



HAL
open science

Développement d'outils d'assimilation de données pour l'estimation augmentée d'écoulements internes

Lucas Villanueva

► **To cite this version:**

Lucas Villanueva. Développement d'outils d'assimilation de données pour l'estimation augmentée d'écoulements internes. Autre. ISAE-ENSMA Ecole Nationale Supérieure de Mécanique et d'Aérotechnique - Poitiers, 2024. Français. NNT : 2024ESMA0019 . tel-04833774

HAL Id: tel-04833774

<https://theses.hal.science/tel-04833774v1>

Submitted on 12 Dec 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 l'obtention du Grade de

DOCTEUR DE L'ÉCOLE NATIONALE SUPÉRIEURE DE MÉCANIQUE ET D'AÉROTECHNIQUE

(Diplôme National - Arrêté du 25 mai 2016 Modifié par l'Arrêté du 26 Août 2022)

École doctorale de Mathématiques, Informatique, Matériaux, Mécanique et Energétique –
MIMME n°651

Institut Pprime

Secteur de Recherche : Mécanique des milieux fluides

Développement d'outils d'assimilation de données pour l'estimation augmentée d'écoulements internes

Par **Lucas VILLANUEVA**

Sous la direction de **Jacques BORÉE**

Co-dirigée par **Marcello MELDI**

Co-encadrée par **Karine TRUFFIN**

Présentée et soutenue publiquement le 02 décembre 2024

Devant un jury composé de :

Eric LAMBALLAIS, *Professeur*, Institut Pprime, Université de Poitiers
Jean-Camille CHASSAING, *Professeur*, d'Alembert, Sorbonne Université
Dominique HEITZ, *Chargé de recherche*, INRAE
Rémi MANCEAU, *Directeur de recherche CNRS*, LMAP, CNRS-UPPA
Jacques BOREE, *Professeur*, Institut Pprime, ISAE-ENSMA
Marcello MELDI, *Professeur*, LMFL, ENSAM
Karine TRUFFIN, *Ingénieure de recherche*, IFP Energies Nouvelles

Invité
Examinateur
Rapporteur
Rapporteur

Remerciements

Je tiens tout d'abord à remercier les membres du jury pour l'intérêt qu'ils ont porté à mon travail de thèse, pour leur écoute et leurs questions durant la soutenance de thèse et en particulier les rapporteurs pour leur travail sur le manuscrit de thèse.

J'aimerais ensuite remercier mon encadrement, Jacques, Marcello et Karine. Jacques n'était pas destiné à l'encadrement de ma thèse, il s'est gentiment proposé pour prendre la relève de Marcello lors de son départ pour une autre école doctorale. Plus accoutumé des souffleries que des techniques numériques, il a su m'aiguiller sur des problèmes physiques complexes. Il a été d'un grand soutien pour les vacances, ou encore pour le conseil scientifique de l'école, toujours avec le sourire ! Marcello a dû quitter son poste à l'Ensmat durant le début de ma thèse. Il a cependant toujours pris le temps de m'écouter et de m'aiguiller sur des problèmes poussés d'assimilation de données, et l'a toujours fait avec humour. Son encadrement s'est avéré très précieux pour le déroulement de ma thèse et je n'en serais sûrement pas là sans lui. Enfin, malgré sa gentillesse, Karine n'a jamais hésité à poser des questions précises et poussées indispensables à la bonne avancée de ma thèse. Elle a également su aiguiller le travail notamment dans le choix du dernier cas d'étude. Je les remercie tous les trois à nouveau.

À mes yeux, une thèse réussie est une somme de contributions de toutes les personnes qui entourent le doctorant. La bonne ambiance qui règne dans le laboratoire, et en particulier dans le département fluides a contribué à cela. Je remercie donc toutes les personnes, qui au détour d'un café ou d'une conversation dans les couloirs ont participé à rendre ces trois années meilleures. Particulièrement, j'aimerais remercier Eric pour son aide avec les heures de calcul et François pour sa gentillesse entre autres lors des projets étudiants qui furent une expérience d'enseignement précieuse. Je remercie également Valérie, pour sa disponibilité et sa gentillesse. C'est une personne indispensable au département qui est toujours prête à aider avec le sourire en cas de problèmes.

