dy + \int_{\Gamma_N} (\mathcal{B}_N v(y) - g(y))^2 dy + \int_{\Gamma_D} (\mathcal{B}_D v(y) - d(y))^2 dy. \end{aligned} \quad (2.54)$$

En utilisant une méthode de Monte-Carlo pour l'approximation des intégrales $\|r(v)\|_{L_2(\Omega)}^2$ peut être approximé par

$$\begin{aligned} \|r(v)\|_{L_2(\Omega)}^2 &\approx \frac{|\Omega^\circ|}{N_{y\Omega^\circ}} \sum_{k=1}^{N_{y\Omega^\circ}} w^{\mathcal{L}}(y_k) \mathbf{v}(y_k) - s(y_k))^2 \\ &+ \frac{|\Gamma_N|}{N_{y\Gamma_N}} \sum_{k=1}^{N_{y\Gamma_N}} (w^{\mathcal{B}_N}(y_k) - g(y_k))^2 + \frac{|\Gamma_D|}{N_{y\Gamma_D}} \sum_{k=1}^{N_{y\Gamma_D}} (w^{\mathcal{B}_D}(y_k) - d(y_k))^2 \end{aligned} \quad (2.55)$$

avec, $|\Omega^\circ| = \int_{\Omega^\circ} 1 dy$, $|\Gamma_N| = \int_{\Gamma_N} 1 dy$ et $|\Gamma_D| = \int_{\Gamma_D} 1 dy$. (2.55) peut être réécrite sous la forme compacte,

$$\|r(v)\|_{L_2(\Omega)}^2 \approx \|\mathbf{D}_{\Omega^\circ} \mathbf{v} - \mathbf{s}\|^2 + \|\mathbf{D}_{\Gamma_N} \mathbf{v} - \mathbf{g}\|^2 + \|\mathbf{D}_{\Gamma_D} \mathbf{v} - \mathbf{d}\|^2. \quad (2.56)$$

La méthode LS-RBF-FD n'utilise pas la collocation puisque l'ensemble d'évaluation du résidu diffère des nœuds d'approximation d'EDP X . Y suréchantillonne l'espace comparé à X de sorte que $\text{card}(Y) \gg \text{card}(X)$. D'après nos observations $\text{card}(Y)/\text{card}(X) \geq 3$. Ainsi \mathbf{u}_a est la solution d'un problème surdéterminé des moindres carrés,

$$\mathbf{u}_a = \arg \min_{\mathbf{v}} \|\mathbf{D}\mathbf{v} - \mathbf{f}\|^2, \quad (2.57)$$

avec \mathbf{u}_a et $\mathbf{v} \in \mathbb{R}^{N_x}$, $\mathbf{f} \in \mathbb{R}^{N_y}$ and $\mathbf{D} \in \mathbb{R}^{N_y \times N_x}$. En utilisant (2.56), \mathbf{D} and \mathbf{f} peuvent être exprimées comme

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{\Omega^\circ} \\ \mathbf{D}_{\Gamma_D} \\ \mathbf{D}_{\Gamma_N} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{s}(Y_{\Omega^\circ}) \\ \mathbf{d}(Y_{\Gamma_D}) \\ \mathbf{g}(Y_{\Gamma_N}) \end{pmatrix}, \quad (2.58)$$

avec $\mathbf{s}(Y_{\Omega^\circ})$ le terme source évalué aux nœuds Y_{Ω° et similairement pour les conditions aux limites $\mathbf{d}(Y_{\Gamma_D})$ et $\mathbf{g}(Y_{\Gamma_N})$. Contrairement à la RBF-FD en collocation, la méthode LS-RBF-FD permet de relaxer les contraintes dures imposées par les conditions aux limites. Ainsi la stabilité de \mathbf{u}_a est renforcée car les points Y qui évaluent des quantités connues sont plus nombreux que le nombre d'inconnues dans le vecteur \mathbf{u}_a .

Le problème de moindres carrés (2.57) traduisant le problème \mathcal{P}_1 décrit en (2.50) est linéaire quadratique. En conjecturant que \mathbf{D} est de rang plein, une solution unique au problème des moindres carrés (2.57) est donnée par,

$$\mathbf{c}_a = \mathbf{D}^+ \mathbf{f}, \quad (2.59)$$

avec $\mathbf{D}^+ = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T$ la pseudo inverse à gauche de \mathbf{D} . En pratique la solution \mathbf{u}_a , peut être calculée efficacement par l'utilisation d'un solveur itératif de moindres carrés comme l'algorithme L-BFGS-B ou le solveur LSQR [Paige1982]. Par ailleurs l'opérateur différentiel approximé \mathbf{D} dans (4.15) reste creux grâce à l'usage d'approximations locales des opérateurs \mathcal{L} , \mathcal{B}_N et \mathcal{B}_D , le niveau de remplissage de \mathbf{D} étant paramétrable par les tailles de stencil.

L'approche LS-RBF-FD reste adaptable pour résoudre des EDP en régime transitoire comme étudié dans ce travail et exposé dans [Tominec2022]. De plus, tout type d'opérateur peut être approximé, comme par exemple, d'autres opérateurs particuliers de conditions limites. Les EDP non linéaires par rapport à la variable d'intérêt (p. ex. Navier-Stokes) peuvent aussi être traitées avec des méthodes RBF-FD [Chinchapatnam2009]. Enfin, les LS-RBF-FD sont particulièrement avantageuses pour traiter les conditions de bords et permettent de se dispenser de l'usage de nœuds fantômes dont la paramétrisation peut être rendue difficile [Tominec2021b].

2.4.7 Avantages et inconvénients des RBF-FD et LS-RBF-FD

L'un des principaux avantages des méthodes RBF-FD et LS-RBF-FD sont leur grande précision malgré des résolutions plus faibles que des méthodes concurrentes comme observé dans [Flyer2012] et [Tominec2021c]. D'autre part le domaine de calcul est flexible grâce à son absence de maillage [Bayona2015, Fornberg2015a] et permet donc de traiter des géométries complexes et diverses tout en économisant le nombre de points de calculs. Les côtés d'obstacles de domaine sont par exemple plus faciles à détourner qu'avec des mailles. L'absence de calcul intégral et de recherche de recherche d'une pondération de fonctions tests est également très économe en ressources de calcul comme le rappelle C.A.J. Fletcher [Fletcher1998, p. 151].

Les méthodes basées sur les RBF-FD possèdent quelques défauts. Comme les Différences Finies classiques elles ne garantissent notamment pas la positivité des solutions et peuvent montrer des erreurs dites d'oscillations : là où le gradient de la solution est fort, on peut observer une oscillation du vecteur de solutions autour d'une "vraie valeur". Cela peut s'observer lors de modélisation de phénomènes fortement discontinus telle qu'une rupture de barrage comme l'a montré N. Cellier [Cellier2018, p.74] où il a remarqué que la méthode des volumes finis (FVM) ne montre pas de telles erreurs. La Méthode des Eléments Finis (FEM) est également connue pour bien traiter ce genre de problèmes et offre de manière générale une précision proche sinon meilleure à la FVM. De plus, les méthodes RBF-FD manquent encore de résultats théoriques pour garantir la convergence des solutions calculées notamment sur le traitement des nœuds fantômes. C'est notamment la motivation du travail de Tominec et al. qui ont développé la méthode LS-RBF-FD considérée comme plus robuste de ce point de vue [Tominec2021b]. Enfin les RBF-FD requièrent

l'ajustement de nombreux paramètres tels que le type de RBF choisie ou encore la taille du stencil tandis que les EDP hyperboliques doivent être stabilisées [Fornberg2015c, p. 151 paragraphe 7.1.3]. Cependant ce n'est pas le cas des problèmes comportant un opérateur de diffusion comme dans le cas de la dispersion de polluants.

2.4.8 Conclusion sur les méthodes RBF-FD et LS-RBF-FD

Nous avons présenté les méthodes de discrétisation d'EDP RBF-FD et LS-RBF-FD. Les RBF-FD gagnent en intérêt dans la littérature et rendent aisée la discrétisation de l'espace grâce à son absence de maillage. Cependant les RBF-FD en collocation peuvent être limitées pour des problèmes très convectifs ou alors pour des problèmes très contraints par des conditions de Neumann [Tominec2021b]. Les RBF-FD en collocation et les LS-RBF-FD ont par ailleurs montré des vitesses de calcul très concurrentielles face à autres méthodes eulériennes comme les Méthodes des Éléments Finis [Tominec2021b] ou encore de Galerkin discontinues [Flyer2012]. Par ailleurs pour les RBF-FD en moindres carrés, le résidu r peut être calculé sur un grand espace dans Y alors que le vecteur de solution peut être défini sur un espace X petit ce qui permet une réduction conséquente de modèle en vue d'appliquer des méthodes inverses.

2.5 Assimilation de données

L'assimilation de données est historiquement développée par les météorologues dans les années 1950. Il s'agissait au départ de faire de la prévision de météo notamment du champ de température en fonction de l'information donnée par un réseau de stations de mesures. Les informations données par les mesures sont mises en relation avec des modèles physiques de dispersion, la plupart du temps décrits par des équations aux dérivées partielles [Asch2016, Zannetti1990] afin d'aboutir à l'estimation optimale d'une inconnue comme une température ou des conditions initiales. Les techniques d'assimilations de données se sont également révélées utiles pour l'identification de paramètres et le contrôle optimal d'appareil par exemple à l'aide du filtrage de Kalman [Kalman1960]. Dans la communauté automatique, l'assimilation de données fait référence à la problématique de l'observation et de la synthèse d'observateurs. Certains scientifiques considèrent «deux écoles» d'assimilation de données. D'un côté, l'assimilation variationnelle consiste en la minimisation d'une fonction de coût : la variable d'intérêt varie jusqu'à minimiser un critère qui quantifie l'écart entre les mesures ponctuelles issues de capteurs et les mesures prédites par simulations de modèles physiques déterministes. Cette méthode est très populaire en météorologie et utilise souvent l'approche adjointe développée dans les années 1980 [Dimet1986]. L'assimilation bayésienne est la deuxième école d'assimilation de données. Elle consiste à maximiser une fonction de vraisemblance dans un contexte de modélisation stochastique (le modèle physique et les mesures sont soumis à des processus stochastiques). Le filtrage de Kalman est une version particulière de cette approche dans le cadre de systèmes linéaires soumis à des bruits gaussiens. La technique BLUE (Best Linear Unbiased Estimator) est également une méthode d'assimilation bayésienne. Ces écoles ne sont pas séparées et plusieurs scientifiques développent des méthodes hybrides, combinant assimilation variationnelle et bayésiennes : on peut citer des exemples tels que l'IEKs [Bocquet2014] ou l'EnsVAR [Jardak2018] qui font partie de la sous-catégorie des filtres d'ensemble très en vogue dans les travaux d'assimilation de données actuellement [Bocquet2014, Defforge2021, Mons2016].

Que ce soit pour générer des simulations réalistes de dispersion atmosphérique ou pour réaliser des procédures d'assimilation de données bayésiennes ou variationnelles, il est nécessaire de disposer d'un modèle physique de dispersion atmosphérique de polluants représentatif du phénomène en jeu. La section suivante a pour objectif de présenter différentes méthodes d'assimilation de

données en donnant quelques exemples bayésiens et quelques exemples variationnels.

2.5.1 Introduction

Les techniques d'Assimilations de Données ont de nombreuses applications en physique dont nombre de modèles sont des Equations aux Dérivées Partielles (EDP) [Asch2016, Zannetti1990] et tout particulièrement dans le domaine de la dispersion de polluants régie par des équations d'advection-diffusion qui nous intéresse dans cette thèse. Ce champ de recherche a été assez étudié depuis la catastrophe de Fukushima en 2011 pour la construction de cartes de pollution nucléaires [Šmídl2013, Winiarek2014b]. De manière plus commune, l'assimilation de données a été utilisée pour estimer des variables météorologiques, telles que la prévision de précipitations [Županski1995] par méthode 4D-Var, ou l'estimation instantanée des précipitations (avec aussi prévision de court terme) [Marécal1999]. Une autre approche intéressante pour les météorologues est d'estimer des paramètres des équations aux dérivées partielles qui régissent un modèle en fonction d'observations mesurées par un réseau de capteurs [Winiarek2014b]. L'approche consistant à résoudre une EDP pour estimer la variable du champ (souvent notée u) est appelée « problème direct ». Celle consistant à considérer la variable de champ connue en des points de mesures et d'estimer des paramètres inconnus est appelée « problème inverse ». Il existe différentes stratégies pour assimiler des données et en déduire des paramètres inconnus d'une EDP.

Mathématiquement l'assimilation variationnelle vise à estimer un vecteur d'état noté $x(t)$ en disposant d'observations notées y^m , appartenant à l'espace des observations noté O . Le vecteur d'état est fonction de l'espace et du temps, et représente des variables physiques d'intérêt comme la température de l'air, la salinité, la concentration massique d'une espèce pour citer quelques exemples. L'évolution de l'état est décrit par un modèle \mathcal{M} consistant en un système (non linéaire en général) d'équations différentielles dans un domaine Ω tel que,

$$\begin{cases} \frac{dx}{dt} = \mathcal{M}(x, \gamma) & \text{dans } [\Omega \times T], \\ x(t=0) = x_0. \end{cases} \quad (2.60)$$

où x est l'état régi par le modèle dynamique \mathcal{M} , et où γ contient des paramètres d'entrée de modèle. Pour la suite, nous posons θ un vecteur de contrôle de dimension n_θ qui est la variable cible à estimer dans un problème d'assimilation de données. Le vecteur de contrôle θ peut être considéré comme séparé de l'état x et contenir seulement une ou plusieurs entrées de modèle \mathcal{M} que nous pouvons noter γ , avec ainsi $\theta = \gamma$ dans le cas où toutes les entrées de modèle sont inconnues. Les entrées de modèles γ peuvent contenir une condition initiale x_0 pour un modèle dynamique, des coefficients de modèle ou encore dans le cadre de cette thèse des paramètres de terme source. Le vecteur de contrôle θ peut encore être considéré comme un état augmenté, l'état enrichi peut alors intégrer l'état x et les entrées de modèle γ tel que $\theta = (x, \gamma)^T$. Un exemple d'état enrichi est visible dans [Kato2013].

2.5.2 Assimilation variationnelle

Principe général

Le problème le plus courant dans une assimilation variationnelle est l'estimation d'un état initial x_0 car bien souvent dans les problèmes dépendant du temps, la condition initiale de l'état à modéliser est mal connue ou inconnue. Si on dispose d'observations $y^m(t) \in O$, l'espace des observations, et d'un opérateur d'observation \mathcal{H} qui décrit les observations disponibles, alors on peut caractériser une distance entre les observations et la prédiction des observations issues du modèle d'observation \mathcal{H} et de la dynamique du système, à savoir une fonction de coût :

$$J^0(x_0) = J^b(x_0) + J^{\text{obs}}(x_0). \quad (2.61)$$

Avec J^0 une fonctionnelle associée à la condition initiale (en considérant un modèle physique transitoire), J^b un terme de régularisation aidant à mieux poser le problème d'estimation [Tikhonov1977]. Elle prend en compte un terme d'erreur d'initialisation. L'initialisation est noté x_b et peut aussi être appelée ébauche. J^{obs} est une fonctionnelle quantifiant un écart avec des mesures observées. La fonctionnelle peut alors être explicitée pour un problème des moindres carrés par l'expression :

$$J^0(x_0) = \frac{1}{2} \|x_0 - x_b\|_{\mathcal{B}}^2 + \frac{1}{2} \int_0^T \|\mathcal{F}(x(x_0, t)) - y^m(t)\|_{\mathcal{O}}^2 dt, \quad (2.62)$$

avec,

$$\mathcal{F} = \mathcal{H} \circ \mathcal{M}, \quad (2.63)$$

l'opérateur en avant, comme composition de l'application de l'opérateur de mesure \mathcal{H} et le modèle physique \mathcal{M} . La norme $2 \|\mathbf{x}\|$ désigne une norme pondérée par des erreurs d'observation réalisées dans \mathcal{O} ou d'initialisation réalisé dans \mathcal{B} sur un vecteur \mathbf{x} . Dans un problème discret avec des vecteurs d'états discrets nous considérons ces normes comme des normes discrètes pondérées notées $\|\mathbf{x}\|_{\mathbf{A}}^2$. Cette norme euclidienne d'un vecteur \mathbf{x} est pondérée par une matrice d'erreur d'initialisation \mathcal{B} ou de mesure \mathcal{O} (qui sont assimilables respectivement à des matrices inverses de covariance de bruit d'initialisation et de mesure \mathbf{B}^{-1} et \mathbf{R}^{-1}) telle que

$$\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}. \quad (2.64)$$

L'état analysé à trouver est celui qui minimise le gradient. Un minimum local est atteint pour un état analysé \mathbf{x}_0^a , sous la condition nécessaire de stationnarité :

$$\nabla J(\mathbf{x}_0^a) = 0. \quad (2.65)$$

La recherche de ce minimum local se réalise par descente de gradient, lorsque le problème est de grande taille ou non linéaire, couramment par descente de gradients conjugués de type L-BFGS-B. Il est possible d'appliquer cette méthode pour des estimations autres que la condition initiale, comme les conditions aux limites, des paramètres divers ou un terme source. Divers exemples d'assimilation variationnelles sont notamment visibles en documentation de *dolphin-adjoint* [Farrell2013]. Les techniques d'assimilation variationnelles sont très populaires et ce dans de nombreux domaines notamment fluide [Mons2016], médical [Funke2019], climatologique [Hersbach2020] ou encore dans le domaine de la pollution atmosphérique [Kumar2015]. L'assimilation variationnelle n'est cependant pas exclusive et il existe de nombreuses manières de résoudre des problèmes inverses. Une des difficultés de la mise en place d'une méthode variationnelle est l'expression du gradient de la fonction de coût : celle-ci ne peut pas être explicitement obtenue lorsque l'état n'est pas linéaire par rapport à l'opérateur en avant \mathcal{F} .

A propos de l'opérateur d'observation : Généralement dans le cadre de l'Estimation de Terme Source (STE) pour la pollution atmosphérique, l'opérateur de mesures est un soit un sélecteur de position qui échantillonne spatialement les sorties de modèles \mathcal{M} soit l'opérateur de mesure applique une fonction de mesure caractéristique à un capteur qui témoigne de la manière dont la mesure est réalisée spatialement et temporellement. En considérant un vecteur d'état \mathbf{x} formellement distinct du vecteur de contrôle θ , nous pouvons traduire ces actions sur un capteur par la relation suivante :

$$\begin{aligned} y_i &= \mathcal{H}_i \circ x(\theta, x, t), & y_i &\in \mathcal{O} \\ &= \int_{\Omega} \Delta(x - x_i) x(\theta, x, t) dx \end{aligned} \quad (2.66)$$

avec \mathbf{x} l'état défini en espace et en temps avec les coordonnées x et t , x_i la position d'un capteur i parmi N_c capteurs, \mathcal{H}_i l'opérateur d'observation spécifique au capteur d'indice i et où $\Delta : x \in$

$\Omega \mapsto (x - x_i)$ est la fonction caractéristique de mesure. Dans le cadre de la STE, l'état considéré est le champ scalaire de concentration en polluants c (c. à. d. $c = x$). Dans la cadre de la STE pour la pollution atmosphérique, la fonction caractéristique de mesure est considérée comme un Dirac pour un opérateur de sélection de position comme dans [Singh2015], ou une fonction porte dans [Kovalets2011] ou encore un noyau gaussien dans [Lopez-Ferber2022]. De manière générale, les opérateurs d'observation peuvent intégrer d'autres informations telle qu'une dérive de la mesure en fonction de la température comme considérée dans des problèmes inverses liés aux feux de forêts [Georges2019].

Méthode 3D-Var

La méthode 3D-Var tient compte des observations faites au temps t_k pour estimer un état au temps t_k . En météorologie, l'ébauche est la variable de champ mesurée la veille tandis qu'en étude de pollution, ce peut être une initialisation considérant des paramètres physiques réalistes ou simplement 0 [Asch2016, Marécal1999]. Cette approche peut être également utilisée pour l'assimilation de données dans un contexte de problème stationnaire, comme nous le verrons dans la suite de cette thèse. La fonction de coût 3D-Var s'écrit pour un état θ à estimer :

$$J(\theta) = \frac{1}{2} \|\theta - \theta^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \|\mathcal{F}(\theta) - \mathbf{y}^m\|_{\mathbf{R}^{-1}}^2 \quad (2.67)$$

où θ , θ^b et \mathbf{y}^m désignent le vecteur de contrôle, l'ébauche du vecteur de contrôle, et le vecteur d'observation (mesures). Cette fonction de coût est un équilibre entre les connaissances *a priori* données par l'ébauche et les connaissances apportées par l'observation. \mathbf{R} et \mathbf{B} sont difficiles à estimer. Dans l'approche bayésienne, ces matrices s'interprètent respectivement comme les matrices de variance-covariance de l'erreur d'estimation de θ et du bruit de mesure. En pratique, elles sont souvent considérées comme diagonales, c'est-à-dire que les erreurs sont supposées non corrélées et leurs coefficients sont ajustés pour obtenir le meilleur résultat possible [Winiarek2014b] p. 40.

A noter, que si les mesures sont obtenues à partir de capteurs dont la variance du bruit est connue, on peut utiliser cette variance comme coefficients de la matrice \mathbf{R} . De manière générale, \mathbf{B} prend des grandes valeurs pour que \mathbf{B}^{-1} aient des termes petits.

Une solution au problème 3D-Var est donnée par la condition de stationnarité de la fonction de coût

$$\nabla \mathcal{J}(\theta) = \mathbf{B}^{-1}(\theta - \theta^b) + \frac{\partial \mathcal{F}^T}{\partial \theta} \mathbf{R}^{-1}(\mathcal{F}(\theta) - \mathbf{y}^m) = 0. \quad (2.68)$$

La fonction de coût 3D-Var admet une solution explicite pour des modèles linéaires de dynamique \mathbf{M} et d'observation \mathbf{H} tels que $\mathcal{M}(\theta) = \mathbf{M}\theta$, et que $\mathcal{H}(\theta) = \mathbf{H}\theta$ et donc que $\mathcal{H} \circ \mathcal{M}(\theta) = \mathbf{F}(\theta)$, où \mathbf{F} est un opérateur linéaire. A partir de la condition (2.68), on obtient la solution explicite

$$\theta^a = (\mathbf{B}^{-1} + \mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} (\mathbf{B}^{-1} \theta^b + \mathbf{F}^T \mathbf{R}^{-1} \mathbf{y}^m). \quad (2.69)$$

Grâce à l'utilisation du lemme d'inversion matricielle, cette solution peut se réécrire sous "filtre de Kalman non récursif" avec la matrice de gain de Kalman \mathbf{K} telle que

$$\theta^a = \theta^b + \mathbf{K}(\mathbf{y}^m - \mathbf{F}\theta^b) \quad (2.70)$$

avec,

$$\mathbf{K} = \mathbf{B}\mathbf{F}^T (\mathbf{R} + \mathbf{F}\mathbf{B}\mathbf{F}^T)^{-1}. \quad (2.71)$$

Lorsque l'opérateur \mathcal{F} est non linéaire, il est nécessaire d'utiliser une méthode de descente pour obtenir une solution. On peut utiliser une méthode de Newton qui utilise la matrice Hessienne \mathbb{H} , c'est-à-dire la matrice exprimant les dérivées partielles secondes de la fonction de coût $\mathbb{H}_{i,j} =$

$\frac{\partial^2 \mathcal{J}}{\partial x_i \partial x_j}$, $(i, j) = \{1, \dots, N_\theta\}^2$. Ainsi il est possible de calculer le vecteur optimal θ^a en se servant de la descente de gradient en effectuant des itérations de descente en actualisant l'état θ :

$$\theta_{j+1} = \theta_j - \mathbb{H}^{-1} \nabla \mathcal{J}(\theta_j), \quad (2.72)$$

où j représente l'indice d'itération.

En pratique, l'inversion de la matrice Hessienne est assez coûteuse pour les problèmes de grande dimension et des algorithmes de descente de type "quasi-Newton" permettent d'approximer la Hessienne, comme par exemple, l'algorithme L-BFGS-B [Blayo2010, p. 26].

La 3D-Var peut également s'exprimer en continu, avec une norme exprimée sous forme d'une intégrale. De plus, la 3D-Var fonctionne aussi pour avancer en temps en se servant de l'état analysé au temps précédent comme initialisation du temps actuel. La 4D-Var de son côté peut prendre en compte les états précédents sur une fenêtre de temps donnée et ainsi tenir compte de l'historique de la dynamique d'état.

4D-Var

La méthode 4D-Var tient compte des données disponibles sur une fenêtre temporelle $[t_{k_0}, t_k]$ avec $k_0 < k$, pour estimer un état au temps t_k . Avec N_C mesures aux erreurs décorréllées, et pour $x \in \Omega$, $t \in T$, sa fonctionnelle s'écrit :

$$\mathcal{J}(\theta) = \sum_{i=1}^{N_C} \frac{1}{2r^2} \int_T (\mathcal{F}_i(\theta, x, t) - y_i^m)^2 dt \quad (2.73)$$

avec,

$$\mathcal{F}_i(\theta, x, t) = \mathcal{H}_i \circ \mathcal{M}(\theta, x, t) = \int_{\Omega} \Delta(x - x_i) c(\theta, x, t). \quad (2.74)$$

où $c : (\theta, x, t) \rightarrow c(\theta, x, t)$ sont les concentrations simulées retournées par le modèle \mathcal{M} . r^2 est ici un terme de variance d'erreur de mesures considéré comme identique pour toute les N_C mesures $\{y_i^m\}_{i=1}^{N_C}$.

En discrétisant et en ajoutant un terme de régularisation, la fonction de coût 4D-Var revient à écrire :

$$\mathcal{J}(\theta) = \frac{1}{2} \|\theta - \theta^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{i=k_0, k_0+k}^k \|\mathcal{F}_i(\theta) - \mathbf{y}_i^m\|_{\mathbf{R}_i^{-1}}^2. \quad (2.75)$$

Comme la fonction de coût 4D-Var tient compte de mesures précédentes en temps, l'expression du gradient de la fonction de coût selon le vecteur de contrôle θ n'est pas triviale.

Sur les difficultés de la minimisation de la fonction de coût

En pratique le modèle physique \mathcal{M} et le modèle d'observation \mathcal{H} et donc le modèle "en avant" (direct) \mathcal{F} sont rarement linéaires par rapport à θ . La fonction de coût n'est pas facilement minimisable et aucune solution explicite de la fonctionnelle ni d'expression explicite de son gradient est accessible. Une autre approche populaire pour minimiser la fonction de coût grâce à son gradient consiste en l'utilisation des différences finies pour le calcul de gradient, de sorte que, pour chaque composante du gradient,

$$\nabla \mathcal{J}(\theta)_i = \frac{\mathcal{J}(\theta + \epsilon_i) - \mathcal{J}(\theta)}{\epsilon_i} \quad (2.76)$$

où ϵ_i est une petite variation dans la direction de la i -ème composante du vecteur de contrôle θ . Cependant l'utilisation des différences finies est déconseillée en raison de son coût prohibitif. En effet, le calcul de sensibilité doit être réalisé composante par composante pour obtenir l'intégralité du gradient, ce qui implique n_θ évaluations de coût et donc n_θ exécutions de modèles pour un vecteur cible de longueur n_θ par itération (voire $2n_\theta$ pour des différences finies centrées).

L'approche adjointe, une solution à la minimisation de la fonction de coût

L'approche adjointe est une manière de calculer le gradient d'une fonction de coût de manière précise et relativement rapide. L'utilisation de l'adjoint permet d'exprimer le gradient d'une fonction de coût \mathcal{J} de manière explicite, et ainsi de procéder à une méthode de descente de gradient afin de minimiser la fonction \mathcal{J} par rapport à la variable de contrôle θ . Formellement nous pouvons définir le concept d'adjoint de la manière suivante,

Définition : Soit H un espace préhilbertien (EV muni d'un produit scalaire) sur un ensemble réel ou complexe muni du produit scalaire $\langle \cdot, \cdot \rangle$ et soient \mathcal{M} et \mathcal{M}^* deux opérateurs de H . L'opérateur \mathcal{M}^* est adjoint de \mathcal{M} si et seulement si,

$$\forall (x, y) \in H^2, \quad \langle \mathcal{M}(x), y \rangle = \langle x, \mathcal{M}^*(y) \rangle$$

L'adjoint permet alors pour le problème de la minimisation d'un critère entre les observations et l'application d'un modèle d'exprimer la sensibilité de l'opérateur de modèle \mathcal{M} par rapport à θ .

Adjoint continu ou adjoint discret ? Il existe deux manières différentes d'exprimer un adjoint :

1. Écrire le calcul des variations et utiliser les équations adjointes des équations de la physique comme modèle adjoint continu puis discrétiser ce modèle adjoint (c'est l'approche dite directe).
2. Construire le modèle adjoint à partir du modèle en avant discrétisé (c'est l'approche dite indirecte).

La première approche implique de déduire en mathématiques continues un modèle adjoint. Cependant écrire formellement le modèle adjoint continu peut être difficile pour des Equations aux Dérivées Partielles et peut être source d'erreur en particulier lorsque les conditions aux limites sont complexes. On trouve de nombreuses applications de cette approche appliquée par exemple à l'estimation des feux de forêt ou du trafic routier [Georges2019], [Nguyen2016a]. La deuxième approche a l'avantage d'être plus rapide à programmer et évite de formellement écrire l'adjoint du modèle en avant. Cependant, le modèle adjoint construit à partir du modèle en avant discret (Discretisation-ADJOINT) est réputé être moins fiable qu'en utilisant un modèle adjoint déduit formellement en mathématiques continues puis discrétisé (Adjoint-Discretisation) [Sirkes1997]. Dans leur exemple, les auteurs de l'étude citée ont observé que la variable adjointe suivant le schéma Adjoint-Discretisation montre une meilleure stabilité temporelle que celle obtenue avec le schéma Discretisation-Adjoint. Les approches Discretisation-Adjoint restent populaires. En effet, la stabilité de la solution adjointe peut être aidée par des schémas numériques stables. La suite Dassflow [Honnorat2007] permet de modéliser les écoulements d'eau de surfaces et offre l'option Adjoint-Discretisation ainsi que l'option Discretisation-Adjoint. Des codes adjoints discrets ont également été intégrés dans la suite openFOAM [Towara2013].

Des calculateurs de codes adjoints alternatifs existent comme la bibliothèque dolfin-adjoint [Farrell2013] qui permet de différencier automatiquement une fonction de coût (approche discrétisation adjoint) et a été appliquée avec succès sur des problèmes fluidiques pour des applications médicales [Funke2019]. La bibliothèque Python *fenics* est notamment capable d'écrire des éléments finis en déclarant formellement les formes variationnelles des EDP, le code *dolfin-adjoint* a la particularité de générer du code adjoint à partir d'écritures symboliques de problèmes variationnels et donc de palier les défauts traditionnels de la différentiation automatique entre autres la stabilité.

Adjoint Continu : L'adjoint exprimé en continu se dérive d'un problème d'optimisation continu comme proposé dans [Georges2019, Nguyen2016a]. Plaçons nous dans l'exemple de la méthode 4D-Var en considérant cette fois-ci les opérateurs \mathcal{H} , \mathcal{M} et \mathcal{F} . A titre d'exemple, nous voulons déterminer le vecteur de paramètres θ de la source qui minimise l'erreur quadratique liée aux observations pour un problème unidimensionnel de transport de polluants :

$$\mathcal{J}(\theta) = \sum_{i=1}^{N_C} \frac{1}{2r^2} \int_{t=t_0}^{t_L} (y_i^m(t) - y_i(\theta, t))^2 dt \quad (2.77)$$

où r désigne l'écart-type du bruit de mesure (ici considéré unique) et les y_i^m représentent les mesures issues de capteurs, les $y_i(t)$ sont les concentrations issues de l'opérateur de mesure décrit par (2.66). Pour reprendre les notations précédentes, il s'agit de l'application de l'opérateur d'observation \mathcal{H} sur les prédictions du modèle \mathcal{M} (c.à.d. $c = \mathcal{M}(\theta, x, t)$ et donc $y_i = \mathcal{F}_i(\theta)$). Par construction de l'opérateur en avant \mathcal{F} , la fonction de coût (2.77) est contrainte par le modèle de transport,

$$\partial_t c = k \partial_{xx}^2 c(x) - u \partial_x c(x) + s(x, \theta), \quad (2.78)$$

où k désigne le coefficient de diffusion $L^2 T^{-1}$, $u(LT^{-1})$, le coefficient d'advection et s , un terme source. Les conditions initiales et aux limites considérées étant,

$$\begin{aligned} c(0, t) = 0, \quad c(x_f = 1, t) = 0 \\ c(x, 0) = 0 \end{aligned} \quad (2.79)$$

Afin de calculer le modèle adjoint, nous utilisons une approche par le lagrangien suivant :

$$\begin{aligned} \mathcal{L}(\theta, c, p) = \sum_{i=1}^{N_C} \frac{1}{2r^2} \int_{t=t_0}^{t_L} [(y_i^m(t) - \int_{x_0}^{x_f} \Delta(x - x_i) c(x, t) dx)]^2 dt \\ + \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} [p(x, t) (\partial_t c + u \partial_x c(x) - k \partial_{xx}^2 c(x) - s(x, \theta))] dx dt. \end{aligned} \quad (2.80)$$

Remarque 1 : Notons qu'un terme contenant le résidu de modèle,

$$\begin{aligned} r(c, \theta) = \mathcal{M}(x, \theta) - s(x, \theta) \\ = \partial_t c + u \partial_x c(x) - k \partial_{xx}^2 c(x) - s(x, \theta), \end{aligned} \quad (2.81)$$

a été ajouté à la fonction de coût. Il correspond au produit scalaire $\langle p, r(c, \theta) \rangle$ entre un multiplicateur de Lagrange p et la contrainte correspondant au modèle physique. Le produit scalaire $\langle f, g \rangle$ est défini par $\int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} f g dx dt$. Le multiplicateur de Lagrange s'interprète comme l'état adjoint du système. Le système adjoint peut être obtenu simplement en utilisant les conditions de stationnarité du lagrangien.

Remarque 2 : Les $y_i(\theta, x, t)$ dans (2.77) correspondent aux mesures prédites à partir d'un opérateur de mesure, par exemple, un opérateur local $\Delta(x - x_i)$ tel que $y_i = \int_{x_0}^{x_f} \Delta(x - x_i) c(x, t) dx$, où $c(x, t)$ est le profil de concentration calculée par simulation directe.

Pour obtenir le système adjoint, nous allons calculer les variations du lagrangien en fonction des variations des θ et c .

$$\begin{aligned} \delta \mathcal{L}(\theta, c, p) = \int_{t=t_0}^{t_L} \sum_{i=1}^{N_C} -\frac{1}{r^2} \int_{x=x_0}^{x_f} \Delta(x - x_i) [y_i^m(x, t) - y_i(\theta, x, t)] \delta c dx dt \\ + \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} [p(x, t) (\partial_t \delta c + u \partial_x \delta c - k \partial_{xx}^2 \delta c - \partial_\theta s(x, \theta) \delta \theta)] dx dt \end{aligned} \quad (2.82)$$

Nous avons trois intégrations par partie à réaliser pour exprimer $\delta \mathcal{L}$ en fonction des δc et $\delta \theta$

uniquement (sans les dérivées de δc) :

$$\begin{aligned}
 I_1 &= \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} p(x, t) \partial_t \delta c \, dx \, dt \\
 &= - \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} \partial_t p(x, t) \delta c \, dx \, dt + \int_{x=x_0}^{x_f} [p(x, t_L) \delta c(x, t_L) - p(x, 0) \delta c(x, 0)] dx \\
 I_2 &= \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} p(x, t) u \partial_x \delta c \, dx \, dt \\
 &= - \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} \partial_x p(x, t) u \delta c \, dx \, dt + \int_{t=t_0}^{t_L} [p(x_f, t) u \delta c(x_f, t) - p(x_0, t) u \delta c(x_0, t)] dx \\
 I_3 &= \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} -p(x, t) k \partial_{xx}^2 \delta c \, dx \, dt \\
 &= - \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} \partial_{xx}^2 p(x, t) k \delta c \, dx \, dt - \int_{t=t_0}^{t_L} [p(x_f, t) k \partial_x \delta c(x_f, t) - p(x_0, t) k \partial_x \delta c(x_0, t)] dt \\
 &\quad + \int_{t=t_0}^{t_L} [\partial_x p(x_f, t) k \delta c(x_f, t) - \partial_x p(x_0, t) k \delta c(x_0, t)] dt
 \end{aligned} \tag{2.83}$$

En réunissant tous les termes de calcul,

$$\begin{aligned}
 \delta \mathcal{L}(\theta, c, p) &= \\
 &\int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} \sum_{i=1}^{N_C} -\frac{1}{r^2} \Delta(x - x_i) [y_i^m(x, t) - y_i(\theta, x, t)] - \partial_t p(x, t) - u \partial_x p(x, t) - k \partial_{xx} p(x, t) \delta c \, dx \, dt \\
 &+ \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} [p(x, t) (\partial_\theta s(x, \theta) \delta \theta)] \, dx \, dt \\
 &+ \int_{x=x_0}^{x_f} [p(x, t_L) \delta c(x, t_L) - p(x, 0) \delta c(x, 0)] dx + \int_{t=t_0}^{t_L} [p(x_f, t) u \delta c(x_f, t) - p(x_0, t) u \delta c(x_0, t)] dx \\
 &- \int_{t=t_0}^{t_L} [p(x_f, t) k \partial_x \delta c(x_f, t) - p(x_0, t) k \partial_x \delta c(x_0, t)] dt \\
 &+ \int_{t=t_0}^{t_L} [\partial_x p(x_f, t) k \delta c(x_f, t) - \partial_x p(x_0, t) k \delta c(x_0, t)] dt
 \end{aligned} \tag{2.84}$$

La concentration initiale ne varie pas dans ce problème, ainsi $\delta c(x, 0) = 0$, Les conditions sont de Dirichlet, et sont nulles, donc $\delta c(x_f, t) = \delta c(x_0, t) = 0$. Finalement, nous obtenons l'expression de la variation du lagrangien, en regroupant tous les termes de calcul en fonction des δc , $\delta \theta$ sous la forme suivante

$$\begin{aligned}
 \delta \mathcal{L}(\theta, c, p) = & \\
 & \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} \sum_{i=1}^{N_C} \frac{1}{r^2} [(y_i^m(x, t) - y_i(\theta, x, t))] - \partial_t p(x, t) - u \partial_x p(x, t) - k \partial_{xx} p(x, t)] \delta c \, dx \, dt \\
 & + \int_{t=t_0}^{t_L} \int_{x=x_0}^{x_f} [p(x, t) (\partial_\theta s(x, \theta) \delta \theta)] \, dx \, dt \\
 & + \int_{x=x_0}^{x_f} p(x, t_L) \delta c(x, t_L) \, dx - \int_{t=t_0}^{t_L} [p(x_f, t) k \partial_x \delta c(x_f, t) - p(x_0, t) k \partial_x \delta c(x_0, t)] \, dt
 \end{aligned} \tag{2.85}$$

Finalement, $\delta \mathcal{L}(\theta, c, p) = 0 \forall (x, t) \in \Omega \times T$ satisfait nécessairement les conditions d'optimalité :

$$\partial_t c = k \partial_{xx}^2 u(x) - u \partial_x c(x) + s(x, \theta) \tag{2.86}$$

$$c(0, t) = 0, \quad c(x_f = 1, t) = 0 \quad \text{et} \quad c(x, 0) = 0 \tag{2.87}$$

$$- \sum_{i=1}^{N_C} \frac{1}{r^2} \Delta(x - x_i) [(y_i^m(x, t) - y_i(\theta, x, t))] - \partial_t p(x, t) - u \partial_x p(x, t) - k \partial_{xx} p(x, t) = 0 \tag{2.88}$$

$$p(x, t_L) = 0, \quad p(x_f, t) = 0, \quad p(x_0, t) = 0 \tag{2.89}$$

$$- \int_{t=t_0}^{t_f} \int_{x=x_0}^{x_f} p(x, t) \partial_\theta s(x, \theta) \, dx \, dt = 0 \tag{2.90}$$

Le système représenté par les équations (2.86)-(2.90) constitue un problème avec conditions aux deux bouts ("two-point boundary value problem"), qui peut être résolu par exemple avec une méthode de quasi-Newton. Pour calculer le gradient de la fonction de coût (2.77), nous nous servons de l'équation (2.90). Cette équation tirée du deuxième terme de (2.85) est équivalente à

$$\langle \nabla_\theta J, \delta \theta \rangle = 0$$

avec,

$$\nabla_\theta J = - \int_{t=t_0}^{t_f} \int_{x=x_0}^{x_f} p(x, t) \partial_\theta s(x, \theta) \, dx \, dt. \tag{2.91}$$

Cette expression du gradient de la fonction de coût vient du deuxième terme de l'équation (2.85). Pour accéder à ce gradient il faut pour chaque mise à jour de gradient :

- Résoudre les équations (2.86) en utilisant (2.87) correspondant à l'exécution du modèle physique de transport.
- Résoudre le modèle adjoint (2.88)- (2.89).
- Calculer le gradient en utilisant (2.90).

L'équation adjointe (2.88), est aussi valable en dimension 2 ou supérieure. Dans tous les cas, des conditions de Dirichlet nulles pour le modèle direct conduisent à des conditions de Dirichlet nulles pour le modèle adjoint [Davoine2007]. Pour des conditions de Neumann dans le cas d'une EDP d'advection diffusion, les conditions aux limites de l'EDP adjointe sont des conditions de Robin.

Adjoint discret : L'adjoint discret est la méthode qui utilise l'adjoint du modèle physique après discrétisation de celui-ci. Ici nous considérons un exemple stationnaire. Nous cherchons donc à trouver la solution optimale de vecteur de contrôle pour une fonctionnelle discrète par exemple,

$$J(\theta) = \| \mathcal{F}(\theta) - \mathbf{y}_i^m \|_{\mathbf{R}^{-1}}^2 \tag{2.92}$$

Le problème d'optimisation est :

$$\begin{aligned} \min J(\mathbf{c}, \theta) \\ \text{sous } r(\mathbf{c}, \theta) = 0, \end{aligned} \quad (2.93)$$

avec $r(c, \theta)$ le résidu de modèle. Pour reprendre l'exemple de la dispersion de polluants, nous considérons,

$$r(c, \theta) = -k\partial_{xx}^2 c(x) + u\partial_x c - s(x, \theta) \quad (2.94)$$

Le résidu discrétisé \mathbf{r} étant alors de la forme,

$$\mathbf{r} = \mathbf{A}\mathbf{c} - \mathbf{s} \quad (2.95)$$

Comme $\mathbf{r}(\mathbf{c}, \theta) = 0$ par construction, posons le Lagrangien,

$$L(\mathbf{c}, \theta, \lambda) = J(\mathbf{c}, \theta) - \lambda^T \mathbf{r}(\mathbf{c}, \theta) \quad (2.96)$$

On obtient la différentiel du lagrangien sous la forme :

$$dL|_{\mathbf{r}=\mathbf{0}} = \left(\frac{\partial J}{\partial \mathbf{c}} - \lambda^T \frac{\partial \mathbf{r}}{\partial \mathbf{c}} \right) d\mathbf{c} + \left(\frac{\partial J}{\partial \theta} - \lambda^T \frac{\partial \mathbf{r}}{\partial \theta} \right) d\theta \quad (2.97)$$

Les conditions nécessaires d'optimalité impose que chaque variation soit nulle. Ainsi nous obtenons l'équation adjointe,

$$\frac{\partial J}{\partial \mathbf{c}} - \lambda^T \frac{\partial \mathbf{r}}{\partial \mathbf{c}} = 0 \quad (2.98)$$

et ainsi nous obtenons la dérivée totale (=gradient) évaluée à $\mathbf{r} = \mathbf{0}$

$$\left. \frac{dL}{d\theta} \right|_{\mathbf{r}=\mathbf{0}} = \left. \frac{dJ}{d\theta} \right|_{\mathbf{r}=\mathbf{0}} = \frac{\partial J}{\partial \theta} - \lambda^T \frac{\partial \mathbf{r}}{\partial \theta}. \quad (2.99)$$

Et ainsi le gradient de la fonctionnelle et accessible, et le problème de contrôle peut être résolu par descente de gradient.

Algorithm 1 Etapes clés d'une méthode variationnelle

- 1: Initialiser θ avec θ^b
 - 2: **while** $\|\nabla \mathcal{J}\| > \text{tolérance}$ ou $k < \text{Nombre Maximum d'itérations}$ **do**
 - 3: Résoudre numériquement l'EDP du modèle physique (direct)
 - 4: Résoudre numériquement l'équation adjointe pour obtenir λ
 - 5: Actualiser la fonction de coût J
 - 6: Actualiser le gradient de $J = \frac{\partial J}{\partial \theta} - \lambda^T \frac{\partial \mathbf{r}}{\partial \theta}$
 - 7: Mettre à jour le vecteur de paramètres θ en utilisant le gradient de J
 - 8: Incrémenter k
-

Conclusion sur l'assimilation variationnelle L'assimilation variationnelle a été présentée. Retenons que les différences finies (dans le cas de la 3D-Var) et les méthodes adjointes forment des approches courantes dans la littérature. C. Defforge a par exemple utilisé une 3D-Var à différences finies pour de l'assimilation de champ de vent [Defforge2019b]. Les méthodes adjointes sont cependant plus précises et moins coûteuses en calcul. Elles restent très populaires notamment dans les études climatiques [Hersbach2020]. Cependant les méthodes adjointes demandent du temps à être mis en œuvre car elles nécessitent de programmer les modèles adjoints. Des générateurs formels de code adjoint existent notamment *Odysee* ou *Tapenade* [Hascoet2013] pour des programmes en *fortran* ou encore *dolphin-adjoint* [Farrell2013] pour une bibliothèque d'éléments finis sur Python. Les principales étapes d'une méthode variationnelle générique sont présentées en algorithme 1. Dans la présente partie, nous n'avons pas pris en compte de mobilité des capteurs. Il est cependant possible de le faire en considérant un opérateur d'observation dépendant du temps \mathcal{H}_k et dont les positions de capteurs $\{x_i\}_{i=1}^{N_c}$ s'actualisent à chaque mise à jour de mesure.

2.5.3 Assimilation Bayésienne

L'assimilation bayésienne repose sur la maximisation d'une fonction de vraisemblance. Chaque variable à estimer est considérée suivant une densité de probabilité (DDP ou PDF en anglais), le mode de la DDP correspond généralement à l'état cible et sa dispersion quantifie l'incertitude. Bien souvent la DDP approximant un état est considérée comme une normale. L'assimilation bayésienne repose fondamentalement sur la relation de Bayes avec

$$\mathbb{P}(\theta | y^m) = \frac{\mathbb{P}(y^m | \theta)\mathbb{P}(\theta)}{\mathbb{P}(y^m)}, \quad (2.100)$$

où θ désigne le vecteur d'inconnues à estimer avec la connaissance des mesures y^m et où $\mathbb{P}(a | b)$ désigne une probabilité conditionnelle sur une grandeur a sachant des données b . Une formulation équivalente de la loi de Bayes en probabilité continue est

$$\mathbb{P}(\theta | y^m) = \frac{\mathbb{P}(y^m | \theta)\mathbb{P}(\theta)}{\int_{\theta} \mathbb{P}(y^m | \theta)\mathbb{P}(\theta) d\theta}. \quad (2.101)$$

Le modèle physique \mathcal{M} est représenté par la probabilité conditionnelle $\mathbb{P}(y^m | \theta)$ appelée également « vraisemblance ». $\mathbb{P}(\theta)$ est une information *a priori* sur θ . $\mathbb{P}(\theta | y^m)$ est appelé « probabilité *a posteriori* ». Finalement l'assimilation bayésienne peut se mémoriser comme,

$$\text{DDP } a \text{ posteriori} \propto \text{DDP modèle} \times \text{DDP } a \text{ priori}. \quad (2.102)$$

Le but est de trouver le vecteur de contrôle θ_* qui maximise la probabilité *a posteriori* pour estimer le jeu de paramètre optimal :

$$\theta_* = \arg \max_{\theta} \mathbb{P}(\theta | y^m). \quad (2.103)$$

Mathématiquement l'assimilation bayésienne considère pour chaque temps un modèle dynamique d'état le faisant passer d'un temps t_k à un temps t_{k+1} en considérant des erreurs de modèle ϵ_k^q :

$$\theta_{k+1} = \mathcal{M}_{k+1}(\theta_k) + \epsilon_k^q, \quad (2.104)$$

où θ_{k+1} est l'état au temps t_{k+1} , \mathcal{M} , le modèle dynamique de transition de l'état, ϵ_k^q , l'erreur de modèle. Les mesures y_k^m sont liées aux observations par l'opérateur en avant \mathcal{F} ,

$$y_k^m = \mathcal{F}_k(\theta_k) + \epsilon_k^o, \quad (2.105)$$

avec ϵ_k^o le bruit de mesures au temps t_k , $\mathcal{F}_k = \mathcal{H}_k \circ \mathcal{M}_k$, l'opérateur en avant comme composition de l'opérateur d'observation \mathcal{H}_k qui ramène les états dans un espace de mesures et de l'opérateur de modèle physique \mathcal{M}_k . Remarquons qu'il est possible de prendre en compte des mesures mobiles en utilisant un opérateur de mesures dépendant de l'indice de temps k .

Remarque : Comme discuté en introduction de l'assimilation variationnelle (2.5.1), il est formellement possible de séparer le vecteur d'état $x \in \mathbb{R}^{N_x}$ du vecteur de contrôle θ de taille N_θ de sorte que,

$$x_{k+1} = \mathcal{M}(x_k, \theta_k) + \epsilon_k^q. \quad (2.106)$$

Les observations disponibles au temps t_k (y_k^m) sont alors liées au vrai état x_k^t par,

$$y_k^m = \mathcal{H}_k(x_k^t) + \epsilon_k^o. \quad (2.107)$$

En pratique, le bruit de mesures ϵ_k^o et de modèle ϵ^q sont souvent considérés comme gaussiens. Cependant l'erreur de modèle peut être difficile à estimer. Le bruit de mesures s'il suit une loi normale peut s'exprimer comme,

$$\epsilon_k^o \sim \mathcal{N}(0, \mathbf{R}_k), \quad (2.108)$$

où \mathbf{R}_k est la matrice de covariance d'erreur de mesures associée au temps t_k . Dans cette thèse, les erreurs sont considérées comme non corrélées et ainsi \mathbf{R}_k est considérée comme diagonale. Enfin remarquons que si les variables en jeu θ , y^m , ϵ^o , ϵ^q suivent une loi normale, le problème de l'estimation de θ consiste à estimer sa moyenne et sa variance.

Filtre de Kalman

Le filtre de Kalman [Kalman1960] a de nombreuses applications et est très populaire dans les méthodes d'estimation. Il est notamment utilisé dans les radars, ou pour donner la position GPS. En connaissance d'un modèle dynamique (de déplacement concernant la navigation pour un GPS par exemple), ce filtre réduit l'influence du bruit de mesure, d'où son intérêt dans l'assimilation de données. La formulation du filtre de Kalman s'exprime traditionnellement avec des opérateurs modèle-observation linéaires. Les équations de dynamique et d'observation (2.104) et (2.105) se reformulent alors,

$$\theta_{k+1} = \mathbf{M}_{k+1}\theta_k + \epsilon_k^m \quad (2.109)$$

$$\mathbf{y}_k^m = \mathbf{H}_k\theta_k + \epsilon_k^o. \quad (2.110)$$

où \mathbf{M}_{k+1} et \mathbf{H}_k sont des opérateurs linéaires de modèle d'observation. $k + 1$ désigne l'indice de temps vers lequel le modèle modifie l'état courant au temps k , θ_k . Pour rester général, l'opérateur d'observation peut considérer des mesures mobiles dont les positions évoluent en temps.

Les erreurs de modèle et de mesures au temps t_k , ϵ_k^m et ϵ_k^o sont considérées comme gaussiennes, non corrélées et de moyenne nulle, c'est-à-dire

$$\epsilon_k^q \sim \mathcal{N}(0, \mathbf{Q}_k) \quad (2.111)$$

$$\epsilon_k^o \sim \mathcal{N}(0, \mathbf{R}_k), \quad (2.112)$$

où \mathbf{Q}_k et \mathbf{R}_k sont les matrices de covariances d'erreurs de modèles et de mesures au temps t_k . Une particularité du filtre de Kalman est la séquentialité. Contrairement à la 4D-Var, le filtre peut se mettre à jour à chaque avancée de temps. Généralement l'état estimé au temps t_k sert alors d'ébauche au temps t_{k+1} .

Déroulement de l'algorithme : Nous définissons les erreurs *a priori* et *a posteriori* de la manière suivante,

$$\epsilon_k^b = \theta_k^b - \theta_k^t \quad (2.113)$$

$$\epsilon_k^a = \theta_k^a - \theta_k^t, \quad (2.114)$$

avec θ_k^b , θ_k^a , les vecteurs d'états d'ébauche et d'analyse au temps t_k et θ_k^t , l'état véritable inconnu. Leurs covariances d'erreurs respectives s'écrivent,

$$\mathbf{P}_k^b = \text{cov}(\epsilon_k^b) = \mathbb{E}[\epsilon_k^b (\epsilon_k^b)^T] \quad (2.115)$$

$$\mathbf{P}_k^a = \text{cov}(\epsilon_k^a) = \mathbb{E}[\epsilon_k^a (\epsilon_k^a)^T]. \quad (2.116)$$

- **Etape 1 :** Initialisation de θ_k^a et \mathbf{P}_k^a . L'algorithme commence par une initialisation de l'état. Comme le filtre de Kalman assimile une variable aléatoire suivant une loi normale, il faut initialiser la moyenne de l'état et sa covariance d'erreur et donc de donner θ_0^a et \mathbf{P}_0^a .

- **Etape 2 :** ensuite le filtre entre dans une étape de prédiction où il estime une ébauche sur la moyenne de l'état au temps suivant (les moyennes étant notées comme l'état) ainsi que sa matrice de covariance d'erreur associée construite à partir des erreurs de modèle,

$$\theta_{k+1}^b = \mathbf{M}_{k+1}\theta_k^a \quad (2.117)$$

$$\mathbf{P}_{k+1}^b = \mathbf{M}_{k+1}\mathbf{P}_k^a\mathbf{M}_{k+1}^T + \mathbf{Q}. \quad (2.118)$$

- **Etape 3 :** Mise à jour, actualisation de l'analyse par pondération entre savoir *a priori* (θ_k^b) et les erreurs entre les mesures et les estimations. Le filtre tient alors compte des observations pour créer

une nouvelle estimation qui est une pondération entre l'ébauche et les écarts aux observations. Le poids de cette pondération est déterminé par la matrice de gain de Kalman d'expression à $t = t_{k+1}$,

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^b \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_{k+1}^b \mathbf{H}^T + \mathbf{R} \right)^{-1}. \quad (2.119)$$

La mise à jour de la moyenne et de la covariance de θ s'écrit alors,

$$\theta_{k+1}^a = \theta_{k+1}^b + \mathbf{K}_{k+1} \left(\mathbf{y}_{k+1}^m - \mathbf{H}_{k+1} \theta_{k+1}^b \right), \quad (2.120)$$

$$\mathbf{P}_{k+1}^a = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \mathbf{P}_{k+1}^b. \quad (2.121)$$

Remarques :

- La mise à jour de l'état analysé (2.120) est solution de la fonction de coût 3D-Var (2.67) quand l'opérateur d'observation se confond avec l'opérateur en avant c.à.d. en considérant une absence d'application de modèle physique où ainsi $\mathbf{H} = \mathbf{F}$.
- Les étapes de prédiction et de mise à jour s'appellent récursivement jusqu'au dernier temps d'assimilation.
- Si des entrées de modèle γ sont considérées, le filtre de Kalman impose le choix d'un vecteur d'état augmenté $\theta = (\mathbf{x}, \gamma)^T$ ce qui nécessite quelques adaptations quant à la construction des opérateurs de modèle (\mathbf{M}) et d'observation (\mathbf{H}). Cette question a été traitée en 4.5 pour un problème de STE.

Bien qu'au fondement de l'assimilation bayésienne, le filtre de Kalman classique est peu utilisé dans les travaux d'assimilation de pollution atmosphérique, notamment à cause de non-linéarités dans les modèles physiques ainsi qu'au coût de stockage de la matrice de gain (2.119) et de l'inversion à effectuer. Il existe des possibilités de linéarisation locale des opérateurs \mathcal{M} et \mathcal{H} pour se replacer dans le formalisme de ce filtre notamment le filtre de Kalman étendu.

Filtre de Kalman étendu

Le filtre de Kalman étendu (Extended Kalman Filter en anglais, sigle : EKF) reprend les formulations du filtre de Kalman en considérant cette fois-ci les relations de dynamiques et d'observations non linéaires écrites en (2.104) et (2.105) pour rappel,

$$\begin{aligned} \theta_{k+1} &= \mathcal{M}_{k+1}(\theta_k) + \epsilon_k^q \\ \mathbf{y}_k^m &= \mathcal{H}_k(\theta_k) + \epsilon_k^o. \end{aligned}$$

Pour traiter la non linéarité, l'EKF linéarise localement autour de l'état estimé les opérateurs de modèle \mathcal{M} et d'observation \mathcal{H} tel que,

$$\mathbf{M}_{k+1} = \frac{\partial \mathcal{M}_{k+1}}{\partial \theta}(\theta_k^a), \quad \mathbf{H}_k = \frac{\partial \mathcal{H}_k}{\partial \theta}(\theta_k^a). \quad (2.122)$$

La prédiction se réalise de manière similaire au filtre de Kalman,

$$\theta_{k+1}^b = \mathcal{M}(\theta_k^a), \quad \mathbf{P}_{k+1}^b = \left(\frac{\partial \mathcal{M}_{k+1}}{\partial \theta}(\theta_k^a) \right) \mathbf{P}_k^a \left(\frac{\partial \mathcal{M}_{k+1}}{\partial \theta}(\theta_k^a) \right)^T + \mathbf{Q} \quad (2.123)$$

et la mise à jour se fait alors ainsi, la matrice de gain de Kalman est calculée,

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^b \left(\frac{\partial \mathcal{H}_{k+1}}{\partial \theta}(\theta_{k+1}^b) \right)^T \left(\left(\frac{\partial \mathcal{H}_{k+1}}{\partial \theta}(\theta_{k+1}^b) \right) \mathbf{P}_{k+1}^b \left(\frac{\partial \mathcal{H}_{k+1}}{\partial \theta}(\theta_{k+1}^b) \right)^T + \mathbf{R} \right)^{-1} \quad (2.124)$$

La moyenne de l'état et sa covariance d'erreur sont mis à jour de la manière suivante

$$\theta_{k+1}^a = \theta_{k+1}^b + \mathbf{K}_{k+1} \left(\mathbf{y}_{k+1} - \left(\frac{\partial \mathcal{H}_{k+1}}{\partial \theta} (\theta_{k+1}^b) \right) \theta_{k+1}^b \right) \quad (2.125)$$

$$\mathbf{P}_{k+1}^a = \left(\mathbf{I} - \mathbf{K}_{k+1} \left(\frac{\partial \mathcal{H}_{k+1}}{\partial \theta} (\theta_{k+1}^b) \right) \right) \mathbf{P}_{k+1}^b \quad (2.126)$$

Dans les faits, le filtre de Kalman étendu est peu populaire dans les travaux d'assimilation pour la pollution atmosphérique ou en géosciences car la méthode rencontre des difficultés à traiter les fortes non-linéarités des modèles physiques en jeu et la linéarisation locale du modèle est coûteuse en calcul [Asch2016, p. 99]. Par ailleurs, des problèmes de stabilité numérique apparaissent dans les problèmes d'estimation de grande dimension. Les auteurs dans ce livre recommandent d'ailleurs davantage les filtres d'ensemble de Kalman comme alternatives.