Que dire de tous les doctorants. Sans eux l'expérience n'aurait pas été la même. J'ai passé trois années formidables à leurs côtés, toujours avec beaucoup d'humour. Je ne me risquerais pas à une liste exhaustive au risque d'en oublier, mais les concernés se reconnaîtront. J'aimerais cependant remercier spécifiquement certains d'entre eux. Je remercie tout d'abord Gabriel pour son accueil et son aide précieuse durant ma première année de thèse. Je remercie également Lucas et en particulier Aurélien, pour toutes ces conversations ces dernières semaines qui m'ont fait me sentir moins seul à affronter la fin de thèse. Enfin, j'aimerais remercier mes deux collègues de bureau doctorants, Antoine et Mickaël. Je n'aurais pas pu rêver mieux pour mes deux dernières années de thèse. Ils ont participé à rendre le quotidien agréable et ont été d'un grand soutien.

Je remercie également la team CONES. C'était un plaisir de travailler avec eux à distance ou lors de mes déplacements. J'ai toujours été bien accueilli et était systématiquement très content de retourner les voir à Lille. Je remercie en particulier Miguel, avec qui j'ai commencé ma thèse, et créé CONES au départ. Tout s'est toujours très bien déroulé et il n'y a jamais eu aucune mésentente entre nous, malgré quelques sacrées prises de têtes durant le développement de CONES !

Enfin, j'aimerais remercier ma famille et mes proches pour leur soutien durant mes trois

années de thèse. Ils ne comprenaient pas toujours ce que je racontais mais suivaient l'évolution de ma thèse avec intérêt. J'aimerais adresser un premier remerciement spécial à ma mère sans qui je ne serais pas là aujourd'hui. Elle m'a toujours soutenu durant toute ma scolarité et c'est grâce à elle que je suis maintenant docteur. Mon deuxième remerciement spécial est bien sûr destiné à ma compagne, Héloïse. Elle a été d'un soutien sans faille durant ces trois années et a contribué sans aucun doute à la réussite de mon doctorat. C'était une personne indispensable pour mon doctorat comme c'est une personne indispensable à ma vie.

Pour finir, je dédie cette thèse à mon père, qui aurait tant aimé me voir devenir docteur.

Table des matières

Table des matières	5
Table des figures	8
Liste des tableaux	16
Liste des algorithmes	17
1 Introduction	19
1.1 Contexte	19
1.2 Les limites des outils expérimentaux et numériques	20
1.3 Le paradigme du jumeau numérique et le pilotage par les données	23
1.4 Annonce de plan	25
2 Approches numériques pour la turbulence	29
2.1 Les équations de Navier–Stokes	29
2.2 Algorithme de résolution et discrétisation	32
2.3 Modélisation de la turbulence	34
2.3.1 La turbulence moyennée	36
2.3.2 Simulation des grandes échelles	38
3 Assimilation de données : méthodes conventionnelles et développements	43
3.1 Inférence bayésienne	44
3.2 Méthodes variationnelles	48
3.2.1 3DVar	49
3.2.2 4DVar	50
3.2.3 Application des méthodes variationnelles en mécanique des fluides	51
3.3 Le filtre de Kalman appliqué à la mécanique des fluides	54
3.3.1 Le filtre de Kalman et les contraintes de linéarité	55
3.3.2 Le filtre de Kalman d’ensemble	57
3.3.3 Corrections des défauts du EnKF	59
3.3.3.1 L’inflation	59
3.3.3.2 La localisation	61
3.3.4 HLEnKF : <i>Hyper-Localized Ensemble Kalman Filter</i>	63
3.3.5 Application des méthodes séquentielles en mécanique des fluides	66