Filtre à particules

Le filtre à particules repose sur un raisonnement bayésien construit à partir de tirages. Il est par conséquent qualifié de « méthode d'ensemble ». Le formalisme du filtre de référence nommé « Bootstrap Particule Filter » (BFT) est visible en détail dans [Gordon1995]. Cette méthode d'ensemble permet d'approximer une Densité De Probabilité (DDP) du vecteur de contrôle à estimer θ . Le filtre à particules est donc envisageable pour estimer des variables dont la DDP n'est pas normale. Cependant, de par son coût de calcul et sa construction mathématique, le filtre à particules est mal adapté pour l'estimation de vecteurs cibles avec de nombreuses variables. L'estimation reste bayésienne et à la manière du filtre de Kalman, cherche à reconstruire la distribution d'un vecteur cible. Un filtre à particules tel que posé dans le BFT se construit dans les grandes étapes de la manière suivante :

A. Tirage : Considérons $\{\theta_0, \dots, \theta_{m-1}\}$ particules avec $\theta_i \in \mathbb{R}^{N_\theta}$, $i \in \{0, \dots, m-1\}$. La DDP au temps t_k est $p_k(\theta) \approx \sum_{i=1}^m \omega_i^k \delta(\theta - \theta_k^i)$ avec δ un Dirac et w_i^k les poids qui pondèrent l'importance de chaque particule dans l'approximation de la DDP. En principe les poids à ce stade sont considérés uniformes, $\{\omega_i^k = \frac{1}{m}, 0 \leq i \leq m-1\}$.

B. Prédiction a priori : Les particules sont propagées au temps suivant par le modèle dynamique sur la distribution $p_{k+1}(\theta) \approx \sum_{i=1}^m \omega_k^i \delta(\theta - \theta_{k+1}^i)$ tel que chaque $\theta_{k+1}^i = \mathcal{M}_{k+1}(\theta_k)$ avec $i \in \{1, \dots, m\}$. Il est possible d'ajouter une considération de bruit de modèle à ce stade.

C. Analyse : L'estimation *a posteriori* (analyse) s'actualise en appliquant la loi de Bayes pour avoir accès à la distribution postérieure en actualisant les poids par rapport aux observations,

$$\omega_{k+1}^{a,i} \propto \omega_{k+1}^{b,i} p(\mathbf{y}_{k+1} | \theta_{k+1}^i), \quad i \in \{0, \dots, m-1\}, \quad (2.127)$$

avec $\omega_{k+1}^{a,i}$ représentant les poids d'analyse *a posteriori* et $\omega_{k+1}^{b,i}$ les poids *a priori*. La fonction de vraisemblance $p(\mathbf{y}_{k+1} | \theta_{k+1}^i)$ caractérise les observations perturbées par le vecteur de contrôle θ . Elle peut par exemple être considérée comme une loi normale multivariée,

$$p(\mathbf{y} | \theta) = \mathcal{N}(\mathbf{y}, \mu = \mathcal{H}(\theta), \Sigma = \mathbf{R}) \quad (2.128)$$

où \mathcal{N} désigne une loi normale multivariée ici évaluée au vecteur \mathbf{y} , de vecteur moyen μ et de covariance d'erreur d'observation Σ (ici \mathbf{R}).

D. Rééchantillonnage : Afin d'éviter que des poids prennent trop le dessus sur d'autres, des techniques de modération des poids existent. La plus populaire semble être le rééchantillonnage à importance séquentielle (Sequential Importance Resampling, SIR). Cette technique normalise les poids en sortie d'analyse et un rééchantillonnage est de nouveau réalisé si le tirage ne permet pas de bien approximer la DDP de l'état. Une étude d'assimilation de panaches de pollution a utilisé un filtre à particules rééchantillonné avec la technique SIR [Šmídl2013].

D'autres types de filtre à particules existent pour bien équilibrer les poids estimés et ainsi obtenir des estimations de DDP fiables. Nous pouvons citer l'algorithme AMIS (Adaptative Multiple Importance Sampling) utilisé dans une étude d'estimation de source ponctuelle de pollution urbaine en amplitude et en position [Septier2020]. Etant une méthode ensembliste, ce genre de méthode requiert en général une initialisation proche de l'état recherché pour converger rapidement.

Filtres d'ensembles de Kalman

Les filtres d'ensemble de Kalman (EnKF) ont une importance croissante dans les études d'assimilations, y compris dans certains modèles météorologiques [Szunyogh2008] qui jusque-là utilisaient traditionnellement les méthodes variationnelles adjointes grâce à leur robustesse et leur faible coût de calcul (ces dernières restent cependant très utilisées). Comme le filtre de Kalman, ces méthodes récursives permettent en effet une grande flexibilité, puisque l'estimation s'améliore au fur et à mesure de la collecte de nouvelles observations (du moins pour l'EnKF stochastique). Ce type de filtre reprend les hypothèses de normalité des DDP de l'état à estimer comme le filtre de Kalman mais au lieu d'estimer directement la moyenne et la covariance d'ébauche, le filtre d'ensemble de Kalman (EnKf) approxime la moyenne et la covariance de l'état à estimer en moyennant des tirages.

Filtre d'ensemble de Kalman (EnKF) Stochastique Prenons en exemple un EnKF de référence tel que développé par Evensen [Evensen2003]. Cette méthode a été proposée pour pallier les problèmes de stockage de la matrice de covariance d'erreur de l'état visible en (2.118). De plus, le filtre de Kalman étendu (EKF) est connu pour présenter des instabilités [Asch2016, page 99]. Cette méthode est ensembliste. Le filtre d'Ensemble de Kalman Stochastique (EnKF stochastique) a été la première méthode de la famille des filtres d'ensemble de Kalman à avoir été développée. Prévu pour traiter les problèmes avec de fortes non-linéarités et adapté avec ses variantes aux traitements de problèmes de grandes dimensions, l'EnKF stochastique va estimer la variance et la moyenne d'un état en faisant des tirages pour estimer la distribution de l'état en assumant des erreurs gaussiennes. Il s'agit en cela d'une vision ensembliste du filtre de Kalman au sens de Monte-Carlo. Comme le KF, l'EnKF passe par :

- Une phase d'initialisation de l'état à estimer en donnant un ensemble d'états initiaux générés à partir d'une loi normale suivant la moyenne de l'état initial et la covariance de l'état initial posées *a priori* par l'utilisateur.
- Une phase de prédiction où le modèle physique fait avancer l'état analysé en temps et sert ensuite d'état initial pour l'analyse au temps suivant.
- Une phase de d'analyse où le gain \mathbf{K} est calculé, et où l'état est mis à jour par une pondération entre l'état prédit et le terme d'écart aux mesures.

Mesures : Les mesures tiennent compte comme dans le KF d'une matrice de covariance d'erreurs \mathbf{R}_k à chaque indice de temps (ou plus généralement d'itération) k . Cette fois-ci, l'erreur est considérée dans l'ensemble de mesures telle que,

$$\mathbf{y}^m = \mathbf{y} + \boldsymbol{\epsilon}^o, \quad \boldsymbol{\epsilon}^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \quad (2.129)$$

où \mathbf{y}^m représente la mesure vraie bruitée, \mathbf{y} représente la mesure idéale et sans erreur, et $\boldsymbol{\epsilon}^o$ est l'erreur aléatoire de mesure dont la covariance d'erreur est contenue dans \mathbf{R} .

Initialisation : Tout d'abord un ensemble de réalisations *a priori* (ébauches) est créé dans une matrices d'ensemble \mathbf{A}_0^b contenant N_{en} réalisations différentes,

$$\mathbf{A}_0^b = \left(\theta_0^{b(1)}, \theta_0^{b(2)}, \dots, \theta_0^{b(N_{\text{en}})} \right). \quad (2.130)$$

Cette matrice étant construite en tirant dans une loi normale multivariée de N_θ dimensions (la taille du vecteur de contrôle θ) telle que $\mathbf{A}_0^b \sim \mathcal{N}(\theta_0^b, \mathbf{P}_0^b)$, avec \mathbf{P}_0^b une matrice de covariance d'état initial de dimension $N_\theta \times N_\theta$ et θ_0^b l'état initial supposé de l'état à estimer.

Prédiction La prédiction est similaire dans son principe à un filtre de Kalman standard. La matrice d'ensemble est actualisée ainsi au temps t_{k+1} , en appliquant le modèle sur chacun des états d'analyse de l'ensemble :

$$\mathbf{A}_{k+1}^b = \left(\mathcal{M} \left(\theta_k^{a(1)} \right), \mathcal{M} \left(\theta_k^{a(2)} \right), \dots, \mathcal{M} \left(\theta_k^{a(N_{en})} \right) \right). \quad (2.131)$$

La matrice de covariance d'erreur d'ébauche devient,

$$\mathbf{P}_{k+1}^b = \frac{1}{N_{en} - 1} \sum_{i=1}^{N_{en}} \left(\theta_{k+1}^{b,(i)} - \bar{\theta}_{k+1}^{b,(i)} \right) \left(\theta_{k+1}^{b,(i)} - \bar{\theta}_{k+1}^{b,(i)} \right)^T = \mathbf{\Theta}_{k+1}^b \left(\mathbf{\Theta}_{k+1}^b \right)^T. \quad (2.132)$$

avec $\mathbf{\Theta}^b$ désignant un terme d'"anomalie" normalisée permettant de quantifier la variance de l'ensemble qui s'écrit,

$$\left[\mathbf{\Theta}^b \right]_{i,k} = \frac{\theta_{i,k}^b - \bar{\theta}_k^b}{\sqrt{N_{en} - 1}}. \quad (2.133)$$

Analyse : Dans cette partie, le but est de calculer la matrice de gain pour pouvoir calculer un état analysé corrigé *a posteriori* par les mesures. Pour ce faire on calcule un terme d'"anomalie" normalisée pour toute perturbation $i = 1, \dots, N_{en}$,

$$\left[\mathbf{Y}^b \right]_{i,k} = \frac{\mathcal{H}_k \left(\theta_{i,k}^b \right) - \epsilon_i^o - \bar{y}_k^b + \bar{\epsilon}^o}{\sqrt{N_{en} - 1}}. \quad (2.134)$$

Ces matrices expriment les écarts de chaque état θ_i membre de l'ensemble des $\{\theta_i\}_{i=1}^{N_{en}}$ au temps k face à la moyenne de l'ensemble $\bar{\theta}^b$ au temps k , similairement pour \mathbf{Y}^b où est exprimé un écart entre les états ramenés aux mesures $\mathcal{H}_k \left(\theta_{i,k}^b \right)$ et la moyenne desdits états ramenés aux mesures $\bar{y}^b = \frac{1}{N_{en}} \sum_{i=1}^{N_{en}} \mathcal{H} \left(\theta_i^b \right)$. La matrice de gain s'obtient alors ainsi,

$$\mathbf{K}_k = \mathbf{\Theta}_k^b \left(\mathbf{Y}_k^b \right)^T \left(\mathbf{Y}_k^b \left(\mathbf{Y}_k^b \right)^T \right)^{-1} \quad (2.135)$$

Ainsi l'ensemble des états est mis à jour comme pour un filtre de Kalman,

$$\theta_{i,k}^a = \theta_{i,k}^b + \mathbf{K}_k \left(\mathbf{y}_{i,k}^m - \mathcal{H}_k \left(\theta_{i,k}^b \right) \right). \quad (2.136)$$

L'algorithme peut ainsi itérer et fournir en sortie la moyenne de l'ensemble des états ($\mathbb{E}\{\theta_i\}_{i=0}^{N_{en}-1}$) et sa dernière covariance obtenu (\mathbf{P}^a) qui permet d'accéder à l'incertitude de l'état estimé. La matrice de covariance d'analyse s'écrit,

$$\mathbf{P}^a = \mathbf{\Theta}^a \left(\mathbf{\Theta}^a \right)^T, \text{ où } \left[\mathbf{\Theta}^a \right]_i = \frac{\theta_i^a - \bar{\theta}_k^a}{\sqrt{N_{en} - 1}}. \quad (2.137)$$

Le principe du filtre d'ensemble de Kalman et son déroulé sont présentés [Asch2016, chapitre 6] ainsi que ses extensions aux problèmes d'estimations sur fenêtre de temps longue à l'image d'une 4D-Var, par exemple la méthode Ensemble Kalman Smoother EnKS. Cette variante a d'ailleurs été appliquée dans des études d'estimations de flux turbulents [Mons2016]. Il existe cependant différentes variations du filtre de Kalman stochastique. Ici nous faisons le choix de présenter à

la manière de Asch et al. [Asch2016], cependant on retrouve une autre formulation qui varie légèrement dans [Labbe2018]. L'EnKF s'exprime aussi parfois dans certains travaux en supprimant le modèle d'observation \mathcal{H} linéaire de θ [Kato2013, Katzfuss2016]. Les filtres d'ensemble de Kalman possèdent de nombreuses variantes comme le LETKF (Local Ensemble Transform Kalman filter) utilisé avec succès en étude de la pollution atmosphérique [Dai2019] ou en climatologie [Bonan2014]. Des variantes hybrides entre le filtre de Kalman d'ensemble et l'assimilation variationnelle comme l'IEnKS ont aussi été utilisées avec succès [Defforge2021]. Les méthodes basées sur les filtres d'ensemble de Kalman gagnent à utiliser des techniques pour améliorer les estimations malgré des tailles d'ensemble petites (par exemple 20 membres d'ensemble) par rapport à la taille de la variable à estimer (par exemple 1000 inconnues). Les plus usuelles sont la *localisation* et l'*inflation*. Pour les deux méthodes il s'agit d'altérer la matrice de covariance d'erreur dans la phase de prédiction.

La localisation consiste à "régulariser" la matrice de covariance, en appliquant une fonction sur toute la matrice pour diminuer l'ampleur des corrélations avec des membres géographiquement éloignés. L'inflation consiste à multiplier la matrice de covariance par une constante légèrement supérieure à 1. Plus de détails sont présentés dans [Katzfuss2016]. L'EnKF stochastique avec l'implémentation de [Asch2016] est présenté par l'algorithme 2.

Algorithm 2 EnKF décrit à la manière de [Asch2016, p.163]

Require: Pour $k = 0, \dots, K$: la matrice de covariance d'erreurs d'observation \mathbf{R}_k , les modèles d'observations \mathcal{H}_k , les modèles directs \mathcal{M}_k , la matrice de covariance d'erreur initiale d'état, \mathbf{P}_0^b . Une intialisation d'état θ_0^b et de matrice de convariance d'erreur \mathbf{P}_0^b

- 1: Initialiser l'ensemble $\{\theta_{i,0}^b\}_{i=1,\dots,N_{en}}$ en tirant la loi multivariée $\mathcal{N}(\theta_0^b, \mathbf{P}_0^b)$.
- 2: Pour $k = 0, \dots, K$
- 3: Construire un ensemble d'observation : pour $i = 1, \dots, N_{en}$: $\mathbf{y}_{i,k}^m = \mathbf{y}_k + \epsilon_i^o$, avec $\epsilon_i^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$
- 4: Calcul de la moyenne des ensembles $\bar{\theta}_k^b = \frac{1}{N_{en}} \sum_{i=1}^{N_{en}} \theta_{i,k}^b$, $\bar{\epsilon}^o = \frac{1}{N_{en}} \sum_{i=1}^{N_{en}} \epsilon_i^o$, $\bar{\mathbf{y}}_k^b = \frac{1}{N_{en}} \sum_{i=1}^{N_{en}} \mathcal{H}_k(\theta_{i,k}^b)$ et de la moyenne des "anomalies"

$$[\Theta_b]_{i,k} = \frac{\theta_{i,k}^b - \bar{\theta}_k^b}{\sqrt{N_{en} - 1}}, \quad [\mathbf{Y}^b]_{i,k} = \frac{\mathcal{H}_k(\theta_{i,k}^b) - \epsilon_i - \bar{\mathbf{y}}_k^b + \bar{\epsilon}^o}{\sqrt{N_{en} - 1}}.$$

- 5: Calcul de la matrice de gain : $\mathbf{K}_k = \Theta_k^b (\mathbf{Y}_k^b)^T \left\{ \mathbf{Y}_k^b (\mathbf{Y}_k^b)^T \right\}^{-1}$
 - 6: Mise à jour de l'ensemble : pour $i = 1, \dots, N_{en}$: $\theta_{i,k}^a = \theta_{i,k}^b + \mathbf{K}_k (\mathbf{y}_{i,k}^m - \mathcal{H}_k(\theta_{i,k}^b))$,
 - 7: Calcul de l'ensemble de prédiction : pour $i = 1, \dots, N_{en}$, $\theta_{i,k+1}^b = \mathcal{M}_{k+1}(\theta_{i,k}^a)$.
 - 8: Fin pour
-

2.5.4 Méthodes hybrides d'Assimilation de Données

Les méthodes hybrides telle que l'IEnKS s'inspirent des méthodes variationnelles sans utiliser l'écriture d'un adjoint. La variation des paramètres pour minimiser une fonctionnelle y est propagée par l'ensemble. V. Mons l'utilise avec l'EnVar [Mons2017] ou M. Bocquet et son équipe avec l'IEnKS [Bocquet2014]. Les études qui l'utilisent peuvent alors se permettre d'utiliser des logiciels tiers (comme des modèles de dispersion de polluants pour le thème de cette thèse) et de ne pas avoir à utiliser l'adjoint. V. Mons a d'ailleurs comparé le temps d'assimilation de la méthode EnVar avec la 4D-Var et une méthode nouvelle combinant krigeage et EnVar, KEnVar [Mons2019]. Pour

son cas de test, l'EnVar est compétitif en temps de calcul par rapport à la 4D-Var et il offre des performances similaires à nombre égal de simulations numériques de mécanique des fluides (CFD).

2.6 Exemples d'assimilation de données

Plusieurs études ont été en pratique menées pour procéder à de l'Estimation de Terme Source de pollution atmosphérique avec diverses méthodes dont quelques exemples sont répertoriés dans le Tableau 2.2 et les abréviations de ce tableau sont explicitées dans le Tableau 2.3. On retrouve de nombreux travaux portant sur l'estimation de sources ponctuelles uniques [[Kovalets2011](#), [Kumar2015](#), [Kovalets2018](#), [Mons2017](#)].

Dans nombre de travaux, les sources sont estimées, soient par des méthodes variationnelles adjointes [[Singh2015](#), [Kovalets2011](#), [Kovalets2018](#), [Kouichi2016](#), [Kumar2015](#)], soient par des méthodes ensemblistes comme l'EnKF ou voisines [[Mons2016](#), [Defforge2019b](#), [Tang2013](#)]. Ces méthodes ne sont pas exclusives, d'autres travaux remontent à la source en utilisant la méthode du maximum de vraisemblance [[Winiarek2014b](#)] ou le filtre à particules [[Septier2020](#)]. Les deux derniers exemples cités utilisent la matrice source-récepteur pour économiser en temps de calcul. D'autres méthodes d'identification existent comme les méthodes à Bases Réduites présentées par J. K. Hammond [[Hammond2019](#)]. La relation source-récepteur est une référence pour Estimer un Terme Source en utilisant le principe de l'adjoint [[Seibert2004](#)].

Bien que le cas de sources ponctuelles isolées reste principalement considéré en STE, des auteurs se sont intéressés à l'estimation de sources de pollution ponctuelles multiples en adaptant une méthode variationnelle basée sur la relation source-récepteur à l'estimation de la position et des amplitudes de 1, 2, et 3 sources à la fois [[Singh2015](#)]. Il est aussi possible d'estimer des sources émissives distribuées sur un domaine de simulation comme dans l'étude de [[Tang2013](#)] où les émissions ont été considérées comme un grand vecteur de conditions aux limites.

De manière plus générale les méthodes d'Assimilation de Données et les méthodes pour répondre à des problèmes inverses sont employées dans d'autres cas d'étude que la STE. En effet, les méthodes adjointes ont par exemple été utilisées pour l'estimation de conditions limites et initiales sur des problèmes de mécanique des fluides [[Mons2016](#)]. Les Filtres d'Ensemble de Kalman et leurs variantes variationnelles ont été aussi employés pour l'estimation de constantes de modèles de mécanique des fluides [[Kato2013](#)] ou aussi pour réaliser des estimations précises de champ de vents [[Defforge2019a](#)]. Dans ce dernier travail, les auteurs ont utilisé une variante variationnelle des filtres d'ensemble de Kalman, l'IEKS (Iterative Ensemble Kalman Filter [[Bocquet2014](#)]) pour mieux estimer les conditions limites régissant le champ de vent dans une zone vallonnée ce qui a eu pour effet d'améliorer les résultats de simulation de modèle météorologiques face aux mesures terrains et ainsi d'estimer de manière plus juste le potentiel éolien d'un site géographique. Dans le même esprit, [[Defforge2021](#)] ont amélioré la précision de l'assimilation de cartes de pollution à l'échelle urbaine améliorant l'estimation de champ de vent et faisant ainsi diminuer l'écart entre les données simulées de concentration et les mesures de concentration en polluants de 40 à 50 %.

TABLEAU 2.2 – Synthèse des travaux d'estimations basé sur des modèles

Référence	Thème	θ	Algorithme	Type de domaine	Modèle vent/trans-port	Modèle source
[Singh2015]	STE	$(q_i, x_i), i = 1, 2, 3$	var-adj s-r	2D plan	uniforme/panache stat.	$s_i = q_i \delta(x - x_i), i = 1, 2, 3$
[Hammond2019]	STE	p_v, q_i	bases réduites	2D Urbain, $125 m \times 75 m$	RANS $k-\epsilon$ / AD (FEM)	Ponctuelle
[Mons2017]	STE	$\alpha, U_H, x_s, y_s, d_s$ ⁴	PODEnVar	urbain 3D, $2 \times 2 km^2 \times 0.6 km$ ⁵	VLES NS/AD	$s_{surf} \equiv (x_s, y_s, d_s)$ ⁶
[Kovalets2011]	STE stat. ⁷	urb. x_s, y_s, z_s, q_s	var-adj s-r	$340^2 m^2 \times 20 m$ MUST (soufflerie)	RANS $k-\epsilon$ /AD stat. ⁸	unique car-rée ⁹
[Kovalets2018]	STE trans.	urb. $x_s, y_s, z_s, t_s, q_s(t)$	var-adj s-r	urb. 3D CUTE 3 exp.	RANS $k-\epsilon$ /AD trans.	unique car-rée ¹⁰
[Kouichi2016]	STE/Plt. opt. cap-teurs	x_s, y_s, q_s	var-adj/S.A.	2D ($\approx km^2$)	panaches	ponctuelle
[Winiarek2014a]	STE ¹¹	q_s (PBq)	V. M., s-r	3D méso-globale	Eulérien/AD (Polair3D) ¹²	ponctuelle
[Septier2020]	STE trans.	$x_s, y_s, q_s(t)$ ¹³	PF (AMIS)	s-r 3D urb., $1.1 \times 0.9 \times 1.6 km^3$	RANS/AD	poncutelle
[Kumar2015]	STE stat. ¹⁴	urb. q_s, x_s, y_s	var-adj s-r	$384 \times 384 m^2 \times 50 m$, MUST	RANS $k-\epsilon$ /AD	poncutelle
[Tang2013]	STE régionale. trans.	(c^{CO}, s_{BC}^{CO})	EnKf	$200 \times 200 km^2 \times 2 km$	NAQPMS	distribuée $T km^{-2} annee^{-1}$

Etudes hors STE

Suite du tableau en prochaine page

4. resp. angle vent, BC vent entrant (inlet), coordonnées de source surfacique, diamètre de source.
5. Shinjuku, Tokyo. $1 \times 1 km^2$ au centre. domaine élargie pour bords.
6. source surfacique, condition de Dirichlet à $C = 10^6$ ppm.
7. polluant passif méthode adjointe source stationnaire. Etude RMSE vs capteurs. Simulation et $\mu_{terrain}$ tunnel à vent, 150 et plus capteurs.
8. $(\partial_t c = 0) + u_i \partial_{x_i} c - \partial_{x_i} (D \partial_{x_i} c) = q^S \delta_\epsilon(x - x_s)(y - y_s)(z - z_s) (= f_s(x, y, z))$, δ_ϵ fonction porte. ADREA-HF
9. $q^S \prod_{i=1}^3 \delta_\epsilon(x - x_i^s)$
10. La source émet à un moment inconnu une quantité fixe inconnue pendant une durée inconnue. $q^S \delta_{\epsilon_1}(t - T) \prod_{i=1}^3 \delta_\epsilon(x - x_i^s)$
11. Reconstruction de la source avec incertitude. Observations : stations de mesures japonaises, mesures de dépôt de ^{137}Cs autour de Fukushima.
12. WRF, Polair3D.
13. $\mathbf{x}_s = (x_s, y_s)$ avec q_s vitesse de rejet (positive et instationnaire) à T_s pas de temps.
14. Expériences must, STE par rétropanache et "renormalisation inversion" (méthode adjointe appuyée sur la relation s-r)

Tableau 2.2 – suite

Ref	Thème	θ	Algorithme	Type de do- maine	Modèles vent/trans- port	Modèle source
[Kato2013]	constantes RANS ¹⁵	$\rho, U, P, \tilde{\nu}$	EnKF	2D "flat plate bound- ary"	RANS sa/-	-
[Mons2016]	Navier- Stokes $C.L.(t), C.I.$ ¹⁶	C. I. et C. L.	4Dvar, En- Var, IEnKs	Flux 2D contre cylindre	DNS (FVM) compressible/-	-
[Defforge2019b]	C.L. vent	U_{bc} ¹⁷	IEnKs	Montagne 2 et 3D, 8×2.5 $\text{km}^2, 4 \times 4 \times$ 2.5 km^3	RANS $k-\epsilon$ /-	-
[Defforge2021]	vent et ou conc. ¹⁸	1. (u, v) 2. c , 3. (c, u, v)	IEnKs	MUST	RANS $k-$ $\epsilon/3D$ AD (FVM)	-

TABLEAU 2.3 – Variables et acronymes utilisés dans Tableau 2.2

Abbréviation	Description
θ	variable de contrôle
STE	estimation de source
var	méthode variationnelle
adj	adjointe
s-r	relation "source récepteur"
stat.	stationnaire
trans.	instationnaire/transitoire
conc.	concentration
S.A.	Simulated Annealing
V. M.	Vraisemblance Maximale
AD	Advection Diffusion
C.I.	Condition Initiale
C.L.	Condition aux Limites

15. $\tilde{\nu} = (C_{b1}, C_{b2}, C_{\omega1}, C_{\omega2}, C_{\omega3}, C_{v1})$ constantes de turbulence RANS Spallart-Almaras. \mathbf{U}, \mathbf{P} champs vitesses et pressions. $\rho, u, v, w, p, \tilde{\nu}$: densité, champs de vitesses, pressions, paramètres de turbulence.

16. 4D Var, EnVar, IEnKs

17. Profil vertical de vent latéral $U(z)$ et $V(z)$ sur 20 points (x_i, y_i) . Un vecteur de contrôle jusqu'à $l = 840$ inconnues. Six profils de vents placés en x_i, y_i pour cas 3D

18. 14 observations sont disponibles dans ce scénario

Sur les filtres de Kalman d'ensemble et leurs variantes appliquées à des problèmes inverses

Les problèmes inverses traités avec des méthodes variationnelles sont nombreux. Cependant les approches basées sur les filtres d'Ensemble de Kalman gagnent en popularité sur des cas de grandes dimensions et non linéaires où le filtrage de Kalman classique montre des limitations. [Kato2013] ont par exemple utilisé un EnKF pour estimer des constantes de modèle CFD qui sont des entrées de modèle. Comme les auteurs n'ont pas utilisé de connaissances *a priori* sur la valeur des constantes, l'initialisation des états a été réalisée par la création d'un ensemble de paramètres de valeurs distribuées dans un hypercube latin. Le principe des méthodes EnKf et de ses variantes ont en effet besoin que l'ensemble initial des états ait quelques membres proches de l'état estimé *a posteriori*. Concernant les études sur la pollution atmosphérique et dans d'autres applications où une variable d'état de large dimension est à estimer, le filtre d'ensemble de Kalman (EnKf) a besoin d'une initialisation d'état (simulation *a priori*) relativement proche en amplitude de l'état optimal sachant les mesures effectuées tandis que la covariance d'erreur initiale doit être en cohérence avec l'ensemble initial. [Tang2013] ont par exemple estimé des concentrations en monoxyde de carbone à l'échelle d'une province chinoises 50 à 150 % supérieures par endroits à la simulation *a priori* ce qui signifie que l'initialisation que les auteurs ont utilisée possédait déjà des ordres de grandeurs proches du vecteur cible. D'autres auteurs ont utilisé un EnKF pour corriger les valeurs d'une simulation numérique - déjà précise au niveau des ordres de grandeurs de concentrations en polluant - grâce à des stations de mesures [Lopez-Restrepo2021]. Le problème du choix d'une initialisation proche de la réalité est clairement mentionné dans [Defforge2019b] où il est commenté qu'une initialisation et une covariance d'erreur réaliste de l'état est à donner pour un premier cycle d'analyse¹⁹. Ainsi, donner une ébauche nulle peut être problématique avec cette méthode ou ses variantes.

2.7 Récapitulatif modèles météorologiques, techniques d'assimilation, étude de cas de la littérature

La méthode d'assimilation variationnelle en particulier les méthodes adjointes, sont des méthodes populaires dans l'estimation d'inconnues issues d'EDP [Mons2016, Nguyen2016b]. Elles ont l'avantage d'être robustes et ont un coût de calcul raisonnable. Elles ont en revanche l'inconvénient de devoir programmer un code adjoint, qui passe par la résolution d'un deuxième système d'équations différentielles. Cela rend son implémentation difficile dans des logiciels de simulation comme *Code_Saturne* ou *OpenFOAM*. Il existe différents logiciels pour générer du code adjoint à partir d'un code direct tel que *Tapenade* [Hascoet2013] développé par l'INRIA pour générer de l'adjoint en Fortran ou *dolfin-adjoint* [Farrell2013] pour des adjoints d'éléments finis sur *fenics* (Python). Cependant ces générateurs imposent un format de syntaxe au code de modèle physique direct. Les techniques d'assimilations dites « non intrusives » sont en plein essor. Elles ne nécessitent pas de code adjoint, et peuvent être couplées avec des modèles physiques très aboutis tels que *Code_Saturne* ou *paraflow* [Defforge2021, Mons2017]. L'algorithme d'assimilation choisi y est souvent une variante d'un filtre d'ensemble de Kalman, ce qui évite la programmation d'un modèle adjoint. D'autres méthodes sans adjoint existent comme les méthodes à bases réduites PBDW telles qu'exposées dans [Hammond2019].

19. "One of the major issues raised by DA methods is that they require the estimation of the background and observation error covariances. For ensemble-based methods, such as the IEnKS, the background error is represented by a limited number of vectors – the ensemble members – and the problem is to chose the background ensemble for the first analysis cycle." [Defforge2019b].

2.8 Conclusion sur l'état de l'art

Dans le présent chapitre (Chapitre 2) nous avons réalisé un état de l'art, premièrement, sur la dispersion atmosphérique de polluants en particulier en milieu urbain, deuxièmement, nous avons développé le principe de quelques méthodes numériques utilisées par les modèles de transport notamment les RBF-FD en moindre carrés (LS-RBF-FD). Enfin, nous avons présenté diverses techniques utilisées dans le domaine de l'assimilation de données pour estimer un état ou un paramètre de modèle physique tel qu'un terme de source. Nous pouvons en retenir quelques enseignements :

- Les obstacles que constitue un environnement urbain exercent une grande influence sur les circulations d'air ayant lieu dans ce milieu. La modélisation des circulations d'air à partir de la simulation numérique directe des équations représentatives de Navier-Stokes n'est pas possible à l'échelle urbaine à cause du coût de calcul prohibitif : il y a besoin de simplifier les équations de Navier-Stokes et de conservation de la masse qui régissent la dispersion de polluants.
- Les équations de Navier-Stokes moyennées (RANS) pour modéliser la micro-météorologie combinées une équation d'Advection Diffusion pour modéliser l'action du transport de polluants par les circulations d'air forment une approximation raisonnable en précision et en temps de calcul.
- Il est possible d'estimer des sources de pollution avec une méthode bayésienne ou variationnelle en précalculant un champ de vent et en faisant varier les entrées d'un modèle de dispersion de polluants. Une grande variété de méthodes d'estimation existe. L'utilisation de la relation de l'adjoint, notamment par la relation source récepteur est assez courante mais pas systématiquement utilisée. Des techniques ensemblistes à vocation plus précise que cette dernière méthode existent telles que le filtre à particules ou des variantes du filtre d'Ensemble de Kalman (p. ex. IEnKS).

Pour réaliser des estimations de terme source basées sur des modèles et par leurs capacités intéressantes de haute précision malgré une faible résolution, nous avons développé un modèle de dispersion de polluants qui résout l'ADPDE par RBF-FD puis avec variante LS-RBF-FD. Nous nous sommes appuyés sur les outils de programmation en langage Python mis à disposition par T. Hines [[Hines2015](#)]. A partir de ces bases, nous avons implémenté les équations discrétisées de l'ADPDE, et avons généré des domaines de simulations représentatifs de quartier urbains. Le chapitre suivant aborde le développement et les tests de ces modèles.

3

Modélisation et simulation par RBF-FD

Ce chapitre aborde le développement de méthodes basées sur les différences finies supportées par des fonctions de base radiales (RBF-FD) pour modéliser/simuler le transport de polluants en régime stationnaire et transitoire. Une première méthode dite de collocation (RBF-FD), puis une variante, la méthode RBF-FD aux moindres carrés (LS-RBF-FD) ont été étudiées. Le chapitre s'intéresse également à la géométrie et à la génération du domaine de simulation ainsi qu'aux unités des termes sources. Dans les cas étudiés, les RBF-FD en collocation manquent de fiabilité pour modéliser des transports de polluants dans des domaines 3D contenant des obstacles avec une météorologie réaliste. Les LS-RBF-FD permettent de relaxer les contraintes de conditions aux limites et ont retourné des résultats de simulation en plus grande cohérence avec la physique. Ce travail met en œuvre pour la première fois une simulation 3D de transport scalaire avec des méthodes de type RBF-FD. Les méthodes développées dans ce chapitre conduisent naturellement à des modèles de dimension finie qui seront utilisés pour la résolution de problèmes d'estimation de sources.

Sommaire

3.1	Introduction	62
3.2	Formalisme des méthodes RBF-FD appliqué à l'ADPDE	62
3.2.1	Mise en œuvre des opérateurs RBF-FD	63
3.2.2	Méthodes RBF-FD et LS-RBF-FD pour l'ADPDE stationnaire	64
3.3	Modélisation en 3D de domaines complexes	65
3.3.1	Le shapefile : des données d'entrée de bâti urbain.	65
3.3.2	Maillage (nuage de points) conforme et non conforme	66
3.3.3	Création d'un maillage grossier conforme aux bâtiments	67
3.3.4	Stratégie de placement de nœuds	69
3.3.5	Domaines d'études retenus	73
3.3.6	Définition formelle des sources et conversion de leurs unités	74

3.3.7	Perspectives	75
3.4	Analyse de performance des méthodes RBF-FD et LS-RBF-FD	75
3.4.1	Une solution explicite de référence en 2D	75
3.4.2	Résultats sur un cas simple en 2D	78
3.4.3	Simulations urbaines RBF-FD 2D à champs météorologiques uniformes	84
3.4.4	Résultats sur un cas simple en 3D, Résultats sur le quartier de Bonne à Grenoble en 3D	85
3.5	Une méthode LS-RBF-FD dans le cas instationnaire	89
3.5.1	Un schéma d'Euler implicite	89
3.5.2	Résultats de simulation	91
3.5.3	Conclusion	92

3.1 Introduction

Les modèles de dispersion de polluants capables de représenter la physique du transport de polluants en milieu urbain sont généralement coûteux en calcul, à l'exception des modèles en boîtes tels que SIRANE ou ADMS. Cependant, ce genre de modèle considère des champs de vélocité de vent moyens au sein des rues d'une ville ainsi que des concentrations moyennes au sein des rues. Ainsi, si une rue mesure 50 ou 100 mètres de long, il ne serait pas possible de localiser précisément une source particulièrement polluante dans cette rue. D'autre part, les méthodes résolvant des problèmes inverses gagnent à utiliser des modèles physiques sur lesquels l'utilisateur du modèle a une bonne maîtrise du dit modèle afin d'en changer le code source pour, par exemple, intégrer du code adjoint. Pour ces raisons, un effort de modélisation de dispersion de polluants en disposant de champs météorologiques pré-calculés a constitué un axe de travail important. Les méthodes de discrétisation de type RBF-FD ont besoin de peu de points de calcul pour fournir une précision satisfaisante. Cela a pour conséquence d'économiser en coût de calcul et de se permettre d'approximer des champs de pollution avec peu de points discrets. En d'autres termes les méthodes RBF-FD et LS-RBF-FD offrent une caractéristique de réduction de modèle intéressantes dans un contexte de problème inverse. De plus ces méthodes peuvent être implémentables en 2D et 3D de manière assez directe. Par conséquent ces approches ont été choisies pour créer un modèle original de dispersion de polluants.

Le chapitre est organisé comme suit : Premièrement, nous commenterons la manière dont l'équation d'Advection-Diffusion (ADPDE) est approximée par des opérateurs linéaires obtenus par RBF-FD. Ensuite, nous rappellerons brièvement le principe des méthodes RBF-FD et commenterons son paramétrage. Nous ferons ensuite de même pour les LS-RBF-FD. Nous nous attarderons ensuite sur la génération de domaine de calcul urbains adapté à notre problème. La performance des méthodes RBF-FD et LS-RBF-FD sera discutée à travers divers exemples bidimensionnels et tridimensionnels. Les deux méthodes seront notamment confrontées dans un cas 2D disposant d'une solution analytique. Nous avons développé un schéma implicite d'Euler pour la méthode LS-RBF-FD. Nous montrerons comment ainsi que quelques résultats de simulations. Enfin nous conclurons.

3.2 Formalisme des méthodes RBF-FD appliqué à l'ADPDE

Dans le cadre de ce travail des modèles de dispersion de polluants (au sens du transport scalaire modélisé par l'équation d'ADPDE) 2D et 3D tenant compte du bâti ont été développés. La météorologie étant calculée au préalable par un modèle tiers.

Pour rappel l'Advection Diffusion simplifiée modélisée ici s'écrit comme dans (2.13),

$$\partial_t c + \sum_{i=1}^3 u_i \partial_i c - \sum_{i=1}^3 \partial_i (k_i \partial_i) c = s. \quad (3.1)$$

avec les conditions initiales et aux limites,

$$\begin{aligned} \sum_{i=1}^3 \mathbf{n}_i \partial_i c(x, t) &= 0 & \mathbf{x} \in \Gamma_N \\ c(x, t) &= 0 & \mathbf{x} \in \Gamma_D \\ c(x, t = t_0) &= c_0 & \mathbf{x} \in \Omega, \end{aligned} \quad (3.2)$$

où u_i et k_i appartiennent chacun à \mathbb{R} comme composantes cartésiennes des champs de vent moyen $\mathbf{u} \in \mathbb{R}^3$ et de diffusion totale $\mathbf{k} \in \mathbb{R}^3$ (turbulence comprise) moyens. La micro-météorologie

est calculée par un logiciel tiers capable de fournir des champs météorologiques moyennés en sorties par exemple des modèles RANS. c , s représentent respectivement le champ scalaire de concentration ($[M V^{-1}]$) et le terme source ($[M V^{-1} T^{-1}]$).

Le modèle (3.1)-(3.2) peut être semi-discrétisé, en conservant l'expression continue des opérateurs de dérivées mais en considérant des vecteurs pour représenter les champs scalaires de concentrations et des sources et des matrices pour représenter les champs vectoriels météorologiques. Ainsi, le modèle est défini par

$$\partial_t \mathbf{c} + \sum_{i=1}^3 \mathbf{U}_i \partial_i \mathbf{c} - \sum_{i=1}^3 \partial_i (\mathbf{K}_i \partial_i) \mathbf{c} = \mathbf{s}. \quad (3.3)$$

avec pour conditions initiales et aux bords,

$$\begin{aligned} \sum_{i=1}^3 \mathbf{n}_i \partial_i \mathbf{c}(\mathbf{x}, t) &= 0 & \mathbf{x} \in \Gamma_N \\ \mathbf{c}(\mathbf{x}, t) &= 0 & \mathbf{x} \in \Gamma_D \\ \mathbf{c}(\mathbf{x}, t = t_0) &= \mathbf{c}_0 & \mathbf{x} \in \Omega, \end{aligned} \quad (3.4)$$

avec $\mathbf{c}, \mathbf{s} \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_x}$ désignant les vecteurs de concentration de terme source et $\mathbf{U}, \mathbf{K} \in \mathbb{R}^{3 \times N_x} \times \mathbb{R}^{3 \times N_x}$. N_x est le nombre de points discret où la solution au problème (3.1)-(3.2) est approximée.

Nous allons détailler la construction des opérateurs discrets de dérivée spatiale par RBF-FD par la suite. Le traitement de l'évolution temporelle du problème (3.1)-(3.2) sera discuté en 3.5.

3.2.1 Mise en œuvre des opérateurs RBF-FD

Nous choisissons ici d'implémenter les opérateurs RBF-FD dans un formalisme hors collocation où on sépare les points d'évaluation Y des points de centres de bases X avec $Y \neq X$. Nous notons les différents ensembles considérés (p. ex. ensemble de points d'évaluation de Dirichlet, Neumann) avec les notations choisies dans 2.4.6.

Afin de discrétiser complètement le modèle (3.3)-(3.4), nous recherchons des opérateurs matriciels de différentiation tels que l'on puisse écrire un opérateur différentiel $\mathcal{L}c$ sous la forme $\mathbf{L}c$ où l'opérateur \mathcal{L} a été discrétisé en $\mathbf{L} \in \mathbb{R}^{N_{y\Omega} \times N_x}$ dans un espace de dimension finie tout comme $c \in \mathbb{R}$ est discrétisé en $\mathbf{c} \in \mathbb{R}^{N_x}$. Les opérateurs différentiels discrets étant construits par RBF-FD ou LS-RBF-FD tels que présentés dans 2.4.4. Ainsi l'opérateur d'advection $\mathbf{U} \cdot \nabla$ est discrétisé pour la composante cartésienne d'indice i en,

$$\mathbf{U}_i \partial_i \equiv \mathbf{U}_i \otimes \mathbf{D}_i \quad (3.5)$$

avec $\mathbf{D}_i \in \mathbb{R}^{N_{y\Omega} \times N_x}$ désignant l'opérateur traduisant une dérivée première pour la composante cartésienne d'indice i . Chaque vecteur de poids de différentiation $\mathbf{w} \in \mathbb{R}^{N_x}$ contenu ligne à ligne dans \mathbf{D}_i est calculé selon la relation (2.47). Le produit de Hardamard \otimes signifie une multiplication élément par élément entre $\mathbf{U}_i \in \mathbb{R}^{N_x}$ et chaque ligne de \mathbf{D}_i contenant les vecteurs de poids de différentiation $\mathbf{w} \in \mathbb{R}^{N_x}$ de dimensions $1 \times N_x$. Pour construire la matrice de différentiation totale \mathbf{D} , il est nécessaire de linéariser les opérateurs d'advection et de diffusion. L'expression de l'opérateur d'advection a été montrée en (3.5). Pour exprimer l'opérateur diffusif de l'ADPDE, nous lui appliquons donc la relation de dérivée,

$$\partial_i (\mathbf{K}_i \partial_i) = \partial_i (\mathbf{K}_i) \partial_i + \mathbf{K}_i \partial_{ii} \quad (3.6)$$

où ∂_{ii} désigne la dérivée spatiale seconde pour la composante cartésienne d'indice i ($i = \{1, 2, 3\}$). Ainsi l'opérateur de diffusion s'écrit,

$$\partial_i (\mathbf{K}_i \partial_i) \equiv (\mathbf{D}_i \mathbf{K}_i) \otimes \mathbf{D}_i + \mathbf{K}_i \otimes \mathbf{D}_{ii}, \quad (3.7)$$

avec $\mathbf{D}_{ii} \in \mathbb{R}^{N_y \Omega \times N_x}$, une matrice de différentiation de même dimension que \mathbf{D}_i construite cette fois ci en considérant un opérateur de dérivée seconde dans la relation (2.47). Remarquons la possibilité d'utiliser une interpolation RBF-FD directement sur l'opérateur gradient-turbulence $\partial_j(\mathbf{K})$ au lieu d'appliquer la relation $\mathbf{D}_i \mathbf{K}_i$, une fois la matrice de différentiation de dérivée première \mathbf{D}_i obtenue. Cette dernière option est choisie pour accélérer les calculs des opérateurs de l'ADPDE pour calculer uniquement deux matrices de différentiation RBF-FD \mathbf{D}_i et \mathbf{D}_{ii} . De plus, d'après nos observations, les calculs ne sont pas pénalisés par cette considération. Par conséquent, l'équation modélisée par le code stationnaire LS-RBF-FD devient sur les points intérieurs $y \in Y_\Omega$,

$$\mathbf{U}_i(Y) \otimes \mathbf{D}_i(Y_\Omega, X) \mathbf{c}(X) = ((\mathbf{D}_i(Y_\Omega, Y) \mathbf{K}_i(Y)) \otimes \mathbf{D}_i(Y_\Omega, X)) \mathbf{c}(X) + (\mathbf{K}_j(Y) \otimes \mathbf{D}_{ii}(Y_\Omega, X)) \mathbf{c}(X) + \mathbf{s}, \quad (3.8)$$

où nous rappelons entre parenthèses les ensemble de points discrets couverts par les différentes matrices et vecteurs du modèle en dimension finie. Remarquons que pour une diffusion uniforme le terme diffusif de dérivée première s'annule, $(\mathbf{D}_i(Y_\Omega, Y) \mathbf{K}_i(Y)) \otimes \mathbf{D}_i(Y_\Omega, X) = 0$. De (3.8), l'opérateur de nœuds intérieur discrétisé devient :

$$\mathbf{D}_\Omega = \sum_{i=1}^3 \mathbf{U}_i(Y) \otimes \mathbf{D}_i(Y_\Omega, X) \mathbf{c}(X) - ((\mathbf{D}_i(Y_\Omega, Y) \mathbf{K}_i(Y)) \otimes \mathbf{D}_i(Y_\Omega, X)) \mathbf{c}(X) + (\mathbf{K}_i(Y) \otimes \mathbf{D}_{ii}(Y_\Omega, X)) \mathbf{c}(X). \quad (3.9)$$

L'opérateur de conditions limites quant à lui s'écrit :

$$\mathbf{D}_{\partial\Omega}(y_k, \mathbf{x}) = \begin{cases} \sum_{i=1}^3 \mathbf{n}_i \mathbf{D}_i(Y_{\Gamma_N}, X) \text{ si } y_k \in \Gamma_N \\ \mathbf{D}_0(Y_{\Gamma_D}, X) \text{ si } y_k \in \Gamma_D \end{cases} \quad (y_k, \mathbf{x}) \in (Y \times X^{N_x}) \subset (\mathbb{R}^3 \times \mathbb{R}^{3 \times N_x}), \quad (3.10)$$

avec Γ_N désignant le domaine de frontière où s'appliquent les conditions de Neumann en (3.2) et similairement pour les conditions de Dirichlet sur Γ_D . \mathbf{n}_i est la composante cartésienne d'indice i du vecteur normal à la surface centrée au point y_k si une condition de Neumann s'applique. Y_{Γ_N} , Y_{Γ_D} sont les sous domaines de frontières selon le type de condition limite. \mathbf{D}_0 est l'opérateur RBF-FD d'interpolation à l'ordre de dérivée 0 (interpolateur) de Y_{Γ_D} vers X .

L'implémentation informatique des opérateurs est explicitée en Annexe A. Dans ce travail, nous faisons le choix d'utiliser comme fonction de base RBF par défaut les PHS5 présentées en (2.33), régularisées à l'aide de monômes d'ordre 2 ou plus, pour approximer les opérateurs de la physique intervenant dans le modèle de dispersion (3.1). Les bases PHS5 ont en effet l'avantage d'être précises avec l'usage de polynômes de régularisation [Flyer2012, Bayona2017, Bayona2019] et ne nécessitent pas de déterminer un facteur de forme. De plus elles sont dérivables deux fois et à dérivées continues (C^2) ce qui permet de construire les opérateurs RBF-FD d'advection (3.5) et de diffusion (3.7).

Remarquons que si les RBF-FD sont en collocation alors $Y = X$ ce qui ne change pas le formalisme des opérateurs de différentiation RBF-FD montrés en (3.9)-(3.10).

3.2.2 Méthodes RBF-FD et LS-RBF-FD pour l'ADPDE stationnaire

Nous avons développé en 3.2.1 l'implémentation des opérateurs RBF-FD considérés pour construire les matrices de différentiation en vue d'appliquer les méthodes RBF-FD et LS-RBF-FD à l'ADPDE. Nous avons donné en 2.4 un exemple d'EDP stationnaire discrétisée puis résolu par RBF-FD. Nous avons fait de même en 2.4.6 pour les LS-RBF-FD où nous avons exposé un exemple de résolution d'une EDP stationnaire par cette méthode. Tel qu'implémenté dans nos programmes informatiques, nous avons décidé de construire une grande matrice de différentiation \mathbf{D} et un terme

de contraintes \mathbf{f} pour représenter l'ADPDE (3.3)-(3.4) considérée en régime stationnaire ($\partial_t c = 0$) avec

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_\Omega \\ \mathbf{D}_{\partial\Omega} \end{pmatrix}, \mathbf{f} = \begin{pmatrix} \mathbf{s} \\ \mathbf{0} \end{pmatrix}, \quad (3.11)$$

où \mathbf{D}_Ω et $\mathbf{D}_{\partial\Omega}$ sont les matrices de différentiation spécifiques à l'ADPDE respectivement exposées en (3.9) et (3.10). \mathbf{s} un vecteur source évalué sur l'intérieur du domaine de simulation, et $\mathbf{0}$, un vecteur de conditions de bords nulles sur le bord du domaine.

Pour réaliser une méthode RBF-FD comme exposé en 2.4, \mathbf{D} est une matrice carrée inversible de dimension $N_x \times N_x$ et le vecteur source \mathbf{s} inclus dans le vecteur de contraintes \mathbf{f} est défini sur l'ensemble d'approximation X de cardinalité N_x . Le vecteur de concentration approximé (\mathbf{c}_a) a alors pour expression,

$$\mathbf{c}_a = \mathbf{D}^{-1}\mathbf{f}. \quad (3.12)$$

Dans ce travail, (3.12) est résolu soit avec un solveur généraliste de systèmes creux tel que le solveur de *scipy* pour des matrices creuses¹ soit avec le solveur GMRES en utilisant le préconditionnement proposé par T. Hines [Hines2015]. La matrice de différentiation contient $N_x \times n$ termes non nuls avec n le nombre de voisins considéré dans les stencils.

Pour appliquer la méthode LS-RBF-FD comme exposé en 2.4.6, \mathbf{D} est une matrice rectangulaire de dimensions $N_y \times N_x$ et le vecteur source \mathbf{s} inclus dans le vecteur de contraintes \mathbf{f} est cette fois évalué dans l'ensemble d'évaluation Y spécifique à la méthode LS-RBF-FD, comme discuté en 2.4. Résoudre l'ADPDE stationnaire est ainsi réduit à résoudre le problème des moindres carrés suivant :

$$\mathbf{c}_a = \arg \min_{\mathbf{v}} \|\mathbf{D}\mathbf{v} - \mathbf{f}\|^2. \quad (3.13)$$

Ce problème est résolu dans nos programmes informatiques avec le solveur LSQR mais a été testé avec succès avec le solveur L-BFGS-B. Cette fois, la matrice de différentiation représentative de l'ADPDE (\mathbf{D}) contient $N_y \times n$ éléments non nuls et le vecteur de contraintes (\mathbf{f}) contient N_y éléments. Dans une méthode LS-RBF-FD, il est de bonne pratique de considérer un ensemble de points d'évaluation Y de cardinalité supérieure à l'ensemble d'approximation X avec, à minima, $N_y \geq 3N_x$. Cela implique donc que la méthode LS-RBF-FD traite au moins 3 fois plus d'entrées non nulles dans le problème (3.13) que la méthode RBF-FD dans le problème (3.12) à nœuds d'approximation X identiques.

3.3 Modélisation en 3D de domaines complexes

A notre connaissance, aucune référence ne traite de la création d'un domaine de simulation 2D ou 3D pour la simulation de dispersion de polluants en milieu urbain par méthode sans maillage. Il a donc été nécessaire de développer des stratégies de placement de nœuds adaptées pour réaliser des simulations de type RBF-FD. Ici la problématique est de produire un nuage de calcul à partir d'une carte de bâtiments géolocalisés.

3.3.1 Le shapefile : des données d'entrée de bâti urbain.

Les fichiers *shapefile* (abréviation SHP) servent à renseigner le bâti urbain. Ils permettent de renseigner l'empreinte au sol des bâtiments sous forme de liste de polygones ainsi que leurs

1. Voir la documentation de la fonction `scipy.sparse.linalg.inv` sur internet.

hauteurs respectives. Les tronçons routiers peuvent également être renseignés dans un SHP et sont renseignés par leur altitude et par des lignes. On peut trouver des données de bâti sur des bases de données ouvertes comme Open Street Maps ou encore l’institut géographique national français, l’IGN. Un exemple d’un tel fichier geolocalisé en visible en Figure 3.1.

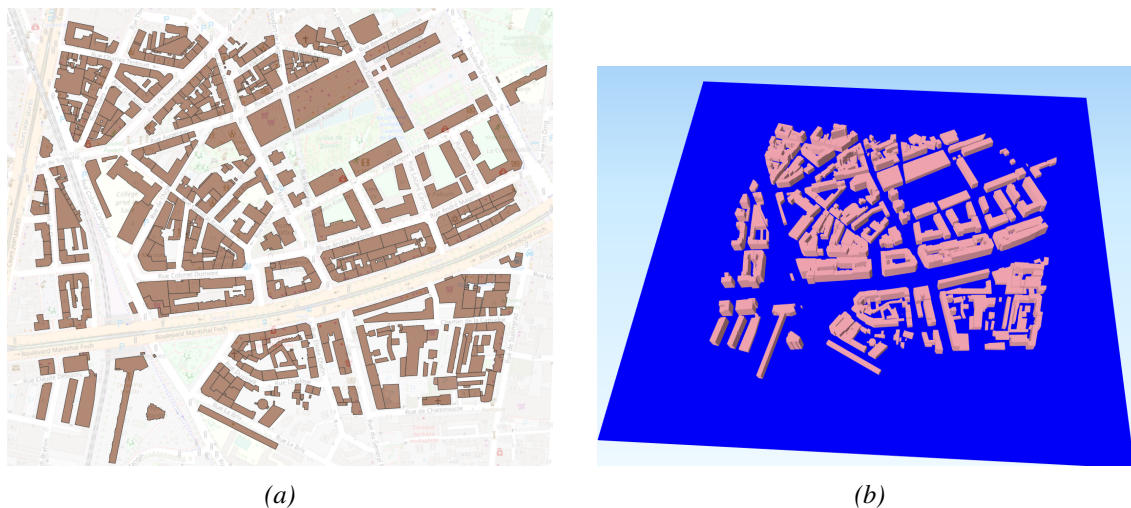


FIGURE 3.1 – (a) Empreinte au sol d’un quartier de Grenoble (Bonne), (b) bâtiments extrudés selon leur hauteur sur un terrain plat.

Les données d’intérêt sont contenues dans des fichiers SHP de bâtiments. Les attributs d’un SHP de bâtiment peuvent être mis sous la forme du tableau 3.1 où les identifiants de bâtiments et l’expression de l’empreinte au sol a été abrégée avec l’ajout de "...". Pour aller à l’essentiel nous n’avons pas développé de modèle de terrain pour adapter la hauteur du sol à la base des bâtiments renseignée par le champ `z_min_sol`. Par conséquent nous considérons cette valeur à 0 pour tous les bâtiments et les construisons sur un modèle de terrain plat en renseignant leur hauteur afin d’aboutir à un domaine 3D dont la forme des obstacles est illustrée en Figure 3.1b.

TABLEAU 3.1 – Attributs pertinents pour la description du premier bâtiment d’un shapefile quelconque.