4	C.O.N.ES : Coupling OpenFOAM with Numerical EnvironmentS	71
4.1	Besoins et objectifs	71
4.2	Fonctionnement du couplage	72
4.3	Formats et possibilités de calcul	75
4.4	Exemple : Cavity test case	77
4.4.1	Optimisation avec un EnKF classique	79
4.4.2	Effets des hyper-paramètres	80
4.4.3	(Hyper)-localisation	82
4.5	Consommation et temps d'exécutions	84
5	Écoulement incompressible stationnaire autour d'un corps non profilé	87
5.1	Introduction	87
5.2	Configuration de l'observation et de la simulation RANS	89
5.2.1	Modèle de bâtiment et essais en soufflerie	89
5.2.2	Observation : données expérimentales de la soufflerie	90
5.2.3	Simulation RANS	91
5.3	Procédure d'assimilation de données	92
5.3.1	Cas A : EnKF classique	96
5.3.2	Cas B : EnKF avec localisation de la covariance	97
5.3.3	Cas C : EnKF avec localisation et inflation	98
5.4	Résultats	100
5.4.1	Champ de vitesse	100
5.4.2	Champ de pression	103
5.5	Conclusions	106
6	Écoulement incompressible statistiquement stationnaire académique	109
6.1	Introduction	109
6.2	Cas test et configuration de la procédure d'AD	111
6.2.1	Écoulement turbulent en canal plan, $Re_\tau \approx 550$	111
6.2.2	Stratégie d'assimilation de données	115
6.2.2.1	Procédure d'AD 1 (DA-LES1) : optimisation du coefficient C_k basé sur la simulation BL-LES	117
6.2.2.2	Procédure d'AD 2 (DA-LES2) : décomposition spatiale du coefficient C_k	117
6.2.2.3	Procédure d'AD 3 (DA-LES3) : optimisation du coefficient C_k basé sur la simulation VD-LES	118
6.3	Prédiction des caractéristiques statistiques avec l'AD séquentielle	118
6.4	Synchronisation des champs de l'écoulement	127
6.5	Conclusion	131
7	Écoulement interne compressible et instationnaire d'une géométrie de moteur	133
7.1	Introduction	133
7.2	Configurations numériques	135
7.2.1	Présentation de la géométrie et validation dans le cas stationnaire	135

7.2.2	Simulation de référence haute fidélité (LES-HF)	137
7.2.3	Simulation basse fidélité (LES-LF)	141
7.3	Calibration de la condition d'entrée	142
7.3.1	Mise en place de l'assimilation de données	142
7.3.2	Structure hyper-paramétrique et conditions initiales	146
7.3.3	Résultats	146
7.4	Synchronisation et reconstruction de structures locales	152
7.4.1	Détails de configuration	152
7.4.2	Capacités de synchronisation	157
7.5	Conclusion	163
8	Conclusion et perspectives	167
A	Annexe : Chapitre 6	173
A.1	Utilisations d'informations physiques multiples dans la procédure d'AD	173
A.2	Détails supplémentaires sur la synchronisation	174
A.3	Consommation des calculs	175
B	Annexe : Validation du filtre de Kalman d'ensemble hyper-localisé	177
C	Annexe : Chapitre 7	183
C.1	Calibration de la condition d'entrée	183
C.1.1	Évolution de l'écart type des paramètres	183
C.1.2	Estimation paramétrique à partir d'un jeu d'observations différent	184
C.2	Reconstruction et synchronisation de structures locales	185
C.2.1	Synchronisation des composantes u_y et u_z et de la pression p	185
C.2.2	Capacités de synchronisation pour un nombre d'observations plus élevé	189
C.3	Temps de calcul et consommation	191
	Bibliographie	193

Table des figures

1.1	Schéma adapté et traduit de [101]. Celui-ci illustre les différentes sources d'erreur menant aux écarts rencontrés entre une réalisation expérimentale et numérique d'un même écoulement. ϵ_{mes} : erreur due à la précision des instruments de mesure / ϵ_{CL} : erreur due au choix des conditions aux limites numériques / ϵ_{mod} : erreur due aux hypothèses de modèle / ϵ_{num} : erreur due aux méthodes numériques employées (discrétisations, schémas...)	21
1.2	Schéma illustrant le paradigme du jumeau numérique.	23
1.3	Schéma illustrant les différentes sources d'erreur menant aux écarts rencontrés entre une réalisation expérimentale et numérique d'un même écoulement avec assimilation de données. Pour une inférence optimale de l'écoulement, ϵ_{CL} et ϵ_{mod} tendent vers 0.	26
2.1	Schéma adapté et traduit de [38] illustrant le mouvement des structures turbulentes à gauche et l'évolution de la première composante de vitesse en un point en fonction du temps à droite pour différents modèles de turbulence.	36
3.1	Schéma adapté et traduit de [37]. Celui-ci montre de manière unifiée les hypothèses conduisant aux algorithmes principaux d'assimilation de données.	45
3.2	Schéma représentant le filtre de Kalman d'ensemble. L'état du modèle u_k^f est mis à jour via la phase d' <i>analyse</i> de l'algorithme EnKF en prenant en compte les observations α_k . Chaque ligne grise représente la trajectoire d'un membre de l'ensemble.	58
3.3	Schéma représentant l'inflation utilisée pour le filtre de Kalman d'ensemble. Chaque ligne grise représente la trajectoire d'un membre de l'ensemble. L'augmentation de la variance de l'ensemble est représentée par le trapèze gris écartant les 5 membres de la moyenne.	62
3.4	Schéma représentant les différentes méthodes de localisation sur une grille cartésienne. Les points rouges représentent les observations. En bleu : l'EnKF est exécuté sur un domaine réduit. En rouge foncé : la localisation de covariance pondère la correction apportée par l'EnKF dans une zone donnée. En vert : Exemple du comportement du LETKF pour une des mailles du domaine.	64
3.5	Schéma représentant le principe de l'hyper-localisation. Les points rouges représentent les observations. L'EnKF local classique s'exécute sur un domaine réduit contenant les observations. Pour l'hyper-localisation, un EnKF est exécuté autour de chaque observation. Les régions sont fusionnées si elles s'intersectent.	65