ID	hauteur	z_min_sol	geometry
bâtiment...742	16.6	212	POLYGON Z ((714001 5006713 228.7, ..., 714001 5006713 228.7))

3.3.2 Maillage (nuage de points) conforme et non conforme

Pour générer le domaine de simulation, nous avons opté pour la génération d’un nuage de points de calcul conforme à l’enveloppe des bâtiments au contraire d’un domaine de calcul non conforme qui viendrait par exemple découper les bâtiments sur une grille carrée comme montré en Figure 3.2a où les bords des bâtiments sont crénelés par l’action de la grille carré tandis que les maillages et nuages de points sont conformes aux obstacles dans les exemples illustrés en Figure 3.2b et Figure 3.2c. La génération d’un espace de calcul non conforme aux bâtiments ferait sens, car dans le cadre de l’estimation de source de pollution en milieu urbain, la connaissance précise de la physique aux interfaces bâtiments-air n’est pas indispensable. De plus l’enveloppe des bâtiments fournie par les fichiers géolocalisés de l’IGN ou de Open Street Maps est déjà

une simplification et une approximation de la forme réelle des bâtiments urbains. Par ailleurs, les modélisations de polluants atmosphériques ne considèrent habituellement pas la forme des toits des immeubles qui sont généralement considérés comme plats. Dans le travail présent, la génération de domaine conforme est choisie pour des raisons de praticité car des flux de travail de placements de nœuds de calcul avaient déjà été générés dans ce sens par Trever Hines [Hines2015].

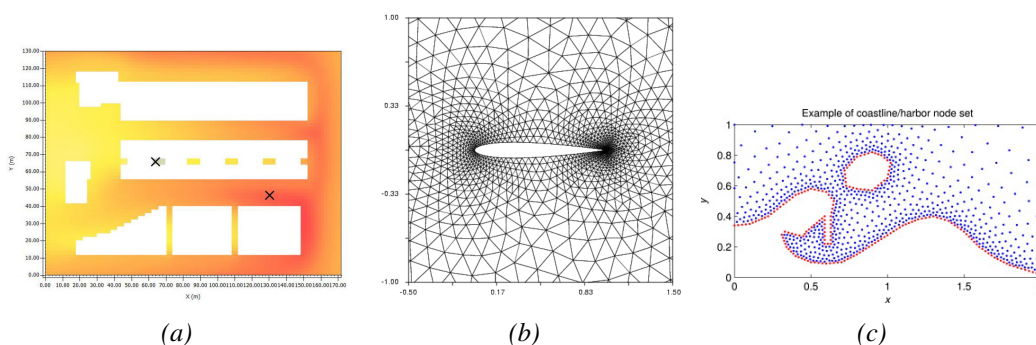


FIGURE 3.2 – (a) Exemple de maillage non conforme aux obstacles sur grille structuré [López-Cabeza2018]. (b) Maille non structurée conforme [Cusdin2003]. (c) Nuage de points RBF-FD conforme [Fornberg2015b].

3.3.3 Création d'un maillage grossier conforme aux bâtiments

Nous utilisons l'algorithme de placement de nœuds programmé par T. Hines [Hines2015]. L'algorithme de placement de nœuds attend une description en grosse maille du domaine avec les obstacles urbains, le sol, ainsi que les frontières de domaine sur le côté et sur le plafond. Entendons par une description en "grosse maille", un maillage assez défini pour représenter les contours de l'enveloppe du domaine de simulation ; ainsi le plafond qui n'a pas d'obstacles est défini seulement par deux triangles constituant sa face.

Pour définir les frontières de domaine, 5 faces élémentaires sont créées : 4 pour décrire les côtés coplanaires aux axes $x - z$ et $y - z$ comme visible sur en Figure 3.3 où la face coplanaire à $x - z$ est illustrée au premier plan. Une face coplanaire à $x - y$ définit le plafond du domaine de simulation en Figure 3.3 à 80 mètres de hauteur. Une face assise sur le plan $x - y$ définit le sol de la simulation. Le maillage des obstacles est construit en calculant l'enveloppe convexe des bâtiments en 3D, en donnant en entrée les coordonnées de la face à hauteur du sol h_0 et de même à hauteur h_f pour la hauteur du toit, pour chaque obstacle. Le sol et les obstacles permettent de renseigner la couche basse du domaine de simulation. Le maillage grossier représenté en perspective est visible en Figure 3.4. Remarquons que le terrain plan n'est pas connecté avec les obstacles, l'algorithme de placement de nœuds se chargeant de faire les liaisons. Mais il serait intéressant plus tard que le maillage grossier des bâtiments coïncide avec le sol c'est à dire que les bâtiments et le sol soient connectés ensemble par des nœuds communs.

Renseignement des conditions limites selon les surfaces de domaine : Des conditions de Neumann de flux nul sont attribuées par défaut au plan $x - y$ pour la frontière à basse hauteur du domaine, composée du sol et des bâtiments. Le plafond de simulation en $x - y$ reçoit également des conditions de Flux nul. Ce choix se justifie par l'hypothèse de faible dispersion des polluants en hauteur. Cependant une condition de Dirichlet nulle imposerait de trop grandes contraintes sur les espèces en dispersion pour un plafond avec une hauteur relativement limitée. Sur les bords, des conditions de Dirichlet nulles sont imposées. Cela traduit une hypothèse de nullité des concentrations en polluants loin des sources. En pratique, cette considération impose le choix de bords de domaine très éloignés d'une zone de calcul utile rendant alors les calcul coûteux. On

appelle la partie de domaine hors de la zone d'intérêt la zone de nudge dont un exemple tiré de [Defforge2021] est illustré en Figure 3.5a. Cette zone impose des contraintes sur les variables mais ne fausse pas les valeurs dans la zone de domaine utile. Parfois des conditions de Neumann en aval du sens du vent peuvent permettre de se dispenser de zone de nudge trop grande.

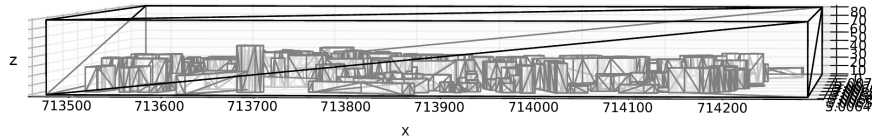


FIGURE 3.3 – Domaine grosse maille en vue $x - z$

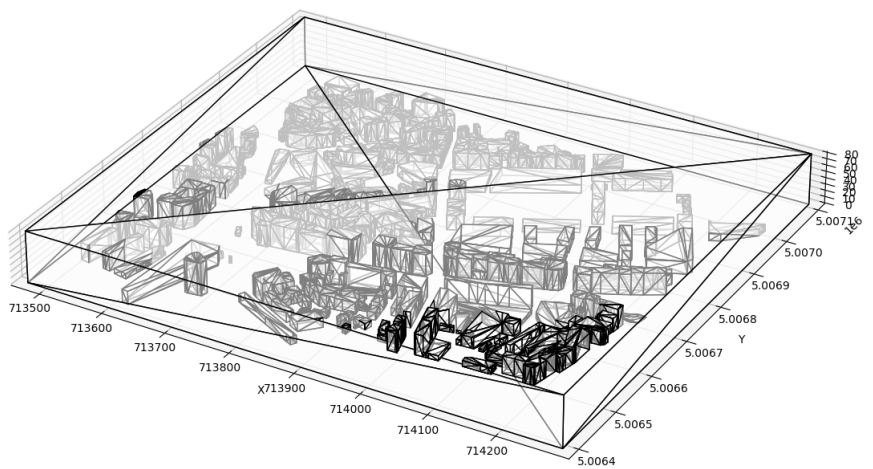


FIGURE 3.4 – Domaine grosse maille vu en perspective.

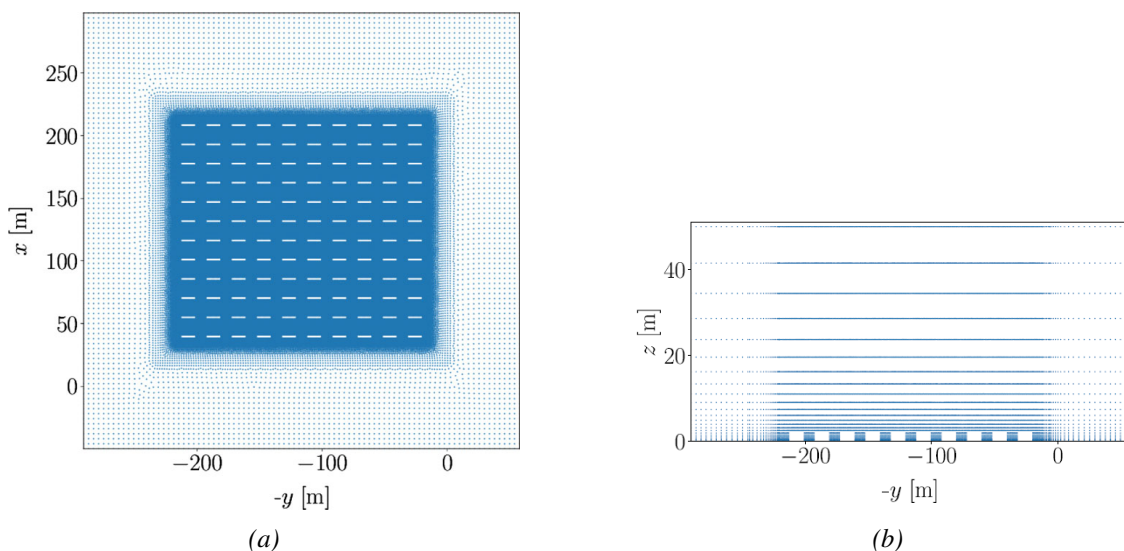


FIGURE 3.5 – Exemple d'un domaine de calcul Eulerien 3D (a) vue du dessus dans le plan $x - y$ avec zone de nudge autour du domaine utile (partie du domaine où les cellules de calcul sont beaucoup moins denses que le domaine d'intérêt au centre). (b) vue de côté. Images tirées de [Defforge2021].

Une fois ledit maillage grossier de domaine créé, nous pouvons le passer à l'algorithme de placement de nœuds de calcul, qui va utiliser l'information de la description des contours du

domaines pour placer les conditions aux limites, et les points de calcul dans la zone de circulation d'air et de polluants. La structure des données de maille grossière est visible en Annexe A.2.

3.3.4 Stratégie de placement de nœuds

Nous avons programmé un algorithme en amont du placement de nœuds de RBF qui effectue un maillage grossier du domaine conforme aux enveloppes 2D de bâtiments cartographiés (contenu dans des *shapefile*, d'extension *.shp*) extrudées selon leurs hauteurs individuelles. La maille conforme aux bâtiments au sol et au niveau de la canopée urbaine est complétée à droite, à gauche, en haut et en bas pour ses côtés ainsi qu'en hauteur au plafond de domaine par des plans où sont renseignés les différents types de conditions limites de bords. La maille conforme composée de triangles élémentaires est passée à un algorithme de placement de nœuds, ici développé par Trever Hines [Hines2015]. Les nœuds sont alors placés par un algorithme de Minimisation d'Énergie (ME) [Fornberg2015c, p. 196]. La ME pour le placement de centre de bases RBF permet de garantir une distribution uniforme de nœuds à petite échelle afin de permettre la bonne approximation des opérateurs spatiaux de l'équation différentielle à discrétiser.

A propos des algorithmes de placement de nœuds

Le placement de nœuds par ME repose sur une analogie physique avec une loi de Coulomb, consistant à attribuer une charge électrique à chaque nœud et à distribuer les nœuds en minimisant l'énergie totale. Les nœuds de frontières sont placés par un algorithme d'échantillonnage par rejet. Les nœuds sont identifiés selon leurs conditions (intérieur, Neumann, Dirichlet etc) et permutés selon leur proximité avec les nœuds voisins pour que des groupes de nœuds proches géographiquement soient proches au niveau des indices de matrices : l'algorithme de Cuthill-MacKee est utilisé à ces fins et permet d'accélérer ensuite les calculs de modèle. Il est par ailleurs possible de forcer une plus grande densité de nœuds ρ à des endroits localisés, par exemple au niveau de la canopée urbaine. Il faut cependant donner une bonne fonction de densité. Au niveau de l'analogie avec une loi de Coulomb, la densité de nœuds ρ est équivalente à l'inverse d'une densité de charge q de sorte que $\rho \equiv 1/q$. Cela traduit le fait que plus la charge de répulsion des nœuds est faible dans un milieu localisé du domaine, plus il peut y avoir un grand nombre de nœuds dans le milieu local.

Précisons comment est réalisée la distribution des nœuds par Minimisation d'Énergie. Tout d'abord, les nœuds de calcul sont préalablement distribués de manière pseudo-aléatoire à l'intérieur du domaine de simulation en utilisant une suite de Halton. A ce stade, les nœuds de calcul sont dispersés par une opération de répulsion que nous allons détailler. Soit un espace de dimension d , d étant égal à 3 dans une modélisation tridimensionnelle. Nous devons disperser un ensemble de nœuds $\mathbf{X} \in \mathbb{R}^{N \times d}$. Chaque nœud $\mathbf{x} \in X$ est entouré de nœuds voisins $\mathbf{S} \in \mathbb{R}^{m \times d}$ inclus dans X . Chaque nœud $\mathbf{x} \in X$ est entouré par au moins $m = 4$ nœuds voisins distincts $(s_1, s_2, s_3, s_4) \in \mathbf{S}^4$ quand $d = 3$, les forces de répulsion des nœuds s'expriment alors de la manière suivante :

$$\mathbf{F}_{\mathbf{S}_j \rightarrow \mathbf{x}_i} = \frac{1}{\rho(\mathbf{X}_i)\rho(\mathbf{S}_j)} \frac{\mathbf{X}_i - \mathbf{S}_j}{\|\mathbf{X}_i - \mathbf{S}_j\|^3} \quad i = \{1, \dots, N\}, j = \{1 \dots m\} \quad (3.14)$$

avec $\mathbf{F}_{\mathbf{S} \rightarrow \mathbf{X}}$ désignant une matrice de dimension $N \times m \times d$ collectant les vecteurs de forces de répulsion appliquées sur tous les nœuds de X . La résultante de ces forces permet alors de déplacer les nœuds de calcul X pour une distribution optimale au sens de la Minimisation d'Énergie.

Le flux de travail du placement de nœuds est illustré par l'algorithme 3 et les étapes de la dispersion de nœuds par minimisation d'énergie sont illustrées en algorithme 4. Les détails de l'implémentation du code de génération de nœuds après extraction du maillage grossier du domaine de simulation sont disponibles publiquement ².

2. <https://github.com/treverhines/RBF/blob/master/rbf/pde/nodes.py>

Algorithm 3 Flux de travail de la fonction de placement de nœuds pour min_energy_nodes

```

1: procedure MIN ENERGY NODES
2:   Entrées : N : int nombre de nœuds, domaine : classe, sommets et triangles (array), rho :
      fonction de densité, start : int, indice de premier terme de la suite de Halton.
3:   Sorties : coordonnees_nœuds : matrice (N,d), dictionnaire_nœuds : dict
4:
5:   procedure REJECTION SAMPLING(taille, rho, domaine, start)
6:     Distribue de manière pseudo aléatoire et en accord avec la densité rho des points de
      Halton à l'intérieur du domaine.
7:   procedure PREPARE-NODES(nœuds, domaine, rho, voisins)
8:     Orientation de normales sortantes par calcul de produit vectoriel de 2 arrêtes incidentes
      à chaque sommet de maille grossière.
9:     Dispersion de nœuds
10:    Assignation de nœuds aux frontières
11:    for nœuds do
12:      if nœud proche du bord then Translation du nœud sur une arrête de frontière.
13:      Assignation des normales aux nœuds de bords.
14:      Création de dictionnaire d'identifiants de nœuds
15:      Création de nœuds fantômes si nécessaire
16:      Réorganisation Cuthill-Mac Kee
17:      Retourne : dictionnaire d'identifiants de nœuds, nœuds
18:    Retourne : dictionnaire d'identifiants de nœuds, nœuds

```

Algorithm 4 Dispersion de nœuds par Minimisation d'Energie

```

1: Entrées : N : int nombre de nœuds, domaine : classe, sommets et triangles (array), rho :
      fonction de densité, delta : pas de dispersion, itérations : int, nombre d'itérations
2: Sorties : nœuds
3: for i in itérations do
4:   procedure ETAPE DE DISPERSION(nœuds, rho)
5:     Calcul vectorisé des forces de répulsion sur tous les nœuds :  $\mathbf{F}_{S_j \rightarrow X_i} =$ 
      
$$\frac{1}{\rho(X_i)\rho(S_j)} \frac{X_i - S_j}{\|X_i - S_j\|^3}$$

6:     Somme des forces résultantes de répulsion pour chaque nœud
7:     Calcul de la direction de la force par normalisation de la force résultante
8:     Déplacement des nœuds par ajout de translation colinéaires aux forces de répulsions
9:     Retourne nouveaux nœuds
10:    Les nœuds repoussés derrière les bords sont ramenées au bord là où leur translation
      intersecte la frontière.
11:  Retourne nœuds

```

Paramétrisation de la fonction de densité

Le choix d'une fonction de densité est un moyen de placer des points de calcul en grand nombre dans des zones d'intérêt et en petit nombre dans des zones moins importantes (p. ex. zone de nudge, hauteur supérieures à 10 m au sein de la canopée urbaine). Nous cherchons ainsi à construire une fonction de densité de nœuds qui est maximisée à 2 mètres de hauteur et décroît plus haut. La densité dépend d'une coordonnée spatiale $x \in \mathbb{R}^3$ afin de construire un rayon

$$r = \|x - (0, 0, 2)\|^2. \quad (3.15)$$

avec $(0, 0, 2)$ la coordonnée de centre du rayon.

Dans la plupart des cas de tests menés, la fonction de densité de nœuds choisie est

$$d(x) = d_0 + \alpha \exp(-\beta r^\gamma) \quad (3.16)$$

avec d_0 , la densité de nœuds basale qui permet d'assurer une couverture minimale du domaine, α est la pondération de la partie variable de la densité, β et $\gamma > 0$ commandent la décroissance de la partie variable de la densité.

Dans le cadre de la modélisation de la dispersion atmosphérique de polluants en milieu urbain et pour des polluants rejetés à une faible hauteur, les polluants ont tendance à se déplacer davantage dans le plan horizontal que dans le plan vertical au-delà de la canopée urbaine. De plus, le domaine physique est peu contraint au delà de la canopée urbaine. Pour ces raisons, il est raisonnable de se donner pour but d'utiliser peu de points de calcul au delà de la canopée urbaine. De même pour la zone de nudge qui ne contient pas de bâtiments, il n'est pas nécessaire d'utiliser autant de nœuds de calcul que dans un domaine contraint par les obstacles. C'est pourquoi nous choisissons par défaut les coefficients $d_0 = 0.02$, $\alpha = 0.98$, $\beta = 3$, $\gamma = 2$ dans la fonction de densité (3.3.4).

Extraction 3D d'un domaine urbain simplifié par enveloppe convexe

Le flux de travail pour construire un maillage 3D à partir de l'enveloppe convexe des bâtiments est utilisé pour générer la maille grossière des obstacles, dont un exemple est illustrée en Figure 3.4. Ces données géométriques sont passées en entrée des algorithmes 4 et 3. Nous obtenons alors un domaine de points de calcul informé des conditions aux limites comme visible en Figure 3.6 où on peut voir les points de calcul marquant les conditions de Dirichlet, les points de nœuds intérieurs ainsi que les conditions de Neumann présentes au sol et au plafond de simulation. Le sol et les obstacles sont davantage visibles en Figure 3.7.

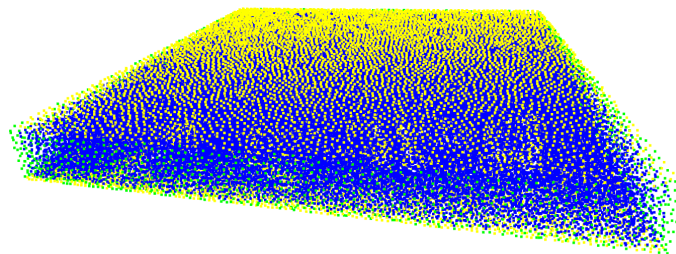


FIGURE 3.6 – Domaine de points de calcul, bleu : nœuds intérieurs, vert : nœuds de condition de Dirichlet, jaune : nœuds de condition de Neumann (flux nul).

Extraction 3D d'un domaine urbain fidèle

Pour ne pas se limiter à une représentation des obstacles par l'enveloppe convexe extrudée des empreintes au sol, nous avons essayé de générer un maillage grossier du domaine en procédant

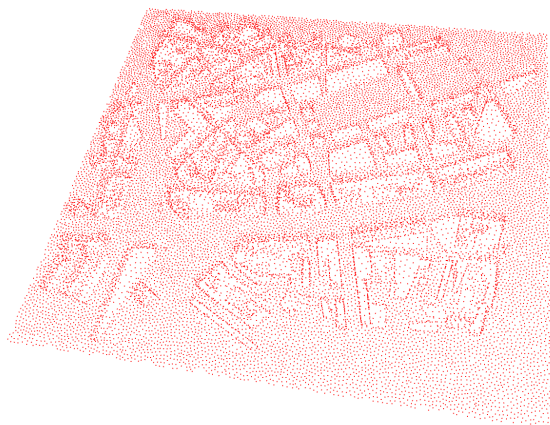


FIGURE 3.7 – Nœuds de calcul modélisant les obstacles (sol et bâtiments) et comptés comme condition de flux nul.

à une triangulation et à une extrusion des obstacles. Les polygones représentant les obstacles sont subdivisés en régions d'influence grâce à des diagrammes de Voronoi (Figure 3.8a) puis une triangulation de ces zones d'influence est opérée (Figure 3.8b). Les triangles de bords servant à l'extrusion en hauteur sont obtenus en reliant les sommets de l'empreinte au sol d'un obstacle avec son empreinte au niveau du toit (Figure 3.8c). En s'appuyant sur une collection d'obstacles complexes, il est possible de dresser une représentation en maille grossière d'un quartier urbain complexe (Figure 3.9). Cependant en donnant en entrée les faces de bâtiments agglomérées avec les faces de sol, les côtés et le plafond de la simulation à la manière de la Figure 3.4, l'algorithme de placement de nœuds 4 ne parvient pas à placer des nœuds de calcul de manière satisfaisante. Ce flux de travail permet en revanche de générer des domaines pour des bâtiments de formes simples : carrés, losanges et non adjacents. Cette incapacité de génération de domaine aux bâtiments complexes et adjacents peut être à trouver dans le fait que le nombre d'arêtes et de sommets est élevé, notamment aux zones d'interfaces bâtiments-sol et au niveau de l'interface entre deux bâtiments distincts attenants. Une procédure resterait donc à développer pour :

- Relier le terrain sans obstacles (sol) aux bâtiments, afin d'unifier la déclaration de la face basse du domaine de simulation.
- Trouver une procédure pour connecter les bâtiments adjacents et ignorer les nœuds d'interface d'obstacles hors du domaine.

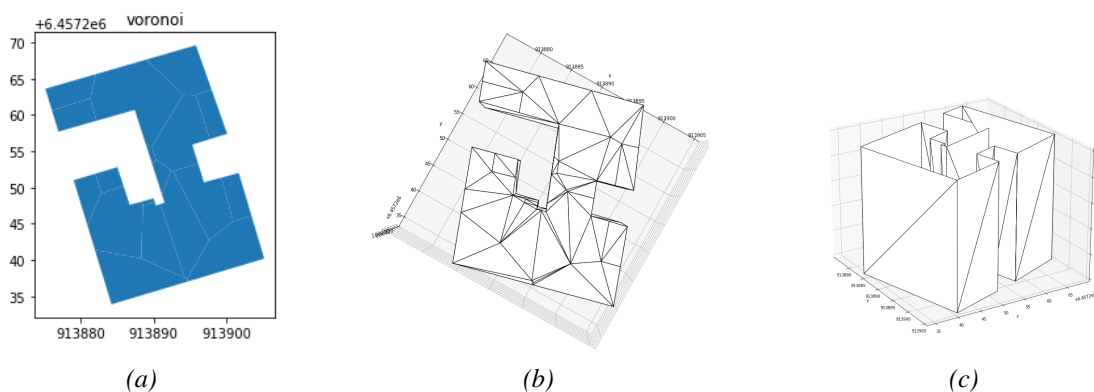


FIGURE 3.8 – (a) Identification des zones d'influence par diagramme de Voronoi, (b) Triangulation de l'empreinte au sol d'un obstacle, (c) Faces de "côté" servant à lier la face haute à la face basse des obstacles.

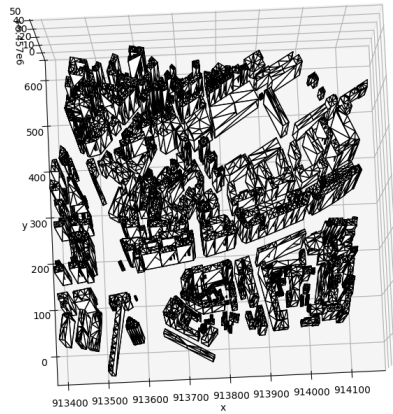


FIGURE 3.9 – Maille grossière de bâtiments sur le quartier de Bonne (Grenoble)

3.3.5 Domaines d'études retenus

Quatre domaines différents constituant quatre cas d'études sont considérés dans ce travail et sont montrés en Figure 3.4 et Figure 3.10.

1. Un domaine "jouet" carré sans obstacles (Figure 3.10, (a)) de dimensions $L_x = 250\text{m} \times L_y = 250\text{m} \times L_z = 25\text{m}$. Notons le domaine carré sans obstacles \mathcal{D}_1 .
2. Un domaine "jouet" rectangulaire avec trois obstacles de formes rectangulaires (Figure 3.10 (b)) de dimensions $L_x = 205\text{m} \times L_y = 120\text{m} \times L_z = 105\text{m}$. Notons le domaine rectangulaire à trois obstacles \mathcal{D}_2 .
3. Un domaine rectangulaire "jouet" aux dimensions identiques au domaine jouet avec obstacles illustré en Figure 3.10 (b) et sans obstacles. Notons le domaine rectangulaire sans obstacles \mathcal{D}_3 .
4. Un domaine urbain complet d'environ 1km de côté et de 80m de haut montré en Figure 3.4. Notons le domaine urbain complet \mathcal{D}_4 .

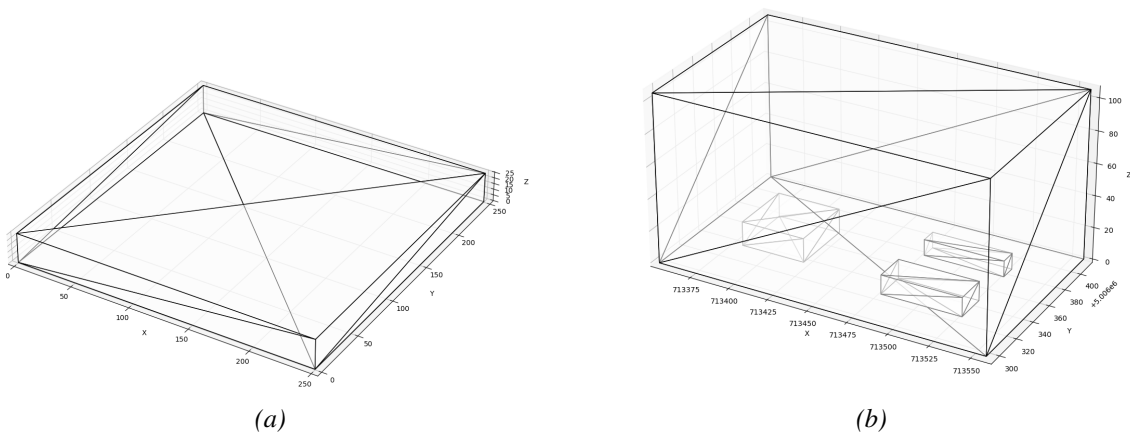


FIGURE 3.10 – (a) Domaine "jouet" sans obstacles, (b) Domaine jouet avec obstacles.

Pour les cas d'études en 2D sans dimension de hauteur, les domaines de calcul exposés sont identiques et les obstacles sont décrits par leur empreinte au sol.

3.3.6 Définition formelle des sources et conversion de leurs unités

Dans l'équation d'ADPDE (3.1), le terme source est homogène à une unité de concentration divisée par un temps $\left[\frac{X}{T}\right]$. Avec, T en s. X peut être choisi en mol m^{-3} ou kg m^{-3} ou encore en parties par million ppm $\text{mg}(\text{concentration})/\text{kg}(\text{air})$ qui est une fraction massique. Nous passons facilement d'une concentration massique à des ppm par la relation

$$c(\text{mg}/\text{m}^3) = c(\text{mg}/\text{kg}) \rho_{\text{air}}. \quad (3.17)$$

Avec ρ , la masse volumique de l'air en kg/m^3 . Les types de sources considérées dans la modélisation de la pollution varient ainsi que leurs unités. Supposons maintenant que les concentrations soient en ppm $[\text{mg}(\text{concentration})/\text{kg}(\text{air})]$ ou en mgm^{-3} . Nous considérons alors les sources :

- ponctuelles s_p (mg s^{-1}),
- linéiques s_l ($\text{mgs}^{-1}\text{m}^{-1}$),
- surfaciques s_s ($\text{mgs}^{-1}\text{m}^{-2}$) et
- volumiques s_v ($\text{mgs}^{-1}\text{m}^{-3}$).

Pour les introduire dans un modèle, par exemple en ppm, il faut que la source soit de la même dimension que la source formelle s en (3.1) à savoir en $\text{mg kg}^{-1} \text{s}^{-1}$. Les sources doivent donc, en toutes cellules de calcul dans lesquelles elles sont définies, être converties dans un premier temps en s^* en mg/s :

$$\begin{aligned} s^* &= s_p \\ &= s_l \delta r \\ &= s_s \delta S \\ &= s_v \delta V \end{aligned} \quad (3.18)$$

où δr désigne la longueur de la source linéique au sein d'une cellule de calcul selon la direction de la source, δS , la surface d'une source surfacique pour une cellule de calcul, δV , le volume d'une cellule de calcul. Ensuite la source est ramenée au volume de chaque cellule où elle est définie par

$$s(\text{ppm}) = \frac{s^*}{\rho_{\text{air}} \delta V}. \quad (3.19)$$

Si la concentration est définie en masse par volume avec $s(\text{mgs}^{-1}\text{m}^{-3})$, alors, $s = s^*/\delta V$. Cette définition formelle se retrouve dans [Bruse2007] où sont par ailleurs développées les modélisations des processus de lessivage, sédimentation etc.

Méthode proposée de conversion de source ponctuelle pour les méthodes RBF-FD

Les RBF-FD étant des méthodes sans maillage, les cellules de calcul ne sont pas explicitement définies. Nous avons donc développé une procédure de conversion des unités de source en commençant par identifier des cellules virtuelles de calcul, centrées en chaque point permettant de définir une source (des points parmi X pour la méthode RBF-FD ou parmi Y pour la méthode LS-RBF-FD, puis d'en extraire les grandeurs caractéristiques δr , δS , δV . Les diagrammes de Voronoi sont utilisés à ces fins. Cet algorithme permet d'identifier des surfaces et des volumes d'influence centrés sur des nuages de points 2D, 3D ou plus. Concernant le traitement de source ponctuelle ($[M/T]$), après avoir acquis les coordonnées de nœuds d'une cellule virtuelle de calcul, le maillage grossier de ladite cellule est calculé par extraction de son enveloppe convexe. Le volume de cette cellule est calculé ensuite avec la fonction `rbf.pde.geometry.volume` de la bibliothèque de T. Hines [Hines2015], qui prend en entrée le maillage grossier décrivant la cellule. Le terme source construit dans l'espace d'évaluation (Y dans le cas d'une méthode LS-RBF-FD, X dans le cas d'une méthode RBF-FD) est ainsi converti en $[M/T]$.

3.3.7 Perspectives

Nous avons présenté les diverses méthodes de générations de domaine qui ont été envisagées dans ce travail. D'autres méthodes de placement de nœuds sont possibles, comme présenté dans [Tominec2021a]. Comme nous l'avons présenté dans le chapitre d'état de l'art, les auteurs ont utilisé un domaine conforme et informé des conditions limites noté Y pour les nœuds d'évaluations des moindres carrés. Ils ont également utilisé un espace d'approximation RBF noté X ne contenant que des nœuds intérieurs qui débordent légèrement des bords du domaine, assurant un traitement plus souple des conditions aux limites évaluées dans Y .

L'utilisation de domaines non conformes aux obstacles aurait l'avantage de simplifier le domaine, notamment pour la génération d'une maille grossière à passer à un algorithme de placement de nœuds par minimisation d'énergie. En effet, le sol pourrait être connecté plus facilement aux bâtiments et le traitement des bâtiments adjacents pourrait être simplifié.

L'ajustement de la fonction de densité de nœuds est aussi un sujet à explorer. En effet, dans le cadre de la modélisation de la pollution atmosphérique, il faudrait que le nombre de points de calcul diminue fortement en hauteur (Figure 3.5b), pour économiser des ressources de calcul. De même, la paramétrisation d'une zone de nudge latérale avec un sous domaine beaucoup moins résolu que le domaine utile est un sujet d'intérêt que nous n'avons pas exploré. Par ailleurs des méthodes de discrétisation adaptative de domaine tel que l'*octree* semblent prometteuses et ont déjà été utilisées avec succès dans des méthodes sans maillage pour mieux modéliser des phénomènes physiques entre interfaces ou pour modéliser des couches limites, tout en disposant d'une densité de point faible au niveau des zones de domaine facilement modélisables [Dobracev2022].

3.4 Analyse de performance des méthodes RBF-FD et LS-RBF-FD

Dans ce paragraphe, la méthode RBF-FD en collocation sa variante de moindres carrés (LS-RBF-FD) a été appliquée au transport de polluants dans des cas bidimensionnels et tridimensionnels. Les méthodes ont été testées sur des cas simples 2D. La méthode RBF-FD a été appliquée sur des cas 2D avec obstacles et champs météorologiques uniformes. Les méthodes RBF-FD et LS-RBF-FD ont été ensuite testées sur des cas 3D simples, sans obstacles à météorologie uniforme et non uniforme. Enfin les LS-RBF-FD ont été comparés à PMSS puis nous avons testé une simulation avec obstacles. Pour prouver les capacités de la méthode LS-RBF-FD à gérer de grands domaines de calcul, nous avons testé une simulation 3D à l'échelle d'un quartier.

3.4.1 Une solution explicite de référence en 2D

Les modèles numériques de dispersion basés sur les méthodes RBF-FD et LS-RBF-FD ont été testés pour modéliser un transport sans obstacles en 2D. Cette partie a pour but de montrer que ces méthodes sont capables de modéliser précisément le phénomène de dispersion sur un cas 2D.

Pour prouver tout d'abord que les méthodes RBF-FD et LS-RBF-FD sont adaptées au traitement des équations de dispersion de polluants telle que l'équation d'Advection-Diffusion, nous allons comparer les résultats d'un transport scalaire de polluants à partir d'une méthode RBF-FD, ceux de la méthode des moindres carrés LS-RBF-FD, ceux d'une méthode classique d'éléments finis (FEM) et enfin ceux d'une formule analytique. Le domaine physique est considéré dans ce cas d'utilisation comme un domaine rectangulaire bidimensionnel de longueur L et de largeur H , écrit $\Omega = \{(x, y) : 0 \leq x \leq L, 0 \leq y \leq H\}$ pour le paragraphe suivant. L'équation d'ADPDE de (2.13) est considérée dans un régime stationnaire avec des conditions de Dirichlet,

$$\begin{aligned} c(0, y) &= c(L, y) = 0 \\ c(x, 0) &= c(x, H) = 0. \end{aligned} \quad (3.20)$$

Le terme source choisi est

$$s(x, y, t) = \frac{1}{4\epsilon^2} \mathbf{1}_{[x_s - \epsilon, x_s + \epsilon] \times [y_s - \epsilon, y_s + \epsilon]} \bar{S}, \quad (3.21)$$

avec la position de source, $\mathbf{x}_s = (x_s, y_s)$ et $\mathbf{1}$ une fonction indicatrice de sous domaine carré telle que,

$$\mathbf{1}_{[\alpha, \beta] \times [\alpha', \beta']} = \begin{cases} 1, & \alpha \leq x \leq \beta \text{ and } \alpha' \leq x \leq \beta' \\ 0, & \text{sinon.} \end{cases} \quad (3.22)$$

Calcul d'une solution explicite de l'ADPDE : Il est possible de calculer une formule analytique pour résoudre une équation d'advection-diffusion sur un domaine rectangulaire. Nous considérons un champ de vent et un champ de diffusion uniformes. Nous pouvons considérer dans le domaine un écoulement incompressible de sorte que $\text{div}(\mathbf{U}) = 0$. Dans ce cas, l'équation ADPDE se réécrit explicitement comme suit,

$$\begin{aligned} \frac{\partial c}{\partial t}(x, y, t) + U_x(x, y, t) \frac{\partial}{\partial x} c(x, y, t) + U_y(x, y, t) \frac{\partial}{\partial y} c(x, y, t) \\ - D_x(x, y) \frac{\partial^2}{\partial x^2} c(x, y, t) - D_y(x, y) \frac{\partial^2}{\partial y^2} c(x, y, t) = s \end{aligned} \quad (3.23)$$

Avec la source s et les conditions aux limites définies respectivement dans (3.21) et (3.20). D_x et D_y désignent les coefficients de diffusion horizontale. En outre, ces coefficients seront considérés comme égaux, de sorte que $D = D_x = D_y$, car la diffusion scalaire est généralement considérée comme un champ vertical et horizontal. Appliquons une transformation de Helmholtz pour convertir le problème d'advection-diffusion (3.23) en un problème de diffusion. Pour ce faire, nous écrivons une nouvelle variable,

$$v(x, y, t) = c(x, y, t) e^{-p_1 x - p_2 y} \quad (3.24)$$

avec,

$$p_1 = \frac{U_x}{2D}, \quad p_2 = \frac{U_y}{2D} \quad (3.25)$$

Le choix de l'expression des coefficients p_1 et p_2 permet de réécrire l'équation (3.23) en un problème diffusif. Par conséquent, en appliquant la transformation de Helmholtz, le problème se transforme en,

$$\partial_t v(x, y, t) + qv(x, y, t) - D \frac{\partial^2}{\partial x^2} v(x, y, t) - D \frac{\partial^2}{\partial y^2} v(x, y, t) = S' \quad (3.26)$$

avec,

$$S' = s e^{-\frac{U_x x + U_y y}{2D}} \quad (3.27)$$

$$q = \frac{1}{4} \left(\frac{U_x^2}{D} + \frac{U_y^2}{D} \right) \quad (3.28)$$

Supposons que nous cherchions une solution où la solution peut être exprimée séparément dans le temps et dans chaque composante de l'espace de sorte que, $v(x, y, t) = \phi_x(x) \phi_y(y) \psi(t)$. L'introduction dans (3.26) de cette expression conduit à

$$\frac{\frac{d}{dt} \psi(t)}{k \psi(t)} = + \frac{\frac{d^2}{dy^2} \phi_y(y)}{\phi_y(y)} + \frac{\frac{d^2}{dx^2} \phi_x(x)}{\phi_x(x)} - \frac{q}{k} \quad \forall (x, y, t) \in \Omega^2 \times T. \quad (3.29)$$

Par conséquent, les termes à gauche et à droite de (3.29) sont nécessairement égaux à une constante ici nommé κ et où

$$\frac{d}{dt}\psi(t) = -\kappa \quad (3.30)$$

$$+\frac{\frac{d^2}{dy^2}\phi_y(y)}{\phi_y(y)} + \frac{\frac{d^2}{dx^2}\phi_x(x)}{\phi_x(x)} - \frac{q}{D} = -\kappa \quad (3.31)$$

Ainsi, la partie temporelle Ψ de la solution v est de la forme,

$$\psi(t) = ce^{-\kappa Dt} \quad c, t \in \mathbb{R} \times T \quad (3.32)$$

Encore par principe de séparation des variables cette fois dans l'espace, nous obtenons de (3.31),

$$\frac{\frac{d^2}{dx^2}\phi_x(x)}{\phi_x(x)} = -\mu \quad (3.33)$$

$$\frac{\frac{d^2}{dy^2}\phi_y(y)}{\phi_y(y)} = -\nu \quad (3.34)$$

$$\mu + \nu = \kappa - \frac{q}{D}, \quad (3.35)$$

avec μ et λ des constantes. Ainsi les valeurs propres et les vecteurs propres peuvent être déduits du problème aux valeurs propres (3.33) et (3.34). Avec, $\mu_n = \left(\frac{n\pi}{L}\right)^2$, $\phi_x^n(x) = \sin\left(\frac{n\pi x}{L}\right)$. Similairement, $\nu_n = \left(\frac{n\pi}{H}\right)^2$, $\phi_y^n(y) = \sin\left(\frac{n\pi y}{H}\right)$. En considérant le produit scalaire $\langle \cdot, \cdot \rangle = \int \int_{\Omega} f(x, y)g(x, y)dx dy$, ces solutions sont orthogonales et vérifient,

$$\langle \phi_x^n, \phi_x^m \rangle = 0, \quad (n, m) \in \mathbb{N}^2 \quad (3.36)$$

$$\langle \phi_y^n, \phi_y^m \rangle = 0, \quad (n, m) \in \mathbb{N}^2 \quad (3.37)$$

Finalement une solution spatiale s'écrit :

$$\phi_{nm}(x, y) = \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi y}{H}\right) \quad (n, m) \in \mathbb{N}^2 \quad (3.38)$$

Déduisons maintenant une solution en espace et en temps. Nous allons chercher une solution v du problème (3.26) constituée de deux termes contenant une solution homogène de la série de Fourier et une solution particulière, telle que

$$v(x, y, t) = v_{hom}(x, y, t) + v_{part}(x, y, t) \quad (3.39)$$

avec,

$$v_{hom}(x, y, t) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_{nm} e^{-(D(\mu_n + \nu_m) + q)t} \phi_{n,m}(x, y) \quad (3.40)$$

$$v_{part} = \int_{s=0}^t v_s(x, y, t) ds \quad (3.41)$$

$$v_s = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} F_{nm}(s) e^{-(D(\mu_n + \nu_m) + q)(t-s)} \phi_{nm}(x, y) \quad (3.42)$$

$$c_{nm} = \frac{\langle v_0, \phi_{nm} \rangle}{\langle \phi_{nm}, \phi_{nm} \rangle} \quad F_{nm} = \frac{\langle S'(x, y, t=s), \phi_{nm} \rangle}{\langle \phi_{nm}, \phi_{nm} \rangle} \quad (3.43)$$

Enfin, une solution c_e peut être explicitement écrite en développant (3.40), et en introduisant l'expression de c utilisée dans (3.24) avec :

$$\begin{aligned}
 c_e(x, y, t) &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} c_{nm} e^{-(D(\mu_n + \nu_m) + q)t} \phi_{nm}(x, y) \\
 &+ \frac{4}{LH} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} e^{p_1 x + p_2 y} \phi_{nm}(x, y) \\
 &\times \int_0^t e^{(-q - D(\mu_n + \nu_m))(t - \tau)} \\
 &\times \left[\int_0^L \int_0^H \phi_{nm}(x', y') \times S'(x', y', \tau) dx' dy' \right] d\tau
 \end{aligned} \tag{3.44}$$

Solution stationnaire : De la même manière, on peut dériver une solution stationnaire d'une manière similaire à celle utilisée pour dériver une solution instationnaire. Par conséquent, une solution stationnaire d'une solution de transport ADPDE simple s'écrit,

$$c_e(x, y) = e^{(p_1 x + p_2 y)} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{D\lambda_{nm} + q} \frac{4}{LH} \langle s_{mod}, \phi_{nm} \rangle \phi_{nm}(x, y) \tag{3.45}$$

avec $\lambda_{nm} = \mu_n + \nu_m$, les coefficients déjà utilisés dans la solution transitoire. Calculons l'expression $\langle s_{mod}, \phi_{nm} \rangle$ avec un terme source carré définie comme,

$$S(x, y) = \frac{1}{4\epsilon^2} \mathbf{1}_{[x_s \pm \epsilon \text{ and } y_s \pm \epsilon]} \bar{S}. \tag{3.46}$$

avec $\bar{S} = 1$, le produit scalaire $\langle s_{mod}, \phi_{nm} \rangle$ s'écrit,

$$\langle s_{mod}, \phi_{nm} \rangle = \int_{x \in \Omega} \int_{y \in \Omega} s_{mod}(x, y) \phi_{nm}(x, y) dx dy \tag{3.47}$$

$$= \frac{1}{4\epsilon^2} I_x I_y \tag{3.48}$$

Les intégrales I_x et I_y sont alors :

$$I_x = \frac{(c_x \cos(c_x x_0) + p_1 \sin(c_x x_0)) e^{-p_1 x_0} - (c_x \cos(c_x x_f) + p_1 \sin(c_x x_f)) e^{-p_1 x_f}}{c_x^2 + p_1^2} \tag{3.49}$$

$$I_y = \frac{(c_y \cos(c_y y_0) + p_2 \sin(c_y y_0)) e^{-p_2 y_0} - (c_y \cos(c_y y_f) + p_2 \sin(c_y y_f)) e^{-p_2 y_f}}{c_y^2 + p_2^2}$$

avec $x_0, x_f, y_0, y_f = [x_s - \epsilon, x_s + \epsilon, y_s - \epsilon, y_s + \epsilon]$ et avec $c_x = n\pi/L$ et $c_y = m\pi/H$. Pour obtenir une bonne approximation de la solution analytique (3.45) qui utilise des sommes infinies nous choisissons un nombre de termes à $n_f = m_f = 200$ termes.

3.4.2 Résultats sur un cas simple en 2D

Nous exploitons la forme analytique de la solution en considérant un champ de vent $(U_x, U_y) = (1 \text{ ms}^{-1}, 4 \text{ ms}^{-1})$ et un champ de diffusion $(D_x, D_y) = (D, D)$ avec $D = 1000 \text{ ms}^2$. Nous générons la solution analytique avec 200 termes pour approximer chaque somme infinie de (3.45). La solution analytique est évaluée sur 3000, 6000, 10000 et 20000 nœuds (points de coordonnées) dispersés. Rappelons que la précision de la solution analytique ne dépend pas du nombre de nœuds où elle est évaluée. La solution analytique évaluée sur 20000 nœuds est visible en Figure 3.11. Nous la considérons comme une solution de référence dans cette sous partie.

Pour mettre en œuvre les RBF-FD et LS-RBF-FD, nous choisissons :

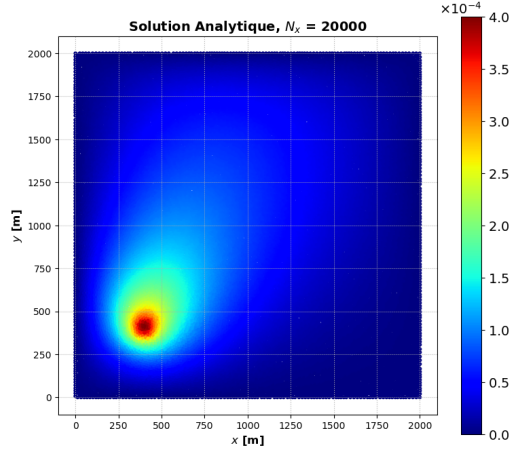


FIGURE 3.11 – Solution analytique évaluée sur 20000 nœuds et construite avec $n_f = m_f = 200$ termes. $(U_x, U_y) = (1\text{ms}^{-1}, 4\text{ms}^{-1})$, $(D_x, D_y) = (D, D)$ avec $D = 1000\text{ms}^2$.

- Des fonctions RBF polyharmoniques PHS5 choisies avec des polynômes de régularisation d'ordre 4. La fonction PHS5 (pour rappel $\phi(r) = r^5$, $r = \|y - x_i\|$) présentée en (2.33) est en effet dérivable deux fois et à dérivée continue, et a montré avec l'usage d'une telle régularisation polynomiale des résultats adéquats. De plus ce choix est en cohérence avec la littérature pour des simulations 2D [Flyer2016a].
- Nous sélectionnons une taille de stencil de 40. Nous avons observé en 2D que cette taille est raisonnable. Par ailleurs cet ordre de grandeur est cohérent avec la littérature pour des simulations en 2D [Flyer2016a].
- Pour les simulations réalisées par LS-RBF-FD, les points d'évaluation Y sont choisis de manière à être trois fois plus nombreux que les points d'approximation dans X . $N_y = 3N_x$. Dans ce travail, les points d'évaluation sont générés indépendamment des points d'approximations X .

Pour cette étude en 2D, nous évaluons les erreurs de deux manières : premièrement nous considérons l'erreur relative par point,

$$\text{erreur relative}(i) = \frac{|\mathbf{c}(i) - \mathbf{c}_e(i)|}{|\mathbf{c}_e(i)|}, \quad i \in \{1, \dots, N_x\}, \quad (3.50)$$

avec $\mathbf{c} \in \mathbb{R}^{N_x}$, la sortie de modèle RBF-FD et $\mathbf{c}_e \in \mathbb{R}^{N_x}$ la solution analytique évaluée sur N_x points discrets de solution. Pour des très basses valeurs de solution analytique, par exemple autour de 10^{-30} , l'erreur relative peut être élevée car les sorties numériques de modèles RBF-FD peuvent avoir des valeurs en ces points proches de zéro mais d'ordre de grandeur supérieur par exemple d'environ 10^{-20} . Nous considérons donc un seuillage dans l'erreur avec,

$$\text{erreur relative} = \kappa \quad \text{si } \mathbf{c}_e(i) \leq \zeta, \quad i \in \{1, \dots, N_x\}. \quad (3.51)$$

Avec ζ , un seuil pertinent que nous sélectionnerons pour visualiser au mieux les erreurs relatives observées et κ une constante. Dans cette étude $\zeta = 5 \times 10^{-30}$. κ est sélectionnée à 0 pour l'étude qui suit. Deuxièmement nous considérons la racine de l'erreur quadratique moyenne (RMSE) pour évaluer l'erreur cumulée sur tout le vecteur de solution de modèle numérique (\mathbf{c}) par rapport à la solution analytique (\mathbf{c}_e) dont l'expression est,

$$\text{RMSE} = \frac{\|\mathbf{c} - \mathbf{c}_e\|_2}{\sqrt{N_x}} \quad \text{RMSE} \in \mathbb{R}^+. \quad (3.52)$$

La méthode RBF-FD tout comme la méthode LS-RBF-FD ont été capables de calculer des solutions s'approchant de la solution analytique avec $N_x = 1000, 3000, 6000, 10000$ et 20000

points d'approximation. La Figure 3.12 montre des erreurs relatives déjà raisonnable en dehors d'une zone en amont du panache de pollution, ailleurs la solution RBF-FD n'excède pas 25 % d'erreur. Cependant la Figure 3.13 montre que la méthode LS-RBF-FD retourne des erreurs plus faibles à nombre de nœuds égal, les erreurs relatives se situant majoritairement autour de 10 %.

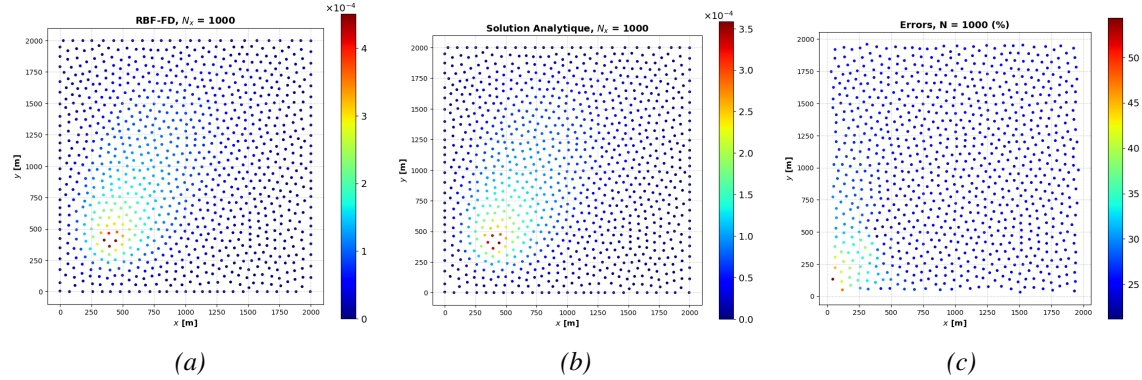


FIGURE 3.12 – (a) : Solution approximée sur 1000 nœuds par RBF-FD, (b) : solution analytique évalué sur 1000 nœuds, (c) : Erreur relative entre la simulation RBF-FD et la solution analytique.

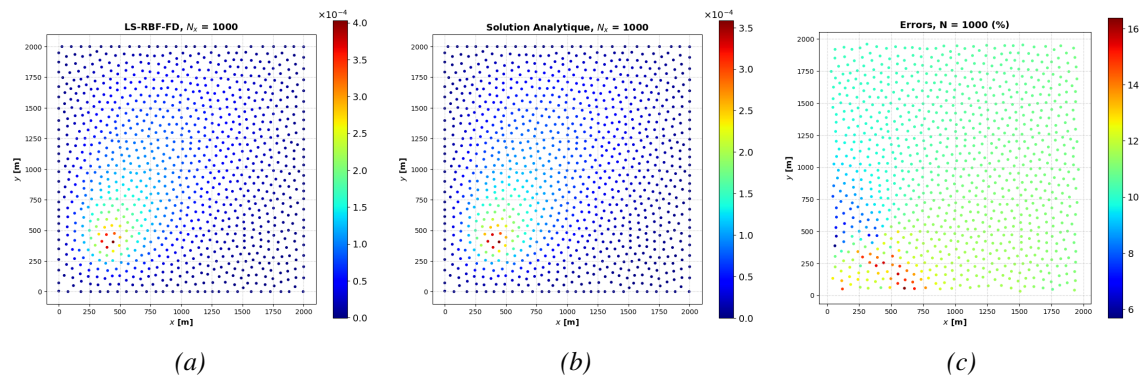


FIGURE 3.13 – (a) : Solution approximée sur 1000 nœuds par LS-RBF-FD, (b) : solution analytique évaluée sur 1000 nœuds, (c) : Erreur relative entre la simulation LS-RBF-FD et la solution analytique.

Nous allons voir maintenant qu'il se confirme que la méthode LS-RBF-FD est plus précise que la méthode RBF-FD lorsque l'on considère $N_x = 3000$, 6000 et 10000 nœuds de calcul.

La Figure 3.14, avec $N_x = 3000$, montre une nette amélioration de la solution RBF-FD, l'erreur n'excède pas 14 %, le pic d'erreur est localisé près du terme source. Ailleurs les erreurs sont proches de 10 %. La Figure 3.15 montre que la méthode LS-RBF-FD donne lieu à des erreurs plus faibles avec une erreur ne dépassant pas 5 % près du centre du panache. Le reste des erreurs est compris entre 2 et 3 %.

La Figure 3.16, avec $N_x = 6000$, montre encore une amélioration de la solution RBF-FD, l'erreur n'excède pas 11 % au pic d'erreur tandis que les autres erreurs relatives sont proches de 7 %. La Figure 3.17 montre que la méthode LS-RBF-FD conduit à des erreurs plus faibles ne dépassant pas 4 % au pic d'erreur. Le reste des erreurs est compris entre 0 et 2 %. Nous remarquons cependant l'apparition d'une erreur sur le bord le plus éloigné du panache (le bord localisé à $(x, y) = (2000, 2000)$). L'apparition d'une telle erreur pourrait se trouver dans la disposition du domaine de calcul : nos outils génèrent des nœuds d'évaluation Y et des nœuds d'évaluation X disposés uniquement sur l'intérieur du domaine de simulation et sur son bord, cela a pour effet de rendre les stencils de conditions aux limites asymétriques et de nuire à l'approximation de solution aux bords. Tominec et al. ont montré qu'un domaine d'approximation (X) non conforme

au domaine de simulation Ω avec des points de X dépassant des bords de domaine $\partial\Omega$ permet d'améliorer la précision de la méthode LS-RBF-FD [Tominec2021a].

La Figure 3.18, avec $N_x = 10000$, ne montre pas une amélioration perceptible de la solution RBF-FD, l'erreur par rapport à la solution analytique demeure encore autour de 9 %. La Figure 3.19 montre que la méthode LS-RBF-FD donne des erreurs très faibles, proches de 0 à 1 % sauf vers bord le plus éloigné du panache (le bord localisé à $(x, y) = (2000, 2000)$) où l'erreur relative atteint 15 %.

La Figure 3.20, avec $N_x = 20000$, montre une nette amélioration de la solution RBF-FD, l'erreur par rapport à la solution analytique est comprise entre 0 et 3 %, ce pic d'erreur étant atteint près de la source. Enfin la Figure 3.21 montre que la méthode LS-RBF-FD donne des erreurs très faibles, proches de 0 à 1 %, sauf vers bord le plus éloigné du panache (le bord localisé à $(x, y) = (2000, 2000)$) où l'erreur relative atteint 30 %. A ce stade, la précision de la solution issue de la méthode RBF-FD est équivalente à la précision de la solution obtenue par la méthode LS-RBF-FD.

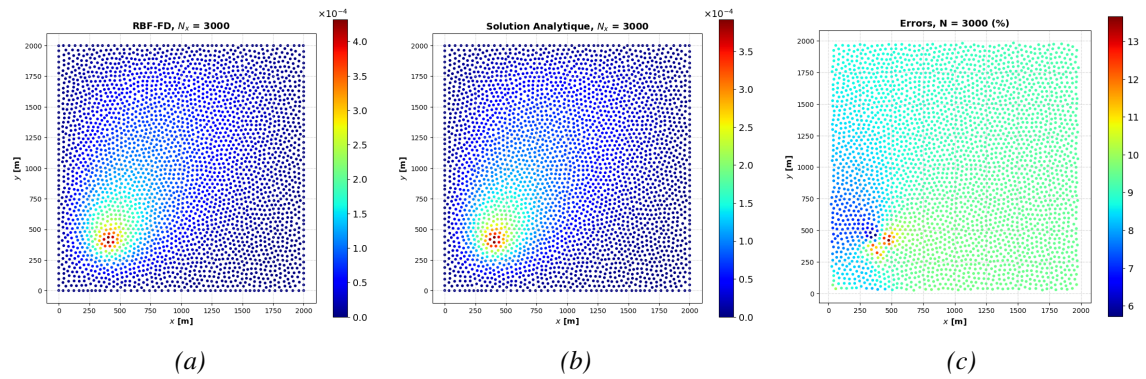


FIGURE 3.14 – (a) : Solution approximée sur 3000 nœuds par RBF-FD, (b) : solution analytique évalué sur 3000 nœuds, (c) : Erreur relative entre la simulation RBF-FD et la solution analytique.