4.1	Schéma de fonctionnement de CONES	73
4.2	Géométrie du cas d'étude de la cavité 2D. L'écoulement est entraîné par la vitesse imposée sur la face supérieure.	78
4.3	Groupes d'observations utilisées. Groupe 1 en rouge, groupe 2 en bleu	79
4.4	Évolution de la vitesse imposée sur la face supérieure. En rouge évolutions pour le groupe 1 et en bleu évolutions pour le groupe 2.	80
4.5	Évolution temporelle du paramètre optimisé. En bleu clair $\sigma_\alpha = 1\%$ et en bleu foncé $\sigma_\alpha = 50\%$	81
4.6	Évolution temporelle du paramètre optimisé. En bleu clair $\sigma_\alpha = 1\%$ et $\lambda_p = 1\%$. En bleu foncé $\sigma_\alpha = 20\%$	82
4.7	Évolution temporelle du paramètre optimisé pour $\sigma_\alpha = 20\%$. En bleu clair aucune localisation, en bleu foncé localisation de la covariance, en violet hyper-localisation.	83
4.8	Évolution temporelle du paramètre optimisé avec hyper-localisation. En bleu clair $\sigma_\alpha = 1\%$, en bleu foncé $\sigma_\alpha = 1\%$ et $\lambda_p = 1\%$, en violet $\sigma_\alpha = 20\%$	83
5.1	Section d'essai de la soufflerie utilisée pour produire les données expérimentales. a) Visualisation du modèle de bâtiment et de son installation dans la soufflerie et (b) forme du profil de vitesse moyenne et de l'intensité de la turbulence dans la direction de l'écoulement x	90
5.2	Géométrie du bâtiment de grande hauteur avec a) les dimensions principales et le système de coordonnées ; b) la vue supérieure avec l'emplacement des prises de pression ; c) les façades avec l'emplacement des prises de pression ; d) les observations de vitesse mesurées au-dessus du toit.	90
5.3	Vue du maillage utilisé pour les calculs RANS. La gauche montre le plan vertical central et un plan horizontal à $H/2$. La droite montre le plan vertical central. La direction de l'écoulement asymptotique est indiquée par la flèche blanche.	91
5.4	Région rectangulaire utilisée pour la localisation : les capteurs de pression sont représentés en rouge et les captures de vitesse sont affichées en vert.	96
5.5	Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF sans inflation et localisation.	97
5.6	Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF sans inflation mais avec localisation.	98
5.7	Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF avec inflation et localisation.	99
5.8	Profils de vitesse verticale et dans le sens de l'écoulement au-dessus des emplacements marqués en rouge sur le toit (Points : 20, 22, 36, 38, 54) : a) comparaison entre les données de soufflerie (WT), le modèle $\mathcal{K} - \varepsilon$ et le modèle $\mathcal{K} - \varepsilon$ augmenté par AD, b) comparaison entre les données de soufflerie (WT), le modèle $\mathcal{K} - \omega$ et le modèle $\mathcal{K} - \omega$ augmenté par AD. Ici, u_{ref} représente la vitesse obtenue à 1 m en amont du bâtiment à la hauteur H dans chaque cas test ($u_{ref} \approx 13$ m/s dans les simulations et $u_{ref} \approx 16$ m/s dans la soufflerie).	101