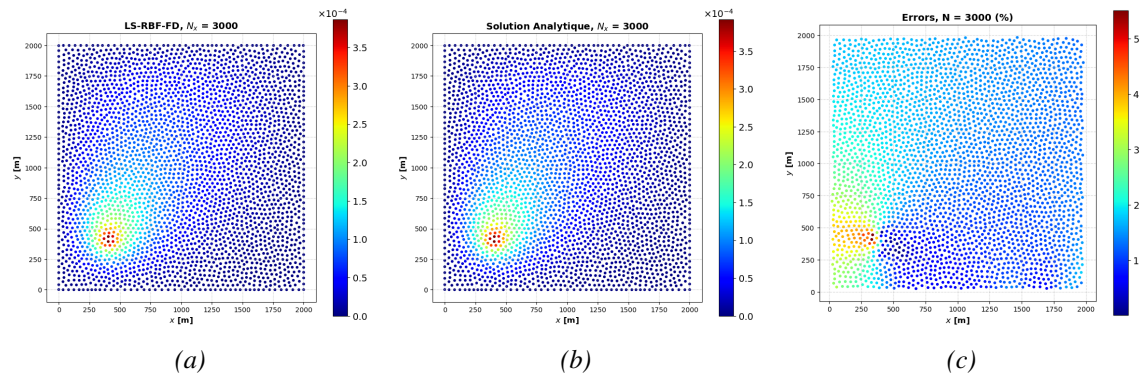


FIGURE 3.15 – (a) : Solution approximée sur 3000 nœuds par LS-RBF-FD, (b) : solution analytique évaluée sur 3000 nœuds, (c) : Erreur relative entre la simulation LS-RBF-FD et la solution analytique.

Nous pouvons conclure que les solutions RBF-FD et LS-RBF-FD convergent rapidement vers la solution analytique et l'erreur globale converge rapidement au-delà de 3000 nœuds vers une RMSE de l'ordre de 5×10^{-7} . Lors de l'utilisation d'un petit nombre de nœuds, nous avons observé que la plus grande contribution à l'erreur a lieu à proximité de la source. Cela peut probablement s'expliquer par la discrétisation de la source, car la fonction de support de la source (3.21) est continue pour $x, y \in [x_s - \epsilon, x_s + \epsilon] \times [y_s - \epsilon, y_s + \epsilon]$. On note que la solution LS-RBF-FD montre une plus grande précision que la solution RBF-FD, à nombre de points d'approximation (point de

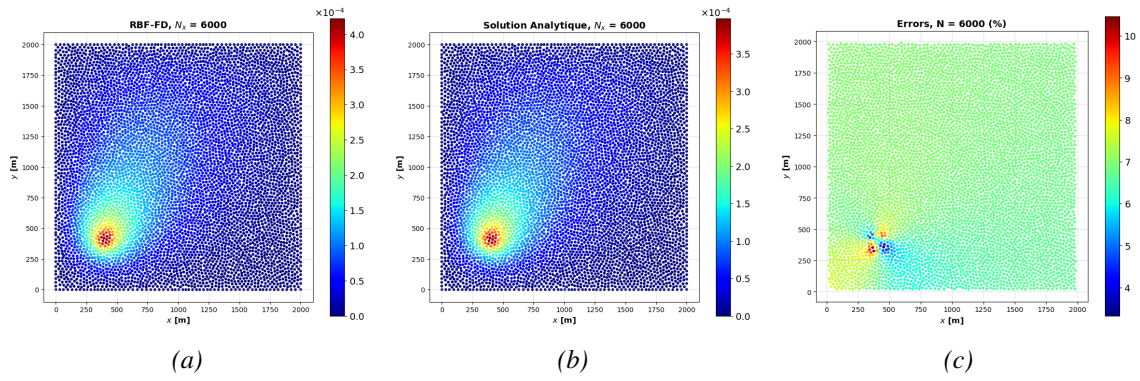


FIGURE 3.16 – (a) : Solution approximée sur 6000 nœuds par RBF-FD, (b) : solution analytique évalué sur 6000 nœuds, (c) : Erreur relative entre la simulation RBF-FD et la solution analytique.

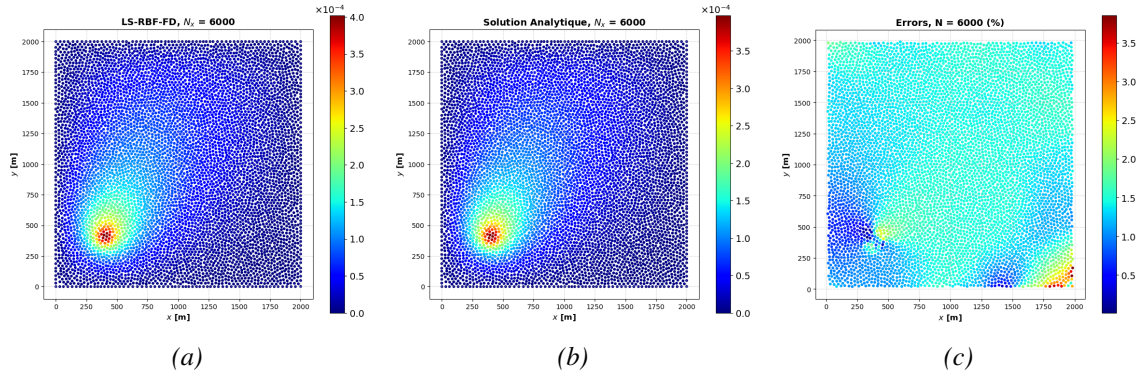


FIGURE 3.17 – (a) : Solution approximée sur 6000 nœuds par LS-RBF-FD, (b) : solution analytique évaluée sur 6000 nœuds, (c) : Erreur relative entre la simulation LS-RBF-FD et la solution analytique.

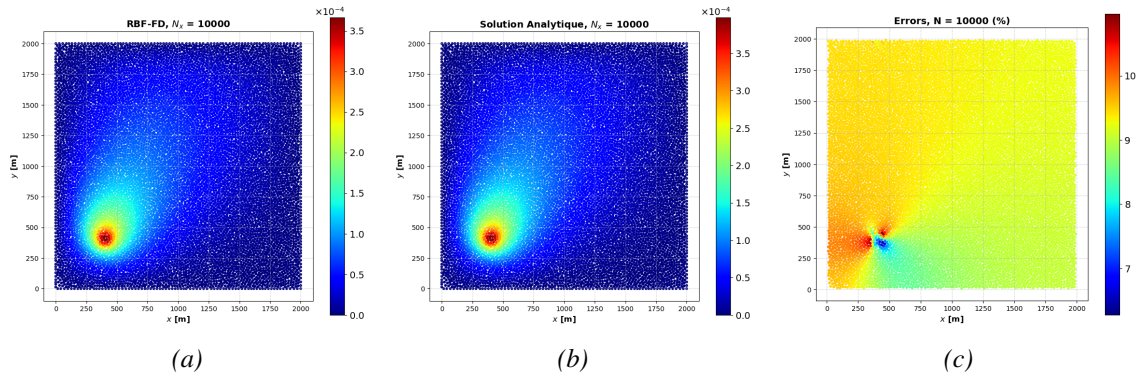


FIGURE 3.18 – (a) : Solution approximée sur 10000 nœuds par RBF-FD, (b) : solution analytique évalué sur 10000 nœuds, (c) : Erreur relative entre la simulation RBF-FD et la solution analytique.

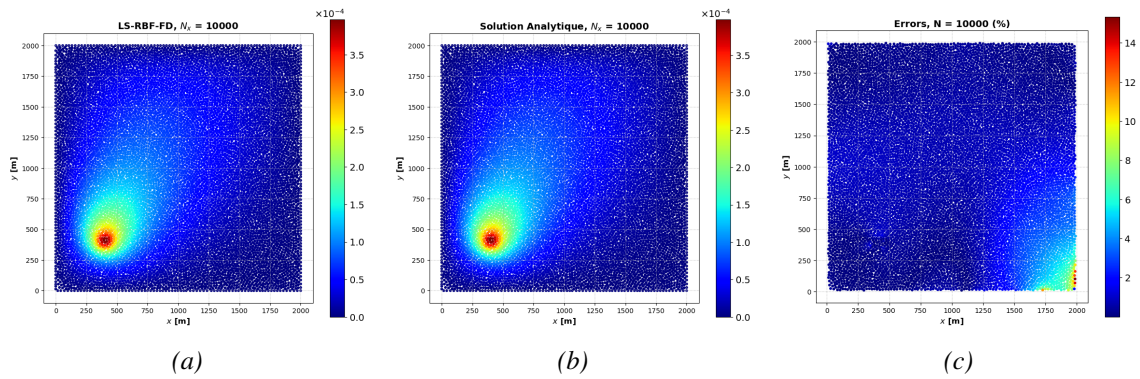


FIGURE 3.19 – (a) : Solution approximée sur 10000 nœuds par LS-RBF-FD, (b) : solution analytique évaluée sur 10000 nœuds, (c) : Erreur relative entre la simulation LS-RBF-FD et la solution analytique.

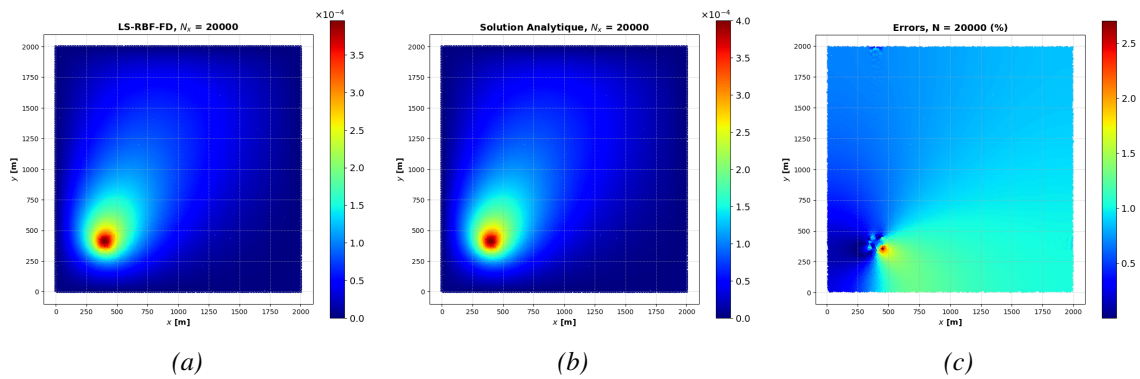


FIGURE 3.20 – (a) : Solution approximée sur 20000 nœuds par RBF-FD, (b) : solution analytique évalué sur 20000 nœuds, (c) : Erreur relative entre la simulation RBF-FD et la solution analytique.

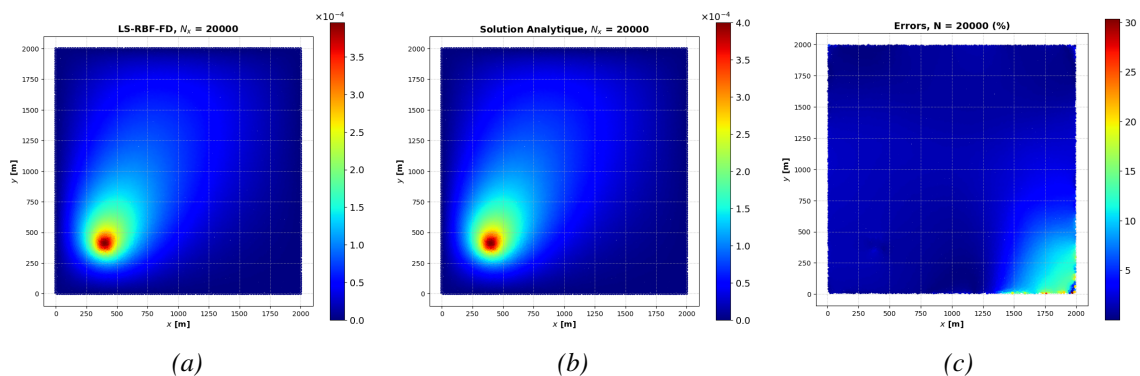


FIGURE 3.21 – (a) : Solution approximée sur 20000 nœuds par LS-RBF-FD, (b) : solution analytique évaluée sur 20000 nœuds, (c) : Erreur relative entre la simulation LS-RBF-FD et la solution analytique.

l'ensemble X) égal, jusqu'à 10000 nœuds de calcul, comme visible en Figure 3.22 qui affiche la RMSE des solutions RBF-FD et LS-RBF-FD en fonctions de nombre de points d'approximation. A 3000 nœuds de calcul, la racine de l'erreur quadratique moyenne (RMSE) est de $6,8 \times 10^{-6}$ pour la solution LS-RBF-FD contre $3,1 \times 10^{-5}$ pour la solution RBF-FD. Cet écart de RMSE de presque un ordre de grandeur est maintenu jusqu'à 10000 nœuds. Ensuite la solution retournée par le modèle RBF-FD obtient une RMSE de $4,1 \times 10^{-7}$ contre $6,2 \times 10^{-7}$ pour celle calculée par LS-RBF-FD. Le modèle LS-RBF-FD pourrait atteindre une précision supérieure en considérant un domaine de points d'approximation non conforme au domaine de simulation comme discuté dans [Tominec2021a].

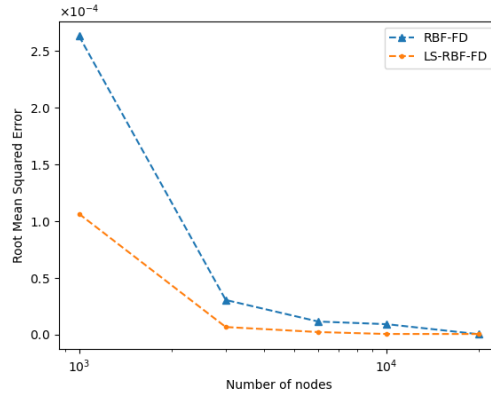


FIGURE 3.22 – Evolution de la racine de l'erreur quadratique moyenne (RMSE) des solutions RBF-FD et LS-RBF-FD en fonction du nombre de nœuds d'approximation (nombres de centres RBF). Les solutions numériques sont calculées avec 1000, 3000, 6000, 10000 et 20000 nœuds.

Nous retenons donc que les méthodes RBF-FD et LS-RBF-FD sont capables de modéliser des phénomènes de transport modélisés par l'équation d'advection-diffusion sur des domaines 2D. Les LS-RBF-FD obtiennent une précision supérieure aux RBF-FD à nombre de nœuds de calcul égal jusqu'à un certain seuil (ici 20000 nœuds de calcul).

3.4.3 Simulations urbaines RBF-FD 2D à champs météorologiques uniformes

Dans cette partie, nous avons testé la méthode RBF-FD en 2D sur des obstacles. Nous avons donné en entrée des champs météorologiques uniformes. Le but est de vérifier le comportement du modèle de transport RBF-FD en 2D dans ses capacités à calculer une dispersion et à traiter des obstacles sur un problème bidimensionnels.

Pour réaliser un transport de polluants, une source "gaussienne" a été choisie afin d'étaler le terme source pour ne pas créer trop de discontinuités dans le champ de solution. Ladite source Gaussienne a pour expression :

$$s(x, y) = a_s \exp\left(-\frac{(x - x_s)^2}{2\sigma_x^2}\right) \exp\left(-\frac{(y - y_s)^2}{2\sigma_y^2}\right) \quad (x, y), (x_s, y_s) \in \Omega \times \Omega \quad (3.53)$$

avec $\mathbf{x} = (x, y) \in \Omega$ et $\mathbf{x}_s = (x_s, y_s) \in \Omega$ les coordonnées planaires de l'intérieur du domaines et la position de source respectivement. Dans cette partie, les champs météorologiques de vent et de diffusion sont considérés comme uniformes pour chacune de leur composante : les composantes de vent sont fixées à $U_x = 10$ et $U_y = 2,8 \text{ ms}^{-1}$ et de diffusion à $D_x = D_y = 20 \text{ m}^2\text{s}^{-1}$. Ces ordres de grandeurs étant cohérent avec des conditions météorologiques urbaine à vent fort observées avec le logiciel PMSS. Comme le montre la Figure 3.23 les RBF-FD traitent de manière satisfaisante des cas 2D avec obstacles. Le modèle RBF-FD-2D a été également testé sur un domaine urbain à l'échelle d'un quartier. Cependant, sur le cas d'exemple de de dispersion en milieu urbain

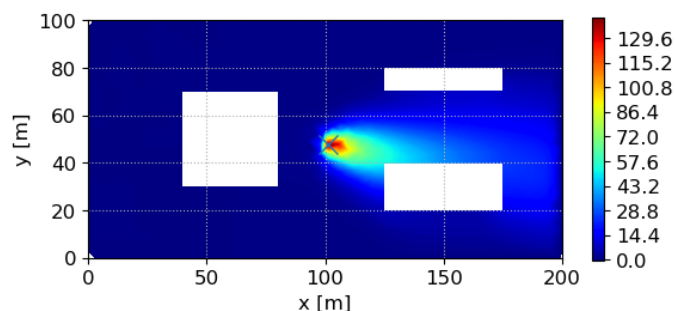


FIGURE 3.23 – Dispersion RBF-FD 2D à faible diffusion et à nuage de calcul grossier. RBF-FD 2D 1200 points, $(U_x, U_y) = (10, -2) \text{ ms}^{-1}$, $(K_x, K_y) = (20., 20) \text{ m}^2\text{s}^{-1}$. Fonctions de base radiales (ϕ) PHS5, stencils de 50 voisins, polynômes de régularisation d'ordre 2.

complexe (Figure 3.24), quand bien même le transport semble plausible, des erreurs de modèle sont rendues, inévitables par la présence de cours intérieures qui créent des indéterminations. Ces indéterminations ont d'ailleurs été également observées dans le modèle d'apprentissage bidimensionnel planaire développé dans [Mendil2022]. Les approches 2D pourraient donner des ordres de grandeurs en concentration de polluants satisfaisants dans l'hypothèse de rejet faible de polluants, d'une température faible de rejet, et d'une hauteur de rejet assez basse, où la composante verticale de vent est faible. Ainsi de tels panaches évolueraient peu verticalement et justifieraient une approche 2D. Cependant, une telle approche 2D appliquée à du bâti urbain réaliste nécessiterait de fermer les cours intérieures en les considérant comme du bâti afin de supprimer des artefacts dus au cours intérieures ou encore à des bâtiments adjacents dont des nœuds de calcul seraient comptés à l'interface. Cette fermeture de bâtiment, ou d'ensemble de bâtiments formant des espace fermés n'est cependant pas triviale et est une des raisons qui complique le développement de tels modèles. Par ailleurs, les RBF-FD en 2D semblent être stables face à des phénomènes très advectifs et seraient donc adaptés d'un point de vue opérationnel, une fois ces indéterminations traitées. Finalement les RBF-FD en collocation sont une méthode prometteuse pour générer rapidement des transports scalaires en 2D en milieu urbain.

Comme nous l'avons mentionné, d'une part, des indéterminations de modèle de dispersion 2D sont créées par les cours intérieures de bâtiments, d'autre part le transport vertical de polluants n'est pas négligeable en aire urbaine [Gonzalez Olivardia2019]. C'est pourquoi nous nous sommes davantage intéressés à la modélisation 3D de la dispersion.

3.4.4 Résultats sur un cas simple en 3D, Résultats sur le quartier de Bonne à Grenoble en 3D

Cas de la méthode RBF-FD

Il a été possible de réaliser des simulations de transport urbains RBF-FD sur des domaines complexes 2D. Comme le montre la Fig 3.24 où une simulation complexe retournée par un modèle RBF-FD avec de nombreux obstacles a été présentée. Dans la pratique Mendil et al. ont observé des zones d'indéterminations causées par les cours intérieures [Mendil2022], de plus, le déplacement vertical des panaches urbains n'est pas vraiment négligeable, notamment à cause du transport causé par la composante verticale de la diffusion qui en pratique dans les cas de test de cette étude est proche de la moitié des composantes horizontales du champ de diffusion.

Les RBF-FD en 3D ont été testées avec des champs météorologiques uniformes et non uniformes. La météorologie uniforme est choisie telle que $(K_x, K_y, K_z) = (20, 20, 10) \text{ m}^2\text{s}^{-1}$ avec $(U_x, U_y, U_z) = (10, -2, 0) \text{ ms}^{-1}$. Ces ordres de grandeurs correspondent aux moyennes des données retournées par PMSS pour un cas jouet sans obstacles. Sur un domaine sans obstacles les RBF-FD échouent à calculer une solution 3D avec des conditions de Neumann au sol et de Dirichlets

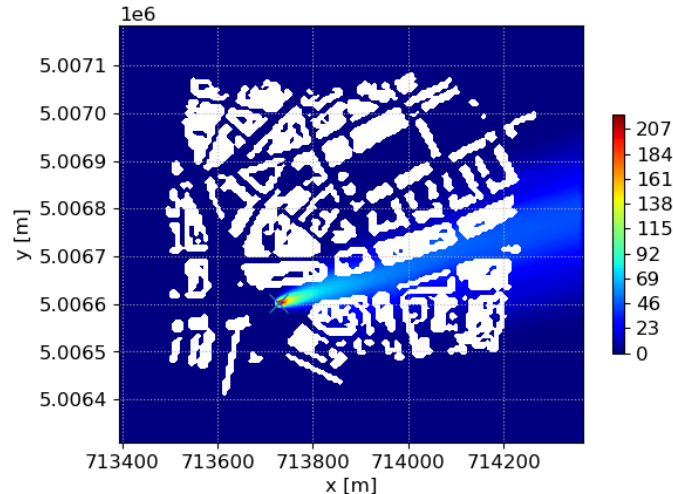


FIGURE 3.24 – Simulation RBF-FD 2D d’une dispersion 2D avec $N_x = 17542$ points d’approximation à faible diffusion sur un quartier de près d’un kilomètre de côté (Grenoble, quartier de Bonne). Champs météorologiques uniformes par composantes avec $(U_x, U_y) = (10, -2) \text{ ms}^{-1}$, $(K_x, K_y) = (20., 20) \text{ m}^2\text{s}^{-1}$. Fonctions de base radiales (ϕ) PHS5, stencil de 60 voisins, polynômes de régularisation d’ordre 2.

partout ailleurs. L’introduction d’un plafond de simulation portant des conditions de Neumann permet en revanche de calculer une solution cohérente avec un champ uniforme (Figure 3.25).

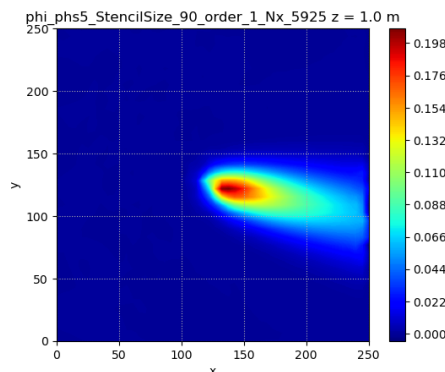


FIGURE 3.25 – Simulation tri-dimensionnelle RBF-FD d’un panache de pollution sans obstacles $(K_x, K_y, K_z) = (20, 20, 10) \text{ m}^2\text{s}^{-1}$ avec $(U_x, U_y, U_z) = (10, -2, 0) \text{ ms}^{-1}$, 5925 points de simulation. PHS5, stencil 90, ordre 1, Neumann sol et plafond avec nœuds fantômes. Domaine de calcul \mathcal{D}_∞ .

Les RBF-FD en 3D atteignent une limite en présence d’une météorologie hétérogène fournie par un logiciel de référence tel que PMSS [Moussafir2004]. L’explication à cet échec pourrait être à trouver en ce que l’advection peut localement dominer largement la diffusion, notamment à basse altitude ce qui perturbe la stabilité numérique de la simulation. Des auteurs ont par ailleurs relevé que la présence de conditions aux limites de Neumann sur le domaine de simulation (comme sur le domaine discuté ici) est aussi une source d’erreur pour les méthodes RBF-FD [Tominec2021b]. Malgré l’application de bonnes pratiques, telles qu’une utilisation des nœuds fantômes semblables à [Jančić2021] et l’utilisation d’un terme d’hyperviscosité d’ordre 2 pour la régularisation (voir [Fornberg2015c] page 151 pour plus de détails sur la notion d’hyperviscosité) utilisé dans tous les exemples en RBF-FD en 3D, nous ne sommes pas parvenus à modéliser de manière satisfaisante un transport scalaire de polluants par RBF-FD en 3D. Une simulation RBF-FD appliquée à un cas sans

obstacles avec la météorologie fournie par PMSS illustrée en Figure 3.26. Le panache sans obstacle calculé avec une météorologie hétérogène en Figure 3.26 montre un panache qualitativement bien moins défini que dans l'exemple en Figure 3.25 où une météorologie uniforme a été utilisée, ce qui traduit une difficulté rencontrée par les RBF-FD en collocation. Quelques artefacts sont notamment présents autour du panache dont les contours ne sont pas nets. De plus, beaucoup de nœuds de calcul sont nécessaires pour que le résultat fasse sens, 5978 pour ce cas jouet. Par ailleurs le résultat des simulations RBF-FD devient très sensible à la disposition des points de calcul et à de nombreux hyperparamètres de simulations tel que l'amplitude du coefficient de régularisation d'hyperviscosité.

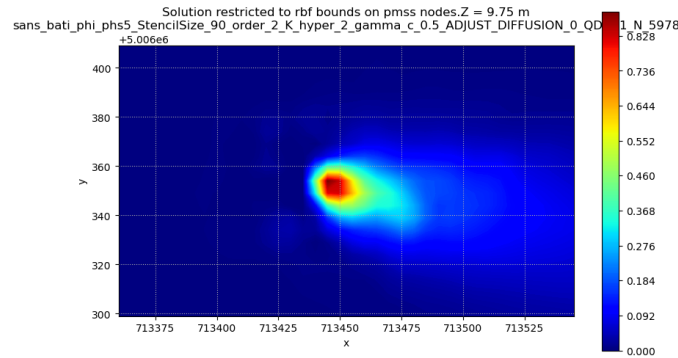


FIGURE 3.26 – Simulation tridimensionnelle RBF-FD d'un panache de pollution sans obstacles, météorologie PMSS, $N_x = 5978$ points de simulation. Base RBF PHS5, taille de stencil 90, ordre 2 de monome de régularisation, conditions de Neumann au sol et au plafond avec des nœuds fantômes. Hyperviscosité d'ordre 2. Constante d'hyper-viscosité $\gamma_c = 0,5$, coefficient d'hyperviscosité $\gamma = \gamma_c N_x^{-2}$. Source ponctuelle de $1000 \mu\text{g s}^{-1}$ placée à 10 mètres de hauteur. Domaine de calcul \mathcal{D}_ϵ .

Pour un cas plus complexe avec une météorologie hétérogène et avec un bâti urbain (cas jouet avec bâtiments), la simulation RBF-FD échoue comme visible en Figure 3.27a. En effet le panache de pollution ne s'étale pas dans la direction du vent à la manière de la simulation de référence réalisé avec PMSS (Figure 3.27b).

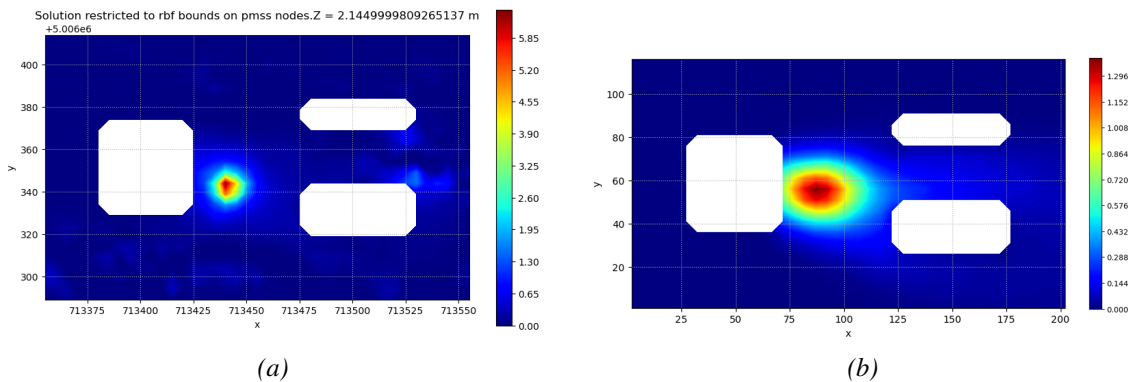


FIGURE 3.27 – (a) Simulation LS-RBF-FD, source ponctuelle à $1000 \mu\text{g s}^{-1}$ avec conversion des unités de terme source, régime stationnaire, $N_x = 2008$, $N_y = 6000$ (b) Résultats PMSS, source ponctuelle à $1000 \mu\text{g s}^{-1}$, régime stationnaire. Sources placées à environ 2 m de haut.

Nous concluons donc que la méthode RBF-FD ne fonctionne pas pour des cas 3D d'advection-diffusion à diffusion faible en régime stationnaire. Nous avons testé sur plusieurs cas d'études : jouet avec bâtiments et jouets sans bâtiments. Les RBF-FD en collocation n'ont pas été concluantes pour résoudre des problèmes d'ADPDE 3D dans ces deux cas avec des données météorologiques hétérogènes réalistes. Pour essayer d'améliorer la stabilité numérique de la solution RBF-FD

une couche de nœuds de calcul a été ajoutée près de la source ou encore la valeur des champs météorologiques aux nœuds fantômes a été configurée pour prendre la valeur entrée pour les nœuds voisins de domaine intérieur les plus proches. Les RBF-FD semblent cependant montrer des franches limites pour la modélisation de transport scalaires en 3D. Pour palier la tendance à l'erreur des RBF-FD dans des simulations aux conditions de flux nul complexes induites par les bâtiments, nous avons donc choisi d'utiliser la méthode LS-RBF-FD.

Cas de la méthode LS-RBF-FD

Vérification des ordres de grandeurs de solution LS-RBF-FD à météorologie hétérogène : Afin de vérifier que les ordres de grandeur de concentration retournés par le modèle de dispersion LS-RBF-FD sont corrects, une simulation sans obstacles a été appliquée avec PMSS (Figure 3.28a) et avec le modèle de transport LS-RBF-FD (Figure 3.28b), les données météorologiques de champ de vélocité du vent et de diffusion ayant été calculées par SWIFT, logiciel météorologique de la suite PMSS. Pour une source située à une hauteur proche de deux mètres. Nous retrouvons des ordres de grandeurs similaires malgré un étalement moindre du centre de panache (zones rouges en Figure 3.28a et Figure 3.28b) pour les LS-RBF-FD comparé à PMSS.

D'autre part, il a été observé que la hauteur de la source a un grand impact sur les concentrations. Si elle est située par exemple à quatre mètres de haut au lieu de deux, les concentrations calculées sont supérieures (autour de $1.1 \mu\text{gm}^{-3}$ au niveau du maximum). Or la "distribution" de nœuds de calcul du domaine LS-RBF-FD est relativement grossière avec environ 5 mètres de résolution spatiale dans toute les directions entre 0 et 10 mètres de hauteur pour le cas jouet avec $N_x = 2008$, $N_y = 6000$ présenté en Figure 3.28a. Cela compromet le placement du terme source. Par exemple si on veut placer une source ponctuelle de coordonnées (x_s, y_s, z_s) , il se peut que la source soit réellement placée à $x_{\text{réel}} = (x_s - 4, y_s + 5, z_s + 6)$. Avec $(-4, 5, 6)$ les écarts des coordonnées réelles d'une source à placer. Un axe d'amélioration possible pour garantir une distribution dense de nœuds à deux mètres de haut sera de forcer une partie des nœuds générés par l'algorithme de génération de domaine RBF (Algorithme 4) à rester à deux mètres de haut. Pour pouvoir disposer de nombreux nœuds candidats à la bonne hauteur de rejet. Un autre axe d'amélioration pour mieux contrôler les coordonnées de source notamment ponctuelles sera de forcer les coordonnées d'émission à rester fixes pendant la génération de domaine par "minimisation d'énergie" (Algorithme 3).

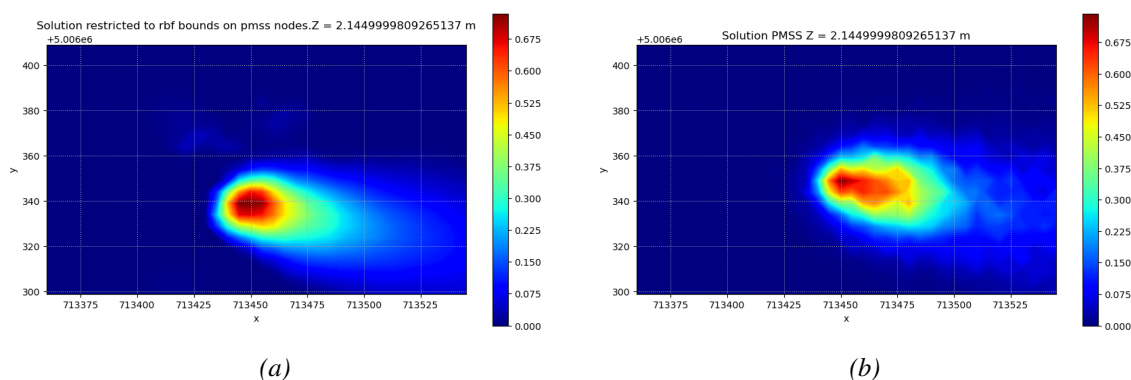


FIGURE 3.28 – (a) Simulation LS-RBF-FD, source ponctuelle à $1000 \mu\text{gs}^{-1}$ avec conversion des unités de terme source, régime stationnaire, $N_x = 2008$, $N_y = 6000$ (b) Résultats PMSS, source ponctuelle à $1000 \mu\text{gs}^{-1}$, régime stationnaire. Sources placées à environ 2 m de haut. Domaine de calcul \mathcal{D}_∞ .

Simulations urbaines par LS-RBF-FD : Contrairement aux RBF-FD en collocation, la méthode LS-RBF-FD est capable de modéliser une dispersion atmosphérique de polluants urbain en régime

stationnaire à travers des bâtiments, avec un nombre de points de calcul restreints, sans utilisation de nœuds fantômes comme visible en Figure 3.29a. Grâce aux matrices de différentiation creuses que permet de construire la méthode LS-RBF-FD, la méthode est capable de s'appliquer sur des grands domaines comme un quartier urbain, comme le montre la Figure 3.29b. Cependant, la rapidité de calcul est compromise avec environ 10 minutes de calcul pour cet exemple. La parallélisation du modèle en tuile, ainsi qu'un meilleur solveur autre que LSQR, par exemple un solveur de quasi Newton de type L-BFGS-B renseigné de la jacobienne du modèle pourrait grandement accélérer le calcul d'une solution. De plus, il serait raisonnable d'utiliser plus de nœuds d'approximation. Par exemple $N_x = 20000$ et de mieux les distribuer spatialement sur le plan $x - y$ ainsi qu'en hauteur. Cependant nous pouvons voir que les LS-RBF-FD sont capables de simuler des transports sur des petits domaines de calcul à météorologie hétérogène ainsi que sur des grands. Notons aussi que la méthode est capable de donner des solutions plausibles sur des grands domaines (Figure 3.29b) y compris avec un très faible nombre de points d'approximation (ici $N_x = 11027$).

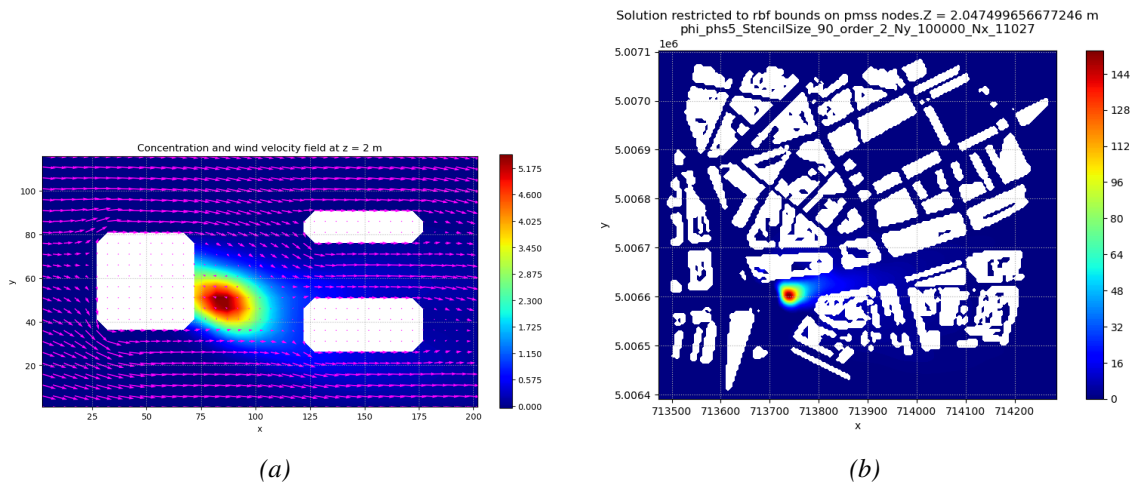


FIGURE 3.29 – Simulations urbaines par LS-RBF-FD. (a) Simulation LS-RBF-FD, source étalée d'amplitude $a_s = 1000 \mu\text{gs}^{-1}\text{m}^{-3}$ et de dispersion $\sigma_s = (8, 8, 8)$ source placée à environ 2 m de haut. $N_y = 50000$, $N_x = 15633$. (b) Simulation LS-RBF-FD de dispersion atmosphérique de polluants sur le quartier de Bonne à Grenoble, source placée à environ 3 m de haut à $1000 \mu\text{gs}^{-1}$, régime stationnaire. $N_y = 100000$, $N_x = 11027$. Les deux simulations utilisent des RBF (ϕ) PHS5, une taille de stencil $n = 90$ et des polynômes de régularisation d'ordre 2.

A notre connaissance, il s'agit de la première mise ne œuvre d'une approche LS-RBF-FD pour une simulation de dispersion de polluants en 3D sur un domaine hétérogène complexe.

3.5 Une méthode LS-RBF-FD dans le cas instationnaire

3.5.1 Un schéma d'Euler implicite

Pour intégrer en temps les solutions issues de la méthode LS-RBF-FD, nous proposons un schéma d'Euler implicite. Le schéma d'Euler implicite est inconditionnellement stable, ce qui veut dire que nous pouvons choisir des pas de temps élevés pour effectuer des simulations en régime transitoire. Ce schéma de calcul est simple, numériquement stable, économise des ressources en calcul par rapport à des schémas d'ordre plus élevés. **De plus ce schéma est innovant car à notre connaissance aucun travail n'a exploité une méthode de ce genre pour réaliser des simulations**

LS-RBF-FD en régime instationnaire.

Nous avons présenté la méthode LS-RBF-FD dans un cadre stationnaire précédemment en 2.4 pour rappel, il s'agissait de résoudre un problème compact des moindres carrés de la forme,

$$\mathbf{c}_a = \arg \min_{\mathbf{v}} \|\mathbf{D}\mathbf{v} - \mathbf{f}\|^2 \quad (3.54)$$

où \mathbf{c}_a est une solution de l'ADEDP stationnaire, \mathbf{D} l'opérateur discrétisé par RBF-FD et représentatif de la physique dans l'intérieur du domaine Ω et sur le bord $\partial\Omega$ et \mathbf{f} étant le terme source aggloméré aux conditions aux limites. Pour considérer un problème en régime transitoire, introduisons une expression semi-discrète exprimant la variation temporelle du vecteur de concentrations \mathbf{c} , le problème de minimisation devient :

$$\min_{\mathbf{c}} \|\partial_t \mathbf{c} + \mathbf{D}\mathbf{c} - \mathbf{f}(t)\|^2. \quad (3.55)$$

Pour avancer en temps dans un schéma de calcul discret on peut considérer un temps suivant $t + 1$, avec dt le pas de temps discret, pour exprimer une dérivée temporelle discrète par un schéma implicite; le problème instationnaire se met alors sous la forme :

$$\min_{\mathbf{c}(t+1)} \left\| \mathbf{D}_0 \frac{\mathbf{c}(t+1) - \mathbf{c}(t)}{dt} + \mathbf{D}\mathbf{c}(t+1) - \mathbf{f}(t) \right\|^2 \quad (3.56)$$

avec, \mathbf{D}_0 désignant l'opérateur RBF-FD à l'ordre de dérivée 0 (interpolateur) de Y_Ω dans X quand il agit sur des nœuds $y_k \in Y_\Omega$ et 0 sinon, de sorte que :

$$\mathbf{D}_0(y_k) = \begin{cases} \mathbf{w}^0(y_k, X) & \text{si } y_k \in Y_\Omega \\ \mathbf{0} & \text{sinon} \end{cases} \quad (3.57)$$

où $\mathbf{w}^0(y_k, X)$ est un jeu de poids RBF-FD pour réaliser une interpolation locale de \mathbf{c} pour un stencil centré en y_k à n voisins dans X . Ainsi on résout,

$$\min_{\mathbf{c}(t+1)} \left\| \frac{\mathbf{D}_0}{dt} [\mathbf{c}(t+1) - \mathbf{c}(t)] + \mathbf{D}\mathbf{c}(t+1) - \mathbf{f} \right\|^2, \quad (3.58)$$

avec,

$$\begin{cases} \mathbf{D}_k = \mathbf{L}_k, & y_k \in Y_\Omega \\ \mathbf{D}_k = \mathbf{B}_k, & y_k \in Y_{\partial\Omega}. \end{cases} \quad (3.59)$$

où \mathbf{L}_k et \mathbf{B}_k sont respectivement les vecteurs de poids permettant la discrétisation RBF-FD des opérateurs spatiaux continus \mathcal{L} , \mathcal{B} respectivement de domaine intérieur (Ω) et de domaine de bords $\partial\Omega$. \mathbf{L}_k , \mathbf{B}_k étant construits par RBF-FD pour chaque centre de stencil $Y = \{y_k\}_{k=1}^{N_y}$. De (3.59) nous pouvons réécrire le problème de régression de la manière suivante,

$$\min_{\mathbf{c}(t+1)} \left\| \left(\frac{\mathbf{D}_0}{dt} + \mathbf{D} \right) \mathbf{c}(t+1) - \left(\frac{\mathbf{D}_0}{dt} \mathbf{c}(t) + \mathbf{f} \right) \right\|^2. \quad (3.60)$$

Cette expression-là est solvable avec l'algorithme LSQR ou encore LBFGS-B, gradients conjugués ou autre solveur de moindres carrés. Pour le moment, le solveur LSQR a donné les meilleurs résultats.

Dans cette étude, l'ADPDE est considérée comme sans terme de réaction. Le schéma d'Euler implicite admet donc une solution analytique des moindres carrés que nous posons à partir d'une formulation implicite séparant les conditions aux limites sur $\partial\Omega$ et sur l'intérieur du domaine Ω . En séparant les opérateurs de conditions aux limites des opérateurs aux nœuds intérieurs au domaine le problème de moindres carrés devient,

$$\min_{\mathbf{c}(t+1)} \left(\left\| \mathbf{D}_{0\Omega} \frac{\mathbf{c}(t+1) - \mathbf{c}(t)}{dt} + \mathbf{D}_\Omega \mathbf{c}(t+1) - \mathbf{s}(t) \right\| + \|\mathbf{D}_{\partial\Omega} \mathbf{u}(t+1)\|^2 \right) \quad (3.61)$$

Avec $\mathbf{D}_{0\Omega}$ une matrice d'interpolation RBF-FD de Y_Ω vers l'ensemble d'approximation X . \mathbf{D}_Ω l'opérateur RBF-FD représentant le modèle d'Advection-Diffusion à l'intérieur du domaine d'évaluation Y^Ω et $\mathbf{D}_{\partial\Omega}$ représente l'opérateur des conditions aux limites, ici pour des conditions de Neumann et de Dirichlet nulles. A partir de (3.61) il est possible de trouver la solution explicite suivante :

$$\mathbf{c}(t+1) = \left(\left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right] + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega} \right)^{-1} \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} \mathbf{c}(t) + \mathbf{s}(t) \right]. \quad (3.62)$$

Il est alors possible d'écrire l'évolution de la solution sous une forme explicite,

$$\mathbf{c}(t+1) = \mathbf{A}\mathbf{c}(t) + \mathbf{B}\mathbf{s}(t), \quad (3.63)$$

avec,

$$\mathbf{A} = \left(\left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right] + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega} \right)^{-1} \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \frac{\mathbf{D}_{0\Omega}}{dt} \quad (3.64)$$

$$\mathbf{B} = \left(\left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right] + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega} \right)^{-1} \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_\Omega \right]^T \quad (3.65)$$

Remarque : Dans ce paragraphe nous avons considéré des matrices de différentiation constantes du temps dans nos schémas d'Euler (3.58) et (3.62). Il est possible, sans modifier les schémas présentés de considérer une météorologie variable du temps et donc des matrices de différentiation variables du temps qu'il s'agira d'actualiser à chaque mise à jour des champs météorologiques.

3.5.2 Résultats de simulation

Il a été possible d'utiliser notre schéma d'Euler implicite instationnaire présenté en 3.5.1 pour faire avancer la simulation LS-RBF-FD en temps. Les simulations résultantes montrent bien une évolution temporelle des solutions ainsi qu'un rapprochement vers des solutions stationnaires (Figure 3.30). Nous avons cependant observé qu'à nombre de nœuds de calcul égal, le code instationnaire converge vers des ordres de grandeurs légèrement supérieurs au code stationnaire. Il serait intéressant de voir si cette différence pourrait être expliquée par l'utilisation d'un schéma d'Euler implicite en considérant à la place un schéma de type Runge-Kutta implicite d'ordre 2 ou plus élevé.

Nous avons aussi mis en oeuvre une simulation multi-sources sans obstacles en régime transitoire notamment pour tester des approches basées sur le filtre de Kalman visible en Figure 4.23. 3 sources de position et de débit fixes sont étalées de manière gaussiennes et sont à l'origine du phénomène de pollution. Des conditions de Neumann sont choisies sur toutes les faces du pavé de simulation à l'exception de la face en aval du panache orthogonale à l'axe $x - z$. Tandis que la météorologie est considérée comme uniforme avec respectivement, les valeurs de champ de diffusion et de champ de vélocité $(K_x, K_y, K_z) = (20, 20, 10) \text{ m}^2\text{s}^{-1}$ avec $(U_x, U_y, U_z) = (10, -2, 0) \text{ ms}^{-1}$.

Perspectives pour la modélisation transitoire : D'autres solutions sont envisageables pour le calcul transitoire. En effet Tominec et Nazarov se sont inspirés des méthodes de Galerkin pour projeter les matrices rectangulaires RBF-FD afin de résoudre *in fine* un système carré avec une méthode des lignes [Tominec2022]. Cela permet alors l'utilisation de solveurs rapides tels que GMRES pour des EDP linéaires.

Enfin il est possible d'envisager d'agglomérer le temps aux opérateurs spatiaux en traitant le temps comme "une dimension comme une autre". L'avantage serait alors de se débarrasser de la relation de récurrence présente dans les schémas d'intégration temporelle classiques afin de paralléliser et d'accélérer les calculs. L'Intégration Temporelle Directe est une méthode référencée dans [Dokainish1989] par exemple.

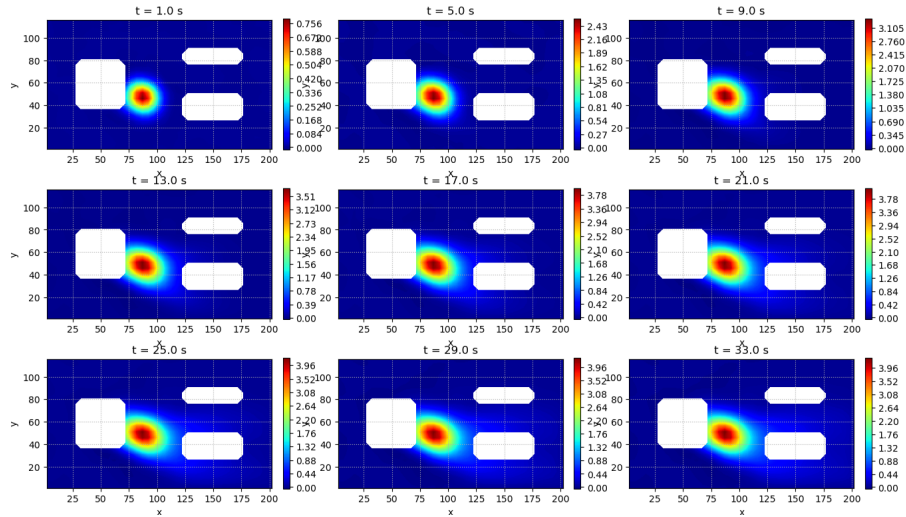


FIGURE 3.30 – Simulation instationnaire avec schéma implicite d’Euler. Cas jouet à 3 obstacles. $N_x = 3529$, $N_y = 12000$ source gaussienne placée à environ 2 mètres, de position $x_s = (713440; 5006340; 1, 8)$, d’amplitude $a_s = 1 \mu\text{g s}^{-1} \text{m}^{-3}$ et de dispersion $\sigma_s = (8, 8, 8) \text{m}$.

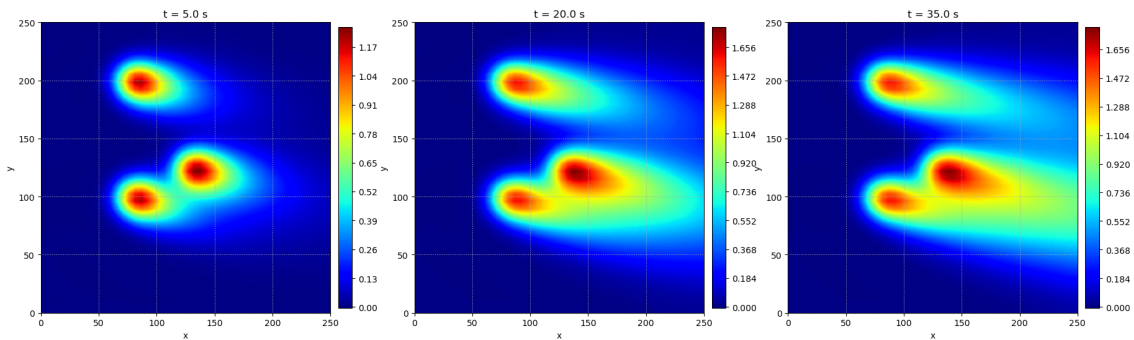


FIGURE 3.31 – Solution stationnaire montrée ici avec à trois différents temps. La solution est approximée en 2883 points $N_x = 2883$, $N_y = 10000$. Pas de temps de calcul $dt = 5$ secondes.

3.5.3 Conclusion

Le présent chapitre a été consacré à la modélisation de la dispersion de polluants (c’est-à-dire la résolution du transport scalaire de polluants par l’ADPDE) par les méthodes LS-RBF-FD et RBF-FD avec des météorologies uniformes et des météorologies précalculées par un logiciel tiers : ici le logiciel P-SWIFT-M qui utilise les RANS à fermeture turbulente d’ordre 0. Nous avons réalisé l’implémentation du modèle numérique basé sur les RBF-FD (et LS-RBF-FD) que nous avons développé dans cette thèse. Nous avons analysé les performances des méthodes RBF-FD et LS-RBF-FD sur différents cas.

- D’abord nous avons considéré un cas simple en 2D sur lequel nous avons comparé les sorties des modèles RBF-FD et LS-RBF-FD à une solution analytique que nous avons recherchée et programmée. Nous avons observé que ces méthodes sont capables de modéliser efficacement le phénomène de dispersion avec de faibles erreurs par rapport à la solution analytique.
- Nous avons ensuite réalisé des simulations sur des domaines 3D simples, sans obstacles, avec des météorologies uniformes ainsi que non-uniformes par RBF-FD et LS-RBF-FD. Le logiciel PMSS a fait l’objet d’études de validation en souffleries et face à des simulations de référence [Tinarelli1994, Oldrini2013]. Nous avons comparé la méthode LS-RBF-FD au logiciel PMSS sur un cas sans obstacle. Nous avons utilisé la procédure de conversion d’unité

de sources décrite en 3.4.4 et avons obtenu les ordres de grandeurs attendus de concentration en polluant ainsi qu'une forme de panache attendu.

- **Nous avons réussi dans un troisième temps à simuler à grande échelle, sur un authentique quartier urbain, une simulation de dispersion de polluants modélisée par LS-RBF-FD. Nous avons au préalable développé des outils de génération du domaine.**
- **Nous avons enfin proposé un schéma d'Euler implicite transitoire pour effectuer des simulations instationnaires basées sur la méthode LS-RBF-FD. Le caractère implicite du schéma assurant la stabilité numérique de la méthode.** Des résultats de simulation utilisant ce schéma ont été discutés.

Les résultats obtenus montrent qu'il est possible d'utiliser une approche LS-RBF-FD pour mettre en œuvre des simulations 3D (y compris en instationnaire) dans des domaines hétérogènes complexes.

Nous pouvons en particulier retenir pour les méthodes de simulation RBF-FD et LS-RBF-FD plusieurs enseignements.

Enseignements sur la méthode RBF-FD : La méthode RBF-FD en collocation a été appliquée au transport de polluants dans des cas bidimensionnels et tridimensionnels. Quelques observations sont à retenir :

1. La méthode RBF-FD en 2D converge vers notre formule analytique pour un cas d'étude sans obstacles avec conditions de Dirichlet et donne des résultats comparables avec celle des éléments finis.
2. La méthode RBF-FD arrive à calculer de manière stable des transports scalaires en 2D dans l'espace y compris pour des phénomènes très convectifs.
3. La méthode RBF-FD en collocation est en échec en 3D pour des valeurs de vent élevées comparées à la diffusion (cas convectif) en particulier pour des météorologies hétérogènes. Les RBF-FD ont également un comportement imprédictible avec une météorologie uniforme : ceci a été révélé sur un cas sans obstacles pour un cas jouet sur un domaine rectangulaire de $250\text{m} \times 250\text{m} \times 50\text{m}$ de haut avec $N = 25000$ nœuds de calcul pour une résolution nodale $h = 0.2\text{m}$, et un champ de vent et de diffusion uniformes par composantes avec $\mathbf{U} = [10, -2, 0] \text{ m s}^{-1}$ et $\mathbf{K} = [20, 20, 10] \text{ m}^2 \text{ s}^{-1}$ où un changement de conditions aux limites sur le plafond de simulation avait suffi à perturber la stabilité des solutions. Pour des cas avec obstacles aux conditions aux limites plus complexes et avec une météorologie réaliste, la méthode RBF-FD en 3D n'arrive pas à calculer correctement un panache de pollution, malgré l'application de bonnes pratiques (utilisation de régularisation, nœuds fantômes, densité de nœuds de calcul élevée).
4. La mise en œuvre des nœuds fantômes pour un calcul en régime stationnaire n'est pas adéquate car elle consiste à imposer une condition physique en dehors du domaine de simulation ce qui manque de sens quand des champs hétérogènes de météorologie sont à donner en entrée. Dans un schéma instationnaire, les nœuds fantômes peuvent être imposés de manière déterministe comme montré par [Barnett2015b] en fonction des valeurs de la variable d'intérêt au temps précédent.

A terme, une meilleure distribution des nœuds avec une densité accrue dans des zones d'intérêt et une faible densité en hauteur, pourrait améliorer le comportement du modèle de dispersion qui utilise cette méthode.

5. Une parallélisation en tuiles permettrait de paralléliser l'utilisation de solveur par la résolution de sous-problèmes plus faciles. Le solveur utilisé ici, LSQR, montre par exemple une plus grande lenteur avec le cas urbain complet.

Il pourra être utile, dans la perspective de travaux futurs, de tester d'autres solveurs de type gradients conjugués comme le solveur L-BFGS-B, en fournissant la jacobienne du modèle pour accélérer la descente ainsi qu'en choisissant un préconditionnement approprié.

Enseignements sur la méthode LS-RBF-FD : La modélisation RBF-FD en moindres carrés (LS-RBF-FD) a été appliquée au transport de polluants dans des cas tridimensionnels suite aux difficultés rencontrées dans une approche RBF-FD à collocation. Quelques observations sont à retenir :

1. Notre travail de modélisation de la pollution par LS-RBF-FD est à notre connaissance une des premières approches de modélisation 3D avec [Tominec2021a] et la première pour les problèmes de pollution de l'air.
2. Les LS-RBF-FD arrivent à calculer de manière stable des transports scalaires en 3D dans l'espace y compris pour des phénomènes très convectifs, hétérogènes et contraints par des conditions aux limites complexes, en régime stationnaire et transitoire grâce à notre propre schéma implicite d'Euler.
3. Le calcul de la dispersion de polluants modélisé par l'ADPDE fonctionne sans utiliser de paramètres complexes tels que des nœuds fantômes ou des termes de régularisation tel que l'hyperviscosité. Des auteurs ont notamment montré que l'usage de nœuds fantômes est dispensable [Tominec2021b].
4. Le calcul fonctionne sur des nuages de points d'approximation très grossiers. Cet avantage a aussi été observé par d'autres auteurs [Tominec2021c].

Suite à nos recherches, nous avons décidé d'utiliser un modèle de dispersion atmosphérique de polluants avec la méthode LS-RBF-FD introduite dans [Tominec2021b] pour la suite de ce travail.

L'utilisation de cette méthode a permis de pallier les défauts rencontrés par la méthode des RBF-FD qui est une méthode de collocation. De plus les LS-RBF-FD restent robustes avec des densités de points de calcul faible. **Il a en effet été possible de calculer une solution d'un transport scalaire dans un quartier urbain d'environ 1 km de côté avec seulement 11027 points d'approximations.**

Le modèle eulérien développé ici n'a pas pour objectif d'égaliser ou de dépasser les performances de calcul de modèles précis comme le modèle lagrangien de dispersion SPRAY de PMSS ou encore le modèle de dispersion de *Code_Saturne*. Cependant la robustesse apparente du modèle LS-RBF-FD face à des domaines de simulation très épars et son temps de calcul raisonnable, malgré un code Python restant certainement à optimiser, sont prometteurs dans la perspective de travaux d'assimilation de données.

Le chapitre 4 va maintenant être consacré à l'Estimation de Terme Source (STE). Nous avons d'abord développé une méthode de STE basée sur les réseaux des neurones profonds dans un cadre 1D. Dans un second temps, nous présenterons plusieurs méthodes de STE basées sur le modèle LS-RBF-FD retenu dans ce travail.

4

Estimation de sources

Ce chapitre discute des méthodes et des résultats d'estimation de sources de pollution. En début de thèse alors que des modèles de transport de polluants 1D étaient à l'étude, nous nous sommes intéressés aux possibilités d'estimation de source par "machine learning" et avons également déployé une méthode variationnelle classique d'inversion, la 3D-Var, ce travail a conduit à une publication [Lopez-Ferber2022]. Nous avons ensuite développé des méthodes d'estimations basées sur un modèle de dispersion utilisant la méthode LS-RBF-FD. Nous avons étudié, dans un contexte de dispersion en régime stationnaire, une méthode adjointe discrète pour reconstruire un terme source sur tout l'espace, une méthode sans modèle adjoint inspirée des réseaux de neurones informés de la physique et enfin un Filtre de Kalman testé en régime transitoire.

Sommaire

4.1	Introduction	98
4.2	Une étude comparative entre approches variationnelle et par apprentissage - cas 1D stationnaire	98
4.2.1	Un modèle 1D de transport-diffusion de polluants	99
4.2.2	Méthodes	101
4.2.3	Étude de cas : estimation de source dans une rue	103
4.2.4	Conclusions	107
4.3	Approche adjointe avec modèle LS-RBF-FD	108
4.3.1	Méthode de STE en utilisant un adjoint discret	110
4.3.2	Etude de cas STE dans sur une zone urbaine simulée	111
4.3.3	Conclusions	116
4.4	Approche de type PIML avec modèle LS-RBF-FD	117
4.4.1	Formalisme de la méthode STE-PIML, cas stationnaire	117
4.4.2	Résultats	118
4.4.3	Conclusions	126
4.5	Filtre de Kalman avec modèle transitoire LS-RBF-FD	127

4.5.1	Introduction	127
4.5.2	Mise en œuvre du filtre de Kalman	128
4.5.3	Résultats	129
4.5.4	Conclusions	134
4.6	Conclusion	134

4.1 Introduction

Plusieurs méthodes d'estimation de sources (STE) ont été développées et mises en œuvre dans cette thèse. Nous nous sommes premièrement intéressés à une STE dans un contexte de dispersion atmosphérique unidimensionnelle de polluants en régime stationnaire. Ce cadre de travail a été posé dans l'intérêt d'explorer des méthodes d'Estimation de Terme Source (STE). Ce premier effort d'exploration de méthodes de STE sur des cas unidimensionnels a fait l'objet d'une publication [Lopez-Ferber2022]. Ce travail a été l'occasion de comparer des performances d'une méthode variationnelle classique la 3D-Var, avec une méthode d'apprentissage. Cette approche trouve son utilité dans un contexte de détection rapide de source très polluante au sein d'une rue. Les détails de cette étude sont discutés en 4.2.

Dans un deuxième temps nous nous sommes intéressés aux problèmes physiques de dimensions supérieures (2D, 3D) pour prendre en compte les contraintes imposées par le bâti urbain. Nous avons notamment développé notre propre modèle de dispersion atmosphérique de polluants pour résoudre l'ADPDE par des méthodes RBF-FD comme discuté en 3. Nous avons ainsi développé une méthode STE basée sur le modèle physique LS-RBF-FD proposé en 3.4.4. Cette méthode de STE utilise un modèle adjoint dans le but d'identifier un terme source 3D quelconque défini sur tout le domaine. Cette approche est choisie grâce à l'hypothèse de disponibilité abondante de mesures (contexte du crowdsensing) qui autorise la recherche d'un vecteur source avec autant d'inconnues que de points d'approximation dans le domaine de simulation (domaine X discuté en 2.4.6). Cette méthode est discutée en 4.3. Nous la nommerons STE-adjointe.

Toujours dans un contexte stationnaire, nous avons mis en place une méthode nouvelle de STE comme une application de l'approche LS-RBF-FD montrée en (3.2) à la STE et inspirée par le PIML, l'apprentissage automatique informé par la physique, notamment comme décrit dans les travaux de [Raissi2019]. Cette méthode est discutée en 4.4. Nous la nommerons STE-PIML.

Enfin des expériences de STE dans un contexte instationnaire ont été réalisées pour estimer des sources quelconques par filtres de Kalman (KF). Cette méthode est discuté en 4.5. Nous la nommerons KF.

Pour les méthodes de STE faisant intervenir notre modèle tridimensionnel LS-RBF-FD (4.3, 4.4 et 4.5), nous utilisons l'erreur relative par point (3.50)-(3.51) et l'erreur absolue par point pour apprécier la précision des estimations. Nous fixons la valeur d'erreur relative intervenant en (3.51) lorsque cette dernière n'est pas calculée à $\kappa = \text{NaN}$. Où NaN est une constante informatique qui précise que κ n'est pas un nombre. Ainsi nous visualisons l'erreur relative uniquement là où elle est définie pour chaque estimation.

4.2 Une étude comparative entre approches variationnelle et par apprentissage - cas 1D stationnaire

Dans cette partie nous avons développé et comparé deux méthodes d'Estimation de Terme Source. Une STE est réalisée par méthode variationnelle 3D-Var, l'autre par un réseau de neurones inspirée de [Georges2020] (sans modèle physique explicite) construit comme un estimateur de source, directement à partir des données d'entrées de mesures et de météorologie.

À notre connaissance, l'apprentissage automatique n'a pas été largement utilisé jusqu'à présent pour la détection des sources. Récemment, une approche d'apprentissage automatique a été utilisée pour résoudre un problème inverse lié à l'estimation du trafic routier [Georges2020]. Cette approche a inspiré le présent travail. Un réseau de neurones de type perceptron multicouche est utilisé avec quelques couches cachées. Cette méthode basée sur les données évite toute utilisation du calcul

des variations comme dans les méthodes variationnelles, et évite ainsi de faire appel à un modèle de dispersion dans un cadre d'optimisation. De plus, les réseaux neuronaux ne sont pas coûteux en termes de calcul une fois l'entraînement effectué, comparés aux méthodes variationnelles, en particulier celles basées sur des modèles numériques complexes. L'entraînement du réseau de neurones est en effet effectué hors ligne, c'est-à-dire avant la phase d'exploitation.

Cette étude propose une méthode pour reconstruire une source de pollution dans un axe unidimensionnel et dans un contexte stationnaire, afin de prendre en compte en premier lieu la dispersion des polluants dans des rues isolées. L'approche d'apprentissage automatique est comparée à une approche variationnelle classique basée sur un modèle (3D-Var), qui est généralement considérée comme une référence en termes de précision d'estimation.

L'intérêt de cette étude est d'évaluer dans quelle mesure des approches d'apprentissage purement basées sur les données et n'utilisant pas de modèle physique de dispersion de polluants se comparent avec une approche d'estimation purement déterministe, la 3D-Var. L'approche d'apprentissage proposée est économe en calcul une fois l'apprentissage effectué, cependant, l'approche, étant uniquement basé sur les données, nécessite la création et l'enrichissement d'une base de données conséquente. Les méthodes classiques telles que la 3D-Var requièrent de leur côté quelques efforts importants : 1. la programmation du modèle de dispersion de polluants 2. le calcul du gradient de la fonction de coût qui n'est pas trivial comme discuté en partie 2.5.2 où nous avons traité des difficultés de l'obtention de la fonction de coût et de l'intérêt du développement de modèles adjoints. Par ailleurs, la méthode 3D-Var est une méthode d'optimisation dont la convergence de sa fonctionnelle vers un minimum global n'est pas garantie lorsque le problème est non convexe (c'est le cas ici), mais surtout elle requiert la mise en œuvre d'une méthode itérative coûteuse en temps de calcul car nécessitant de nombreuses itérations et simulations des modèles direct et adjoint, comme nous l'avons vu au chapitre 2.

Cette étude a constitué des travaux préliminaires au développement de la modélisation de la pollution atmosphérique par les méthodes RBF-FD. Le but étant alors de mettre œuvre des méthodes de STE appliquées au domaine de la pollution atmosphérique et d'explorer le potentiel des méthodes d'apprentissage supervisé. Le cas d'étude est unidimensionnel, en régime stationnaire et avec une météorologie uniforme ramenée à des coefficients scalaires d'advection et de diffusion.

L'étude est organisée de la manière suivante : La partie 4.2.1 présente le modèle de transport des polluants et son approximation en dimension finie. Elle justifie également les choix de la vitesse du vent et du coefficient de diffusion utilisés dans le modèle. La partie 4.2.2 détaille la méthode d'apprentissage automatique proposée et présente brièvement l'approche variationnelle. Les résultats des simulations numériques sont présentés dans la partie 4.2.3, tandis que des remarques conclusives sont discutées dans la partie 4.2.4.

4.2.1 Un modèle 1D de transport-diffusion de polluants

Dans ce cas d'étude, nous utilisons un modèle physique stationnaire pour représenter la concentration des polluants. Dans un contexte unidimensionnel, où une source stationnaire a émis suffisamment longtemps pour atteindre un état stationnaire, la dynamique de la concentration des polluants est modélisée par une équation différentielle d'advection-diffusion, notée \mathcal{M} :

$$u d_x c(x) - k d_{xx}^2 c(x) = s(x), \quad (4.1)$$

avec les paramètres scalaires u [m s^{-1}] et k [$\text{m}^2 \text{s}^{-1}$] correspondant respectivement au champ de vent et à la diffusion totale. c représente les concentrations de polluants en g m^{-3} ou en mol m^{-3} selon le débit de source considéré. d_x et d_{xx}^2 désignent respectivement les premières et deuxièmes dérivées spatiales. s représente la source en [$\text{g/m}^{-3} \text{s}^{-1}$] (ou [$\text{mol m}^{-3} \text{s}^{-1}$] en fonction des unités de concentration de polluants) et est définie sur le domaine physique. Le profil de la source s est supposé être modélisé par une distribution gaussienne définie comme suit :

$$s(x) = a_s \Delta_s(x - x_s), \quad (4.2)$$

$$\text{avec } \Delta_s(x - x_s) = \exp\left(-\frac{(x - x_s)^2}{2\sigma_s^2}\right),$$

où a_s , x_s et σ_s [$\text{g s}^{-1} \text{m}^{-3}$, m , m^{-2}] représentent respectivement l'amplitude, la position et la dispersion d'une unique source de pollution. Δ_s représente la fonction de dispersion choisie pour la source autour de son centre. Notons $\theta = [a_s, x_s, \sigma_s]$ le vecteur des paramètres de la source pour les sections suivantes. L'utilisation d'une approche en une dimension est pertinente, même dans un contexte urbain, si l'on considère le réseau des principales artères urbaines suffisamment longues et étroites et si l'on néglige les phénomènes de dispersion verticale. Dans ce paragraphe, nous nous concentrons sur le problème de détection d'une source de pollution sur l'une de ces artères.

Approximation en dimension finie du modèle

Le modèle de dimension infinie \mathcal{M} est approximé en dimension finie en utilisant un schéma de différences finies basé sur une approximation décentrée en amont (schéma upwind) de l'opérateur d'advection et une approximation standard de première ordre de l'opérateur du Laplacien sur une grille régulière du domaine $\Omega = [0, 1, \text{km}]$ composée de N points. Des conditions de Dirichlet imposent les concentrations aux limites du domaine. Le schéma de discrétisation par différences finies conduit à :

$$\begin{aligned} \mathbf{c} &= \mathcal{M}_a(\theta) \\ &= \mathbf{A}^{-1}(\mathbf{B}_1 \mathbf{s}(\theta) + \mathbf{B}_2 \mathbf{C}), \end{aligned} \quad (4.3)$$

où \mathbf{c} et \mathbf{s} représentent respectivement le vecteur des concentrations de polluants et le vecteur de terme source à chaque nœud de la grille du domaine, \mathbf{A} et \mathbf{C} sont les matrices résultant de la discrétisation spatiale et de l'introduction des deux conditions aux limites en tant qu'entrées, respectivement. \mathbf{B}_1 , \mathbf{B}_2 sont les matrices qui relient respectivement la source discrétisée \mathbf{s} et les conditions aux limites \mathbf{C} à \mathbf{c} . θ désigne les paramètres de la source tels que son débit d'émission, sa position et sa dispersion. Ce modèle est utilisé pour créer un ensemble de données synthétiques pour entraîner le réseau neuronal décrit dans la section 4.2.2 et pour effectuer une estimation basée sur le modèle décrite dans la section 4.2.2.

Choix des paramètres et du bruit de mesure

Comme indiqué dans le praragraphe introductif en début de 4.2, cette étude vise à reconstruire une source de pollution dans une zone urbaine où la vitesse du champ de vent ne dépasse généralement pas $\sim 10 \text{m s}^{-1}$, comme observé dans [Wang2020] où les auteurs ont étudié une technique d'évaluation du champ de vent. Le coefficient de diffusion est fixé à $k = 25000 \text{m}^2 \text{s}^{-1}$ dans nos simulations, comme trouvé dans [Winiarek2012] pour les composantes horizontales de k . Cependant, ce coefficient de diffusion reste élevé, une étude devrait être réalisée pour trouver un coefficient de diffusion unidimensionnel représentatif en ordre de grandeur de la turbulence au sein d'une rue canyon. Les ordres de grandeur en 3D étant typiquement de l'ordre de 5 à 30 $\text{m}^2 \text{s}^{-1}$ par composante. De plus à l'échelle urbaine, le choix des ordres de grandeur de la turbulence est souvent imposé par le modèle de turbulence choisi [Kikumoto2020].

Pour tenir compte du fait qu'un capteur i , $1 \leq i \leq N_s$ (N_s étant le nombre de capteurs) mesure une concentration moyenne dans son voisinage, nous appliquons au capteur $i^{\text{ème}}$ une fonction de mesure comme déjà présenté dans (2.66) que nous réécrivons avec les notations de ce travail,

$$\mathcal{H}_i : (c, x_i) \mapsto \int_{\Omega} \phi_i(x - x_i) c(x) dx, \quad 1 \leq i \leq N_s, \quad (4.4)$$

où $\phi_i : x \mapsto \phi(x - x_i)$ est une fonction d'échantillonnage associée à un capteur i à une position x_i . Dans ce travail, ϕ est choisie comme une fonction noyau gaussienne. L'opérateur d'observation global \mathcal{H} qui fait correspondre les sorties du modèle physique aux mesures est donc un opérateur défini par :

$$\mathcal{H} : c \mapsto [\mathcal{H}_1(c), \dots, \mathcal{H}_{N_s}(c)]. \quad (4.5)$$

Dans cette étude, les capteurs sont répartis uniformément à l'intérieur du domaine et 2 capteurs supplémentaires mesurent les concentrations aux limites. Du bruit gaussien est également considéré pour plusieurs expériences dans cet article. Un bruit additif de mesure est appliqué à chacun des N_s capteurs :

$$y_i^m = \mathcal{H}_i(c) + \epsilon_i^o \quad \text{avec} \quad \epsilon_i^o \sim \mathcal{N}(0, \sigma^{o2}), \quad (4.6)$$

où ϵ_i^o est un bruit gaussien affectant le $i^{\text{ème}}$ capteur. Dans cet article, nous utilisons une variance constante des mesures comme indiqué dans (4.6). Pour chaque mesure d'un profil de concentration c_s , un écart-type constant avec $\sigma_s^o = 5\% \times \max(\mathcal{H}(c_s))$ est choisi. En réalité, cela constitue une hypothèse conservatrice sur la quantité de bruit pour les valeurs mesurées les plus faibles.

4.2.2 Méthodes

L'objectif de cette étude d'estimation de source est de reconstruire le terme de source de paramètres $\theta = (a_s, x_s, \sigma_s)^T$ [$\text{g s}^{-1} \text{m}^{-3}$, m, m^{-2}] tel que défini dans (4.2) en utilisant les mesures de concentration fournies par les N_s capteurs. Avec $N_s \ll N_x$ où N_x est le nombre de points d'approximation du vecteur de concentrations \mathbf{c} solution du modèle discret (4.3). Détaillons les deux méthodes d'estimation présentées : la méthode basée sur l'apprentissage automatique qui repose sur un modèle de réseau neuronal à propagation avant entraîné avec une base de données synthétique, et la méthode classique basée sur le modèle d'assimilation de données (3D-Var) basée sur le modèle physique.

Approche par apprentissage automatique

Un perceptron multicouche avec une ou plusieurs couches cachées est construit pour apprendre les paramètres de la source à partir des profils de concentration et des données de champ de vent en entrée. Ce type d'architecture a montré sa fiabilité pour résoudre un problème inverse [Georges2020]. La structure typique du réseau neuronal utilisé dans ce travail est illustrée dans la figure 4.1 : un réseau neuronal est utilisé pour approximer la fonction inverse de l'opérateur en avant $\mathcal{F} = \mathcal{H} \circ \mathcal{M}$ déjà commenté en (2.63) qui fait correspondre les mesures capturées par les capteurs au vecteur de paramètres θ à l'origine de la source à reconstruire. Dans les expériences présentées dans la section 4.2.3, notre réseau neuronal contient 2 couches cachées, toutes deux activées avec la fonction tanh. Le réseau neuronal est entraîné avec une fonction de coût de l'erreur quadratique moyenne, et la descente de gradient L-BFGS-B est utilisée dans le processus d'apprentissage. Appelons n_s le nombre d'échantillons. Les entrées et les sorties sont décrites comme suit :

Entrées : La matrice $\mathbf{U}^m \in \mathbb{R}^{n_s \times N_s}$ et le vecteur $\beta \in \mathbb{R}^{n_s}$, représentent respectivement n_s exemples de N_s mesures et les données de vent correspondantes qui sont donnés en entrée au réseau neuronal. La matrice \mathbf{U}^m est construite en effectuant n_s simulations du modèle avec différentes vitesses de vent c et différents paramètres de source θ en utilisant le modèle physique (4.1) sur N points. Ensuite, les résultats des simulations sont échantillonnés en utilisant l'opérateur de mesure décrit dans les équations (4.4), (4.5) et (4.6).

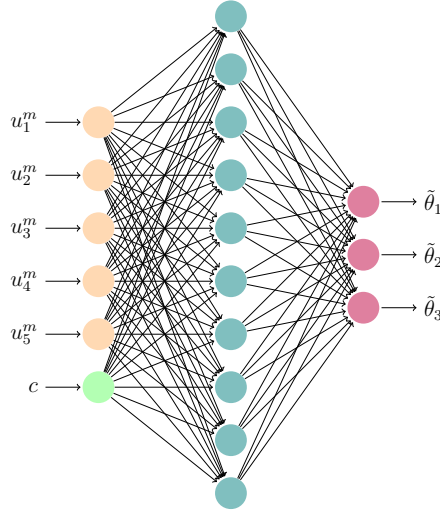


FIGURE 4.1 – Exemple de l’architecture du réseau neuronal pour l’inversion de source. Ici, avec 5 mesures d’entrée échantillonnées à partir de la même simulation (en orange) et une entrée scalaire de vent (en orange), une couche cachée de 10 neurones activés avec la fonction tanh (en bleu) et 3 sorties $\tilde{\theta}_j$ $j \in \{1, 2, 3\}$ correspondant aux paramètres de source remis à l’échelle tels que son amplitude, sa position et sa dispersion.

Sorties : $\Theta \in \mathbb{R}^{n_s \times 3}$ est une matrice contenant n_s vecteurs différents de paramètres de source, où chaque ligne de Θ est de la forme $\Theta_k = [\theta_k(1), \theta_k(2), \theta_k(3)]$, $1 \leq k \leq n_s$. Pour éviter les biais dans le processus d’apprentissage causés par la différence d’ordre de grandeur de chaque paramètre de source, les paramètres de source doivent être mis à l’échelle pour avoir le même ordre de grandeur. Pour cela, nous construisons une matrice mis à l’échelle $\tilde{\Theta}$ où chaque paramètre est mis à l’échelle en utilisant la transformation suivante :

$$\tilde{\theta}(j) = \frac{\theta(j) - \min(\Theta(j))}{\max(\Theta(j))}, \quad 1 \leq j \leq 3. \quad (4.7)$$

Ainsi, le modèle renvoie une estimation des paramètres adimensionnels en sortie. Une mise à l’échelle est appliquée à l’estimation de sortie en utilisant $\min(\Theta(j))$ et $\max(\Theta(j))$ obtenus lors du processus d’apprentissage.

Données d’entraînement : $n_{\text{train}} \in \mathbb{N}$ simulations de dispersion de polluants avec différentes vitesses de vent c et différents paramètres de source θ doivent être réalisées pour générer un ensemble d’entraînement et un ensemble de validation. Pour couvrir l’ensemble réaliste des échantillons d’entraînement possibles, les vitesses de vent et les paramètres de source d’entraînement sont générés en utilisant une séquence pseudo-aléatoire de Sobol avec une faible discrédance. Les séquences à faible discrédance sont connues pour assurer "un bon remplissage du domaine" au sens de la meilleure approximation possible des intégrales de fonction pour un nombre donné d’échantillons (voir [Niederreiter1988]).

Approche variationnelle

Comme nous l’avons vu au chapitre 2, Les méthodes variationnelles reposent sur la minimisation de l’erreur au sens des moindres carrés entre les prédictions de sortie basées sur un modèle connu et les données de mesure comme nous l’avons montré en 2.5.2. Ce type de méthode est largement utilisée, et considéré comme une méthode standard pour estimer des paramètres à partir de mesures, notamment avec l’utilisation d’un modèle adjoint [Georges2019]. Ici, nous proposons

d'appliquer une méthode variationnelle à un problème stationnaire où u peut être exprimé comme une fonction explicite de θ grâce au modèle aux différences finies, évitant ainsi l'utilisation explicite d'un modèle adjoint. La fonction de coût 3D-Var présentée en 2.5.2, qui consiste à minimiser l'erreur de prédiction de sortie au sens des moindres carrés, est choisie. Un terme de régularisation (voir [Tikhonov1977]), dépendant d'une estimation initiale est ajoutée, au terme d'évaluation de l'erreur de prédiction.

Le problème de minimisation de la fonctionnelle (2.67) est résolu pour notre problème présent, de manière directe sans utiliser l'adjoint en se basant sur l'approximation en dimension finie du modèle (4.1) donnée par (4.3), c'est-à-dire,

$$\min_{\theta} \mathcal{J}_a(\theta) = \frac{1}{2}(\theta - \theta^b)^T \mathbf{B}^{-1}(\theta - \theta^b) + \frac{1}{2} \|\mathbf{H}_a \mathbf{A}^{-1}(\mathbf{B}_1 q(\theta) + \mathbf{B}_2 \mathbf{C}) - \mathbf{y}^m\|_{\mathbf{R}^{-1}}^2, \quad (4.8)$$

où \mathbf{H}_a est la matrice obtenue à partir de l'approximation de l'opérateur intégral à N_s capteurs \mathcal{H} donné par (4.5), basée sur une méthode des trapèzes. Les matrices \mathbf{B} et \mathbf{R} représentent respectivement les matrices de covariance d'erreur de l'estimation initiale et de mesure. Nous supposons ici que les bruits sont non corrélés, de sorte que \mathbf{B} et \mathbf{R} sont diagonales et avec une variance identique. σ^b et σ^o sont les écarts-types associés à l'estimation initiale et aux observations. La solution de ce problème non convexe nécessite le calcul du gradient de $\mathcal{J}_a(\theta)$ par rapport à θ , c'est-à-dire

$$\nabla \mathcal{J}_a(\theta) = \mathbf{B}^{-1}(\theta - \theta^b) + \left(\mathbf{H}_a \mathbf{A}^{-1} \mathbf{B}_1 \frac{\partial q}{\partial \theta}(\theta) \right)^T \mathbf{R}^{-1} \left(\mathbf{H}_a \mathbf{A}^{-1}(\mathbf{B}_1 \mathbf{s}(\theta) + \mathbf{B}_2 \mathbf{C}) - \mathbf{y}^m \right). \quad (4.9)$$

Ici, la solution du problème d'optimisation (4.8) est réalisée en utilisant une méthode de descente quasi-Newton basée sur l'algorithme L-BFGS-B (voir [Byrd1995]) et le gradient (4.9).

4.2.3 Étude de cas : estimation de source dans une rue

L'étude de cas proposée ici concerne la détection d'un véhicule polluant arrêté quelque part sur un boulevard avec son moteur allumé. On suppose que les conditions de stationnarité de la distribution de la pollution sont obtenues sous des vitesses de vent constantes. Nous pourrions également envisager une situation où une autre source ponctuelle, telle qu'une fuite de gaz, émet dans un régime stationnaire.

Paramètres

La base de données d'entraînement est construite en créant l'ensemble de paramètres suivant :

$$\{p_k = [a_s^k, x_s^k, \sigma_s^k, u^k] ; 1 \leq k \leq n_{\text{train}}\}, \quad (4.10)$$

avec $[a_s, x_s, \sigma_s, u] \in [10^{-2}, 4] \times [0, 1; 0, 9] \times [1, 6 \times 10^{-2}; 1 \times 10^{-1}] \times [-10^{-3}; 10^{-3}]$ générés par une séquence de Sobol et avec les unités respectives $\text{g km}^{-3} \text{s}^{-1}$, km, km, km s^{-1} . L'ensemble de validation est créé de la même manière avec n_{test} échantillons, en choisissant $[a_s, x_s, \sigma_s, u] \in [5 \times 10^{-2}, 3] \times [0, 1; 0, 9] \times [1, 8 \times 10^{-2}; 8, 5 \times 10^{-2}] \times [-10^{-3}; 10^{-3}]$. Les données de l'ensemble de test sont générées par des séquences quasi-aléatoires de Sobol. La position de la source est générée de manière à ce que $x_s \in [0, 1\text{km}; 0, 9\text{km}]$ afin d'éviter de générer des sources qui seraient situées trop près des limites du domaine, ce qui pourrait conduire à une solution non physique de notre modèle de dispersion donné par (4.1). Notre base de données contient $n_{\text{train}} = 15000$ échantillons d'entraînement, tandis que la validation est effectuée sur $n_{\text{test}} = 200$ échantillons. Comme mesure de performance, l'erreur relative au sens des moindres carrés (RRSE) est choisie, définie par :

$$\text{RRSE} = \frac{\|\hat{\mathbf{s}} - \mathbf{s}_{\mathbf{T}}\|_2}{\|\mathbf{s}_{\mathbf{T}}\|_2}, \quad (4.11)$$

où $(\mathbf{s}_T, \hat{\mathbf{s}}) \in \mathbb{R}^{N_x \times 2}$ représentent respectivement le profil réel de la source et son estimation. L'erreur RRSE exprime une erreur relative entre la prédiction \hat{q} et la vérité terrain \mathbf{s}_T . Plus la prédiction $\hat{\mathbf{s}}$ est correcte, plus la RRSE tend vers 0. Pour faciliter la convergence de la méthode variationnelle, l'estimation initiale est donnée par :

$$\theta_k^b = \left[0, \arg \max_x (\mathbf{y}_k^m), 0 \right], \quad 1 \leq k \leq n_{\text{test}}; \quad (4.12)$$

ce qui signifie que la position initiale de la source estimée est située à la concentration mesurée la plus élevée. L'estimation initiale est utilisée à la fois dans la fonction de coût et comme initialisation de la méthode de descente L-BFGS-B. La meilleure architecture de réseau neuronal a été obtenue après quelques essais. Ainsi, le réseau neuronal choisi contient 2 couches cachées entièrement connectées de 50 neurones chacune, avec la fonction d'activation *tanh*. L'optimisation est réalisée en utilisant l'algorithme L-BFGS-B. La régularisation L_2 est fixée à une faible valeur de 10^{-12} . L'approche d'apprentissage automatique est réalisée à l'aide des bibliothèques *keras* et *tensorflow* disponibles avec le langage Python.

Comparaison sans bruit

Expérience avec 12 capteurs : Des simulations sont réalisées sans bruit pour déterminer les performances du RN par rapport à la méthode 3D-Var. Plusieurs estimations très aberrantes renvoyées par le RN (certainement dues à des conditions non physiques lors du processus d'entraînement) et des mauvaises estimations de la méthode 3D-Var (en raison de la non-convergence de l'algorithme BFGS) sont exclues : 16 sur 200, ce qui donne une étude de 184 estimations. Sur la base de cette expérience, le réseau neuronal est capable d'estimer une source avec une RRSE du même ordre de grandeur que celle obtenue avec la méthode 3D-Var standard, avec une moyenne et un écart-type de RRSE de $\mu_{\text{RRSE}}^{\text{RN}} = 0.22$ et $\sigma_{\text{RRSE}}^{\text{RN}} = 0.12$, et $\mu_{\text{RRSE}}^{\text{var}} = 0.31$ et $\sigma_{\text{RRSE}}^{\text{var}} = 0.21$ pour le RN et la 3D-Var respectivement. Dans l'ensemble, le réseau neuronal donne des résultats légèrement meilleurs. Cela est probablement dû au fait que la méthode variationnelle est un problème d'optimisation non convexe. Par conséquent, certains minima locaux peuvent conduire à des solutions quasi-optimales avec des différences significatives dans les valeurs des paramètres.

Les RRSE obtenus avec les approches d'apprentissage automatique et variationnelle suivent une loi log-normale, confirmée par le test de Kolmogorov-Smirnov (p -value = $0.39 > 0.05$ et $0.44 > 0.05$ respectivement). Les distributions de RRSE sont présentées à la figure 4.3. La méthode d'apprentissage automatique est basée sur les données et ne contient donc aucune information a priori sur le modèle physique, tandis que la méthode variationnelle repose sur la connaissance de la physique et sur une bonne initialisation donnée par (4.12). Nous remarquons que la métrique RRSE est sensible à une légère erreur de position estimée. Par exemple, dans le cas du 65^{ème} échantillon de validation illustré à la figure 4.2, le réseau neuronal estime bien l'amplitude et la dispersion, mais une légère erreur de position conduit à une RRSE de $\text{RRSE}^{\text{RN}} = 0.16$ pour le réseau neuronal contre $\text{RRSE}^{\text{var}} = 0.04$ pour la méthode variationnelle, respectivement.

Dans l'ensemble, sans bruit et en utilisant un total de 12 capteurs (10 régulièrement espacés dans le domaine et 2 qui mesurent les conditions aux limites), l'approche d'apprentissage automatique donne des résultats légèrement meilleurs que la méthode variationnelle en ce qui concerne la métrique RRSE. Le RN (Réseau de Neurones) renvoie parfois des estimations non physiques telles que des sources négatives ou des sources non gaussiennes. La figure 4.4 montre que les performances des deux méthodes sont similaires.

Expérience avec 50 capteurs : Des simulations avec 50 capteurs sont réalisées afin d'observer la sensibilité des estimations à la densité des mesures. Plusieurs estimations non physiques renvoyées par le réseau neuronal ou par la méthode variationnelle sont exclues du réseau : 9 sur 200, ce qui donne une étude de 191 estimations. Les deux méthodes renvoient des résultats similaires à ceux

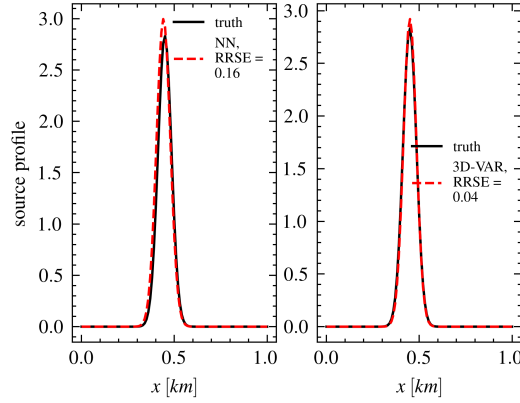


FIGURE 4.2 – Exemple de reconstruction de source : à gauche, avec un réseau neuronal ; à droite, avec une méthode variationnelle.

obtenus précédemment, avec une moyenne et un écart-type de $\mu_{RRSE}^{RN} = 0.13$ et $\sigma_{RRSE}^{RN} = 0.08$, comparés à $\mu_{RRSE}^{var} = 0.24$ et $\sigma_{RRSE}^{var} = 0.15$ respectivement pour le RN et la 3D-Var. Cela signifie que l'erreur diminue d'environ la moitié pour le RN et d'un tiers pour la 3D-Var par rapport à l'utilisation de 12 capteurs. Une fois de plus, le réseau neuronal donne de meilleurs résultats que la 3D-Var. Comme avec 12 capteurs, les distributions des RRSE suivent une loi log-normale avec un test de Kolmogorov-Smirnov (p-value = 0.19 > 0.05 pour le RN et p-value = 0.90 > 0.05 pour la 3D-Var). Dans l'ensemble, la reconstruction de la source en utilisant 50 capteurs sans bruit (48 régulièrement espacés dans le domaine et 2 mesurant les conditions aux limites) permet d'obtenir des améliorations de 30 à 50 % par rapport au cas avec 12 capteurs sans bruit. La figure 4.4 montre que les performances du réseau neuronal sont nettement meilleures que celles de la 3D-Var, sa distribution étant plus étroite et plus proche de 0 sur l'axe des RRSE. En revanche, les deux méthodes s'améliorent avec l'utilisation de 50 capteurs au lieu de 12.

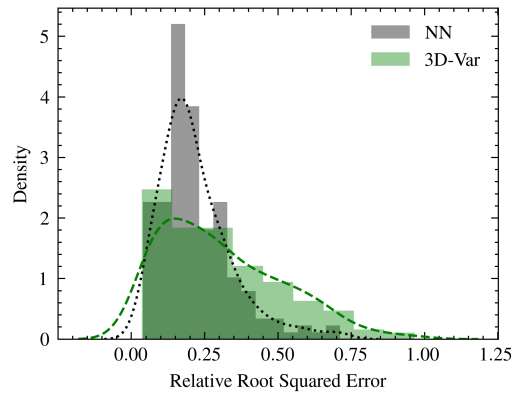


FIGURE 4.3 – Comparaison des distributions des RRSE des estimations de la 3D-Var et du RN avec 12 capteurs, sans bruit.

Comparaison avec du bruit

Dans cette expérience, un bruit gaussien ϵ_s^o est appliqué à chaque échantillon s ($1 \leq s \leq n_{test}$ ou n_{train}) de l'ensemble de données des profils de concentration renvoyés par les N_s capteurs, en suivant la procédure décrite dans la section 4.2.1. Ainsi, le bruit gaussien d'un ensemble de mesures a une variance constante qui dépend du maximum du profil de concentration mesuré

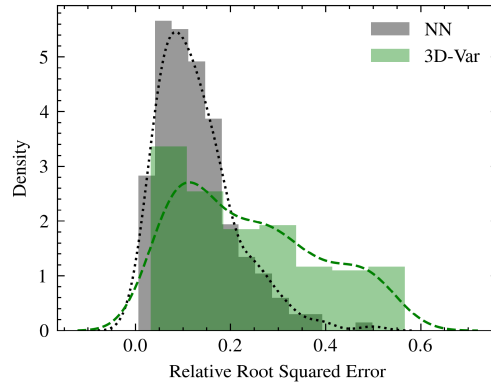


FIGURE 4.4 – Comparaison des distributions des RRSE des estimations de la 3D-Var et du RN avec 50 capteurs, sans bruit.

sans bruit. Cette hypothèse est raisonnablement conservatrice car les valeurs mesurées les plus basses d'un profil de concentration seront relativement plus bruitées que sa valeur maximale. Le réseau neuronal est entraîné ici avec un ensemble de données de concentration bruitées. L'erreur de supposition initiale utilisée dans la 3D-Var a été modifiée de $\sigma^b = 1 \times 10^6$ à $\sigma^b = 5 \times 10^4$ pour rendre le problème plus convexe.

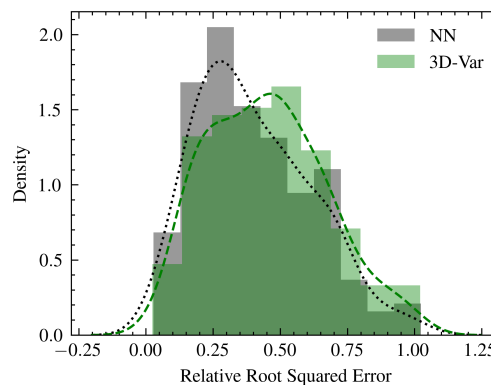


FIGURE 4.5 – Comparaison des distributions des RRSE des estimations de la 3D-Var et du RN avec 50 capteurs, avec bruit.

Expérience avec 12 capteurs : Des simulations sont réalisées avec du bruit de la même manière qu'auparavant. Plusieurs estimations non physiques renvoyées par le RN ou la 3D-Var sont exclues des résultats : 9 sur 200, ce qui donne un ensemble de 191 estimations. Les deux méthodes ont des performances similaires en présence de bruit avec une moyenne et un écart-type de $\mu_{RRSE}^{RN} = 0.48$ et $\sigma_{RRSE}^{RN} = 0.26$ en utilisant le RN et $\mu_{RRSE}^{var} = 0.52$ et $\sigma_{RRSE}^{var} = 0.23$ en utilisant la 3D-Var. La figure 4.6 montre que les distributions des RRSE des deux méthodes sont proches l'une de l'autre.

Expérience avec 50 capteurs : Avec 50 capteurs, dont 48 régulièrement espacés et deux valeurs de concentration mesurées aux limites, les résultats s'améliorent par rapport à l'utilisation de 12 capteurs. Cela conduit à des métriques de moyenne et d'écart-type de $\mu_{RRSE}^{RN} = 0.40$ et $\sigma_{RRSE}^{RN} = 0.21$ en utilisant le RN et $\mu_{RRSE}^{var} = 0.45$ et $\sigma_{RRSE}^{var} = 0.22$ en utilisant la 3D-Var. Ainsi, μ_{RN}^{var} et μ_{RRSE}^{var} diminuent respectivement de 17 et 13 % avec l'utilisation de 50 capteurs au lieu de 12. Par conséquent, le réseau neuronal présente des performances légèrement meilleures que l'approche

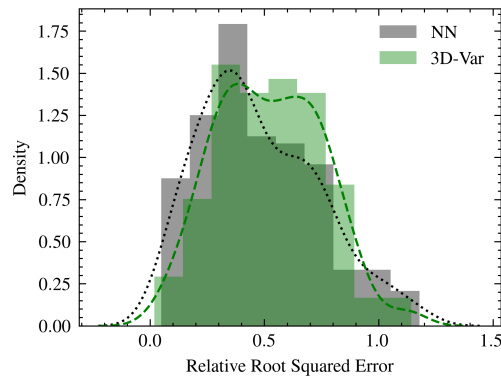


FIGURE 4.6 – Comparaison des distributions des RRSE des estimations de la 3D-Var et du RN avec 12 capteurs, avec bruit.

variationnelle. Cependant, on peut voir dans la figure 4.5 que les distributions des erreurs sont toujours assez proches entre les deux méthodes. Encore une fois, les distributions des erreurs suivent une loi log-normale. Les résultats de ces 4 cas sont illustrés dans le tableau 4.1.

TABLEAU 4.1 – Moyenne mesurée et écart-type des RRSE en utilisant les échantillons de validation pour chaque cas d'étude.

Cas			RN		3D-Var	
N	N° mesures	bruit	moyenne	ecart type	moyenne	écart type
184	12	non	0.22	0.12	0.31	0.21
191	50	non	0.13	0.08	0.24	0.15
191	12	oui	0.48	0.26	0.52	0.23
191	50	oui	0.40	0.21	0.45	0.22

4.2.4 Conclusions

Avec l'utilisation d'un petit nombre de capteurs bruités, les deux méthodes ont des performances similaires, bien que l'approche du RN renvoie souvent des estimations de source plus précises. L'approche par apprentissage automatique montre une amélioration plus importante de la précision que l'approche variationnelle à mesure que la densité des mesures augmente. Le manque relatif de performance de l'approche variationnelle dans certaines situations peut être lié à la non-convexité du problème d'optimisation associé. Nous avons montré que l'approche par apprentissage automatique est capable de reconstruire les sources même en présence de mesures bruitées. Le bruit des capteurs diminue considérablement les performances de l'estimation de source pour les deux méthodes. Globalement, les performances du réseau neuronal sont aussi bonnes, voire supérieures, à celles de l'approche variationnelle, comme le montre le tableau 4.1. Ces résultats révèlent que les approches d'apprentissage automatique sont prometteuses pour la détection des sources dans le contexte de la pollution de l'air urbain.

Cependant, même si la méthode semble démontrer des performances très satisfaisantes sur un cas très simple en 1D (voir tableau 4.1), le passage en 2D voire 3D, plus réaliste dans un contexte urbain complexe, rendrait nécessaire le développement d'une approche par Machine Learning plus coûteuse en terme de mise en œuvre compte-tenu des dimensions, comme, par exemple,

la mise en œuvre de réseaux de neurones à convolution, nécessitant des ressources de calculs haute performance pour l'apprentissage non compatibles avec nos hypothèses de travail. La taille des champs météorologiques pourrait notamment excéder des centaines de milliers d'éléments¹ compromettant alors l'entraînement du RN voire sa convergence.

De plus, il est nécessaire de faire un nouvel apprentissage à chaque changement des conditions atmosphériques. Dans ce cas unidimensionnel simple, une estimation par 3D-Var prend environ une minute tandis qu'un entraînement du RN sur CPU prend environ deux heures en utilisant un ordinateur avec 32 Go de RAM et un processeur à 12 fils de calculs. Étendre cette approche à des dimensions 2D ou 3D pourrait augmenter significativement le nombre de paramètres du RN et donc augmenter son temps d'apprentissage voire nuire à la convergence d'apprentissage. C'est la raison pour laquelle, nous avons fait le choix de ne pas poursuivre dans cette voie.

4.3 Approche adjointe avec modèle LS-RBF-FD

Après avoir développé et testé notre modèle numérique de dispersion basé sur la méthode sans maillage de Différences Finies supportées par des fonctions de base radiales au sens des moindres carrées (LS-RBF-FD) comme montré au chapitre 3 et en particulier en 3.4, nous avons pu envisager d'utiliser le modèle de dispersion LS-RBF-FD dans le but de développer des méthodes pour réaliser le problème inverse d'Estimation de Terme Source (STE) en milieu urbain. Cette étude considère un problème de STE dans un milieu tri-dimensionnel urbain dans le contexte de la pollution atmosphérique. Le problème d'STE est résolu en utilisant une méthode adjointe reposant sur le modèle LS-RBF-FD de dispersion en 3.4 pour estimer des sources de pollution en disposant de mesures. L'approche proposée est illustrée par des cas d'étude où une unique source de pollution est estimée sur un domaine 3D de $205\text{ m} \times 120\text{ m} \times 105\text{ m}$ incluant trois bâtiments déjà discuté en 3.3.5 et montré en Figure 3.10b. Comme mentionné en introduction de la thèse (1.1), les capteurs bas-coût (< 1000 euros) de particules fines sont en plein essor [Alfano2020] et rendent envisageables à terme l'utilisation de mesures en grand nombre. L'abondance de mesures permet alors d'imaginer des exemples d'estimations de sources quelconques possiblement multiples et distribuées sur tout le domaine d'étude. C'est dans ce contexte que la source de pollution est estimée sur tout le domaine de simulation.

La Figure 3.29a illustre une concentration en polluants affichée à environ 2 mètres de hauteur et calculée sur un domaine 3D à trois bâtiments. A cause des bâtiments, les champs météorologiques tels que le vent ne sont pas uniformes, le panache de pollution est ainsi très dépendant des conditions météorologiques. Par soucis de clarté, nous illustrons de nouveau la Figure 3.29a ci dessous, associée à l'enveloppe des bâtiments du domaine de simulation.

Nous avons discuté aux chapitres 2 et 3 de la modélisation de la dispersion de polluants par LS-RBF-FD. Nous choisissons cette modélisation dans cette étude. Nous allons présenter quelques hypothèses de simulation, ainsi que le modèle de dispersion considéré d'où nous tirerons des conditions d'optimalité utilisés dans la méthode de STE.

Hypothèses et modèle de pollution atmosphérique : Comme discuté dans le chapitre 2, à l'échelle urbaine nous pouvons considérer que le fluide dans lequel évolue un polluant est incompressible et que le polluant ne change pas localement la densité de l'air. Nous utilisons donc le modèle d'ADPDE exposé en (3.3)-(3.4) en donnant en entrée des champs météorologiques

1. Les champs météorologiques de vitesse de vent (\mathbf{U}) et de diffusion (\mathbf{K}) pourraient facilement être de taille $100\,000 \times 3$ chacun sur un domaine de calcul 3D de 1 km de côté.

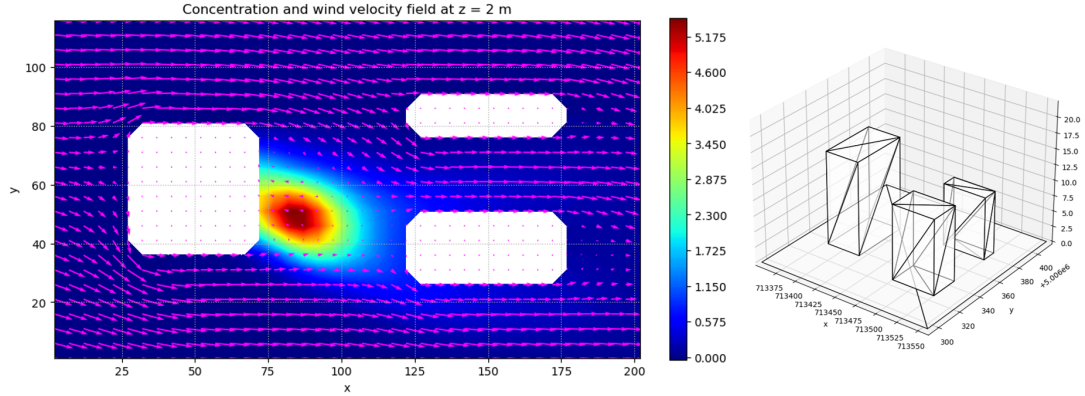


FIGURE 4.7 – (Gauche) Champ de concentration de polluants et champs de vent (flèches roses) à $z = 2$ mètres de haut. (Droite) domaine 3D composé de trois bâtiments de tailles différentes.

moyens et en considérant un régime stationnaire $\partial_t \mathbf{c} = 0$. Comme hypothèse supplémentaire, le terme source est constant dans le temps, ne se déplace pas ou lentement, et émet suffisamment longtemps pour atteindre un état stable. Finalement, les polluants sont considérés comme passifs dans cette étude sans terme de réaction chimique ou de photo-réaction. Les polluants sont par exemple des particules fines : PM2.5 ou PM10. Les concentrations \mathbf{c} et la source \mathbf{s} ont pour unités respectives $[\mu\text{g m}^{-3}]$ et $[\mu\text{g m}^{-3} \text{ s}^{-1}]$.

Modélisation LS-RBF-FD de la dispersion de polluants : Nous avons déjà présenté aux chapitres 2 puis 3 les RBF-FD, et la variante de moindres carrés LS-RBF-FD pour résoudre des EDP. Nous rappellerons seulement comment se construit le domaine de simulation et montrerons l'origine de l'établissement de conditions d'optimalité que nous utiliserons dans notre méthode d'estimation de source en 4.3.1. Rappelons que pour appliquer la méthode LS-RBF-FD, le domaine d'étude est divisé en deux sous-ensembles comme considéré en 2.4.6, pour rappel deux ensembles de nœuds de calcul partitionnant l'espace de simulation Ω :

- Un ensemble X de nœuds qui couvre le domaine Ω , avec $N_x = \text{card}(X)$ et sert d'espace d'approximation de la solution de modèle. L'ensemble X ne tient pas compte des conditions aux limites.
- Un ensemble Y de nœuds d'évaluation composé de deux sous-ensembles Y_Ω de cardinal $N_{y_\Omega} = \text{card}(Y_\Omega)$ correspondant aux nœuds choisis à l'intérieur du domaine et Ω , et $Y_{\partial\Omega}$ de cardinal $N_{y_{\partial\Omega}} = \text{card}(Y_{\partial\Omega})$, correspondant aux nœuds sur la frontière. Le nombre de nœuds total d'évaluation étant $N_y = \text{card}(Y) = N_{y_\Omega} + N_{y_{\partial\Omega}}$.

Comme montrée en 2.4.6 la méthode LS-RBF-FD permet d'évaluer le carré de la norme d'un résidu d'une EDP sur les points de Y qui peut alors être approximé par somme d'écartés quadratiques et s'écrire de manière compacte,

$$\|r(v)\|_{L_2(\Omega)}^2 \approx \|\mathbf{D}_\Omega \mathbf{v} - \mathbf{s}\|^2 + \|\mathbf{D}_{\partial\Omega} \mathbf{v}\|^2. \quad (4.13)$$

Comme vu en 2.4.6, \mathbf{c}_a est solution d'un problème d'optimisation des moindres carrés construit de la manière suivante,

$$\mathbf{c}_a = \arg \min_{\mathbf{v}} \|\mathbf{D}\mathbf{v} - \mathbf{f}\|^2, \quad (4.14)$$

où \mathbf{c}_a et $\mathbf{v} \in \mathbb{R}^{N_x}$, $\mathbf{f} \in \mathbb{R}^{N_y}$ et $\mathbf{D} \in \mathbb{R}^{N_y \times N_x}$. Nous pouvons empiler les matrices de différentiation et le terme de forçage \mathbf{f} regroupant le terme source et les conditions aux limites tels que \mathbf{D} et \mathbf{f} peuvent être exprimés comme,

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_\Omega \\ \mathbf{D}_{\partial\Omega} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{s}(Y_\Omega) \\ \mathbf{0} \end{pmatrix}, \quad (4.15)$$

où $\mathbf{s}(Y_\Omega)$ représente le vecteur de terme source évalué aux nœuds de Y_Ω .

Comme discuté dans le chapitre 2, le problème (4.14) admet une solution explicite unique montrée en (2.59) si \mathbf{D} est de rang plein. Cependant il est intéressant de tirer de (4.14) les conditions d'optimalité suivantes :

$$\begin{aligned} \mathbf{D}^T \mathbf{D} \mathbf{c}_a - \mathbf{D}^T \mathbf{f} &= 0 \\ \Leftrightarrow \mathbf{E} \mathbf{c}_a - \mathbf{D}_\Omega^T \mathbf{s}(Y_\Omega) &= 0, \end{aligned} \quad (4.16)$$

où $\mathbf{E} = \mathbf{D}_\Omega^T \mathbf{D}_\Omega + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega}$.

La relation (4.16) sera utilisée comme *modèle direct* de la méthode de STE basée sur l'adjoint proposée dans la section qui suit. Le reste de cette étude est organisé de la manière suivante : La partie 4.3.1 décrit la formulation et la résolution du problème de STE construit à partir de l'adjoint du modèle direct et dans le cadre de travail d'une modélisation LS-RBF-FD tridimensionnelle . Une étude de cas est présentée en partie 4.3.2, et la partie en 4.3.3 présente des conclusions et des perspectives.

4.3.1 Méthode de STE en utilisant un adjoint discret

Le problème est l'estimation simultanée du vecteur de source inconnue $\mathbf{s} \in \mathbb{R}^{N_x}$ et le vecteur d'état à savoir le vecteur de concentration $\mathbf{c} \in \mathbb{R}^{N_x}$. \mathbf{s} et \mathbf{c} sont tous deux définis dans X . $\mathbf{y}^m \in \mathbb{R}^{N_s}$ est le vecteur contenant N_s mesures dispersées dans le domaine Ω . La STE est pratiquée sur un lot de N_s mesures grâce à une méthode utilisant l'adjoint discret. Notons que l'approche basée sur les RBF-FD permet l'interpolation du vecteur de source \mathbf{s} (défini sur les coordonnées de X) en tout point y dans le domaine Y_Ω . Pour ce faire, la matrice d'interpolation Ψ_s est créée en générant une matrice de poids RBF-FD à la manière de (2.47) en ne considérant aucune application d'opérateur de dérivées. Il est alors possible d'écrire la relation d'interpolation suivante :

$$\mathbf{s}(Y_\Omega) = \Psi_s \mathbf{s}, \quad \mathbf{s}(Y_\Omega) \in \mathbb{R}^{N_{y_\Omega}}. \quad (4.17)$$

L'opérateur de mesure est noté ici, \mathbf{H} , et est défini comme un opérateur d'interpolation très similaire à celui montré en (4.17) :

$$\mathbf{y} = \mathbf{H}(Y_{\text{mes}}) \mathbf{c} \quad (4.18)$$

où \mathbf{y} représente le vecteur de sortie de modèle direct ramené aux mesures et Y_{mes} est l'ensemble des mesures localisées dans Ω . Notons par ailleurs que si les coordonnées de l'ensemble des mesures Y_{mes} sont incluses dans les coordonnées de sorties de modèle X , alors \mathbf{H} devient une matrice de sélection de mesures prenant des valeurs nulles ou égales à 1.

Une approche adjointe

L'approche de STE proposée consiste à résoudre un problème d'optimisation des moindres carrés (un problème régularisé basé sur un modèle LS-RBF-FD de minimisation de l'écart quadratique de la prédiction de la sortie) :

$$\begin{aligned} \min_{\mathbf{s}, \mathbf{c}} \mathcal{J}(\mathbf{s}, \mathbf{c}) &= \frac{1}{2} \|\mathbf{H} \mathbf{c} - \mathbf{y}^m\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\mathbf{s}\|_{\mathbf{B}^{-1}}^2 \\ &\text{sous } \mathbf{E} \mathbf{c} - \mathbf{D}_\Omega^T \Psi_s \mathbf{s} = 0, \end{aligned} \quad (4.19)$$

où \mathbf{R} et \mathbf{B} sont respectivement les matrices de covariance d'erreur et de régularisation. Nous pouvons énoncer le résultat suivant :

Theorem 4.3.1 *La solution optimale (\mathbf{s}^* , \mathbf{c}^*) du problème d'STE (4.19) est solution du système linéaire suivant,*

$$\begin{pmatrix} \mathbf{K} & \mathbf{E} \\ \mathbf{E} & -\mathbf{Q} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^m \\ \mathbf{0} \end{pmatrix}, \quad (4.20)$$

où $\mathbf{K} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ et $\mathbf{Q} = \mathbf{D}_\Omega^T \Psi_s \mathbf{B} \Psi_s^T \mathbf{D}_\Omega$. De plus, $\mathbf{s}^* = \mathbf{B} \Psi_s^T \mathbf{D}_\Omega \lambda^*$, où λ^* est l'adjoint de \mathbf{c}^* . Si la matrice \mathbf{E} est définie positive, $\mathbf{c}^* = [\mathbf{E} + \mathbf{Q} \mathbf{E}^{-1} \mathbf{K}]^{-1} \mathbf{Q} \mathbf{E}^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^m$, et λ^* est donné par $\lambda^* = -\mathbf{E}^{-1} \mathbf{C}^T \mathbf{R}^{-1} (\mathbf{H} \mathbf{c}^* - \mathbf{y}^m)$.

Preuve : Introduisons le Lagrangien du problème (4.19) : $L(\mathbf{s}, \mathbf{c}, \lambda) = \frac{1}{2} \|\mathbf{H} \mathbf{c} - \mathbf{y}^m\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\mathbf{s}\|_{\mathbf{B}^{-1}}^2 + \lambda^T (\mathbf{E} \mathbf{c} - \mathbf{D}_\Omega^T \Psi_s \mathbf{s})$. La théorie de Lagrange fournit les conditions nécessaires d'optimalité :

$$\nabla_{\mathbf{c}} L = 0 \Leftrightarrow \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \mathbf{c} - \mathbf{y}^m) + \mathbf{E} \lambda = 0, \quad (4.21)$$

$$\nabla_{\mathbf{s}} L = 0 \Leftrightarrow \mathbf{B}^{-1} \mathbf{s} - \Psi_s^T \mathbf{D}_\Omega \lambda = 0, \quad (4.22)$$

$$\nabla_{\lambda} L = 0 \Leftrightarrow \mathbf{E} \mathbf{c} - \mathbf{D}_\Omega^T \Psi_s \mathbf{s} = 0, \quad (4.23)$$

où (4.21) est l'équation adjointe du modèle. Sous l'hypothèse que \mathbf{E} est définie positive (et ainsi inversible), λ^* peut être calculé comme, $\lambda^* = -\mathbf{E}^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \mathbf{c}^* - \mathbf{y}^m)$. De (4.22), on obtient $\mathbf{s}^* = \mathbf{B} \Psi_s^T \mathbf{D}_\Omega \lambda^*$. De (4.23) et en utilisant l'expression de \mathbf{s}^* , nous obtenons $\mathbf{E} \mathbf{c}^* - \mathbf{Q} \lambda^* = 0$. Finalement, $\mathbf{c}^* = [\mathbf{E} + \mathbf{Q} \mathbf{E}^{-1} \mathbf{K}]^{-1} \mathbf{E}^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^m$ en utilisant l'expression de λ^* . Notons que $\mathbf{E} + \mathbf{Q} \mathbf{E}^{-1} \mathbf{K}$ est toujours inversible si \mathbf{E} est définie positive puisque la matrice symétrique $\mathbf{Q} \mathbf{E}^{-1} \mathbf{K}$ est au moins définie positive.

L'inversibilité de la matrice \mathbf{E} est garantie si et seulement si la matrice \mathbf{D} dans (4.15) est de rang plein. Cette condition est facilement satisfaite par un choix appropriée de polynôme de régularisation associé à des fonctions de bases PHS et avec la disposition des nœuds de X et de Y . Le système linéaire $2N_x \times 2N_x$ (4.20) peut être efficacement résolu en utilisant un solveur des moindres carrés ou une méthode de Krylov, bien adaptée pour les problèmes de grandes échelles sans inversion explicite \mathbf{E} .

4.3.2 Etude de cas STE dans sur une zone urbaine simulée

Paramètres de l'étude

Le domaine physique consiste en un domaine 3D rempli avec 3 bâtiments de différentes hauteurs de 10, 15 et 12 mètres de haut déjà présenté en 3.3.5. Les dimensions du domaine sont $L_x = 205 \text{ m} \times L_y = 120 \text{ m} \times L_z = 105 \text{ m}$. Le domaine 3D est illustré en Figure 4.7 (à droite). La Figure 4.8 illustre la distribution des nœuds d'évaluation de Y dans le domaine.

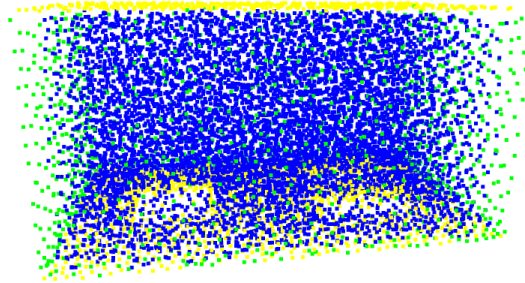


FIGURE 4.8 – Vue en perspective du domaine tridimensionnel Y , avec $N_y = 15000$. Bleu : nœuds intérieurs, jaune : nœuds de Neumann, vert : nœuds de Dirichlet.

Les fonctions de base radiales PHS5 suivantes sont utilisés dans ce travail,

$$\phi(y) = r^5, \text{ avec } r = \|y - x_i\|. \quad i \in \{1, \dots, N_x\} \quad (4.24)$$

La Fonction Radiale de Base (RBF) spline polyharmonique d'ordre PHS5 est utilisée dans cette étude. Les PHS5 ont été présentées en (2.32)-(2.33) et leur usage a été justifié pour le modèle de dispersion (2.13) en 3.2.1. Le modèle direct LS-RBF-FD utilisé dans (4.19) a les caractéristiques suivantes : $N_y = 15000$ nœuds d'évaluation sont utilisés pour assurer une résolution spatiale suffisante. Pour l'ensemble X , la distribution de nœuds est bien plus parcellaire avec $N_x = 2057$ nœuds ce qui permet naturellement une réduction du nombre de variables à estimer pour les vecteurs inconnus \mathbf{s} et \mathbf{c} . Les conditions de bords de flux nul (de type Neumann) sont assignées au sol ainsi qu'aux bords des bâtiments et au plafond du domaine tandis que des conditions de Dirichlet nulles sont imposées aux reste des bords du domaine. La taille des stencils est fixée à $n = 90$.

Les matrices \mathbf{R} and \mathbf{B} montrées dans (4.19) sont choisies comme $\mathbf{R} = (\sigma^o)^2 \mathbf{I}_{N_s}$ et $\mathbf{B} = b \mathbf{I}_{N_x}$, où \mathbf{I}_N représente la matrice identité de dimensions $N. \times N$. et $(\sigma^o)^2$ et b sont des coefficients de variance d'erreur de mesures et d'initialisation/régularisation, les covariances d'erreurs de mesures et d'initialisation étant considérées comme uniformes. Dans cette section $b \in [0, 1; 0, 15]$, les meilleures estimations ayant été observées dans cette plage de valeurs. L'écart type de bruit de mesure σ^o est testé pour différentes valeurs par la suite.

Calcul des champs de vent et de turbulence : Les champs tridimensionnels de vent et de diffusion sont calculés en utilisant le logiciel Micro-SWIFT-M [Oldrini2013] et sont commentés en 2.3.1 (paragraphe des RANS).

Mesures : Nous décidons de simuler des mesures \mathbf{y}^m en utilisant le modèle direct LS-RBF-FD (qui simule les données de terrain) avec un bruit blanc additif tel que

$$\mathbf{y}^m = \mathbf{y}^0 + \epsilon^o \text{ avec } \epsilon^o \sim \mathcal{N}(0, \mathbf{R}), \quad (4.25)$$

où $\mathbf{y}^0 \in \mathbb{R}^{N_s}$ est issu d'un champ de concentrations en polluants échantillonnées sur un modèle LS-RBF-FD de haute résolution avec $N_y = 50000$, $N_x = 15648$ (voir Figure 4.7). ϵ^o est un vecteur de bruit blanc Gaussien dans \mathbb{R}^{N_s} généré par une loi normale multinormée de covariance $\mathbf{R} = \sigma^{o2} \mathbf{I}_{N_s}$ où \mathbf{I}_{N_s} est la matrice identité de taille $(N_s \times N_s)$. Ainsi pour chaque mesure, le bruit de mesure est généré par le tirage d'un bruit blanc d'écart type σ^o . Plusieurs cas d'écart type de bruit sont testés par la suite. Remarquons que nous utilisons des mesures simulées issues d'un modèle LS-RBF-FD finement résolu ($N_y = 50000$, $N_x = 15648$) tandis que notre modèle utilisé dans la formulation STE-adjointe (4.19) est plus réduit ($N_y = 15000$, $N_x = 2057$), ainsi nous introduisons volontairement un bruit de modèle dans l'expérience de STE.

Terme source : Le terme source "réel" utilisé pour générer le champ de concentration est une fonction Gaussienne paramétrée sur Ω :

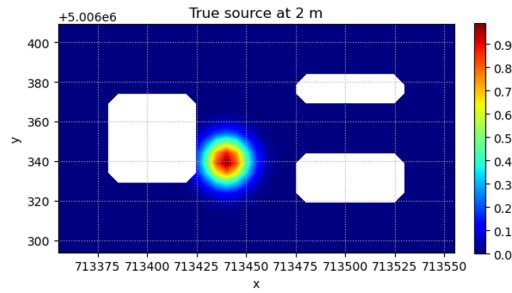
$$s : x \mapsto a_s \exp \left((x - x_s) / (\sqrt{2}\sigma_s) \right)^2, \quad x \in \Omega \quad (4.26)$$

où $a_s, x_s, \sigma_s = 1$, $[0.42L_x, 0.39L_y, 0.02L_z]$ et $[8, 8, 8]$. Le terme source est illustré en 2D en Figure 4.9.

Résultats

Six estimations ont été testées avec du bruit de mesures et avec 1500, 200, 300 et 50 mesures distribuées sur tout le domaine 3D y compris en hauteur du domaine de simulation (Ω).

Sans surprise, l'estimation avec $ns = 1500$ capteurs et un bruit de mesure faible ($\sigma^o = 0,002$) montre des performances correctes bien que la source estimée montre une erreur de localisation Δy d'environ 5m comme visible en Figure 4.10a. Le maximum de la source estimée y atteint

FIGURE 4.9 – Vraie source à $z = 2$ mètres de haut.

environ 0,60 contre 1 pour la source réelle. L'erreur relative pour le cas ($ns = 1500, \sigma^o = 0,002$) est visible en Figure 4.10c. Les erreurs relatives sont calculées comme montré en (3.51) avec un seuil ζ pris à 0,05 unités de source, de même pour les erreurs relatives de concentration. Pour le cas ($ns = 1500, \sigma^o = 0,002$) l'erreur relative au voisinage de la source véritable est comprise entre 0 et 260 %. Les plus grandes valeurs d'erreurs apparaissent au dessus (dans la direction $O - y$) de la source véritable, cela s'explique par l'erreur de position de source de 5 m sur l'axe y . De manière générale la position de source ainsi que ses ordres de grandeurs demeurent bien acquis. Conséquemment le champ de concentration en polluants est correctement estimé avec une forme et des ordres de grandeurs proche du champ de concentration de référence comme visible en Figure 4.11a et en Figure 4.11b. L'erreur relative au voisinage du panache de polluants est comprise entre 0 et 120 % (Figure 4.11c). Les erreurs au sein du panache (là où les concentrations véritables ont les plus grandes valeurs) atteignent 10 à 40 %. Remarquons que les meilleures précisions d'estimation sont atteintes là où la précision d'estimation est recherchée, c'est à dire au voisinage du terme source et au sein du panache de concentration.

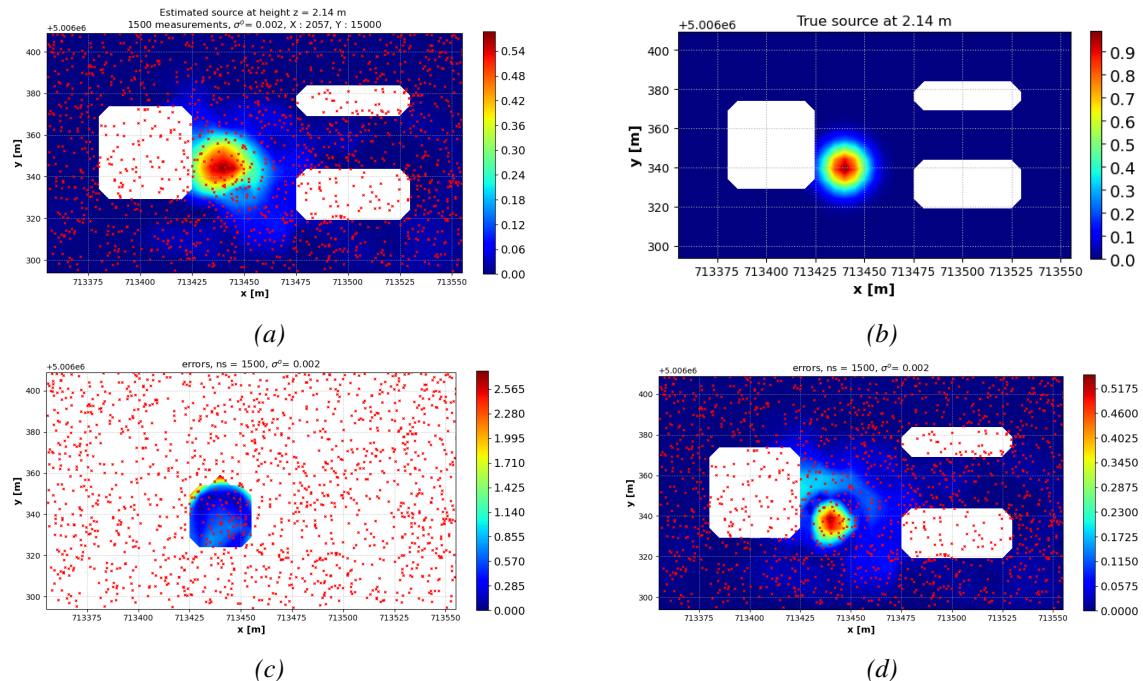


FIGURE 4.10 – (a) : Source estimée avec $n_s = 1500$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,002$, (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

L'estimation avec $ns = 200$ capteurs et un bruit de mesure élevé ($\sigma^o = 0,2$), montre des

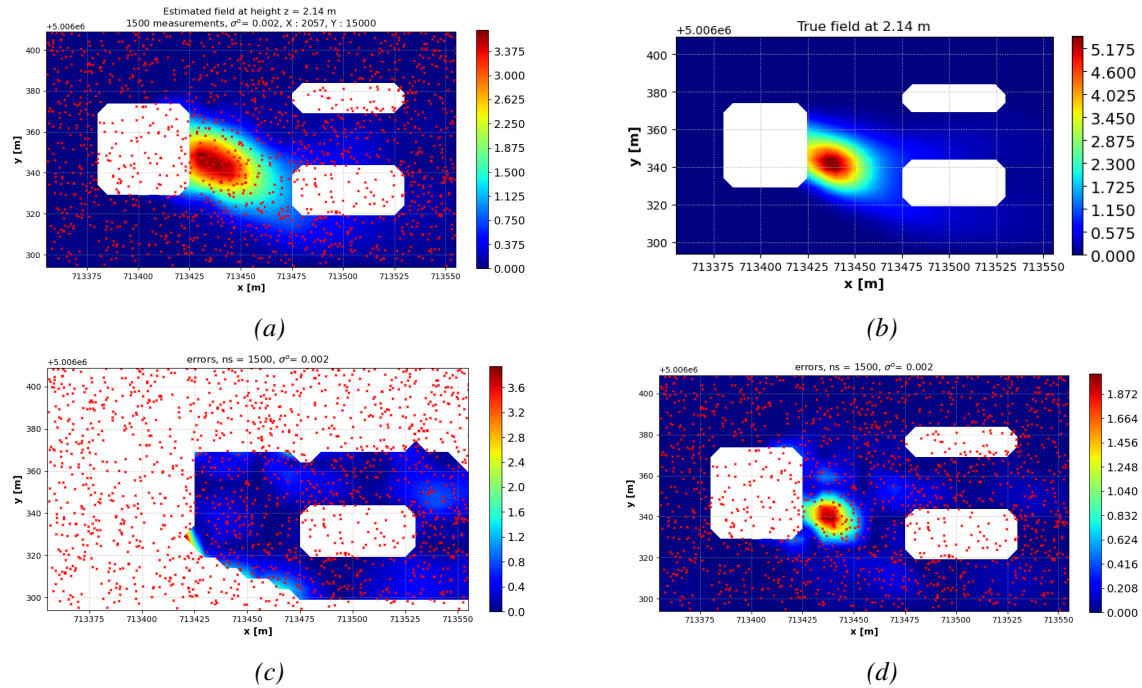


FIGURE 4.11 – (a) : Champ de concentration estimé avec $n_s = 1500$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,002$, (b) : vrai champ de concentration (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

performances d'estimation dégradées. Les erreurs de localisation latérale de la source estimée atteignent $\Delta x = 15$ m et $\Delta y = 5$ m (Figure 4.12a). Le maximum de la source estimée atteint environ 0,42 unités de source contre 1 pour la source réelle. Notons en particulier la présence d'un artefact d'estimation pour la source, ici près du troisième bâtiment au coin droit en Figure 4.12a vers les coordonnées $(x, y) = (713\ 470, 330 + 5,006 \times 10^6)$. L'artefact de source atteint environ 0,18 unités de source ce qui n'est pas négligeable par rapport au maximum estimé de la "vraie source" qui atteint 0,42 unités de source. L'erreur de localisation et d'estimation des amplitudes de source se traduit par une erreur relative élevée au voisinage de la vraie source (Figure 4.12c). Les erreurs relatives au voisinage de la "vraie source" atteignent environ 100 %. Conséquentemnt le champ estimé de concentration en polluants perd en qualité comme visible en Figure 4.13a. L'erreur relative au voisinage du panache de polluants y atteint environ 100 % en de nombreux endroits (Figure 4.13c).

Les erreurs sur la localisation du terme source ainsi que sur la localisation du champ de concentration ont un fort impact sur les erreurs affichées. En effet, les erreurs d'estimation calculées font intervenir la valeur absolue de la différence entre un champ estimé (θ) et un champ véritable θ_T ($|\theta - \theta_T|$) "pixel par pixel". Cette mesure est très sensible à la localisation des estimations. D'autres métriques existent pour évaluer les performances d'estimations géolocalisées. Nous pouvons notamment citer l'usage de la distance de Wasserstein. Son implémentation et son interprétabilité ne sont en revanche pas triviales [Farchi2016].

Plusieurs paramètres peuvent avoir un impact sur la qualité de la STE. Parmi eux, le nombre de nœuds dans X et Y et la valeur du coefficient de régularisation b affectent le modèle de source estimé et son ordre de grandeur. N_y doit être choisi dans le modèle réduit d'inversion de manière à conserver $N_y/N_x \leq 10$ afin de ne pas perturber le conditionnement de la matrice \mathbf{E} montré en (4.19). Le temps de calcul de la STE par la méthode adjointe (4.20) est raisonnablement rapide. Une fois l'opérateur \mathbf{D} de la relation (4.15) calculé, les estimations du terme source et du champ de pollution, solutions du problème (4.19) sont calculées en environ 4 minutes en utilisant un processeur Intel[®] Xeon[®] E-2276M à 6 coeurs couplé avec 32 Go de RAM et en utilisant un code

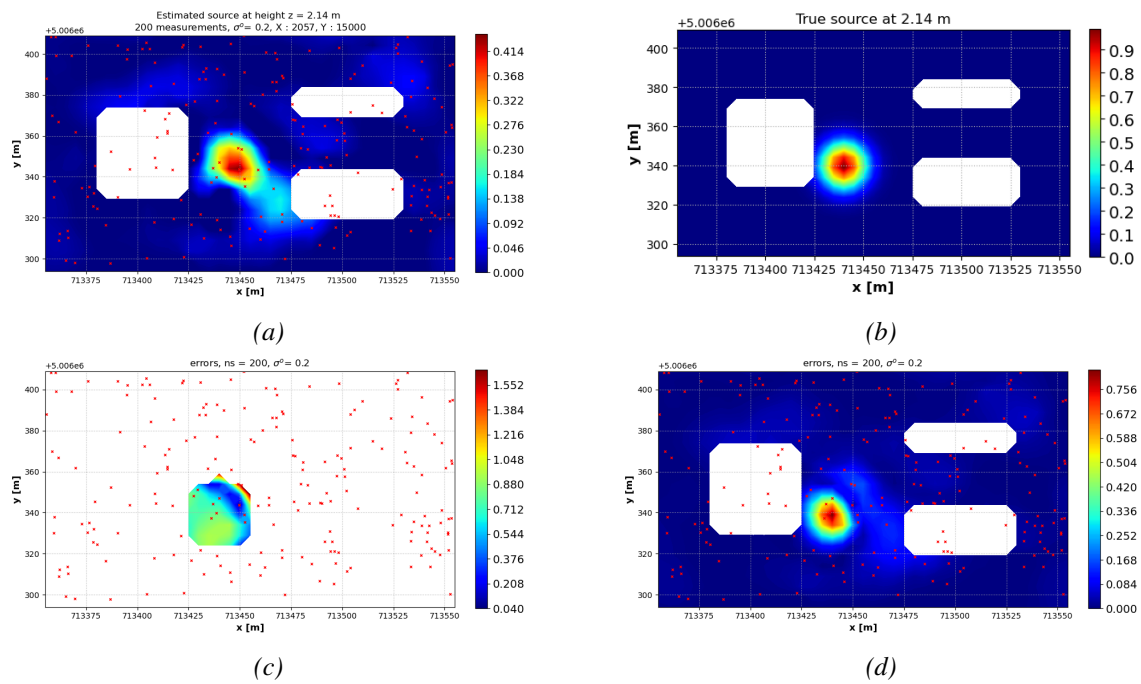


FIGURE 4.12 – (a) : Source estimée avec $n_s = 200$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,2$, (b) : vraie source (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

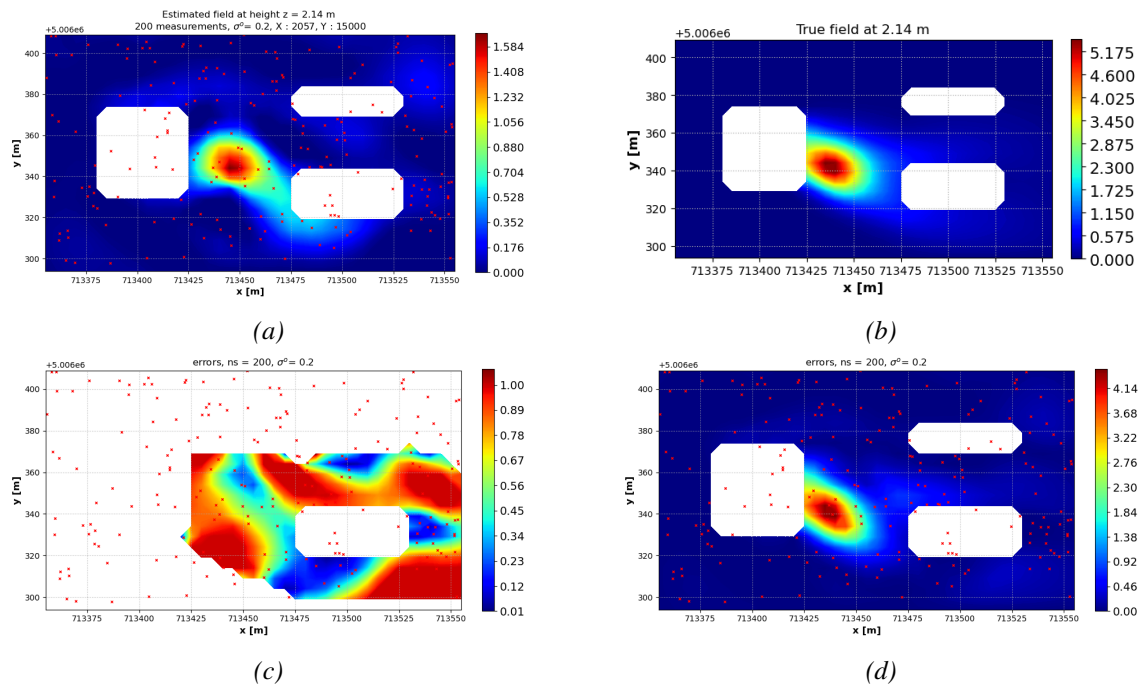


FIGURE 4.13 – (a) : Champ de concentration estimé avec $n_s = 200$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0.002$, (b) : vrai champ de concentration (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

Python. Des optimisations notamment de parallélisation et de calcul matriciels restent encore à faire.

Les erreurs RRSE telles que définies en (4.11) calculées pour tous les scénarios de tests de la méthode STE-adjointe présentée sont compilées en Tableau 4.2. $RRSE_X$, $RRSE_interp$ y sont respectivement les erreurs RRSE d'estimation du terme source sur tout le domaine de simulation X et sur une tranche 2D interpolée à 2,14 mètres de hauteur. Les erreurs sont calculées de la même manière pour les estimations de concentrations avec $RRSE_X_conc$ et $RRSE_interp_conc$. Nous remarquons que la méthode est assez robuste au bruit, les erreurs RRSE étant très proches entre des expériences utilisant des faibles niveau de bruit et des niveaux de bruit élevés. Ce comportement peut par exemple s'observer pour les cas ($n_s = 200, \sigma^o = 0,002$) et ($n_s = 200, \sigma^o = 0,2$). Nous remarquons cependant une forte augmentation des erreurs lorsque le nombre de mesures (n_s) diminue. L'erreur RRSE sur les concentrations est par exemple plus que doublée entre le cas utilisant 1500 mesures ($n_s = 1500, \sigma^o = 0,002$) et le cas utilisant 200 mesures ($n_s = 200, \sigma^o = 0,002$). L'usage de 50 mesures très bruitées ($n_s = 50, \sigma^o = 0,2$) fait encore augmenter les erreurs RRSE mais parvient tout de même à renseigner sur la localisation de la source.

TABLEAU 4.2 – Erreurs RRSE d'estimation du terme source et du champ de concentration. $RRSE_X$ et $RRSE_interp$: erreurs respectives d'estimation de source sur le domaine de simulation X et sur une tranche 2D interpolée à une hauteur de 2,14 mètres de haut. $RRSE_X_conc$ et $RRSE_interp_conc$: erreurs respectives d'estimation de champ de concentration sur le domaine X et une tranche 2D interpolée à 2,14 mètres. n_s : nombre de mesure, bruit : écart type du bruit gaussien de mesure (σ^o).

ns	bruit	RRSE_X	RRSE_interp	RRSE_X_conc	RRSE_interp_conc
1500	0.002	0.71	0.56	0.39	0.26
200	0.002	0.83	0.75	0.88	0.80
200	0.2	0.85	0.78	0.85	0.77
300	0.002	0.84	0.78	0.88	0.78
300	0.05	0.84	0.77	0.85	0.77
50	0.2	0.93	0.92	0.97	0.96

4.3.3 Conclusions

Dans cette partie nous avons présenté des premiers résultats encourageants d'une estimation de source de pollution atmosphérique à partir d'un modèle physique de transport. A notre connaissance, ce travail constitue la première application d'une méthode LS-RBF-FD pour une simulation tri-dimensionnelle de l'EDP d'Advection-Diffusion ainsi que de l'utilisation d'une méthode basée sur l'adjoint du modèle discret basée sur la méthode LS-RBF-FD. D'après nos observations, la méthode est efficace pour produire de bonnes estimations de sources isolées et répond aux besoins d'estimer des sources de pollution dans des environnements urbains 3D complexes. La méthode montre de bonnes capacités à localiser une source, elle produit cependant des artefacts d'estimation et montre des difficultés à estimer un champ de concentration en polluants dont les ordres de grandeurs sont proches du champ véritable.

4.4 Approche de type PIML avec modèle LS-RBF-FD

Dans l'esprit de l'apprentissage par réseaux de neurones informées par la physique [Raissi2019], une méthode des moindres carrés combinant résidu de modèle et erreur de prédiction de sortie a été développée pour réaliser une identification de source, et une estimation du champ de polluant correspondant.

Sa mise en œuvre ne nécessite pas le calcul d'un adjoint contrairement à l'approche décrite dans le paragraphe 4.3. L'absence d'adjoint relaxe les contraintes du problème de STE, l'estimation a ainsi plus de degrés de liberté et n'impose pas de contrainte forte d'égalité par l'usage d'un Lagrangien comme dans la formulation STE-adjointe (4.19). La relaxation des contraintes peut alors améliorer la convergence du problème de STE vers une estimation proche de la réalité. Aussi, l'absence d'adjoint permet d'améliorer le temps de calcul. De plus, la formulation du problème de STE par la présente méthode dispense de l'écriture de produits de matrices de différentiation avec leur transposées. D'une part, cela a pour avantage d'éviter les problèmes de conditionnement que nous avons mentionné en partie 4.3.2 sur la matrice $\mathbf{E} = \mathbf{D}_{\Omega}^T \mathbf{D}_{\Omega} + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega}$; d'autre part, l'approche de type PIML exploite ainsi le caractère creux des matrices de différentiation ce qui est avantageux en temps de calcul.

Cette approche est une méthode d'estimation PIML (Physics-Informed Machine Learning), dans laquelle l'approximation par réseaux de neurones utilisée dans Raissi et al. [Raissi2019] est remplacée par une approximation de type RBF-FD (que l'on pourrait relier à la notion de réseaux de neurones RBF à une couche cachée).

Notre méthode STE-PIML est discutée ici en régime stationnaire mais peut s'adapter directement dans le cas instationnaire en considérant le temps comme une dimension d'espace supplémentaire.

4.4.1 Formalisme de la méthode STE-PIML, cas stationnaire

Dans cette méthode nous nous proposons de résoudre directement un problème LS-RBF-FD intégrant un terme supplémentaire d'erreur de sortie au sens des moindres carrés. L'approche calcule à la fois par LS-RBF-FD une estimation du champ de polluant modélisé par (3.3)-(3.4) et une estimation de la source en conditions stationnaires ($\partial_t c = 0$) en intégrant un terme d'erreur quadratique entre les concentrations prédites par un modèle de sortie et les mesures. Pour rappel, le modèle stationnaire de dispersion considéré est :

$$\sum_{i=1}^3 \mathbf{U}_i \partial_i c - \sum_{i=1}^3 \partial_i (\mathbf{K}_i \partial_i) c = s. \quad (4.27)$$

avec les conditions aux limites,

$$\begin{aligned} \sum_{i=1}^3 \mathbf{n}_i \partial_i c(x) &= 0 & \mathbf{x} \in \Gamma_N \\ c(x) &= 0 & \mathbf{x} \in \Gamma_D \end{aligned} \quad (4.28)$$

\mathbf{U} , \mathbf{K} étant les champs météorologiques tridimensionnels de vitesse du vent et de diffusion, \mathbf{n} , le vecteur normal à une surface et Γ_N et Γ_D , les sous domaines de bords sur lesquels s'appliquent respectivement des conditions de Neumann et de Dirichlet. Les opérateurs du modèle stationnaire sont discrétisés par LS-RBF-FD à la manière de ce qui a été présenté en section 2.4.6. Le problème s'écrit alors comme une méthode LS-RBF-FD "augmentée" de la manière suivante :

$$\min_{\mathbf{c}, \mathbf{s}} \frac{1}{N_s} \|\mathbf{H}\mathbf{c} - \mathbf{y}^m\|^2 + \frac{1}{N_{\Omega}} \|\mathbf{D}_{\Omega}\mathbf{c} - \mathbf{\Psi}_s \mathbf{s}\|^2 + \frac{1}{N_{\partial\Omega}} \|\mathbf{D}_{\partial\Omega}\mathbf{c}\|^2 \quad (4.29)$$

où N_s , N_{Ω} , $N_{\partial\Omega}$ désignent respectivement le nombre de mesures, de points de domaine intérieur et de points de bord de domaine dans le domaine d'évaluation $Y \subset \Omega$, $\mathbf{c} \in \mathbb{R}^{N_x}$, $\mathbf{s} \in \mathbb{R}^{N_y}$, les vecteurs

de concentration et de terme source défini dans \mathbb{R}^{N_x} , \mathbf{y}^m , le vecteur des valeurs des mesures, \mathbf{H} , l'opérateur de sortie, $\Psi_s \in \mathbb{R}^{N_y \times N_x}$ une matrice d'interpolation de la source dans le domaine Ω . Enfin, \mathbf{D}_Ω et $\mathbf{D}_{\partial\Omega}$ désignent respectivement les opérateurs RBF-FD de l'ADPDE à l'intérieur du domaine et aux bords (prise en compte des conditions aux limites).

De manière équivalente, le problème de minimisation se met sous la forme :

$$\min_{\mathbf{c}, \mathbf{s}} = \left\| \mathbf{V} \begin{pmatrix} \mathbf{c} \\ \mathbf{s} \end{pmatrix} - \mathbf{W} \right\|^2 \quad (4.30)$$

avec,

$$\mathbf{V} = \begin{pmatrix} \frac{1}{\sqrt{N_s}} \mathbf{H} & \mathbf{0}_{N_s \times N_x} \\ \frac{1}{\sqrt{N_\Omega}} \mathbf{D}_\Omega & -\frac{1}{\sqrt{N_\Omega}} \Psi_s \\ \frac{1}{\sqrt{N_{\partial\Omega}}} \mathbf{D}_{\partial\Omega} & \mathbf{0}_{N_{\partial\Omega} \times N_x} \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \frac{1}{\sqrt{N_s}} \mathbf{y}^m \\ \mathbf{0}_{(N_\Omega + N_{\partial\Omega}) \times 1} \end{pmatrix} \quad (4.31)$$

Il est possible de régulariser le problème avec l'ajout d'une pénalité d'initialisation comme ci-dessous,

$$\min_{\mathbf{c}, \mathbf{s}} \frac{1}{N_s} \|\mathbf{H}\mathbf{c} - \mathbf{y}^m\|^2 + \frac{1}{N_\Omega} \|\mathbf{D}_\Omega \mathbf{c} - \Psi_s \mathbf{s}\|^2 + \frac{1}{N_{\partial\Omega}} \|\mathbf{D}_{\partial\Omega} \mathbf{c}\|^2 + \frac{d}{N_x} \|\mathbf{s} - \mathbf{s}_0\|^2 \quad (4.32)$$

avec d , un coefficient de régularisation fixé a priori. dans ce cas, le problème se formalise de la manière suivante :

$$\mathbf{V} = \begin{pmatrix} \frac{1}{\sqrt{N_s}} \mathbf{H} & \mathbf{0}_{N_s \times N_x} \\ \frac{1}{\sqrt{N_\Omega}} \mathbf{D}_\Omega & -\frac{1}{\sqrt{N_\Omega}} \Psi_s \\ \frac{1}{\sqrt{N_{\partial\Omega}}} \mathbf{D}_{\partial\Omega} & \mathbf{0}_{N_{\partial\Omega} \times N_x} \\ \mathbf{0} & \mathbf{I}_{N_x, N_x} \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \frac{1}{\sqrt{N_s}} \mathbf{y}^m \\ \mathbf{0}_{(N_\Omega + N_{\partial\Omega}) \times 1} \\ \mathbf{s}_0 \end{pmatrix} \quad (4.33)$$

avec $\mathbf{s}_0 = \mathbf{0}_{N_x \times 1}$ dans le cas d'une initialisation inconnue de la source.

4.4.2 Résultats

Avant de présenter quelques résultats de notre méthode STE-PIML, précisons les conditions appliquées pour construire le champ de mesures ainsi que la paramétrisation retenue pour les sources. Les mesures sont construites avec l'introduction d'un bruit blanc additif gaussien appliqué à une sortie modèle de référence \mathbf{y}^0 en utilisant le modèle de mesures présenté en (4.25). La méthode est évaluée dans plusieurs cas avec des domaines de simulation tridimensionnels.

Le premier cas considère un scénario à météorologie uniforme et sans obstacles. Dans ce scénario, la sortie de référence \mathbf{y}^0 est obtenue dans les mêmes conditions (nombre de nœuds compris) que le modèle direct utilisé pour résoudre le problème inverse ce qui permet d'étudier un scénario sans bruit de modèle.

Le deuxième cas considère un scénario avec météorologie hétérogène et obstacles dans les mêmes conditions que le cas de test avec obstacles étudié en 4.3. Comme dans cette étude précédente, le champ de concentration de référence \mathbf{y}^0 , intervenant dans la construction des mesures simulées (4.25), est issu d'une simulation LS-RBF-FD réalisée sur un domaine de calcul haute résolution (où $N_x = 15648$ et $N_y = 50000$) tandis que le modèle LS-RBF-FD intervenant dans la procédure STE-PIML (4.32) est réduit à peu de points de calcul (avec $N_x = 2057$ et $N_y = 15000$) ce qui introduit un bruit de modèle. Les termes sources considérés sont construits en se basant sur une paramétrisation gaussienne des sources déjà discutée en (4.26). Les sources considérées peuvent être multiples, comme présenté par la suite.

Cas 1 : sans obstacles, météo uniforme

Dans ce cas d'étude, nous considérons le domaine sans obstacles de dimensions $L_x = 250 \text{ m} \times L_y = 250 \text{ m} \times L_z = 25 \text{ m}$ montré en 3.3.5 et en Figure 3.10a. 3 sources distinctes étalées de manière gaussienne comme en (4.26) de paramètres $(a_s, \sigma) = (1 \mu\text{gs}^{-1}\text{m}^{-3}, 10 \text{ m})$ situées à deux mètres de haut et illustrées en Figure 4.14b sont à l'origine du phénomène de pollution atmosphérique (Figure 4.14b).

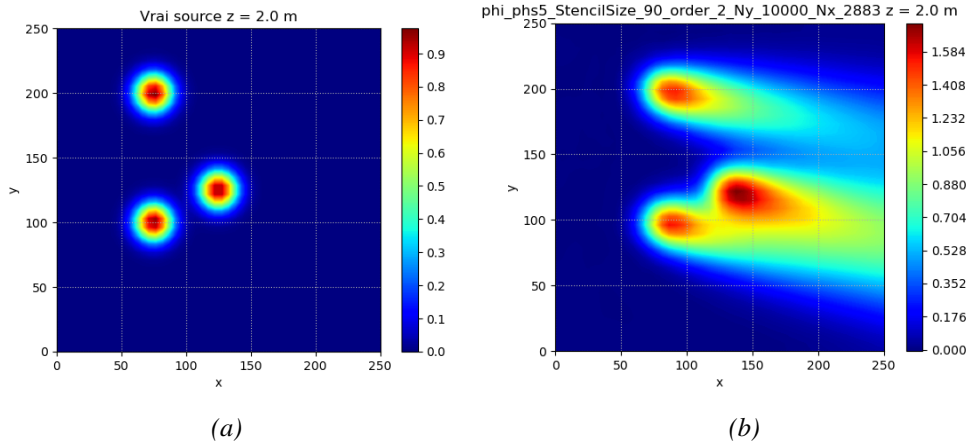


FIGURE 4.14 – (a) Sources véritables. (b) Champ de pollution stationnaire résultant des trois sources. Sources situées à un mètre de hauteur.

Les performances d'inversion d'abord avec autant de capteurs que de points de calcul et sans bruit de mesure ($n_s = 2883, \sigma^o = 0, 2$) permettent une inversion de grande qualité comme en Figures 4.15a et 4.16a. La Figure 4.15c montre dans ce cas que l'estimation des sources est très précise. En effet, au voisinage des sources véritables, l'erreur relative ne dépasse pas les 4 % et se situe majoritairement à 0,3 %. Les erreurs relatives d'estimation du champ de concentration sont également très faibles (Figure 4.16c) et n'excèdent pas 0,7 %, les erreurs relatives atteignant pour la plupart environ 0,07 %.

Nous avons sélectionné un scénario d'estimation à $n_s = 300$ mesures avec un bruit élevé de mesures de $\sigma^o = 0, 2$. Les 300 mesures ayant été générées en sous-échantillonnant les points d'approximations de la solution du champ de concentration de référence montrée en Figure 4.14b. L'estimation perd alors en qualité comme visible en Figure 4.17a. La source estimée saisit encore les positions de source d'une bonne manière, en revanche leur étalement et leur débit sont mal estimés par rapport à la "vraie source" (Figure 4.17b). Notons aussi dans l'estimation l'apparition de sources "fausses" d'amplitudes allant jusqu'à 0,15. Ces artefacts d'estimation sont causés par le bruit exercé sur les mesures et leur faible nombre. Les erreurs d'estimations réalisées au voisinage de la vraie source atteignent environ 50 et 100 % d'erreur (Figure 4.17c). Cependant le graphique d'erreur en Figure 4.17c est à mettre en perspective avec les illustrations adjacentes à cause des erreurs sur l'estimation des positions et de l'étalement des sources. Une régularisation (d) importante a également été ajoutée à la technique STE-PIML (4.32) dans ce scénario ($d = 1, 2$) en donnant un vecteur d'initialisation s_0 nul. Cela a pour effet de diminuer l'amplitude des débits de source estimés. L'estimation du champ de concentration se détériore à l'instar de celle du terme source. Les ordres de grandeur du champ de concentration sont cependant bien restitués sur la Figure 4.18a. L'erreur relative observée au voisinage du champ de pollution véritable (Figure 4.18c) atteint pour la plupart des points d'affichage environ 40 %. Des erreurs élevées sont cependant présentes en amont des panaches.

Au niveau de l'implémentation, la méthode exploite directement le coté creux du modèle direct LS-RBF-FD, permettant des calculs rapides et aussi un passage à l'échelle pour traiter un

problème de grande dimension. L'ajout de mesures bruitées ralentit la convergence vers l'estimation notamment lorsque la régularisation d choisie est faible. L'ajout d'une régularisation permet de mieux poser le problème et d'accélérer la convergence. Ici, d est nul pour le cas $n_s = 2883$, $\sigma^o = 0$ et $d = 1, 2$ dans le cas à 300 mesures bruitées ($n_s = 300$, $\sigma^o = 0, 2$).

300 mesures semblent être le cas limite pour une estimation de source et de champ de concentration convenable avec des mesures dispersées et bruitées pour ce scénario à 3 sources sans obstacles et à $N_x = 2883$ points d'approximation.

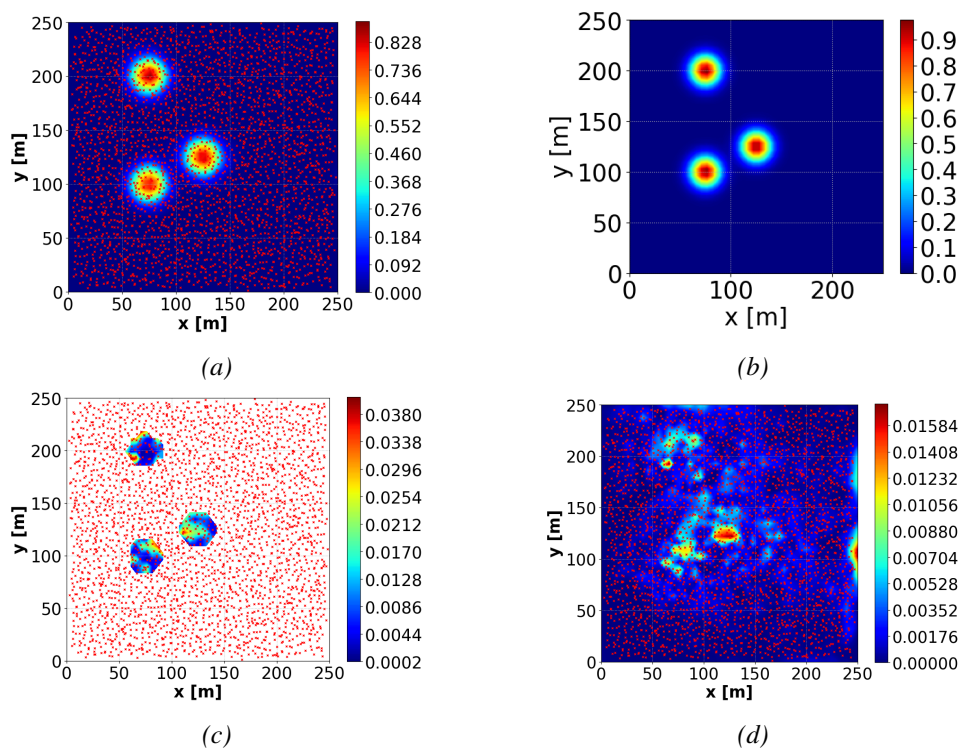


FIGURE 4.15 – (a) : Source estimée avec $n_s = 2883$ capteurs et un bruit de mesure gaussien nul ($\sigma^o = 0.0$), (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

Les erreurs RRSE telles que définies en (4.11) calculées pour tous les scénarios de tests de la méthode STE-PIML présentée sont compilées en Tableau 4.3. RRSE_X, RRSE_interp y sont respectivement les erreurs RRSE d'estimation du terme source sur tout le domaine de simulation X et sur une tranche 2D interpolée à 2 mètres de hauteur. Les erreurs sont calculées de la même manière pour les estimations de concentrations avec RRSE_X_conc et RRSE_interp_conc. Nous remarquons que la méthode est assez sensible au bruit. En effet, les erreurs RRSE pour le cas ($n_s = 300$, $\sigma^o = 0, 2$) sont environ doublées par rapport au cas ($n_s = 300$, $\sigma^o = 0$) pour l'estimation de la source comme pour l'estimation de concentration. Cette baisse de performance est à mettre en perspective car les estimations restent de qualité y compris en présence de bruit. Nous pouvons par exemple deviner assez précisément la localisation des sources dans l'estimation du champ de pollution illustrée en Figure 4.18a et les ordres de grandeur du champ de concentration demeurent réalistes par rapport à la référence. Comme prévu, la précision d'estimation du terme source diminue avec la diminution du nombre de mesure. Ainsi l'erreur sur la source RRSE_X augmente sensiblement de $n_s = 2883$ capteurs à $n_s = 300$. Avec RRSE_X passant de 0,04 à 0,36. L'ajout de bruit d'amplitude faible augmente aussi l'erreur. Un bruit blanc gaussien d'amplitude 0,05 sur les mesures augmente l'erreur RRSE_X à 0,62. Enfin, l'ajout d'un fort bruit blanc de mesures à 0,2 d'amplitude augmente encore la RRSE_X à 0,89. Les erreurs RRSE calculées

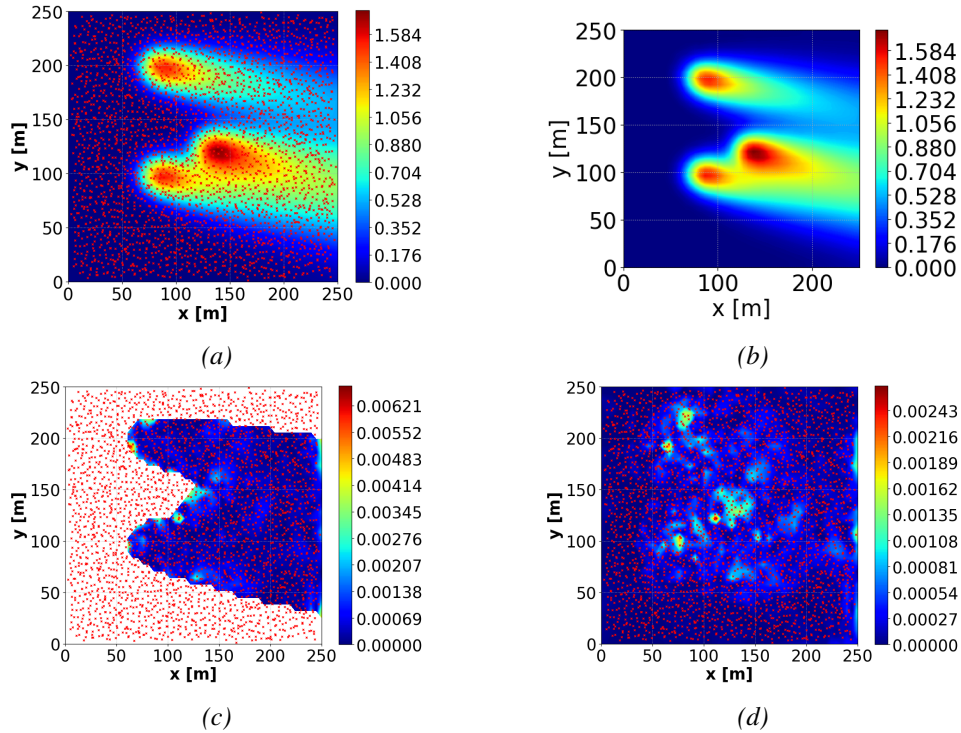


FIGURE 4.16 – (a) : Champ de concentration estimé avec $n_s = 2883$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0$, (b) : vrai champ de concentration (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

sur le champ de source estimé interpolé sur une tranche de visualisation 2D (RRSE_interp) sont similaires, ce qui montre que l'interpolation utilisée pour l'affichage ne fausse pas les résultats.

TABLEAU 4.3 – Erreurs sur les estimations de source et de concentrations par STE-PIML. n_s : nombre de mesures, bruit : écart type du bruit gaussien de mesure (σ^o), RRSE_X et RRSE_interp : erreurs respectives d'estimation de source sur le domaine de simulation X et sur une tranche 2D interpolée à une hauteur de 2,14 mètres de haut. RRSE_X_conc et RRSE_interp_conc erreurs relatives concernant les estimations de champ de concentration.

ns	bruit	RRSE_X	RRSE_interp	RRSE_X_conc	RRSE_interp_conc
2883	0.0	0.04	0.10	0.00	0.00
300	0.0	0.36	0.35	0.16	0.15
300	0.05	0.62	0.53	0.08	0.08
300	0.2	0.89	0.84	0.29	0.26

Influence du placement de mesures à basse altitude : Nous avons testé la méthode STE-PIML avec les mêmes conditions que précédemment en changeant uniquement la position des capteurs. Par défaut dispersés dans tout le domaine, nous avons disposé des capteurs uniquement entre 0 et 9 mètres de haut. Les RRSE pour les estimations de source et les estimations de champ de concentration sont listées dans le Tableau 4.4. Nous remarquons que des capteurs dispersés à basse altitude n'améliorent pas les estimations de source et de concentration dans ce scénario. Un placement astucieux comme préconisé par Kouichi et al. en aval des panaches des pollutions améliorera probablement les estimations [Kouichi2016]. A la vue de ces résultats, des mesures

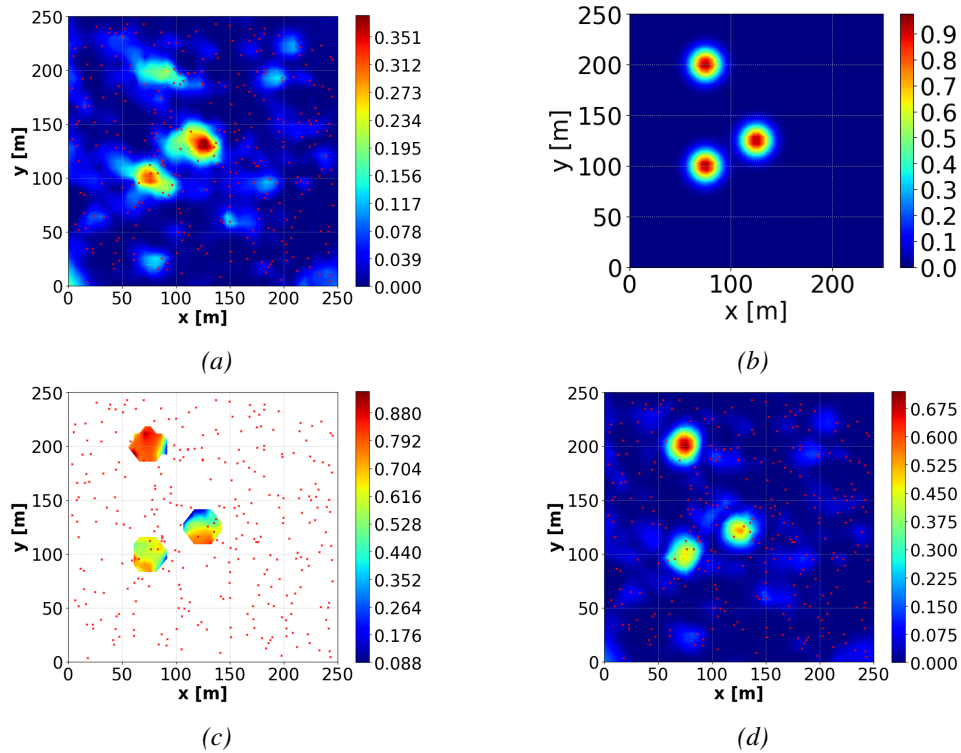


FIGURE 4.17 – (a) : Source estimée avec $n_s = 300$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0.2$, (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

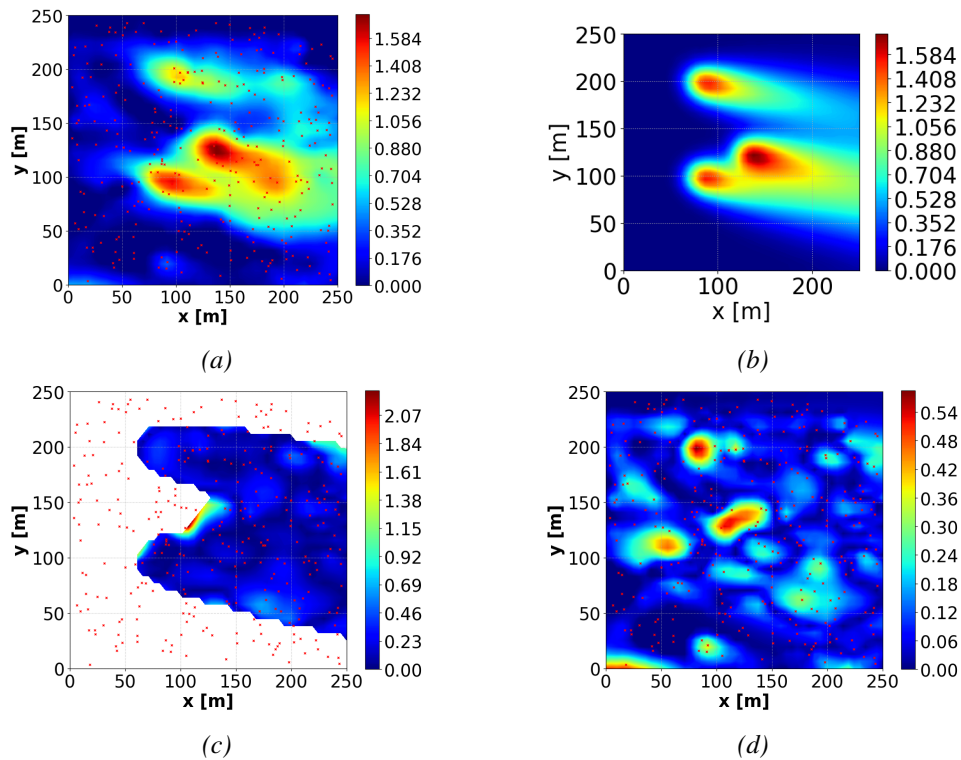


FIGURE 4.18 – (a) : Champ de concentration estimé avec $n_s = 300$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,2$, (b) : vrai champ de concentration (c) : Erreur relative. (d) erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

localisées de manière aléatoire à basse altitude, par exemple sur des murs d'immeuble et au sol, pourront fournir une estimation de qualité similaire à des mesures dont la disposition est plus étalée à des altitudes basses et hautes.

TABLEAU 4.4 – Erreurs sur les estimations de source et de concentrations par STE-PIML. Mesures dispersées à basse altitude.

ns	bruit	RRSE_X	RRSE_interp	RRSE_X_conc	rrse_interp_conc
521	0,0	0,23	0,14	0.10	0.01
300	0.0	0,34	0,25	0.15	0.04
300	0,05	0,71	0,57	0.31	0.11
300	0.2	0,93	0,88	0.45	0,26

Cas 2 : domaine jouet avec obstacles

La méthode STE-PIML est également testée avec un domaine plus contraint que notre premier cas et à météorologie hétérogène pour en vérifier la performance. L'étude est à rapprocher de la partie 4.3, nous utilisons des mesures, des conditions météorologiques et un domaine de simulation rigoureusement identiques à ceux utilisés en partie 4.3 dans le cadre de l'étude de la STE-adjointe sur un cas jouet avec obstacles. Comme en 4.3 six scénarios d'estimations sont évalués en considérant divers bruits de mesures ($\sigma^o = 0,002, 0,05, 0,2$), et divers nombre de mesures (1500, 300, 200 et 50) distribués sur tout le domaine tridimensionnel de simulation (Ω).

Avec $n_s = 1500$ mesures et un faible bruit ($\sigma^o = 0,002$), l'estimation de la source est de qualité (Figure 4.19a). Les erreurs de localisation latérales sur le plan $x - y$ sont nulles. L'étalement de la source est relativement bien estimé, même si on peut observer de trop grandes amplitudes au bord du bâtiment de gauche (le bord situé en $x = 425$). Remarquons que l'amplitude de source est sous-estimée de moitié avec un maximum de source atteint à environ 0,52 unités de source. Conséquemment les erreurs relatives observées au voisinage de la "vraie source" atteignent environ 40 % (Figure 4.19c). Par la suite, le champ de concentration est reconstruit avec grande fidélité (Figure 4.20a). Le champ estimé ne restitue cependant pas parfaitement les ordres de grandeur du "vrai" champ de concentration visible en Figure 4.20b. En effet le champ de concentration estimé atteint un maximum à environ 4,2 unités de concentration contre environ 5,2 unités pour le "vrai" champ de concentration. Cet écart s'explique par le fait que dans le cas présent (cas du domaine jouet avec obstacles) tout comme en partie 4.3, les mesures simulées proviennent d'un simulateur haute résolution à $N_x = 15000$ points d'approximation tandis que le modèle utilisé dans la procédure d'estimation est très réduit avec seulement $N_x = 2057$ points de d'approximation. L'erreur relative au voisinage du panache de pollution (Figure 4.20c) est comprise entre 0 et 400 % environ. Ces grandes erreurs étant commises en bord du panache. Au centre, du panache, les erreurs relatives se situent entre 15 et 30 %. Le champ est globalement bien estimé dans ces conditions.

Avec $n_s = 200$ mesures et bruit de mesures élevé ($\sigma^o = 0,2$), l'estimation de la source perd en qualité (Figure 4.21a). Des erreurs de localisation latérales de la source sont observées et s'élèvent respectivement dans le sens de l'axe x et de l'axe y à $(\Delta x, \Delta y) \approx (3\text{m}, 4\text{m})$. Notons également la présence de "fausses sources" identifiées à tort par la méthode d'estimation. Leurs amplitudes peuvent atteindre 0,13 unités de source. Les erreurs relatives observées au voisinage de la source véritable atteignent environ 60 %. Le panache de polluants estimé (Figure 4.22a) permet de deviner une position de source et restitue d'une assez bonne manière les ordres de grandeur du champ de concentration véritable mesurée illustré en Figure 4.22b, bien que le champ de concentration estimé sous-estime les valeurs véritables de concentration. De plus, le champ de concentration estimé est bruité et est par exemple non nul là où il devrait l'être. Ce phénomène est du à la

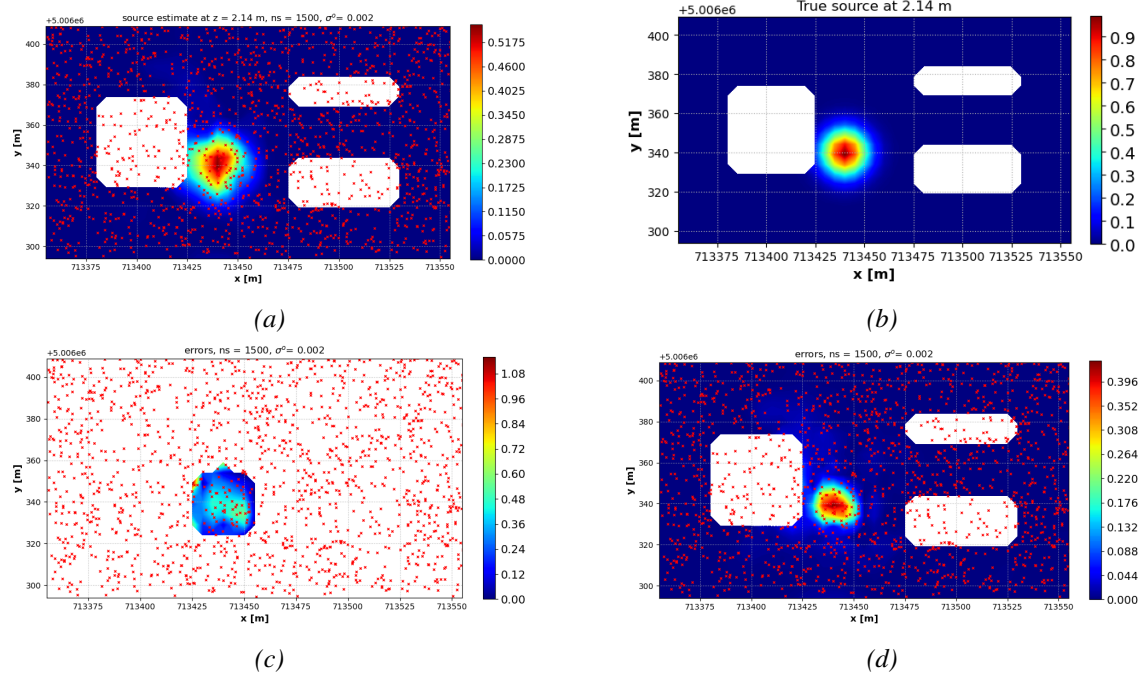


FIGURE 4.19 – (a) : Source estimée avec $n_s = 1500$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,002$, (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

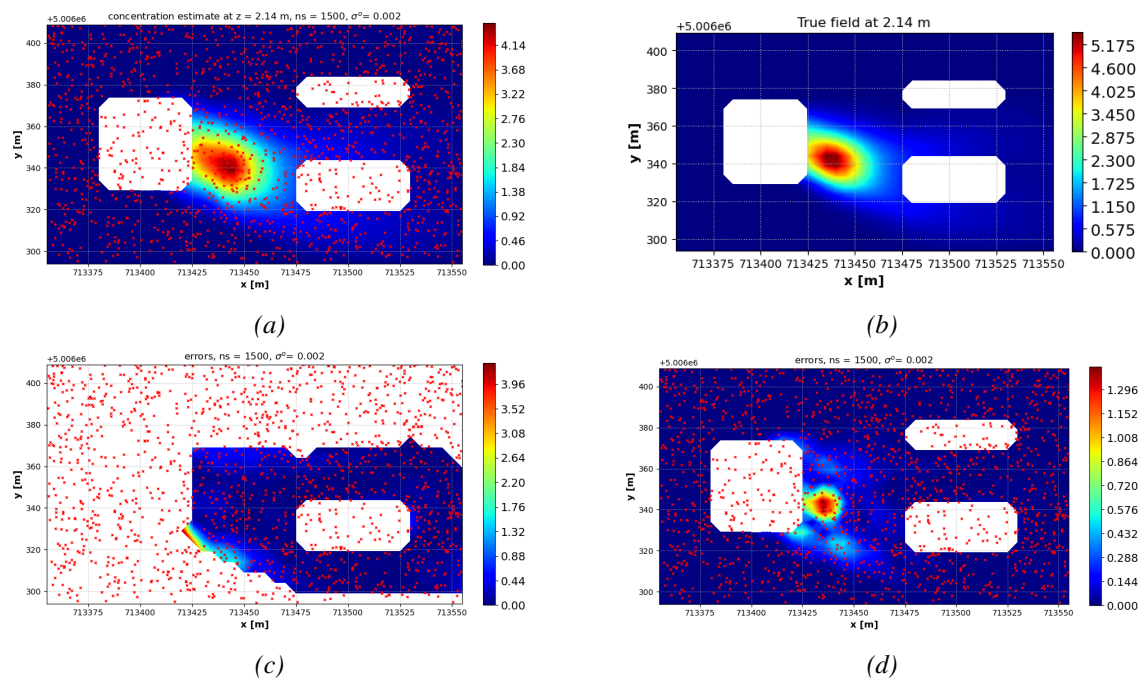


FIGURE 4.20 – (a) : Champ de concentration estimé avec $n_s = 1500$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,002$, (b) : vrai champ de concentration (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

présence de "fausses" sources de pollution identifiées dans tout le domaine. L'erreur relative au voisinage du panache de pollution est comprise entre 0 et 420 % (Figure 4.22c). Les plus grandes erreurs sont effectuées au bord du panache là où les valeurs de concentration véritables sont peu élevées. A l'intérieur du panache l'erreur est d'environ 50 %. En aval du panache où les valeurs de concentration sont faibles, les erreurs peuvent atteindre 130 %.

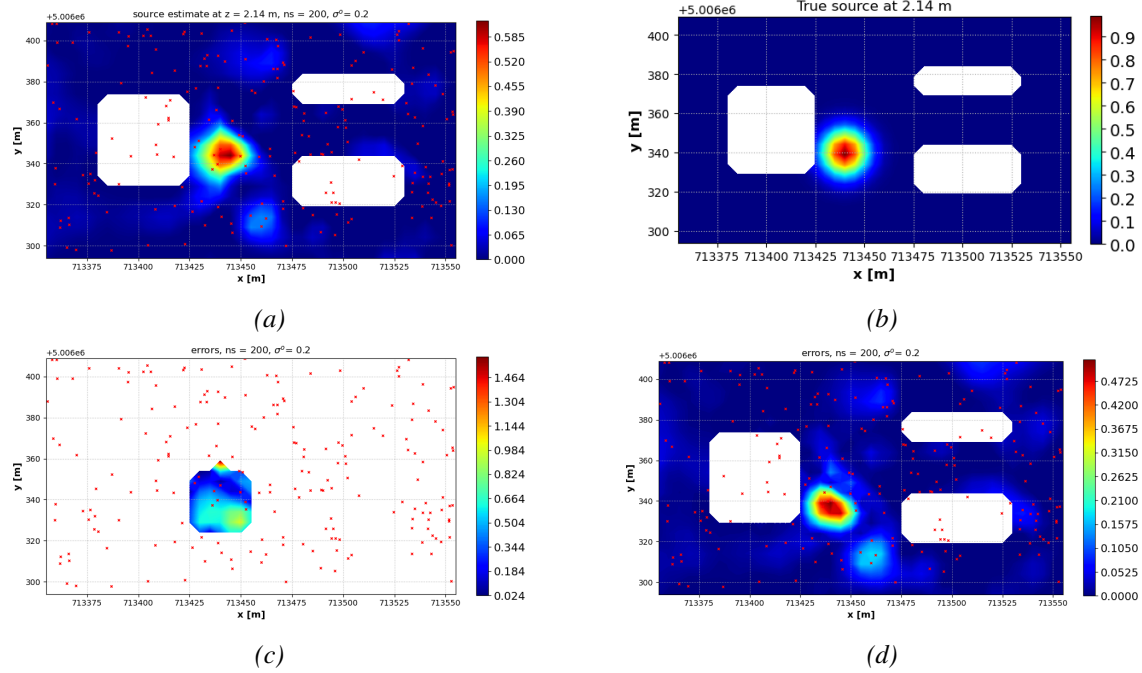


FIGURE 4.21 – (a) : Source estimée avec $n_s = 200$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0, 2$, (b) : vraie source (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut.

Les performances de la méthode STE-PIML au sens de la RRSE pour le scénario de test à 3 obstacles et à météorologie hétérogène sont visibles en Tableau 4.5. $RRSE_X$, $RRSE_interp$ y sont respectivement les erreurs RRSE d'estimation du terme source sur tout le domaine de simulation X et sur une tranche 2D interpolée à 2 mètres de hauteur. Les erreurs RRSE sont calculées de la même manière pour les estimations de concentrations avec $RRSE_X_conc$ et $RRSE_interp_conc$.

Remarquons que la présente méthode STE-PIML montre de meilleures performances au sens de la RRSE que la méthode STE-adjointe (discutée en 4.3) notamment pour les cas ($n_s = 1500, \sigma^o = 0, 002$) et pour le cas ($n_s = 200, \sigma^o = 0, 2$). La méthode STE-adjointe estime cependant mieux la source dans X dans le cas ($n_s = 200, \sigma^o = 0, 2$) ($RRSE_X^{adjointe} = 0, 85$ contre $RRSE_X^{PIML} = 1, 42$) tandis que tous les autres champs sont mieux estimés par la méthode STE-PIML. Dans ce cas d'étude précis nous avons remarqué que les estimations de source par STE-PIML ont une tendance à étaler le terme source verticalement contrairement à la méthode STE-adjointe qui confine le terme source à une basse hauteur. Cela explique la mesure d'une RRSE élevée de source estimée par la méthode STE-PIML dans X dans ce cas. Nous pensons que la méthode STE-adjointe souffre moins de ces erreurs grâce à la contrainte forte imposée par l'usage d'un lagrangien, en revanche, cette contrainte pourrait être trop forte et ainsi expliquer ses moindres performances dans les estimations des champs de source et de concentration interpolés à 2,14 mètres de hauteur ainsi que dans l'estimation du champ de concentration sur le domaine discrétisé de simulation X .

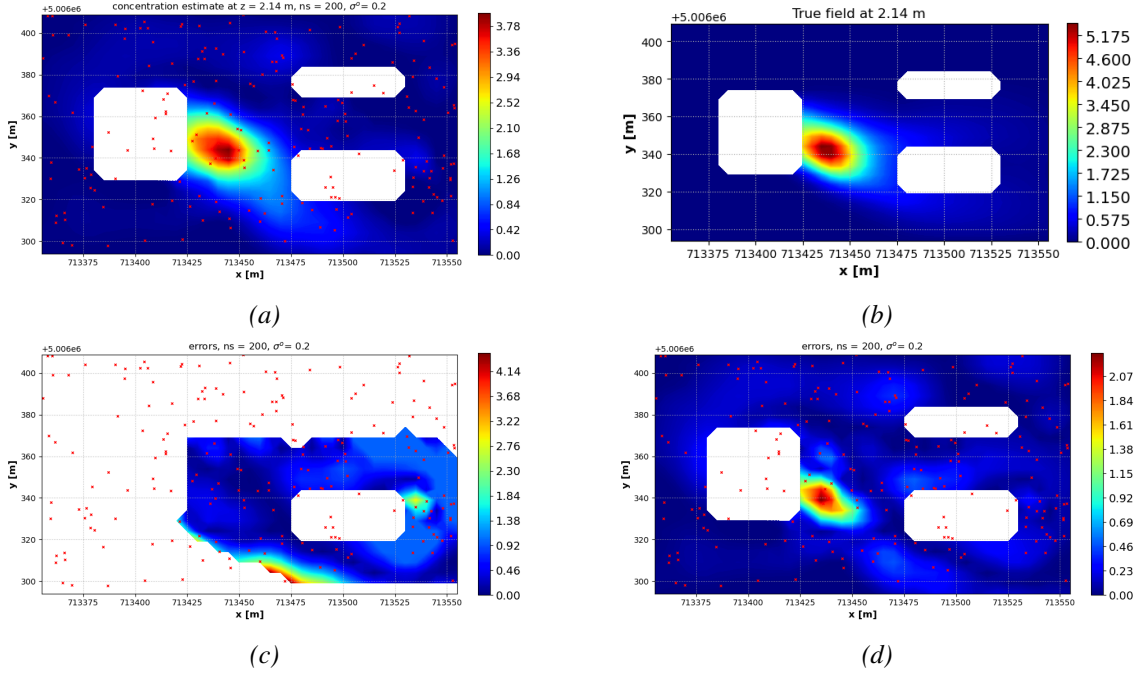


FIGURE 4.22 – (a) : Champ de concentration estimé avec $n_s = 200$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0.002$, (b) : vrai champ de concentration (c) : erreur relative, (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Résultats affichés sur une tranche à 2,14 mètres de haut.

4.4.3 Conclusions

L'application de la méthode PIML sur les cas 3D précédents montre une précision élevée avec peu d'artefacts et une robustesse au bruit remarquable. La méthode exploite tout à fait le côté creux du modèle LS-RBF-FD, elle est ainsi très rapide, avec un temps d'exécution proche du modèle direct LS-RBF-FD (3.13), voire moindre si le problème de STE est fortement régularisé. Pour le cas jouet avec 3 obstacles abordé précédemment, la phase d'estimation dure environ 20 secondes contre 4 minutes en utilisant la méthode adjointe montrée en 4.3.

Cette méthode pourrait servir d'initialisation à des méthodes réputées comme plus précises, comme les méthodes adjointes ou le filtre de Kalman et ses variantes. Bien que nous n'ayons pas eu le temps de la tester dans cette thèse, il serait possible d'adapter la méthode dans le cas instationnaire, en considérant la variable temps comme une dimension spatiale ou bien en intégrant le modèle LS-RBF-FD implicite en temps (3.5) dans la formulation PIML, de la manière suivante

$$\begin{aligned}
 \min_{\mathbf{c}(t), \mathbf{s}(t)} \frac{1}{TN_s} \sum_{t=1}^T \|\mathbf{H}\mathbf{c}(t) - \mathbf{y}^m(t)\|^2 + \frac{1}{(T-1)N_\Omega} \sum_{t=1}^{T-1} \left\| \left(\frac{\mathbf{D}_0}{dt} + \mathbf{D}_\Omega \right) \mathbf{c}(t+1) - \frac{\mathbf{D}_0}{dt} \mathbf{c}(t) - \mathbf{\Psi}_s \mathbf{s}(t) \right\|^2 \\
 + \frac{1}{(T-1)N_{\partial\Omega}} \sum_{t=1}^{T-1} \|\mathbf{D}_{\partial\Omega} \mathbf{c}(t+1)\|^2.
 \end{aligned} \tag{4.34}$$

Dans ce cas, il s'agit de déterminer la suite temporelle discrète, définie sur un intervalle de T périodes d'échantillonnage dt , d'estimés de la concentration de polluant $\mathbf{c}(t)$ et de la source $\mathbf{s}(t)$. Chacun de ces estimés temporels résulte d'une approximation RBF-FD. Ce problème peut être résolu efficacement par des algorithmes d'optimisation quadratique de grande taille exploitant le caractère creux [Byrd1995].

Une autre méthode pour appliquer la méthode STE-PIML pour des phénomènes en régime

TABLEAU 4.5 – Erreurs d’estimation de terme source et de champ de concentration par la méthode STE-PIML, cas jouet avec obstacles. ns : nombre de mesure, bruit : écart type du bruit gaussien de mesure (σ^o), $RRSE_X$ et $RRSE_interp$: erreurs respectives d’estimation de source sur le domaine de simulation X et sur une tranche 2D interpolée à une hauteur de 2,14 mètres de haut. $RRSE_X_conc$ et $RRSE_interp_conc$ erreurs concernant les estimations de champ de concentration.

ns	bruit	RRSE_X	RRSE_interp	RRSE_X_conc	RRSE_interp_conc
1500	0,002	0,45	0,40	0,06	0,17
200	0,002	0,57	0,43	0,18	0,23
200	0,2	1,42	0,57	0,47	0,33
300	0,002	0,55	0,45	0,18	0,23
300	0,05	0,69	0,47	0,21	0,25
300	0,2	1,85	0,61	0,44	0,30
50	0,2	0,81	0,82	0,43	0,55
70	0,2	1,09	0,85	0,52	0,58

transitoire serait d’abandonner le schéma temporel récursif (3.5) pour traiter le temps comme "une dimension comme une autre" dans le problème des moindres carrés posé par la méthode STE-PIML. Une méthode de référence existe d’ailleurs déjà pour intégrer des EDP discrétisées en temps et en espace : l’Intégration Temporelle Directe (DTI) [Dokainish1989].

Par ailleurs, dans l’esprit de la méthode STE-PIML, nous pourrions appliquer un réseau de neurones informé par la physique pour résoudre directement un problème inverse (méthode Inverse-PINN) en condition stationnaire et aussi en régime transitoire comme réalisé par exemple par Garay et al. dans un récent travail où les auteurs ont estimé un champ de vélocité et des paramètres d’entrées d’un modèle élaboré de mécanique des fluides [Garay2023]. Il sera pertinent de développer une telle approche pour notre problème de STE et de comparer les résultats face à notre méthode STE-PIML basée sur la méthode LS-RBF-FD. Les approches Inverse-PINN sont prometteuses.

4.5 Filtre de Kalman avec modèle transitoire LS-RBF-FD

4.5.1 Introduction

Nous avons testé une estimation multi-sources sur un domaine carré 3D sans obstacles (Domaine \mathcal{D}_1) avec météorologie uniforme et avec le modèle de dispersion d’ADPDE (3.3) considéré dans un régime transitoire sur une fenêtre de temps de 0 à 35 secondes par pas de temps de 5 secondes ($dt = 5s$) soit un vecteur temps \mathbf{t} correspondant à $(0, 5, \dots, 35s)$ contenant N_t termes. Les conditions aux limites sont de type Neumann sur toutes les faces sauf en amont du panache de pollution sur la face de côté gauche du domaine. La solution transitoire présentée en Figure 4.23 converge par ailleurs vers la solution stationnaire visible présentée en Figure 4.14 (qui par ailleurs a été générée avec N_x points d’approximation identiques). Nous utilisons trois sources gaussiennes localisées dans le domaine conformément à la figure 4.14a. Nous avons déjà utilisé ce domaine de simulation avec ces termes source pour réaliser des tests de STE en régime stationnaire en utilisant la méthode "STE-PIML" en partie 4.4. Nous choisissons volontairement un nombre élevé de points

de calcul pour ne pas introduire de biais dans les inversions et pour avoir une base commune avec la méthode "STE-PIML".

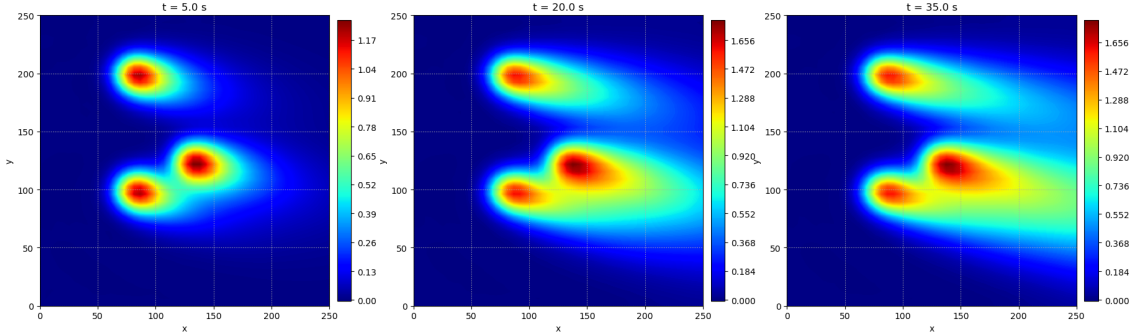


FIGURE 4.23 – Solution stationnaire montrée ici avec à trois différents instants. La solution est approximée en 2883 points $N_x = 2883$, $N_y = 10000$.

4.5.2 Mise en œuvre du filtre de Kalman

Détaillons ici la manière dont est construit le filtre de Kalman utilisé. Le filtre de Kalman nécessite un modèle physique spatio-temporel dont l'expression explicite est possible. A ces fins nous utilisons la forme explicite du schéma d'Euler implicite présenté dans le paragraphe 3.5.1 en utilisant les relations (3.63)-(3.65). Pour rappel,

$$\mathbf{c}(t+1) = \mathbf{A}\mathbf{c}(t) + \mathbf{B}\mathbf{s}(t), \quad (4.35)$$

avec,

$$\mathbf{A} = \left(\left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right] + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega} \right)^{-1} \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right]^T \frac{\mathbf{D}_{0\Omega}}{dt} \quad (4.36)$$

$$\mathbf{B} = \left(\left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right]^T \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right] + \mathbf{D}_{\partial\Omega}^T \mathbf{D}_{\partial\Omega} \right)^{-1} \left[\frac{\mathbf{D}_{0\Omega}}{dt} + \mathbf{D}_{\Omega} \right]^T \quad (4.37)$$

Dans ce cas d'étude la source est considérée comme stationnaire en débit comme en position ainsi,

$$\mathbf{s}(t+1) = \mathbf{s}(t) \quad (4.38)$$

de plus les sorties de modèles ramenées dans l'espace des observations s'écrivent,

$$\mathbf{y}(t) = \mathbf{H}\mathbf{c}(t) \quad (4.39)$$

où \mathbf{H} est un opérateur linéaire de mesures tel qu'un sélecteur de position. Le système représenté par les équations (4.35), (4.38) et (4.39) peut s'écrire sous la forme,

$$\theta(t) = \begin{pmatrix} \mathbf{c}(t) \\ \mathbf{s}(t) \end{pmatrix}, \quad \mathbf{M}(t) = \begin{pmatrix} \mathbf{A} & \mathbf{B}\mathbf{D}_{0\Omega} \\ \emptyset & \emptyset \end{pmatrix}. \quad (4.40)$$

Ces équations sont directement utilisées en entrée d'un filtre de Kalman exposé en section 2.5.3. Les autres paramètres tel que le bruit ou les initialisations données sont listés dans le Tableau 4.7. Dans ce travail, nous n'avons pas traité de dépendance du temps dans l'hypothèse d'une météorologie dépendante du temps ou encore dans l'hypothèse de mesures mobiles. Mais les matrices \mathbf{A} , \mathbf{B} et \mathbf{H} et $\mathbf{D}_{0\Omega}$ pourraient dépendre du temps dans un autre contexte.

4.5.3 Résultats

Dans cette expérience, les mesures simulées sont construites de manière analogue à (4.25) en considérant cette fois une dépendance du temps telle que pour chaque temps t_k ($k = 1, \dots, N_t$),

$$\begin{aligned} \mathbf{y}_k^m &= \mathbf{y}_k^0 + \boldsymbol{\epsilon}_k^o, \\ \boldsymbol{\epsilon}_k^o &\sim \mathcal{N}(0, \mathbf{R}), \end{aligned} \quad (4.41)$$

avec \mathbf{y}_k^m de taille N_s le vecteur contenant les mesures simulées pour N_s capteurs au temps t_k , \mathbf{y}_k^0 de taille N_s le vecteur contenant la simulation du champ de concentration de référence échantillonnée aux N_s coordonnées de mesures au temps t_k (la Figure 4.23 l'illustrant pour les temps 5, 20 et 35 s) et enfin $\boldsymbol{\epsilon}_k^o$ le vecteur des erreurs de taille N_s également. Les erreurs sont générées par une loi normale multivariée pour chaque temps t_k de covariance choisie comme constante et diagonale $\mathbf{R} = \sigma^o \mathbf{I}$ où \mathbf{I}_{N_s} est la matrice identité de taille $(N_s \times N_s)$. Par la suite nous nous intéresserons aux performances de la méthode sur la dernière estimation d'état au temps $t = 35$ s.

Les performances d'inversion, d'abord avec autant de capteurs que de points de calcul et sans bruit de mesure ($n_s = 2883$, $\sigma^o = 0, 2$), permettent une inversion de grande qualité comme en Figure 4.24a et Figure 4.25a. 2883 capteurs permettent d'estimer fidèlement le vecteur d'état θ de taille $n_\theta = 5766$ qui agglomère le vecteur de concentration (\mathbf{c}) et le terme source (\mathbf{s}) tel que $\theta = (\mathbf{c}, \mathbf{s})^T$. La Figure 4.24c montre que l'estimation des sources est très précise avec des erreurs relatives faibles. En effet, au voisinage des sources véritables, l'erreur relative ne dépasse pas 12 % et se situe majoritairement à entre 0 et 5 %. Les plus grandes erreurs sont faites en périphérie des sources pour des faibles valeurs proches de 0. Les erreurs relatives d'estimation du champ de concentration sont également très faibles (Figure 4.25c) et n'excèdent pas 10 %, les erreurs relatives atteignant pour la plupart environ 1 à 2 %. Les plus grandes erreurs au sein des panaches sont commises en aval près des bords du domaine et aux extrémités des panaches de pollution, là où les valeurs de champs de concentration sont peu élevées.

Comme dans la partie précédente avec la méthode STE-PIML en régime stationnaire (Partie 4.4) nous avons testé un cas avec $n_s = 300$ capteurs avec un bruit élevé de mesures de $\sigma^o = 0, 2$. Les 300 mesures ont été générées en sous-échantillonnant les N_x points d'approximations de la simulation du champ de concentration de référence montrée en Figure 4.23 pour chaque pas de temps dt avant d'appliquer le bruit de mesures tel que décrit par (4.41). L'estimation perd alors en qualité comme visible en Figure 4.26a. Le filtre de Kalman utilisé dans ce cas conduit à des erreurs d'estimations de position de source ainsi qu'à une erreur dans l'estimation de l'étalement des 3 véritables sources illustrées en Figure 4.26b. Notons également que la source excentrée localisée en $(x, y) = (75, 200)$ est sous-estimée par rapport aux deux autres sources. Des sources estimées à tort sont présentes. La présence de ces "fausses sources" est causée le bruit de mesure important. Le débit de ces sources atteint des amplitudes d'environ 0,1 unités de source.

Les erreurs relatives au voisinage des sources véritables (Figure 4.26c) atteignent environ 60 à 80 %. Ce graphique d'erreur est cependant à mettre en perspectives avec les Figures 4.26a et 4.26c du fait des erreurs dans la localisation et l'estimation des sources. Le champ de concentration est assez bien estimé pour pouvoir deviner la position des sources véritables (Figure 4.27a) bien que la présence de bruit de mesure conduit à tort à l'estimation de zones polluées, en particulier en amont des panaches.

Nous pouvons également remarquer que notre filtre de Kalman (KF) a tendance à surestimer le champ de concentration notamment au niveau de la source centrale (Figure 4.27b). Les ordres de grandeurs du champ de concentration estimé sont proches du champ de référence. Le champ estimé de concentration affiche ainsi des valeurs maximales des concentration vers environ $1, 7 \mu\text{gs}^{-1}\text{m}^{-3}$ ce qui est très proche de la valeur maximale observée pour le champ de référence à environ $1, 6 \mu\text{gs}^{-1}\text{m}^{-3}$ (Figure 4.27a). Les erreurs relatives d'estimation du champ de concentration sont globalement faibles et se situent pour la plupart entre 0 et 40 % (Figure 4.27c). Cela prouve que la méthode est robuste au bruit.

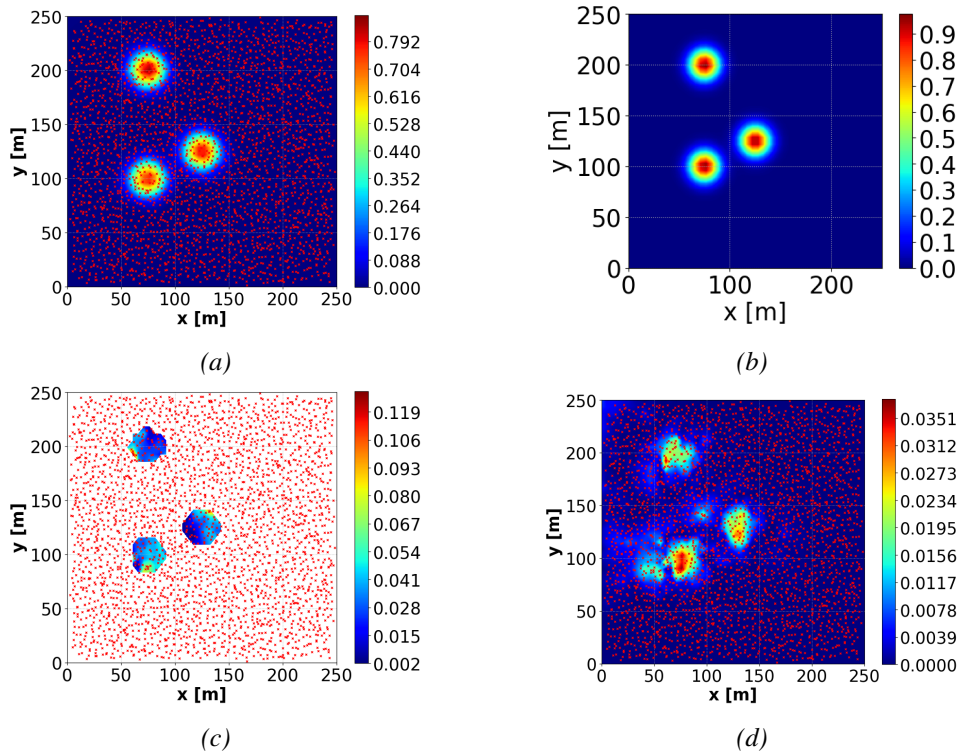


FIGURE 4.24 – (a) : Source estimée avec $n_s = 2883$ capteurs et un bruit de mesure gaussien nul ($\sigma^o = 0.0$), (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2,14 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

Pour réaliser la meilleure estimation possible avec le KF pour un tel niveau de bruit, nous avons utilisé une régularisation importante par le choix de la matrice d'erreur d'initialisation \mathbf{P}_0 . Nous avons imposé $\mathbf{P}_0 = b\mathbf{I}_{N_\theta}$ avec $b = 2$ et \mathbf{I}_{N_θ} une matrice identité de taille $(N_\theta \times N_\theta)$ où $N_\theta = \text{card}(\theta)$.

Nous pouvons conclure que le KF testé ici montre une grande robustesse au bruit et de bonnes performances pour estimer le champ de concentration. Les sources sont assez bien localisées en présence de bruit de mesure mais leurs estimations sont étalées par rapport à la réalité et leurs ordres de grandeurs ne sont pas fidèlement capturés.

Un facteur d'amélioration des estimations retournées par le KF pourrait être un meilleur choix des covariances d'erreur d'initialisation et de covariance d'erreur de modèle contenues dans les matrices \mathbf{P} et \mathbf{Q} dans la formulation du KF. Ici \mathbf{P} et \mathbf{Q} sont considérées comme diagonales et remplies d'un coefficient propre (respectivement b et q) de valeur unique pour tout élément diagonal. Pour le cas présent, le passage d'une covariance d'erreur d'initialisation b à une valeur voisine de 2 permet de réaliser des estimations correctes avec un nombre limité de mesures. Une covariance diagonale de bruit de modèle q comprise entre 0 et 10^{-2} conduit à des résultats similaires. Cela laisse penser que le choix du coefficient diagonal de \mathbf{Q} (q), fixé à 10^{-3} pour les expériences présentées, est raisonnable.

Les performances du filtre de Kalman face à celles de la méthode STE-PIML 4.4.2 sont cependant à relativiser car la méthode STE-PIML exposée considère un régime permanent alors que le filtre de Kalman (KF) ici implémenté considère un régime transitoire. Or nous avons fait l'hypothèse dans les présents tests d'une stationnarité du terme source, le test présent bénéficie donc de plus d'observations par rapport au cas de test utilisant la méthode STE-PIML. Un défaut notable du filtre de Kalman reste sa difficulté à traiter des problèmes de très grande dimension à cause du calcul explicite de la matrice de gain (2.119) très exigeant en calcul et en mémoire. Ce

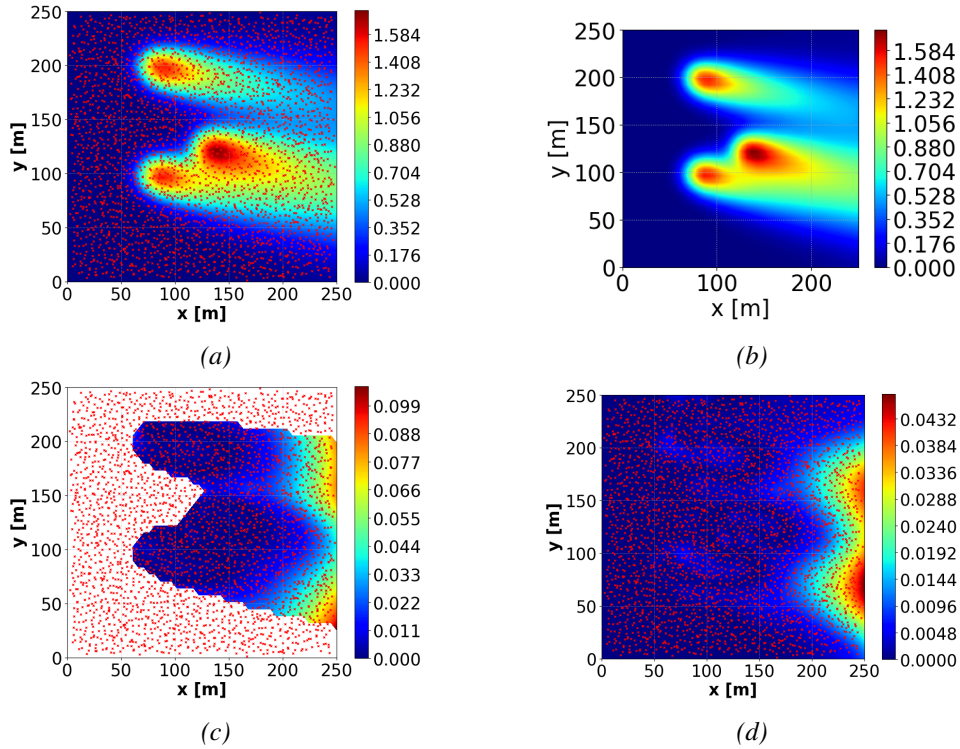


FIGURE 4.25 – (a) : Champ de concentration estimé avec $n_s = 2883$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0$, (b) : vrai champ de concentration (c) : erreur relative, (d) erreur absolue. Résultats affichés sur une tranche à 2 mètres de haut.

point est discuté dans [Asch2016, p. 51]. Les filtres d'ensemble de Kalman ont la réputation de pouvoir traiter des problèmes de grandes dimensions tout comme les méthodes variationnelles.

Malheureusement il n'a pas été possible d'obtenir de solution satisfaisante avec un filtre d'ensemble de Kalman (EnKF) pour ce cas de test. Nous avons pourtant utilisé les mêmes paramètres d'initialisation, d'incertitude de mesures, et d'erreur d'initialisation que pour le filtre de Kalman (KF) ainsi que des grandes tailles d'ensemble à 100, 200 ou 600 termes. Par ailleurs l'implémentation actuelle de L'EnKF dans notre code ne permet pas le traitement de problèmes de grande dimension. Il faudra pour cela utiliser une implémentation différente. Cependant pour un domaine de calcul petit, le KF retourne des résultats satisfaisants au contraire de notre EnKF. L'implémentation actuelle de l'EnKF est peut-être à mettre en cause. De plus l'EnKF utilisé ici (EnKF stochastique) n'utilise pas de fenêtre temporelle prenant en compte les mesures antérieures, et n'utilise pas les techniques dites de localisation et d'inflation aidant à rendre les ensembles de la méthode statistiquement représentatifs.

Les paramètres de Filtre de Kalman et du modèle d'inversion LS-RBF-FD ayant servi à générer les résultats discutés dans cette section sont visibles dans le Tableau 4.7.

Les performances de notre implémentation du KF basé sur notre modèle de dispersion LS-RBF-FD sont visibles en Tableau 4.5. $RRSE_X$, $RRSE_interp$ y sont respectivement les erreurs RRSE d'estimation du terme source sur tout le domaine de simulation X et sur une tranche 2D interpolée à 2 mètres de hauteur. Les erreurs RRSE sont calculées de la même manière pour les estimations de concentrations avec $RRSE_X_conc$ et $RRSE_interp_conc$.

Remarquons les bonnes performances du KF présent face à la méthode STE-PIML (dont les performances sont visibles en Tableau 4.6). Par exemple pour $N_s = 300$ capteurs avec écart type de bruit de mesure $\sigma^o = 0,2$, les RRSE obtenues avec le KF sur le terme source sont $(RRSE_X^{KF}, RRSE_interp^{KF}) = (0,89; 0,74)$ tandis que les erreurs sur le champ de concentration atteignent $(RRSE_X_conc^{KF}, RRSE_interp_conc^{KF}) = (0,30; 0,27)$. Ces résultats sont

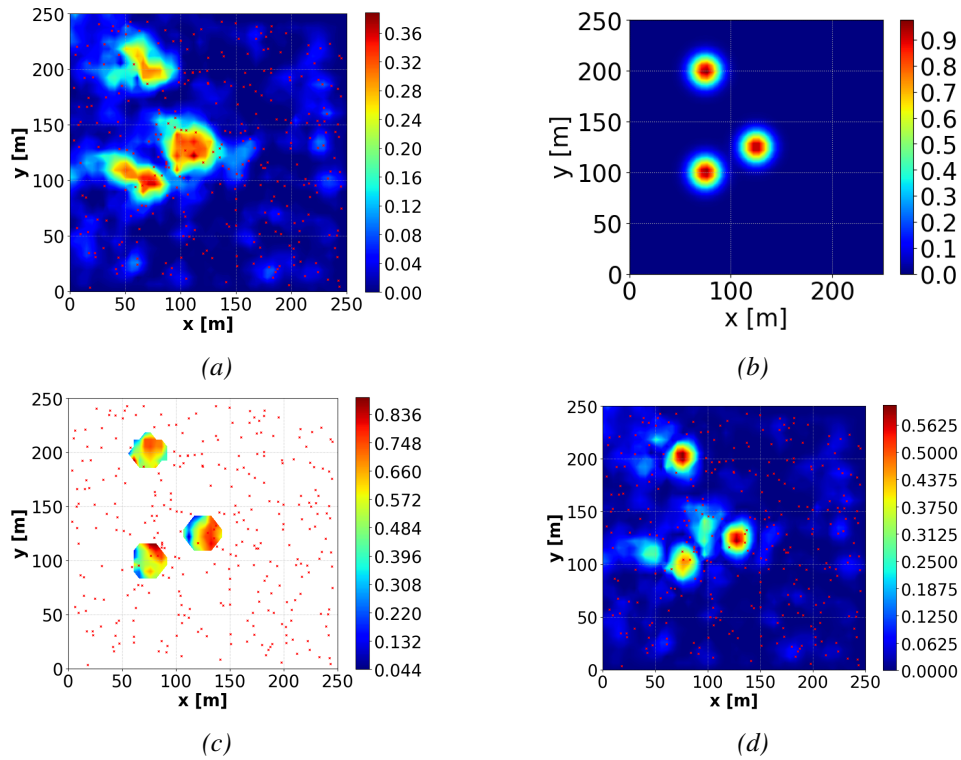


FIGURE 4.26 – (a) : Source estimée avec $n_s = 300$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,2$, (b) : vraie source, (c) : erreur relative (à multiplier par 100 pour obtenir un pourcentage d'erreur), (d) : erreur absolue. Résultats affichés sur une tranche à 2 mètres de haut. Les croix rouges montrent les coordonnées des capteurs projetées sur le plan d'affichage.

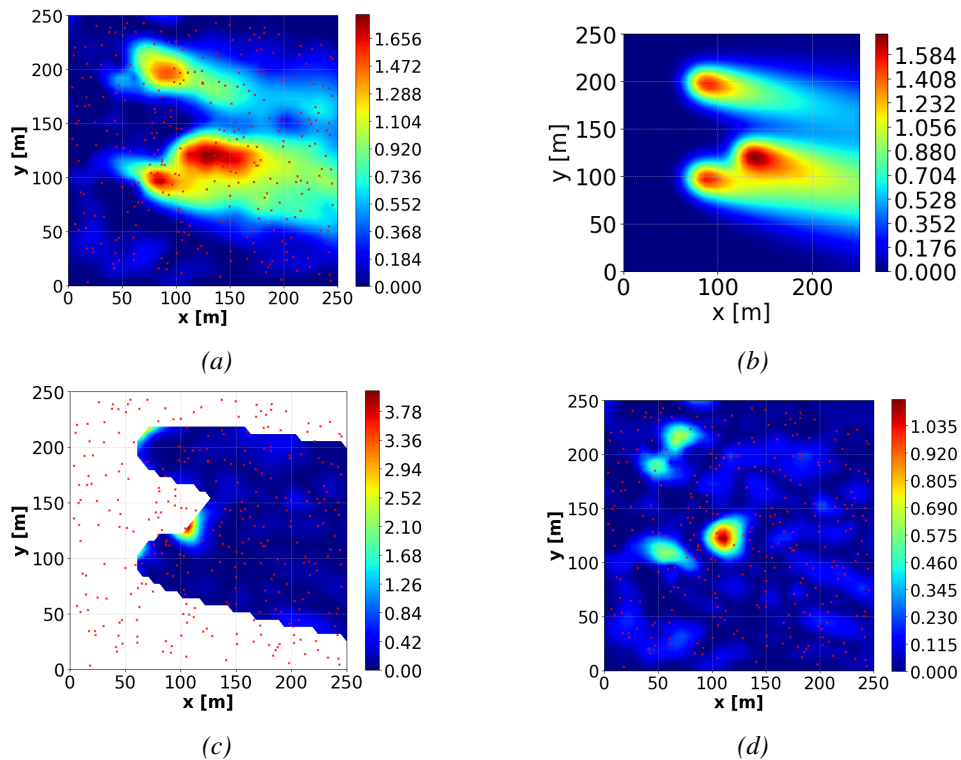


FIGURE 4.27 – (a) : Champ de concentration estimé avec $n_s = 300$ capteurs et un bruit de mesure gaussien d'écart type $\sigma^o = 0,2$, (b) : vrai champ de concentration (c) : erreur relative, (d) erreur absolue. Résultats affichés sur une tranche à 2 mètres de haut.

légèrement meilleurs qu'avec la méthode STE-PIML, où $(RRSE_X^{PIML}, RRSE_interp^{PIML}) = (0, 89; 0, 84)$ et $(RRSE_X_conc^{PIML}, RRSE_interp_conc^{PIML}) = (0, 29; 0, 26)$. Rappelons que la méthode STE-PIML traite des sources stationnaires (comme le KF dans cet exemple) mais considère un régime stationnaire de concentration en polluants.

TABLEAU 4.6 – Erreurs sur l'estimation de terme source et de champ de concentration par Filtrage de Kalman, cas jouet sans obstacles à météorologie uniforme. ns : nombre de mesure, bruit : écart type du bruit gaussien de mesure (σ^o), $RRSE_X$ et $RRSE_interp$: erreurs respectives d'estimation de source sur le domaine de simulation X et sur une tranche 2D interpolée à une hauteur de 2 mètres de haut. $RRSE_X_conc$ et $RRSE_interp_conc$ erreurs relatives concernant les estimations de champ de concentration.

ns	bruit	RRSE_X	RRSE_interp	RRSE_X_conc	RSSE_interp_conc
2883	0,0	0,05	0,12	0,02	0,02
300	0,0	0,85	0,73	0,28	0,25
300	0,05	0,85	0,73	0,28	0,25
300	0,2	0,89	0,74	0,30	0,27
600	0,0	0,76	0,60	0,18	0,13

TABLEAU 4.7 – Paramètres de l'expérience d'ETS avec un Filtre de Kalman

Paramètre	Réglage
Bruit	nul
Initialisation	nulle
t_0, dt, t_f (temps de début, pas de temps, temps final de mesure)	(0, 5, 35)
b (coefficient de variance d'erreur d'initialisation)	3 ou 100
\mathbf{P}_0 (matrice de covariance d'erreur d'initialisation)	$b\mathbf{I}_{N_t}$
r (coefficient de variance d'erreur de mesure)	1
\mathbf{R} (matrice de covariance d'erreur de mesure)	$r\mathbf{I}_{N_s}$
q (coefficient de variance d'erreur de modèle)	0
\mathbf{Q} (matrice de covariance d'erreur de modèle)	$q\mathbf{I}_{N_t}$
Paramètres modèle LS-RBF-FD	
N_x (nombre de nœuds d'approximation)	2883
N_y (nombre de nœuds d'évaluation)	10000
ϕ (type de RBF)	phs5
n (taille de stencil)	90
ordre polynôme (polynôme de régularisation pour RBF-FD)	2

4.5.4 Conclusions

Dans cette partie nous avons présenté des résultats d'une application du filtre de Kalman. Un champ de concentration 3D en régime transitoire a permis de construire des mesures en temps et en espace tandis que le terme source véritable a été considéré comme stationnaire. A notre connaissance, ce travail constitue la première application d'un filtre de Kalman dans le cadre de méthode LS-RBF-FD notamment dans le contexte de la pollution atmosphérique. La méthode est efficace et très robuste au bruit dans les cas testés. Nous obtenons notamment avec l'usage de $n_s = 300$ mesures et un écart type de bruit de mesure de $\sigma^o = 0,2 \mu\text{gm}^{-3}$ des RRSE comparables à la méthode STE-PIML (Partie 4.4.2) sur les estimations du terme source et du champ de concentration.

4.6 Conclusion

Le chapitre 4 a présenté plusieurs méthodes d'Estimation de Terme Source :

- Une première étude, qui a précédé le développement des méthodes RBF-FD, a été présentée dans le paragraphe 4.2 : **il s'agissait de comparer une approche STE de type 3D-Var avec une méthode basée sur un réseau de neurones dans un contexte unidimensionnel et stationnaire. C'est la une première contribution de la thèse en matière de méthodologie d'estimation de source.** Nous avons montré qu'un réseau de neurones est capable d'estimer précisément un terme source avec une précision plus élevée qu'avec la méthode 3D-Var non convexe proposée.

Cependant, nous avons observé quelques limitations opérationnelles, notamment le besoin de réaliser des entraînements de réseau de neurones à chaque nouvelle condition météorologique, qui s'avèrent particulièrement coûteux. De plus le passage à l'échelle en 2D et en 3D n'est pas simple car les vecteurs de champs météorologiques donnés en entrée auraient alors des milliers d'éléments, ce qui nécessiterait certainement le passage à des architectures neuronales plus complexes (de type réseaux de neurones à convolution). Nous avons choisi de ne pas explorer cette voie dans la thèse. Nous notons toutefois que des approximations basées sur les données à base notamment de Réseaux de neurones ont été utilisées pour approximer des modèles directs de météorologie et de dispersion dans l'esprit de quelques recherches existantes [Ding2021, Mendil2022].

Dans le cas des réseaux de neurones, l'introduction de modèles physiques pour réaliser des méthodes PIML (Physics-Informed Machine Learning) améliore grandement leur convergence tout en réduisant le nombre de données d'entraînement nécessaire [Raissi2019]. L'approche PIML à base de réseaux de neurones présentée dans cette dernière référence nous a inspiré pour proposer l'approche STE-PIML utilisant des approximants RBF-FD au lieu d'un réseau de neurones.

- Dans un deuxième temps, nous avons proposé dans le paragraphe 4.3 **une méthode adjointe discrète basée sur un modèle de dispersion LS-RBF-FD tridimensionnel (méthode STE-adjointe). A notre connaissance, l'approche n'existe pas dans la littérature.** La méthode de STE a été testée sur un cas d'étude stationnaire comportant 3 bâtiments. Des mesures simulées extraites de sorties de modèle à résolution fine ont été utilisées. Des mesures abondantes en nombre ont été utilisées pour estimer un vecteur source défini sur tout le domaine. Nous avons observé que la méthode de STE présentée était capable de localiser une source quelconque. Les intensités de la source sont bien restituées avec un grand nombre de mesures bruitées. Cependant, un faible nombre de mesures bruitées conduit à des erreurs

de localisation de la source et des erreurs d'estimation dans son intensité. Bien que les ordres de grandeurs de la source et sa position restent correctement estimés. Ce travail a été soumis à l'European Control Conference (ECC) 2024.

- Dans un troisième temps, nous avons présenté dans le paragraphe 4.4 **une méthode originale STE-PIML basée sur la méthode LS-RBF-FD pour estimer un vecteur source quelconque** défini dans le domaine de simulation $X \subset \Omega$. La méthode montre une robustesse remarquable face au bruit et exploite le caractère creux des opérateurs LS-RBF-FD. Elle est ainsi très rapide en temps d'exécution et retourne des estimations d'un ordre de grandeur proche du modèle direct numérique basé sur les LS-RBF-FD. Bien que n'ayant pas eu le temps de la tester, nous avons formalisé une méthode d'estimation STE-PIML en régime transitoire décrite par (4.34) en utilisant le schéma d'Euler implicite proposée dans le paragraphe 3.5. La méthode de STE-PIML basée sur la LS-RBF-FD est déjà très prometteuse en précision, en robustesse au bruit et en temps de calcul.

Des développements récents dans le domaine du PIML ont montré qu'il est intéressant de traiter des problèmes inverses avec des réseaux de neurones informés de la physique [Lu2021]. Avec cette méthode, Lu et al. ont étudié des problèmes d'estimation de paramètres de l'équation de Lorenz [Lu2021]. D'autres auteurs ont utilisé la méthode inverse-PINN (Inverse Physics-Informed Neural Network) pour estimer des champs de vitesse et des inconnues de modèles de mécanique des fluides [Garay2023]. Il est donc pertinent de poursuivre après cette thèse le développement d'une telle approche pour réaliser des STE et la comparer avec la méthode STE-PIML actuelle basée sur les LS-RBF-FD.

- Enfin, nous avons mis en œuvre avec succès au paragraphe 4.5 **un filtre de Kalman utilisant notre modèle LS-RBF-FD en régime transitoire dans le contexte d'une STE tridimensionnelle à source stationnaire**. La méthode a montré des capacités intéressantes à traiter le problème instationnaire, bien que dans un contexte opérationnel, la mise à jour de la matrice de gain ne soit pas raisonnable pour des domaines de simulation trop denses (par exemple, 20000 points d'approximation et plus ou encore un vecteur d'état de 20000 points et plus). Un filtre d'ensemble de Kalman a également été testé mais sans succès à ce stade. Au vu des résultats encourageants dans la littérature, les STE basées sur le filtre d'Ensemble de Kalman conservent de l'intérêt en évitant le calcul de modèle adjoint, pour des problèmes de grandes dimensions. Nous observons cependant que son implémentation est complexe.

5

Conclusion générale et perspectives

Ce chapitre résume le travail effectué et discute des perspectives de la thèse. Une partie des perspectives de travail reposent sur la consolidation et l'approfondissement des méthodes déjà étudiées dans cette thèse. Le placement optimal de capteurs fixes ou mobiles constitue une perspective intéressante faisant l'objet de nombreux travaux. En particulier, le développement d'estimation de sources à l'aide de capteurs mobiles (de type drones) combiné à des procédures de navigation optimale pour garantir la meilleure estimation possible est une perspective importante.

Sommaire

5.1 Conclusion générale	138
5.2 Perspectives en lien avec les méthodes abordées	139
5.3 Perspectives et contraintes d'application en temps réel et de prise en compte de capteurs mobiles	141
5.4 Placement optimal de capteurs et estimation avec capteurs mobiles	143

Cette partie rassemble d'une part les conclusions et enseignements tirés des travaux entrepris dans le cadre de la thèse, et d'autre part des perspectives de travaux futurs pour améliorer les méthodes existantes et démarrer de nouvelles études.

5.1 Conclusion générale

Dans le contexte de la pollution atmosphérique urbaine chronique, nous avons évoqué dans l'introduction l'intérêt d'identifier les sources de pollution les plus intenses de manière rapide et précise pour la prévention de telles émissions de polluants. Des systèmes de chauffage et des véhicules dysfonctionnels ou vieillissants peuvent s'avérer particulièrement émissifs en polluants tels que les particules fines et les oxydes d'azote. Par ailleurs, les capteurs de polluants, notamment de particules fines se démocratisent. Il est alors intéressant de considérer l'estimation de source de pollution en s'appuyant sur des mesures abondantes (crowdsensing).

Dans ce contexte, cette thèse a proposé différentes méthodes de modélisation directes et inverses basées sur un modèle eulérien régi par une équation aux dérivées partielles d'advection-diffusion :

- Dans le Chapitre 2, un état de l'art a été réalisé. Il porte d'une part sur la dispersion atmosphérique de polluants en particulier en milieu urbain, d'autre part sur les méthodes connues dans le domaine de la simulation et de l'assimilation de données pour estimer un état ou un terme de source. Nous avons choisi dans cette thèse de mettre l'accent sur les approches RBF-FD, qui sont encore moins connues que les méthodes traditionnelles de type volumes finis, différences finies ou éléments finis. Nous pouvons retenir quelques enseignements de cet état de l'art :
 - Les obstacles qui constituent un environnement urbain exercent une grande influence sur les circulations d'air ayant lieu dans ce milieu. La modélisation des circulations d'air à partir des équations représentatives de Navier-Stokes et du transport de polluants par résolution de la conservation de la masse ne sont pas réaliste : il y a besoin de simplifier les équations de Navier-Stokes et de conservation de la masse qui régissent la dispersion de polluants.
 - Les équations de Navier-Stokes moyennées (RANS) pour modéliser la micro-météorologie, combinées à une équation d'Advection-Diffusion pour modéliser l'action du transport de polluants par les circulations d'air, constituent une approximation raisonnable en précision et en temps de calcul.
 - Il est possible d'estimer de manière efficace des sources de pollution avec une méthode bayésienne ou variationnelle en précalculant un champ de vent.
- Dans le Chapitre 3, nous avons exposé les contributions effectuées sur le développement d'un modèle direct de dispersion qui traite l'équation d'Advection-Diffusion (ADPDE) uniquement. Les champs de vitesse de vent et de diffusion sont calculés en amont et considérés à ce stade comme connus. Nous en tirons quelques observations :
 - La méthode de résolution d'Equation aux Dérivées Partielles des Différences Finies générées par des Fonctions de Base Radiale (RBF-FD) par collocation arrive à modéliser rapidement et efficacement des transports de polluants sur des domaines bidimensionnels avec et sans obstacles avec des données météorologiques uniformes et avec des ordres de grandeurs réalistes en 2D.
 - Les RBF-FD à collocation ne parviennent pas à modéliser correctement des phénomènes de transport 3D à météorologie réaliste. Des difficultés sont déjà rencontrées avec une météorologie uniforme sur des cas sans obstacles. Avec des obstacles nous ne parvenons pas à modéliser une dispersion de polluants réaliste.

- Le modèle LS-RBF-FD arrive à modéliser un transport de polluants en 2D et 3D, y compris en présence d'obstacles et avec une météorologie complexe. Le cas de test en 2D montre que la LS-RBF-FD est plus précise que les RBF-FD en collocation à nombre égal de point d'approximation.
- Un schéma instationnaire d'Euler a été formalisé et développé pour calculer une solution en régime transitoire. Celui-ci semble prometteur.
- Le Chapitre 4 a porté sur l'Estimation de Terme Source (STE). Des méthodes d'apprentissage, variationnelles et bayésiennes ont été testées sur des cas 1D et 3D. Nous retenons les informations principales :
 - Les approches variationnelles peuvent conduire à une convergence vers des minima locaux lorsque la paramétrisation de la source est non linéaire (problème de non-convexité) et nécessite généralement une bonne estimation initiale de la source pour donner des résultats de grande précision.
 - L'approche par réseaux de neurones en 1D générant une approximation de la fonction mesures-paramètres de source ponctuelle donne des résultats encourageants. Cependant, les données météorologiques de vitesse de vent et de diffusion constituant des paramètres d'entrée en grande dimension, particulièrement en 3D, il s'avère nécessaire de développer des architectures neuronales complexes, peu adaptées à une mise en œuvre raisonnable en temps réel. En effet, il paraît nécessaire de prévoir un ré-apprentissage de la fonction particulièrement coûteux à chaque fois que les conditions météorologiques changent.
 - Une méthode adjointe reposant sur un modèle réduit LS-RBF-FD (STE-adjointe) a été proposée et validée sur un cas 3D avec obstacles dans lequel une source quelconque a été reconstruite en présence de bruit, dans un contexte stationnaire.
 - La méthode STE-PIML basée sur le modèle de dispersion LS-RBF-FD a été proposée et testée sur des cas d'exemples dans un contexte stationnaire. Il a été possible d'estimer des sources multiples quelconques, en présence d'un bruit de mesure élevé. La méthode est robuste au bruit de mesure, et son temps d'exécution est proche de celui du modèle direct LS-RBF-FD.
 - Une approche basée sur le filtrage de Kalman dans un contexte de dispersion en régime transitoire à sources stationnaires a été testée. Les estimations produites montrent une bonne précision et sont à rapprocher de celles obtenues par la méthode STE-PIML dans un contexte de dispersion en régime stationnaire. A ce stade, une implémentation du filtre d'Ensemble de Kalman sur le même cas d'exemple n'a pas fonctionné.

5.2 Perspectives en lien avec les méthodes abordées

1. **Réalisme des mesures** : De manière générale, l'utilisation de mesures simulées par un modèle de référence comme PMSS ou de vraies mesures de terrain est indispensable pour une validation en vraie grandeur. Dans cette thèse, nous ne disposions pas de mesures réelles et nous avons donc fait le choix de retenir des mesures simulées par un modèle LS-RBF-FD plus raffiné que le modèle utilisé pour la STE, de manière à maîtriser parfaitement le bruit dans nos expériences et de nous mettre dans une situation la plus réaliste possible. Nous avons cependant réalisé avec succès quelques tests préliminaires de STE avec les méthodes STE-adjointe (4.3) et STE-PIML (4.4) sur des mesures simulées avec PMSS. Le travail de réaliser des expériences de STE avec des mesures simulées par PMSS pourra être abordé

dans des travaux futurs. D'autre part, à notre connaissance, il n'existe pas de jeu de données mesurées dans un contexte de mesures abondantes et la création d'un tel jeu de données pour réaliser des expériences de STE est encore difficile faute d'une mise en place de réseaux de capteurs adéquats.

2. **Choix des points de calcul des méthodes RBF-FD** : Concernant le domaine du modèle de transport par LS-RBF-FD, de nombreux nœuds de calcul sont utilisés à des endroits non cruciaux comme en hauteur au-dessus de la canopée urbaine et dans la zone extérieure aux obstacles, zone utilisée pour imposer au loin des conditions limites. Un travail sur une génération plus adroite du domaine pourrait considérablement diminuer le nombre de points de calcul et par conséquent accélérer le temps de calcul du modèle ainsi qu'augmenter sa résolution à basse altitude. Un meilleur placement de points de domaine pourrait être réalisé par une meilleure manipulation de la fonction de densité de nœuds de calculs, ou par l'utilisation de procédure de raffinement de nœuds de calcul. La parallélisation du code de calcul en tuiles pourrait également être intéressante car le solveur ici utilisé LSQR exploite peu les processeurs disponibles pour des grands domaines de calcul. On pourrait également se poser la question de développer un algorithme de placement optimal des points de calcul pour garantir la meilleure estimation de source possible. Ce problème rejoint la problématique du placement optimal de capteurs pour améliorer l'estimation.
3. **Amélioration de la simulation en régime transitoire** : D'autres solutions que celles utilisées dans la thèse sont envisageables pour le calcul du modèle de transport en régime transitoire. Par exemple, on pourrait utiliser l'approche proposée par Tominec et Nazarov pour inverser les systèmes obtenus à chaque pas de temps [Tominec2022]. L'idée serait d'utiliser des solveurs rapides tels que GMRES pour des EDP linéaires. La recherche d'un bon préconditionnement peut aussi favoriser la convergence rapide des solveurs. Le schéma numérique implicite d'Euler développé dans ce travail pourrait aussi être comparé avec par exemple un schéma implicite de type Runge Kutta d'ordre 2 ou supérieur. Le compromis précision/temps de calcul serait intéressant à analyser.
4. **Validation et amélioration de la STE par STE-PIML** : La méthode STE-PIML basée sur notre modèle LS-RBF-FD montre des résultats prometteurs et exploite le côté creux du modèle direct. Simple et raisonnablement rapide, cette méthode pourrait servir d'initialisation à des méthodes reconnues dans la littérature comme les méthodes adjointes ou de Kalman et ses variantes. En tout état de cause, il serait intéressant de comparer cette nouvelle approche avec les méthodes classiques. Bien que nous n'ayons pas eu le temps de l'étudier, il est possible d'adapter la méthode STE-PIML dans le cas instationnaire proposée dans le paragraphe 3.5, avec la formulation (4.34). La validation de la méthode STE-PIML doit également être conduite dans un cadre plus complexe comme un quartier urbain.
5. **Approfondissement de l'approche par filtrage de Kalman** : L'application du filtre de Kalman a prouvé qu'une méthode bayésienne, offrant l'avantage d'une actualisation de l'estimation au fil de la disponibilité des mesures est envisageable pour le cas de l'Estimation de Terme Source étudié avec un modèle réduit de type RBF-FD. Cependant l'application du filtre de Kalman s'est limitée, faute de temps, à un cas jouet. Il faut de plus noter que la mise à jour de la covariance d'erreur et de l'état est très coûteuse en mémoire et le conditionnement numérique de ce calcul se dégrade dans le cas de problèmes de grande taille. Le développement d'un filtre d'Ensemble de Kalman (EnKF) a fait l'objet d'un premier test sans succès. Les approches basées sur les filtres d'ensemble de Kalman combinent des avantages de méthodes bayésiennes avec une possibilité prometteuse de passage à grande échelle, validées dans de nombreuses applications. C'est pourquoi les approches basées sur les EnKF méritent certainement un approfondissement dans le contexte de l'estimation de pollution.
6. **Étude d'une approche PINN** : L'utilisation de modèles physiques, directs et adjoints

capables de traiter des obstacles, a un coût élevé. En vue de faciliter et d'accélérer les traitements de problèmes inverses, les approches d'apprentissage informées de la physique (Physics-Informed Neural Networks) sont prometteuses. L'avantage principal de ce genre d'approches semble être une bonne capacité d'approximation avec un faible nombre de données d'apprentissage. Il serait intéressant de comparer ce type d'approches avec la méthode STE-PIML développée dans la thèse. Les sorties du modèle LS-RBF-FD, celles d'un modèle plus abouti comme PMSS, ou encore des mesures réelles, pourraient servir à créer un jeu de données pour entraîner un réseau de neurones informé par la physique (PINN) qui approximerait le modèle de dispersion. Les approches PINN ont récemment été directement utilisées pour résoudre des problèmes inverses, y compris pour l'estimation de paramètres de modèles de mécanique des fluides [Lu2021, Garay2023]. La méthode STE-PIML proposée dans cette thèse est une méthode inverse dans laquelle il serait intéressant de remplacer les approximations RBF-FD par des réseaux de neurones (on obtient alors une méthode PINN), afin d'évaluer un éventuel accroissement de performance en matière de précision notamment. Dans l'approche PINN, la physique est notamment imposée de manière élégante par auto-différentiation. On trouvera une implémentation intéressante de ce principe dans la bibliothèque Python DeepXDE [Lu2021].

7. **Analyse comparée de performances** : Dans un effort de consolidation des résultats et de comparaison de performances en précision et temps de calcul, il serait souhaitable d'implémenter une méthode de référence pour l'estimation de source ponctuelle paramétrée (par sa position, et son débit) telle qu'une méthode adjointe [Issartel2007, Kumar2015, Singh2015, Kouichi2019]. Cela offrirait l'occasion de comparer les performances de la méthode d'inversion en utilisant le modèle direct LS-RBF-FD, en s'appuyant sur des cas réels mesurés comme la campagne FFT-07 ou MUST.

5.3 Perspectives et contraintes d'application en temps réel et de prise en compte de capteurs mobiles

Plusieurs options sont à envisager pour une stratégie d'estimation de source en temps réel. Soit les capteurs envoient l'information des mesures par technologie de type 5G à un poste central qui fait les calculs et renvoie les résultats aux appareils de mesures. Soit les calculs sont embarqués sur les appareils de mesures eux mêmes. Par exemple des capteurs "maîtres" reçoivent les informations de mesures de capteurs esclaves pour pouvoir faire les calculs. Le protocole de communication ne nécessiterait alors pas une portée plus grande que le kilomètre.

Si les capteurs sont mobiles, il faudrait également un algorithme embarqué peu coûteux afin d'optimiser en temps réel leur navigation si elle est pilotable. On peut imaginer par exemple une dizaine de drones évoluant au-dessus de la canopée urbaine en complément de capteurs installés sur des bus, des trams et d'autres capteurs fixes.

Le modèle physique est aussi une question d'intérêt ; il faut que le modèle utilisé pour la STE, dans notre travail, le modèle de transport LS-RBF-FD, soit le plus réduit possible en trouvant un compromis entre complexité (volume de données à stocker et à traiter, ainsi que temps de calcul) et précision des résultats. Il a notamment été remarqué qu'un nombre d'approximants RBF trop réduit réduisait sensiblement la précision de l'estimation de source (amplitude dégradée et phénomènes de diffusion accentués).

Dans un contexte de STE, disposer de modèles physiques rapides (à complexité algorithmique réduite) constitue un intérêt. Le modèle direct LS-RBF-FD ou bien PMSS pourraient être adaptés à ces fins pour construire un modèle appris basé sur une approche d'apprentissage informé

par la physique (PIML : Physics-Informed Machine Learning). Le modèle LS-RBF-FD pourrait également utiliser des approximations construites par apprentissage automatique dans certaines parties de sa chaîne de calcul, par exemple pour le calcul des opérateurs discrets d'advection et de diffusion tout en conservant assez de généralité dans les modèles. L'inversion du système pourrait également utiliser des algorithmes de type machine learning. Des auteurs ont par exemple déjà développé des solveurs appris deux fois plus rapides que l'algorithme GMRES [Luna2021]. Dans un contexte instationnaire avec une météorologie évolutive, les opérateurs de diffusion et d'advection devront être recalculés en opérant : une interpolation des champs météorologiques sur le domaine LS-RBF-FD grâce à une matrice d'interpolation précalculée, puis une multiplication des champs météorologiques avec les opérateurs élémentaires du modèle (laplaciens, gradients). Or cela peut prendre du temps. L'implémentation d'un code optimisé pour accélérer ces opérations constitue donc un intérêt.

Les méthodes d'observation récursives comme le filtre de Kalman ou de ses variantes comme l'EnKF, peuvent aussi être intéressantes dans un cadre opérationnel en régime transitoire car l'estimation de l'inconnu peut s'améliorer en ajoutant incrémentalement des observations évitant ainsi la réalisation d'estimation par "lot". On notera cependant qu'un estimateur par lot, comme le STE-PIML proposé dans cette thèse, peut être rendu récursif en utilisant un principe de fenêtre glissante des mesures pour aboutir à un observateur dit à horizon glissant [Michalska1995].

Enfin de manière générale, le code utilisé dans ce travail est un code Python peu parallélisé et non distribué. Sur le cas du PIML avec $N_y = 10000$, $N_x = 2883$, le calcul des matrices de poids discrétisant les opérateurs d'advection et de diffusion prennent environ 350 secondes en utilisant le code de Trever Hines [Hines2015]. Une parallélisation sur CPU basée sur la bibliothèque *joblib* a permis de descendre à environ 100 s le calcul des matrices de poids. Dans la construction des matrices de poids, le calcul des opérateurs élémentaires gradients et laplacien sont les étapes les plus coûteuses en calcul. Dans l'hypothèse d'un domaine de calcul inchangé, ces opérateurs peuvent cependant être précalculés. Le code de calcul des matrices de poids pourrait être amélioré en cherchant une alternative à l'utilisation de *joblib*. En effet cette bibliothèque Python de parallélisation CPU semble peu optimisée pour des usages sur plusieurs cœurs CPU dans notre cas. L'utilisation de solveurs mieux configurés pourrait également améliorer le temps de calcul. Le solveur pourrait en particulier bénéficier d'un préconditionnement tel que la mise à l'échelle des données d'entrées ou encore une initialisation pertinente. D'autre part le code de modélisation pourrait bénéficier facilement d'une accélération d'un facteur 2 à 5 typiquement observée dans l'usage en transposant les codes en un langage plus rapide permettant de la programmation typée et parallélisée comme Julia ou le C++.

5.4 Placement optimal de capteurs et estimation avec capteurs mobiles

Un système de surveillance de la pollution urbaine pourrait reposer dans le futur sur un réseau de capteurs fixes et des capteurs mobiles. Les capteurs fixes (capteurs institutionnels ou citoyens) devraient idéalement être installés de manière à garantir la meilleure observation possible du phénomène surveillé. L'observabilité est une propriété structurelle du système considéré. Dans notre problème de STE, elle est fondamentalement liée à la notion de sensibilité des sorties du système (les sorties en question fournissant les valeurs de concentration de polluant issues des différents capteurs) par rapport aux composantes de la source de pollution. L'objectif d'un placement optimal de capteurs fixes sera, par exemple, de maximiser la sensibilité des sorties par rapport aux composantes de la source à l'aide d'une métrique particulière. La littérature actuelle en Automatique propose des métriques fondamentalement de deux natures : des mesures directes de l'observabilité structurelle, basée par exemple sur le grammien d'observabilité du système, ou bien encore, des mesures de la performance de l'estimateur : par exemple, dans le cas du filtrage de Kalman, on pourra chercher à minimiser une norme de la matrice de variance-covariance de l'erreur d'estimation [Zhang2018].

On peut distinguer deux types de capteurs mobiles :

- Ceux dont la navigation est pilotable, par exemple des drones ou des robots terrestres mobiles.
- Et ceux dont la navigation n'est pas pilotable (par exemple, des capteurs fixés sur des véhicules particuliers des Bus ou Tramways).

Dans ce deuxième cas, la trajectoire de capteurs demeure cependant prévisible car elle utilise les transports en commun. Il serait enfin possible d'envisager des capteurs mobiles dont la trajectoire n'est a priori pas prévisible comme des capteurs fixés sur des voitures. Dans tous les cas, il sera nécessaire de disposer de la position des capteurs (par exemple, par GPS) afin de géolocaliser les mesures.

Des auteurs ont imaginé des scénarios où le pilotage de la mobilité des capteurs est exploité pour améliorer l'estimation [Šmídl2013, Georges2013]. Peu de références existent sur le traitement d'expériences à capteurs mobiles dont la navigation est non pilotable. Cependant, dans le contexte réel instationnaire, la mobilité des capteurs peut certainement enrichir l'information de concentration dans l'espace. Ainsi, pour des dynamiques de dispersion de polluants très lentes, quelques capteurs rapides pourraient fournir beaucoup de mesures en de nombreux points d'espace pour identifier des sources de pollutions avec très peu d'appareils de mesures. L'enrichissement en information apporté par la mobilité des capteurs a été observé dans des expériences en vue de générer des cartes de pollution urbaine [Hasenfratz2015].

La question de la navigation optimale de capteurs est aussi en lien avec la localisation optimale des mesures autant dans un contexte de capteurs mobiles pilotables que de capteurs fixes. Plusieurs travaux se sont intéressés à la question avec des capteurs fixes [Mons2017, Kouichi2019] et avec des capteurs mobiles [Šmídl2013, Georges2013]. La considération de mesures mobiles dans un contexte d'Estimation de Terme Source (STE) en milieu contraint, tel qu'un milieu urbain, est à notre connaissance encore peu étudiée. Il serait intéressant de comparer les performances d'une STE reposant sur de telles mesures à une STE utilisant des stations de mesures fixes.

Bibliographie

- [Alfano2020] Brigida Alfano, Luigi Barretta, Antonio Del Giudice, Saverio De Vito, Girolamo Di Francia, Elena Esposito, Fabrizio Formisano, Ettore Masera, Maria Lucia Miglietta et Tiziana Polichetti. *A Review of Low-Cost Particulate Matter Sensors from the Developers' Perspectives*. Sensors (Basel), vol. 20, no. 23, page 6819, November 2020.
- [Archambeau2004] Frédéric Archambeau, Namane Méchitoua et Marc Sakiz. *Code Saturne : A Finite Volume Code for the computation of turbulent incompressible flows - Industrial Applications*. page 63, 2004.
- [Armand2023] Patrick Armand et Jérémie Tâche. *3D modelling and simulation of the impact of wearing a mask on the dispersion of particles carrying the SARS-CoV-2 virus in a railway transport coach*. Sci Rep, vol. 13, no. 1, page 8929, June 2023. Number : 1 Publisher : Nature Publishing Group.
- [Arya1988] S. Pal Arya. Introduction to Micrometeorology. Academic Press, 1988. Google-Books-ID : tA08IQEACAAJ.
- [Asch2016] Mark Asch, Marc Bocquet et Maëlle Nodet. Data assimilation : methods, algorithms, and applications. SIAM, 2016. Pages : xviii + 306.
- [Bahlali2018] Meissam L Bahlali, Eric Dupont et Bertrand Carissimo. *A hybrid CFD RANS/Lagrangian approach to model atmospheric dispersion of pollutants in complex urban geometries*. page 16, 2018.
- [Bahlali2019] Meissam L. Bahlali, Eric Dupont et Bertrand Carissimo. *Atmospheric dispersion using a Lagrangian stochastic approach : Application to an idealized urban area under neutral and stable meteorological conditions*. Journal of Wind Engineering and Industrial Aerodynamics, vol. 193, page 103976, October 2019.
- [Bailly du Bois1995] P. Bailly du Bois, J. C. Salomon, R. Gandon et P. Guéguéniat. *A quantitative estimate of English Channel water fluxes into the North Sea from 1987 to 1992 based on radiotracer distribution*. Journal of Marine Systems, vol. 6, no. 5, pages 457–481, November 1995.
- [Bailly du Bois2012] P. Bailly du Bois, P. Laguionie, D. Boust, I. Korsakissok, D. Didier et B. Fiévet. *Estimation of marine source-term following Fukushima Dai-ichi accident*. Journal of Environmental Radioactivity, vol. 114, pages 2–9, December 2012.
- [Barnett2015a] G A Barnett. *A Robust RBF-FD Formulation based on Polyharmonic Splines and Polynomials*. Thèse de Doctorat, University of Colorado Boulder, 2015.
- [Barnett2015b] Gregory A. Barnett, Natasha Flyer et Louis J. Wicker. *An RBF-FD polynomial method based on polyharmonic splines for the Navier-Stokes equations : Comparisons on different node layouts*. arXiv :1509.02615 [physics], September 2015. arXiv : 1509.02615.

- [Bayona2015] V. Bayona, N. Flyer, G. M. Lucas et A. J. G. Baumgaertner. *A 3-D RBF-FD solver for modeling the atmospheric global electric circuit with topography (GEC-RBFFD v1.0)*. Geosci. Model Dev., vol. 8, no. 10, pages 3007–3020, October 2015.
- [Bayona2017] Victor Bayona, Natasha Flyer, Bengt Fornberg et Gregory A. Barnett. *On the role of polynomials in RBF-FD approximations : II. Numerical solution of elliptic PDEs*. Journal of Computational Physics, vol. 332, pages 257–273, March 2017.
- [Bayona2019] Víctor Bayona, Natasha Flyer et Bengt Fornberg. *On the role of polynomials in RBF-FD approximations : III. Behavior near domain boundaries*. Journal of Computational Physics, vol. 380, pages 378–399, March 2019.
- [Bayona2021] Víctor Bayona, Mario Sánchez-Sanz, Eduardo Fernández-Tarrazo et Manuel Kindelan. *Micro-combustion modelling with RBF-FD : A high-order meshfree method for reactive flows in complex geometries*. Applied Mathematical Modelling, vol. 94, pages 635–655, June 2021.
- [Bernard2023] Jérémy Bernard, Fredrik Lindberg et Sandro Oswald. *URock 2023a : An open source GIS-based wind model for complex urban settings*. EGU sphere, pages 1–45, April 2023. Publisher : Copernicus GmbH.
- [Blayo2010] Eric Blayo et Maelle Nodet. *Introduction a l’assimilation de donn ées variationnelle*. page 72, 2010.
- [Bocquet2014] M. Bocquet et P. Sakov. *An iterative ensemble Kalman smoother : An Iterative Ensemble Kalman Smoother*. Q.J.R. Meteorol. Soc., vol. 140, no. 682, pages 1521–1535, July 2014.
- [Bollig2012] E.F. Bollig, N. Flyer et G. Erlebacher. *Solution to PDEs using radial basis function finite-differences (RBF-FD) on multiple GPUs*. Journal of Computational Physics, vol. 231, no. 21, pages 7133–7151, 2012.
- [Bonan2014] B. Bonan, M. Nodet, C. Ritz et V. Peyaud. *An ETKF approach for initial state and parameter estimation in ice sheet modelling*. Nonlinear Processes in Geophysics, vol. 21, no. 2, pages 569–582, April 2014. Publisher : Copernicus GmbH.
- [Brown2015] Michael J. Brown, Michael D. Williams, Matthew A. Nelson et Kenneth A. Werley. *QUIC Transport and Dispersion Modeling of Vehicle Emissions in Cities for Better Public Health Assessments*. Environmental Health Insights, vol. 9s1, page EHI.S15662, January 2015.
- [Bruse2007] Michael Bruse. *ENVI-met implementation of the gas/ particle dispersion and deposition model PDDM*. 2007.
- [Businger1971] J. A. Businger, J. C. Wyngaard, Y. Izumi et E. F. Bradley. *Flux-Profile Relationships in the Atmospheric Surface Layer*. Journal of the Atmospheric Sciences, vol. 28, no. 2, pages 181–189, March 1971. Publisher : American Meteorological Society Section : Journal of the Atmospheric Sciences.
- [Byrd1995] Richard H. Byrd, Peihuang Lu, Jorge Nocedal et Ciyou Zhu. *A Limited Memory Algorithm for Bound Constrained Optimization*. SIAM J. Sci. Comput., vol. 16, no. 5, pages 1190–1208, September 1995. Publisher : Society for Industrial and Applied Mathematics.
- [Cellier2018] Nicolas Cellier. *Optimisation d’échangeurs à films ruisselants*. Thermique [physics.class-ph], Université Grenoble Alpes, 2018.

-
- [Chandramouli2020] Pranav Chandramouli, Etienne Memin et Dominique Heitz. *4D large scale variational data assimilation of a turbulent flow with a dynamics error model*. Journal of Computational Physics, vol. 412, page 109446, July 2020.
- [Chinchapatnam2009] P.P. Chinchapatnam, K. Djidjeli, P.B. Nair et M. Tan. *A compact RBF-FD based meshless method for the incompressible Navier-Stokes equations*. Proceedings of the Institution of Mechanical Engineers Part M : Journal of Engineering for the Maritime Environment, vol. 223, no. 3, pages 275–290, 2009.
- [Cusdin2003] P Cusdin et Jens-Dominik Mueller. *Automatic differentiation and sensitivity analysis methods for computational fluid dynamics*. April 2003.
- [Dai2019] Tie Dai, Yueming Cheng, Kentaroh Suzuki, Daisuke Goto, Maki Kikuchi, Nick A. J. Schutgens, Mayumi Yoshida, Peng Zhang, Letu Husi, Guangyu Shi et Teruyuki Nakajima. *Hourly Aerosol Assimilation of Himawari-8 AOT Using the Four-Dimensional Local Ensemble Transform Kalman Filter*. Journal of Advances in Modeling Earth Systems, vol. 11, no. 3, pages 680–711, 2019. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001475>.
- [Davoine2007] X. Davoine et M. Bocquet. *Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport*. Atmospheric Chemistry and Physics, vol. 7, no. 6, pages 1549–1564, March 2007. Publisher : Copernicus GmbH.
- [Defforge2019a] Cécile Defforge. *Data assimilation for micrometeorological applications with the fluid dynamics model Code_Saturne*. Thèse de Doctorat, Paris-Est, Paris, 2019.
- [Defforge2019b] Cécile L Defforge, B. Carissimo, Michael Bocquet, R. Bresson et P. Armand. *Improving CFD atmospheric simulations at local scale for wind resource assessment using the iterative ensemble Kalman smoother*. Journal of Wind Engineering and Industrial Aerodynamics, vol. 189, pages 243–257, June 2019. Publisher : Elsevier.
- [Defforge2021] Cécile L. Defforge, Bertrand Carissimo, Marc Bocquet, Raphaël Bresson et Patrick Armand. *Improving Numerical Dispersion Modelling in Built Environments with Data Assimilation Using the Iterative Ensemble Kalman Smoother*. Boundary-Layer Meteorol, vol. 179, no. 2, pages 209–240, May 2021.
- [Dehghan2017] Mehdi Dehghan et Mostafa Abbaszadeh. *The use of proper orthogonal decomposition (POD) meshless RBF-FD technique to simulate the shallow water equations*. Journal of Computational Physics, vol. 351, pages 478–510, December 2017.
- [Depolli2022] Matjaž Depolli, Jure Slak et Gregor Kosec. *Parallel domain discretization algorithm for RBF-FD and other meshless numerical methods for solving PDEs*. Computers & Structures, vol. 264, page 106773, May 2022.
- [Dimet1986] François-Xavier Le Dimet et Olivier Talagrand. *Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects*. Tellus A, vol. 38A, no. 2, pages 97–110, March 1986.
- [Ding2021] Shan Ding et Rui Yang. *Reduced-order modelling of urban wind environment and gaseous pollutants dispersion in an urban-scale street*
-

- canyon*. Journal of Safety Science and Resilience, vol. 2, no. 4, pages 238–245, December 2021.
- [Dobravec2022] Tadej Dobravec, Boštjan Mavrič et Božidar Šarler. *Acceleration of RBF-FD meshless phase-field modelling of dendritic solidification by space-time adaptive approach*. Computers & Mathematics with Applications, vol. 126, pages 77–99, November 2022.
- [Dokainish1989] M.A. Dokainish et K. Subbaraj. *A survey of direct time-integration methods in computational structural dynamics—I. Explicit methods*. Computers & Structures, vol. 32, no. 6, pages 1371–1386, January 1989. Publisher : Pergamon.
- [Evensen2003] Geir Evensen. *The Ensemble Kalman Filter : theoretical formulation and practical implementation*. Ocean Dynamics, vol. 53, no. 4, pages 343–367, November 2003.
- [Faragó2008] István Faragó, Krassimir Georgiev et Zahari Zlatev. *Parallelization of Advection-Diffusion-Chemistry Modules*. In Ivan Lirkov, Svetozar Margenov et Jerzy Waśniewski, éditeurs, Large-Scale Scientific Computing, Lecture Notes in Computer Science, pages 28–39, Berlin, Heidelberg, 2008. Springer.
- [Farchi2016] Alban Farchi, Marc Bocquet, Yelva Roustan, Anne Mathieu et Arnaud Quérel. *Using the Wasserstein distance to compare fields of pollutants : application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident*. Tellus B : Chemical and Physical Meteorology, vol. 68, no. 1, page 31682, December 2016. Publisher : Taylor & Francis _eprint : <https://doi.org/10.3402/tellusb.v68.31682>.
- [Farrell2013] P. E. Farrell, D. A. Ham, S. W. Funke et M. E. Rognes. *Automated Derivation of the Adjoint of High-Level Transient Finite Element Programs*. SIAM J. Sci. Comput., vol. 35, no. 4, pages C369–C393, January 2013.
- [Fletcher1998] Clive A. J. Fletcher. *Computational Techniques for Fluid Dynamics 1*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.
- [Flyer2012] Natasha Flyer, Erik Lehto, Sébastien Blaise, Grady B. Wright et Amik St-Cyr. *A guide to RBF-generated finite differences for nonlinear transport : Shallow water simulations on a sphere*. Journal of Computational Physics, vol. 231, no. 11, pages 4078–4095, June 2012.
- [Flyer2016a] Natasha Flyer, Gregory A. Barnett et Louis J. Wicker. *Enhancing finite differences with radial basis functions : Experiments on the Navier–Stokes equations*. Journal of Computational Physics, vol. 316, pages 39–62, July 2016.
- [Flyer2016b] Natasha Flyer, Bengt Fornberg, Victor Bayona et Gregory A. Barnett. *On the role of polynomials in RBF-FD approximations : I. Interpolation and accuracy*. Journal of Computational Physics, vol. 321, pages 21–38, September 2016.
- [Fornberg2015a] Bengt Fornberg. *Fast generation of 2-D node distributions for mesh-free PDE discretizations*. page 14, 2015.
- [Fornberg2015b] Bengt Fornberg et Natasha Flyer. *Fast generation of 2-D node distributions for mesh-free PDE discretizations*. Computers & Mathematics with Applications, vol. 69, no. 7, pages 531–544, April 2015.
- [Fornberg2015c] Bengt Fornberg et Natasha Flyer. *A Primer on Radial Basis Functions with Applications to the Geosciences*. Society for industrial and applied mathematics edition, 2015.

-
- [Funke2019] Simon Wolfgang Funke, Magne Nordaas, Øyvind Evju, Martin Sandve Alnaes et Kent Andre Mardal. *Variational data assimilation for transient blood flow simulations : Cerebral aneurysms as an illustrative example*. Int J Numer Meth Biomed Engng, vol. 35, no. 1, page e3152, January 2019.
- [Garay2023] Jeremias Garay, Jocelyn Dunstan, Sergio Uribe et Francisco Sahli Costabal. *Physics-informed neural networks for blood flow inverse problems*. arXiv preprint arXiv :2308.00927, August 2023.
- [Georges2013] Didier Georges. *Optimal location of mobile sensors for environmental monitoring*. In 2013 European Control Conference (ECC), pages 1280–1285, Zurich, July 2013. IEEE.
- [Georges2019] Didier Georges. *A Variational Calculus Approach to Wildfire Monitoring Using a Low-Discrepancy Sequence-Based Deployment of Sensors*. In 2019 IEEE 58th Conference on Decision and Control (CDC), pages 5912–5917, Nice, France, December 2019. IEEE.
- [Georges2020] Didier Georges. *Machine Learning for Receding Horizon Observer Design : Application to Traffic Density Estimation*. IFAC, vol. 2020 - 21st IFAC World Congress 2020, page 7, July 2020.
- [Girard2020] Sylvain Girard, Patrick Armand, Christophe Duchenne et Thierry Yalamas. *Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations*. Atmospheric Environment, vol. 224, page 117313, March 2020.
- [Gonzalez Olivardia2019] Franchesca G. Gonzalez Olivardia, Qi Zhang, Tomohito Matsuo, Hikari Shimadera et Akira Kondo. *Analysis of Pollutant Dispersion in a Realistic Urban Street Canyon Using Coupled CFD and Chemical Reaction Modeling*. Atmosphere, vol. 10, no. 9, page 479, September 2019. Number : 9 Publisher : Multidisciplinary Digital Publishing Institute.
- [Gordon1995] Neil Gordon, David Salmond et Craig Ewing. *Bayesian state estimation for tracking and guidance using the bootstrap filter*. Journal of Guidance, Control, and Dynamics, 1995.
- [Guerreiro2018] Cristina Guerreiro, A. Ortiz, Frank de Leeuw et Mar Viana. *Air quality 2018 - EEA report 12 2018*. November 2018.
- [Guerrero2021] Joel Guerrero. *A Crash Introduction to the Finite Volume Method and Discretization Schemes in OpenFOAM®*, June 2021.
- [Guyer2009] Jonathan E. Guyer, Daniel Wheeler et James A. Warren. *FiPy : Partial Differential Equations with Python*. Comput. Sci. Eng., vol. 11, no. 3, pages 6–15, May 2009.
- [Hammond2019] J.K. Hammond, R. Chakir, F. Bourquin et Y. Maday. *PBDW : A non-intrusive Reduced Basis Data Assimilation method and its application to an urban dispersion modeling framework*. Applied Mathematical Modelling, vol. 76, pages 1–25, December 2019.
- [Hascoet2013] Laurent Hascoet et Valérie Pascual. *The Tapenade automatic differentiation tool : Principles, model, and specification*. ACM Trans. Math. Softw., vol. 39, no. 3, pages 1–43, April 2013.
- [Hasenfratz2015] David Hasenfratz. *Deriving high-resolution urban air pollution maps using mobile sensor nodes*. Pervasive and Mobile Computing, page 18, 2015.
-

- [Hersbach2020] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Giannata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume et Jean-Noël Thépaut. *The ERA5 global reanalysis*. Q.J.R. Meteorol. Soc., vol. 146, no. 730, pages 1999–2049, July 2020.
- [Hines2015] Trever Hines. *RBF : Une bibliothèque Python*. 2015.
- [Honorat2007] Marc Honorat, Joel Marin, Jérôme Monnier et Xijun Lai. *Dassflow v1.0 : a variational data assimilation software for 2D river flows*. inria-00137447, 2007.
- [Huang2021] Lilan Huang, Hongze Leng, Xiaoyong Li, Kaijun Ren, Junqiang Song et Dongzi Wang. *A Data-Driven Method for Hybrid Data Assimilation with Multilayer Perceptron*. Big Data Research, vol. 23, page 100179, February 2021.
- [Issartel2007] Jean-Pierre Issartel, Maithili Sharan et Manish Modani. *An inversion technique to retrieve the source of a tracer with an application to synthetic satellite measurements*. Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences, vol. 463, no. 2087, pages 2863–2886, August 2007. Publisher : Royal Society.
- [Jančič2020] Mitja Jančič, Jure Slak et Gregor Kosec. *GPU accelerated RBF-FD solution of Poisson’s equation*. In 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), pages 214–218, September 2020. ISSN : 2623-8764.
- [Jančič2021] Mitja Jančič, Jure Slak et Gregor Kosec. *Monomial Augmentation Guidelines for RBF-FD from Accuracy Versus Computational Time Perspective*. J Sci Comput, vol. 87, no. 1, page 9, April 2021.
- [Jardak2018] Mohamed Jardak et Olivier Talagrand. *Ensemble variational assimilation as a probabilistic estimator – Part 1 : The linear and weak non-linear case*. Nonlin. Processes Geophys., vol. 25, no. 3, pages 565–587, August 2018.
- [Kalman1960] Rudolph Emil Kalman. *A New Approach to Linear Filtering and Prediction Problems*. Transactions of the ASME—Journal of Basic Engineering, vol. 82, no. Series D, pages 35–45, 1960.
- [Kansa1990a] E. J. Kansa. *Multiquadrics—A scattered data approximation scheme with applications to computational fluid-dynamics—II solutions to parabolic, hyperbolic and elliptic partial differential equations*. Computers & Mathematics with Applications, vol. 19, no. 8, pages 147–161, January 1990.
- [Kansa1990b] E.J. Kansa. *Multiquadrics—A scattered data approximation scheme with applications to computational fluid-dynamics—I surface approximations and partial derivative estimates*. Computers & Mathematics with Applications, vol. 19, no. 8-9, pages 127–145, 1990.

-
- [Kato2013] Hiroshi Kato et Shigeru Obayashi. *Approach for uncertainty of turbulence modeling based on data assimilation technique*. *Computers & Fluids*, vol. 85, pages 2–7, October 2013.
- [Katzfuss2016] Matthias Katzfuss, Jonathan R. Stroud et Christopher K. Wikle. *Understanding the Ensemble Kalman Filter*. *The American Statistician*, vol. 70, no. 4, pages 350–357, October 2016.
- [Khodayi-mehr2019] Reza Khodayi-mehr, Wilkins Aquino et Michael M. Zavlanos. *Model-Based Active Source Identification in Complex Environments*. *IEEE Trans. Robot.*, vol. 35, no. 3, pages 633–652, June 2019. arXiv : 1706.01603.
- [Kikumoto2020] Hideki Kikumoto. *Turbulent diffusivity limiter with travel time for eulerian-CFD eulerian analysis of point-source pollutant dispersion*. *Wind Engineering Research*, vol. 26, page 9, 2020.
- [Kouichi2016] H. Kouichi, G. Turbelin, P. Ngae, A. A. Feiz, E. Barbosa et A. Chpoun. *Optimization of sensor networks for the estimation of atmospheric pollutants sources*. pages 11–21, Crete, Greece, June 2016.
- [Kouichi2019] Hamza Kouichi, Pierre Ngae, Pramod Kumar, Amir-Ali Feiz et Nadir Bekka. *An optimization for reducing the size of an existing urban-like monitoring network for retrieving an unknown point source emission*. *Geoscientific Model Development*, vol. 12, no. 8, pages 3687–3705, August 2019. Publisher : Copernicus GmbH.
- [Kovalets2011] Ivan V. Kovalets, Spyros Andronopoulos, Alexander G. Venetsanos et John G. Bartzis. *Identification of strength and location of stationary point source of atmospheric pollutant in urban conditions using computational fluid dynamics model*. *Mathematics and Computers in Simulation*, vol. 82, no. 2, pages 244–257, October 2011.
- [Kovalets2018] Ivan V. Kovalets, George C. Efthimiou, Spyros Andronopoulos, Alexander G. Venetsanos, Christos D. Argyropoulos et Konstantinos E. Kakosimos. *Inverse identification of unknown finite-duration air pollutant release from a point source in urban environment*. *Atmospheric Environment*, vol. 181, pages 82–96, May 2018.
- [Krysta2007a] Monika Krysta et Marc Bocquet. *Source reconstruction of an accidental radionuclide release at European scale*. page 16, 2007.
- [Krysta2007b] Monika Krysta et Marc Bocquet. *Source reconstruction of an accidental radionuclide release at European scale*. *Quarterly Journal of the Royal Meteorological Society*, vol. 133, no. 623, pages 529–544, 2007. _eprint : <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3>.
- [Kumar2015] Pramod Kumar, Amir-Ali Feiz, Sarvesh Kumar Singh, Pierre Ngae et Grégory Turbelin. *Reconstruction of an atmospheric tracer source in an urban-like environment*. *Journal of Geophysical Research : Atmospheres*, vol. 120, no. 24, pages 12589–12604, December 2015. Publisher : John Wiley & Sons, Ltd.
- [Labbe2018] Roger R Jr Labbe. *Kalman and Bayesian Filters in Python*. 2018.
- [Lauder1974] Brian Launder et D.B. Spalding. *The Numerical Computation of Turbulent Flow Computer Methods*. *Computer Methods in Applied Mechanics and Engineering*, vol. 3, pages 269–289, March 1974.
- [LeVeque2002] Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 1 edition, August 2002.
-

- [Liu2013] Hai-Ying Liu, Alena Bartonova, Martin Schindler, Mukesh Sharma, Sailesh N. Behera, Kamlesh Katiyar et Onkar Dikshit. *Respiratory Disease in Relation to Outdoor Air Pollution in Kanpur, India*. Archives of Environmental & Occupational Health, vol. 68, no. 4, pages 204–217, October 2013. Publisher : Routledge _eprint : <https://doi.org/10.1080/19338244.2012.701246>.
- [Liu2019] Jianming Liu, Xinkai Li et Xiuling Hu. *A RBF-based differential quadrature method for solving two-dimensional variable-order time fractional advection-diffusion equation*. Journal of Computational Physics, vol. 384, pages 222–238, May 2019.
- [Logg2012] Anders Logg, Kent-Andre Mardal et Garth Wells, éditeurs. Automated Solution of Differential Equations by the Finite Element Method, volume 84 of *Lecture Notes in Computational Science and Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [Lopez-Ferber2022] Roman Lopez-Ferber, Sylvain Leirens et Didier Georges. *Source term estimation : variational method versus machine learning applied to urban air pollution*. Bulgarie, June 2022.
- [Lopez-Restrepo2021] Santiago Lopez-Restrepo, Andres Yarce, Nicolás Pinel, O. L. Quintero, Arjo Segers et A. W. Heemink. *Urban Air Quality Modeling Using Low-Cost Sensor Network and Data Assimilation in the Aburrá Valley, Colombia*. Atmosphere, vol. 12, no. 1, page 91, January 2021. Number : 1 Publisher : Multidisciplinary Digital Publishing Institute.
- [Lu2021] Lu Lu, Xuhui Meng, Zhiping Mao et George Em Karniadakis. *DeepXDE : A Deep Learning Library for Solving Differential Equations*. SIAM Rev., vol. 63, no. 1, pages 208–228, January 2021. Publisher : Society for Industrial and Applied Mathematics.
- [Luna2021] Kevin Luna, Katherine Klymko et Johannes P. Blaschke. *Accelerating GMRES with Deep Learning in Real-Time*. Rapport technique arXiv :2103.10975, arXiv, March 2021. arXiv :2103.10975 [physics].
- [López-Cabeza2018] V. P. López-Cabeza, C. Galán-Marín, C. Rivera-Gómez et J. Roa-Fernández. *Courtyard microclimate ENVI-met outputs deviation from the experimental data*. Building and Environment, vol. 144, pages 129–141, October 2018.
- [Martin2017] Bradley Martin et Bengt Fornberg. *Seismic modeling with radial basis function-generated finite differences (RBF-FD) – a simplified treatment of interfaces*. Journal of Computational Physics, vol. 335, pages 828–845, April 2017.
- [Marécal1999] Marécal. *Variational retrieval of temperature and humidity profiles from TRMM precipitation data, 1999*. Publisher : ECMWF.
- [Mathews2022] Nathaniel H. Mathews, Natasha Flyer et Sarah E. Gibson. *Solving 3D Magnetohydrostatics with RBF-FD : Applications to the Solar Corona*. Journal of Computational Physics, page 111214, April 2022.
- [Mendil2022] Mouhcine Mendil, Sylvain Leirens, Patrick Armand et Christophe Duchenne. *Hazardous atmospheric dispersion in urban areas : A Deep Learning approach for emergency pollution forecast*. Environmental Modelling & Software, page 105387, April 2022.
- [Michalska1995] H. Michalska et D.Q. Mayne. *Moving horizon observers and observer-based control*. IEEE Transactions on Automatic Control, vol. 40, no. 6,

-
- pages 995–1006, June 1995. Conference Name : IEEE Transactions on Automatic Control.
- [Mons2016] V. Mons, J.-C. Chassaing, T. Gomez et P. Sagaut. *Reconstruction of unsteady viscous flows using data assimilation schemes*. Journal of Computational Physics, vol. 316, pages 255–280, July 2016.
- [Mons2017] V. Mons, L. Margheri, J. C. Chassaing et P. Sagaut. *Data assimilation-based reconstruction of urban pollutant release characteristics*. Journal of Wind Engineering and Industrial Aerodynamics, vol. 169, pages 232–250, October 2017.
- [Mons2019] Vincent Mons, Qi Wang et Tamer A. Zaki. *Kriging-enhanced ensemble variational data assimilation for scalar-source identification in turbulent environments*. Journal of Computational Physics, vol. 398, page 108856, December 2019.
- [Mortensen2011] Mikael Mortensen, Hans Petter Langtangen et Garth N. Wells. *A FEniCS-based programming framework for modeling turbulent flow by the Reynolds-averaged Navier–Stokes equations*. Advances in Water Resources, vol. 34, no. 9, pages 1082–1101, September 2011.
- [Moussafir2004] Jacques Moussafir, Olivier Oldrini, Gianni Tinarelli et John Sontowski. *A new operational approach to deal with dispersion around obstacles : the MSS (Micro Swift Spray) software suite*. page 5, 2004.
- [Nguyen2016a] van Tri Nguyen. *Adjoint-based approach for estimation & sensor location on 1D hyperbolic systems with applications in hydrology & traffic*. Automatic, Université Grenoble Alpes, 2016.
- [Nguyen2016b] Van Tri Nguyen, Didier Georges et Gildas Besançon. *State and parameter estimation in 1-D hyperbolic PDEs based on an adjoint method*. Automatica, vol. 67, pages 185–191, May 2016.
- [Niederreiter1988] Harald Niederreiter. *Low-discrepancy and low-dispersion sequences*. Journal of Number Theory, vol. 30, no. 1, pages 51–70, September 1988.
- [Nielinger2004] Jost Nielinger, Rainer Röckle, Hans-Christian Höfl et Werner-Jürgen Kost. *Lagrange versus Eulerian dispersion modeling comparison for investigations concerning air pollution caused by traffic*. page 4, Garmisch-Partenkirchen, Germany, June 2004.
- [Oldrini2013] Olivier Oldrini, Maxime Nibart, Patrick Armand, Christophe Olry, Jacques Moussafir et Armand Albergel. *Introduction of Momentum Equations in MICRO-SWIFT*. page 5, 2013.
- [Oliver2022] Matthieu Oliver, Didier Georges et Clémentine Prieur. *Spatialized epidemiological forecasting applied to Covid-19 pandemic at departmental scale in France*. Systems & Control Letters, page 105240, April 2022.
- [Paige1982] Christopher C. Paige et Michael A. Saunders. *LSQR : An Algorithm for Sparse Linear Equations and Sparse Least Squares*. ACM Trans. Math. Softw., vol. 8, no. 1, pages 43–71, March 1982.
- [Pudykiewicz1998] Janusz A. Pudykiewicz. *Application of adjoint tracer transport equations for evaluating source parameters*. Atmospheric Environment, vol. 32, no. 17, pages 3039–3050, September 1998.
- [Quélo2004] Denis Quélo. *Simulation numérique et assimilation de données variationnelle pour la dispersion atmosphérique de polluants*. Thèse de Doctorat, Ecole des Ponts ParisTech, Paris, 2004.
-

- [Raissi2019] M. Raissi, P. Perdikaris et G.E. Karniadakis. *Physics-informed neural networks : A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*. Journal of Computational Physics, vol. 378, pages 686–707, February 2019.
- [Raoult2019] Cécile Raoult, Michel Benoit et Marissa L. Yates. *Development and validation of a 3D RBF-spectral model for coastal wave simulation*. Journal of Computational Physics, vol. 378, pages 278–302, February 2019.
- [Seibert2004] P. Seibert et A. Frank. *Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode*. Atmospheric Chemistry and Physics, vol. 4, no. 1, pages 51–63, January 2004. Publisher : Copernicus GmbH.
- [Septier2020] François Septier, Patrick Armand et Christophe Duchenne. *A bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments*. Atmospheric Environment, vol. 242, page 117733, December 2020.
- [Singh2015] Sarvesh Kumar Singh et Raj Rani. *Assimilation of concentration measurements for retrieving multiple point releases in atmosphere : A least-squares approach to inverse modelling*. Atmospheric Environment, vol. 119, pages 402–414, October 2015.
- [Sirkes1997] Ziv Sirkes et Eli Tziperman. *Finite Difference of Adjoint or Adjoint of Finite Difference ?* Mon. Wea. Rev., vol. 125, no. 12, pages 3373–3378, December 1997.
- [Skamarock2008] C. Skamarock, B. Klemp, Jimy Dudhia, O. Gill, Dale Barker, G. Duda, Xiang-yu Huang, Wei Wang et G. Powers. *A Description of the Advanced Research WRF Version 3*. 2008.
- [Slak2021] Jure Slak et Gregor Kosec. *Medusa : A C++ Library for solving PDEs using Strong Form Mesh-Free methods*. ACM Trans. Math. Softw., vol. 47, no. 3, pages 1–25, June 2021. arXiv : 1912.13282.
- [Soulhac2011] Lionel Soulhac, Pietro Salizzoni, F.-X. Cierco et Richard Perkins. *The model SIRANE for atmospheric urban pollutant dispersion ; part I, presentation of the model*. Atmospheric Environment, vol. 45, no. 39, pages 7379–7395, December 2011.
- [Soulhac2016] L. Soulhac, G. Lamaison, F. X. Cierco, N. Ben Salem, P. Salizzoni, P. Mejean, P. Armand et L. Patryl. *SIRANERISK : Modelling dispersion of steady and unsteady pollutant releases in the urban canopy*. Atmospheric Environment, vol. 140, pages 242–260, September 2016.
- [Soulhac2017] L. Soulhac, C.V. Nguyen, P. Volta et P. Salizzoni. *The model SIRANE for atmospheric urban pollutant dispersion. PART III : Validation against NO2 yearly concentration measurements in a large urban agglomeration*. Atmospheric Environment, vol. 167, pages 377–388, October 2017.
- [Spalart1992] Philippe Spalart et Steven Allmaras. *A One-Equation Turbulence Model for Aerodynamic Flows*. AIAA, vol. 439, January 1992.
- [Szunyogh2008] Istvan Szunyogh, Eric J. Kostelich, Gyorgyi Gyarmati, Eugenia Kalnay, Brian R. Hunt, Edward Ott, Elizabeth Satterfield et James A. Yorke. *A local ensemble transform Kalman filter data assimilation system for the*

-
- NCEP global model*. *Tellus A : Dynamic Meteorology and Oceanography*, vol. 60, no. 1, pages 113–130, January 2008. Publisher : Taylor & Francis _eprint : <https://doi.org/10.1111/j.1600-0870.2007.00274.x>.
- [Tang2013] X. Tang, J. Zhu, Z.F. Wang, M. Wang, A. Gbaguidi, J. Li, M. Shao, G.Q. Tang et D.S. Ji. *Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter*. *Atmospheric Environment*, vol. 81, pages 676–686, December 2013.
- [Tikhonov1977] Andreï Nikolaevich Tikhonov. *Solution of Ill-posed Problems*. Vh winston & sons edition, 1977.
- [Tinarelli1994] G. Tinarelli, D. Anfossi, G. Brusasca, E. Ferrero, U. Giostra, M. G. Morselli, J. Moussafir, F. Trombetti et F. Tampieri. *Lagrangian Particle Simulation of Tracer Dispersion in the Lee of a Schematic Two-Dimensional Hill*. *Journal of Applied Meteorology and Climatology*, vol. 33, no. 6, pages 744–756, June 1994. Publisher : American Meteorological Society Section : *Journal of Applied Meteorology and Climatology*.
- [Tominec2021a] Igor Tominec et Eva Breznik. *An unfitted RBF-FD method in a least-squares setting for elliptic PDEs on complex geometries*. *Journal of Computational Physics*, vol. 436, page 110283, July 2021.
- [Tominec2021b] Igor Tominec, Elisabeth Larsson et Alfa Heryudono. *A least squares radial basis function finite difference method with improved stability properties*. Rapport technique arXiv :2003.03132, arXiv, March 2021. arXiv :2003.03132 [cs, math].
- [Tominec2021c] Igor Tominec, Pierre-Frederic Villard, Elisabeth Larsson, Victor Bayona et Nicola Cacciani. *An unfitted radial basis function generated finite difference method applied to thoracic diaphragm simulations*. March 2021.
- [Tominec2022] Igor Tominec et Murtazo Nazarov. *Residual Viscosity Stabilized RBF-FD Methods for Solving Nonlinear Conservation Laws*. *J Sci Comput*, vol. 94, no. 1, page 14, December 2022.
- [Towara2013] M. Towara et U. Naumann. *A Discrete Adjoint Model for OpenFOAM*. *Procedia Computer Science*, vol. 18, pages 429–438, 2013.
- [Wang2020] Jang-Woon Wang, Ho-Jin Yang et Jae-Jin Kim. *Wind speed estimation in urban areas based on the relationships between background wind speeds and morphological parameters*. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 205, page 104324, October 2020.
- [Winiarek2012] Victor Winiarek, Marc Bocquet, Olivier Saunier et Anne Mathieu. *Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant : Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant*. *J. Geophys. Res.*, vol. 117, no. D5, March 2012.
- [Winiarek2014a] Victor Winiarek. *Dispersion atmosphérique et modélisation inverse pour la reconstruction de sources accidentelles de polluants*. Thèse de Doctorat, Université Paris-Est, Paris, 2014.
- [Winiarek2014b] Victor Winiarek, Marc Bocquet, Nora Duhanyan, Yelva Roustan, Olivier Saunier et Anne Mathieu. *Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations*. *Atmospheric Environment*, vol. 82, pages 268–279, January 2014.
-

- [Xue2017] Yu Xue et Zhiqiang John Zhai. *Inverse identification of multiple outdoor pollutant sources with a mobile sensor*. Build. Simul., vol. 10, no. 2, pages 255–263, April 2017.
- [Zaini2022] Nur'atiah Zaini, Lee Woen Ean, Ali Najah Ahmed, Marlinda Abdul Malek et Ming Fai Chow. *PM2.5 forecasting for an urban area based on deep learning and decomposition method*. Sci Rep, vol. 12, no. 1, page 17565, October 2022. Number : 1 Publisher : Nature Publishing Group.
- [Zannetti1990] Paolo Zannetti, editeur. *Air Pollution Modeling : Theories, Computational Methods and Available Software*. Springer US, 1990.
- [Zhang2018] Minxin Zhang et Kirsten Morris. *Sensor Choice for Minimum Error Variance Estimation*. IEEE Trans. Automat. Contr., vol. 63, no. 2, pages 315–330, February 2018.
- [Zhiyin2015] Yang Zhiyin. *Large-eddy simulation : Past, present and the future*. Chinese Journal of Aeronautics, vol. 28, no. 1, pages 11–24, February 2015.
- [Šmídl2013] Václav Šmídl. *Tracking of atmospheric release of pollution using unmanned aerial vehicles*. Atmospheric Environment, page 12, 2013.
- [Županski1995] Dušanka Županski et Fedor Mesinger. *Four-Dimensional Variational Assimilation of Precipitation Data*. Monthly Weather Review, vol. 123, no. 4, pages 1112–1127, April 1995. Publisher : American Meteorological Society Section : Monthly Weather Review.

A

Annexe 1 : Suppléments sur la simulation numérique par RBF-FD

Cette annexe apporte des informations supplémentaires sur la simulation numérique par RBF-FD notamment sur l'implémentation informatique.

A.1 Expression numérique des opérateurs RBF-FD

Le code LS-RBF-FD mis en œuvre dans cette thèse s’appuie sur les travaux de T. Hines [Hines2015] qui a développé une bibliothèque pour la construction d’opérateurs élémentaires RBF-FD en collocation, que nous avons adapté aux techniques de moindres carrés en dehors des méthodes de collocation. Par ailleurs la documentation de la bibliothèque python *RBF* est disponible et donne des exemples de modélisation RBF-FD à collocation¹. Ce code a servi de fondation pour le développement de bibliothèques informatiques Python intitulés *src_utils* et *src_roman*. Ces bibliothèques ont été installées avec l’utilitaire *pip* fourni par le flux de travail de Python. Les deux bibliothèques numériques ont été installées au sein d’un environnement Python 3.9. Un exemple d’implémentation pratique est illustré pour l’opérateur aux nœuds intérieurs dans \mathbf{D}_Ω en A.1. L’opérateur des conditions aux limites de Neumann \mathbf{D}_{Γ_N} est implémenté comme dans A.2 tandis l’opérateur des conditions aux limites de Dirichlet \mathbf{D}_{Γ_D} est implémentée à la manière du code illustré en A.2. Ces extraits de code illustrent les parties essentielles permettant de programmer les opérateurs différentiels discrétisés explicité en 3.2.1.

```

1 #  $\mathbf{D}_\Omega$ 
2 # Construction de la matrice de différentiation pour les
3 # noeuds intérieurs au domaines  $\Omega$ .
4
5 # Operateur d'advection  $\mathbf{U} \cdot \nabla$ 
6 U_grad_u = rbf_fast.weight_matrix(
7     y_nodes[y_groups['interior']], x_nodes, stencil_size,
8     [[1, 0, 0], [0, 1, 0], [0, 0, 1]],
9     coeffs=[U_x[idx_int], U_y[idx_int], U_z[idx_int]],
10    phi=phi, order=order,
11    n_jobs = n_jobs,
12    chunk_size = chunk_size
13    )
14
15 # Calcul de l'operateur de diffusion :
16 #  $\mathcal{L}_{diff} : \mathbf{c} \mapsto \nabla \cdot (\mathbf{K} \nabla \mathbf{c})$ 
17 # Expression de cet operateur sous la forme,
18 #  $\partial_i (\mathbf{K}_i \partial_i) \equiv (\mathbf{D}_i \mathbf{K}_i) \otimes \mathbf{D}_i + \mathbf{K}_i \otimes \mathbf{D}_{ii}$ 
19
20 # Calcul de  $(\mathbf{D}_i \mathbf{K}_i) \otimes \mathbf{D}_i$ 
21 nabla_x = rbf_fast.weight_matrix(
22     y_nodes[y_groups['interior']],
23     y_nodes,
24     stencil_size,
25     [1, 0, 0],
26     phi=phi,
27     order=order,
28     n_jobs = n_jobs,
29     chunk_size = chunk_size
30     )
31 nabla_y = rbf_fast.weight_matrix(
32     y_nodes[y_groups['interior']],
33     y_nodes,
34     stencil_size,
35     [0, 1, 0],
36     phi=phi,
37     order=order,
38     n_jobs = n_jobs,
39     chunk_size = chunk_size
40     )
41 nabla_z = rbf_fast.weight_matrix(

```

1. Visible sur <https://rbf.readthedocs.io/en/latest/fd.html>, consulté le 30 Août 2023

```

42     y_nodes[y_groups['interior']],
43     y_nodes,
44     stencil_size,
45     [0, 0, 1],
46     phi=phi,
47     order=order,
48     n_jobs = n_jobs,
49     chunk_size = chunk_size
50         )
51     nabla_x_D_x = nabla_x @ sp_Diff_x
52     nabla_y_D_y = nabla_y @ sp_Diff_y
53     nabla_z_D_z = nabla_z @ sp_Diff_z
54
55     W_diff_gauche = rbf_fast.weight_matrix(
56         y_nodes[y_groups['interior']],
57         x_nodes,
58         stencil_size,
59         [[1, 0, 0], [0, 1, 0], [0, 0, 1]],
60         coeffs = [np.squeeze(nabla_x_D_x.toarray()),
61                   np.squeeze(nabla_y_D_y.toarray()),
62                   np.squeeze(nabla_z_D_z.toarray())],
63         phi=phi,
64         order=order,
65         n_jobs = n_jobs,
66         chunk_size = chunk_size
67     )
68
69     # Calcul de  $\mathbf{K}_i \otimes \mathbf{D}_{ii}$ 
70     W_diff_droite = rbf_fast.weight_matrix(
71         y_nodes[y_groups['interior']],
72         x_nodes,
73         stencil_size,
74         [[2, 0, 0], [0, 2, 0], [0, 0, 2]],
75         coeffs = [self.Diff_x[idx_int],
76                   self.Diff_y[idx_int],
77                   self.Diff_z[idx_int]],
78         phi=phi,
79         order=order,
80         n_jobs = n_jobs,
81         chunk_size = chunk_size
82     )
83     W_diff = W_diff_gauche + W_diff_droite
84
85     # Expression de l'opérateur d'advection-diffusion aux noeuds intérieurs
86     L_interior = U_grad_u - W_diff

```

Listing A.1 – Exemple d'implémentation de \mathbf{D}_Ω

Nous pouvons voir que la programmation des appels est relativement explicite par rapport aux expressions mathématiques. Les paramètres d'appels sont en revanche relativement nombreux. Essentiellement, les nœuds d'évaluation inclus dans Y sont donnés en premier argument de la fonction `weight_matrix` et les centres de bases inclus dans X nommés ici `x_nodes` sont spécifiés en deuxième position. La taille de stencil règle le nombre de voisins pour réaliser les approximations locales basées sur les RBF. Le quatrième paramètre par exemple visible dans le calcul de l'opérateur d'advection `U_grad_u` spécifie l'ordre des dérivées de l'opérateur à approximer, ici l'opérateur gradient ∇ appliqué aux coordonnées cartésiennes x , y et z . Ainsi la liste `[[1, 0, 0], [0, 1, 0], [0, 0, 1]]` spécifie la nécessité d'approximer un opérateur différentiel de dérivé $\partial/\partial x$ pour la première coordonnée $\partial/\partial y$ pour la seconde et $\partial/\partial z$ pour la troisième. Ainsi la liste des différentiations construit un opérateur de dérivées tel que $(\partial/\partial x, \partial/\partial y, \partial/\partial z) \equiv [[1, 0, 0], [0, 1, 0], [0, 0, 1]]$.

```

1 #  $\mathbf{D}_{\Gamma_N}$ 
2 A_boundary_interior = rbf_fast.weight_matrix(

```

```

3     y_nodes[y_groups['boundary:interior']],
4     x_nodes,
5     stencil_size,
6     [[1, 0, 0], [0, 1, 0], [0, 0, 1]],
7     coeffs=[
8         y_normals[y_groups['boundary:interior'], 0],
9         y_normals[y_groups['boundary:interior'], 1],
10        y_normals[y_groups['boundary:interior'], 2]
11    ],
12    phi=phi,
13    order=order,
14    n_jobs = n_jobs,
15    chunk_size = chunk_size
16 )

```

Listing A.2 – Exemple d’implémentation de l’opérateur de conditions de Neuman \mathbf{D}_{Γ_N} .

```

1 #  $\mathbf{D}_{\Gamma_D}$ 
2 A_boundary_exterior = rbf_fast.weight_matrix(
3     y_nodes[y_groups['boundary:exterior']],
4     x_nodes, stencil_size, [0, 0, 0],
5     phi=phi, order=order,
6     n_jobs = n_jobs,
7     chunk_size = chunk_size
8 )

```

Listing A.3 – Exemple d’implémentation de l’opérateur de conditions de Dirichlet \mathbf{D}_{Γ_D} .

A.2 Structure des données de mailles triangulaires grossières

Cas tridimensionnel Les données de maillage grossier sont organisés en matrices listant d’une part les sommets de maillages et d’autre part les simplexes, ici des triangles qui permettent de les relier. La déclaration de telles formes géométriques est faite par exemple pour un cube comme dans l’exemple en listing A.4.

```

1 # Sommets et indices de sommets de triangles :
2 vertices = np.array([
3     [0, 0, 0], # Vertex 0
4     [1, 0, 0], # Vertex 1
5     [1, 1, 0], # Vertex 2
6     [0, 1, 0], # Vertex 3
7     [0, 0, 1], # Vertex 4
8     [1, 0, 1], # Vertex 5
9     [1, 1, 1], # Vertex 6
10    [0, 1, 1] # Vertex 7
11 ])
12
13 simples = np.array([
14    [0, 1, 2], # Face 0 (bottom, triangle 1)
15    [0, 2, 3], # Face 0 (bottom, triangle 2)
16    [4, 5, 6], # Face 1 (top, triangle 1)
17    [4, 6, 7], # Face 1 (top, triangle 2)
18    [0, 1, 5], # Face 2 (front, triangle 1)
19    [0, 4, 5], # Face 2 (front, triangle 2)
20    [1, 2, 6], # Face 3 (right, triangle 1)
21    [1, 5, 6], # Face 3 (right, triangle 2)
22    [2, 3, 7], # Face 4 (back, triangle 1)
23    [2, 6, 7], # Face 4 (back, triangle 2)

```



```
24 [3, 0, 4], # Face 5 (left, triangle 1)
25 [3, 7, 4] # Face 5 (left, triangle 2)
26 ]]
```

Listing A.4 – Déclaration d'un cube à l'aide d'une matrice de coordonnées de sommets "vertices", et d'une matrices de triangles. Les triangles contiennent les indices de sommet.

Cas bidimensionnel

Pour le cas bidimensionnel, nous cherchons à relier les sommets de bord par des segments au lieu d'arêtes décrivant une surface en 3D. Un exemple de coordonnées de sommets et d'indices de segments servant à relier les sommets est listé en listing A.5. Pour des formes géométriques quelconques comme des bâtiments, les indices de nœuds de segments sont créés en reliant le premier sommet au deuxième puis le deuxième au troisième etc. Cela se traduit par le code en listing A.6.

```
1 # Sommets et indices de sommets de triangles :
2 vertices = np.array([
3     [ 0, 0],
4     [200, 0],
5     [200, 100],
6     [ 0, 100],
7     [ 0, 0]
8 ])
9
10 simplices = np.array([
11     [0, 1],
12     [1, 2],
13     [2, 3],
14     [3, 4]
15 ]]
```

Listing A.5 – Déclaration d'un rectangle $200 \times 100\text{m}^2$ à l'aide d'une matrice de coordonnées de sommets "vertices", et de segments. Les segments contiennent les indices à connecter.

```
1 # Construction segments 2D
2 simplices = np.column_stack(
3     (np.arange(0, len(vertices) - 1), np.arange(1, len(vertices)))
4     )
```

Listing A.6 – Construction des segments 2D