5.9	Lignes de courant colorées par le champ de vitesse moyenné dans le temps obtenu sur un plan $x - z$ au centre du bâtiment. Des comparaisons sont effectuées entre une simulation des grandes échelles (LES validée) [142], un modèle $\mathcal{K} - \omega$ SST standard et augmenté par AD, $\mathcal{K} - \varepsilon$ optimisé par Zhao et al. [154], $\mathcal{K} - \varepsilon$ standard et augmenté par AD. Ici, u_{ref} représente la vitesse obtenue 1 m en amont du bâtiment à la hauteur H dans la soufflerie.	103
5.10	Zoom sur les lignes de courant au-dessus du toit. La flèche rouge indique la position de rattachement de la bulle de recirculation pour le modèle $\mathcal{K} - \varepsilon$ standard et la flèche verte fournit la même information pour la simulation LES validée. La ligne bleue indique la position de rattachement pour la simulation présentée dans chaque image. Ici u_{ref} représente la vitesse obtenue 1 m en amont du bâtiment à la hauteur H dans la soufflerie.	104
5.11	Énergie cinétique turbulente \mathcal{K} obtenue sur un plan $x - z$ au centre du bâtiment. Des comparaisons sont effectuées entre une simulation des grandes échelles (LES validée) [142], un modèle $\mathcal{K} - \omega$ SST standard et augmenté par AD, $\mathcal{K} - \varepsilon$ optimisé par Zhao et al. [154], $\mathcal{K} - \varepsilon$ standard et augmenté par AD. Ici, $u'_{ref} = I_u u_{ref}$ est obtenu à partir de l'intensité turbulente dans le sens de l'écoulement $I_u = 13\%$ et u_{ref} mesurée à 1 m en amont du bâtiment à la hauteur H	105
5.12	Nuage de points du coefficient de pression moyen C_p pour les simulations numériques avec comparaison des mesures de performance avec les données expérimentales. Rangée supérieure : modèle $\mathcal{K} - \varepsilon$, rangée inférieure : modèle $\mathcal{K} - \omega$ SST. Colonne de gauche : modèle RANS <i>a priori</i> , colonne de droite : RANS augmenté par AD.	106
5.13	Histogrammes représentant l'erreur dans la prédiction du coefficient de pression C_p par rapport aux essais. Les résultats numériques sont échantillonnés aux emplacements des prises de pression. Rangée supérieure : modèle $\mathcal{K} - \varepsilon$, rangée inférieure : modèle $\mathcal{K} - \omega$ SST. Colonne de gauche : modèle RANS <i>a priori</i> , colonne de droite : modèle RANS augmenté par AD.	107
5.14	Coefficient de pression moyen aux points de prise de pression le long des lignes rouges marquées sur le toit : a) comparaison entre les données de la soufflerie (WT), du modèle $\mathcal{K} - \varepsilon$ et du modèle $\mathcal{K} - \varepsilon$ augmenté par AD ; b) comparaison entre les données de la soufflerie (WT), du modèle $\mathcal{K} - \omega$ SST et du modèle $\mathcal{K} - \omega$ SST augmenté par AD.	108
6.1	Taille du domaine d'étude.	112
6.2	Profil de vitesse moyenné dans le sens de l'écoulement u^+ pour les simulations de base Smagorinsky LES (\star), avec van Driest (\bullet) et DNS ($-$) comparé aux données de la littérature ($--$).	113
6.3	Distribution du rapport $\bar{\nu}_{SGS}/\nu$ dans la direction verticale pour les LES avec (\bullet) et sans van Driest (\star).	114
6.4	Composantes du tenseur de Reynolds pour les simulations de base Smagorinsky LES (\star), avec van Driest (\bullet) et DNS ($-$) comparé aux données de la littérature ($--$).	114

6.5	Position des capteurs utilisés pour obtenir l'observation utilisée dans l'assimilation de données. La région en rouge correspond au découpage physique pour l'EnKF, c'est-à-dire la région dans l'espace où l'augmentation d'état est effectuée.	116
6.6	En bleu : profils des cinq fonctions gaussiennes utilisées pour représenter C_k . En rouge : distribution de C_k . Gauche : distribution initiale <i>a priori</i> . Droite : distribution optimisée par AD	119
6.7	Profils de vitesse sur le demi-canal pour DA-LES1 (▶), DA-LES2 (+), BL-LES (★) et DNS (−). L'adimensionnement est réalisée en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l'AD.	121
6.8	Composantes du tenseur de contrainte de Reynolds pour DA-LES1 (▶), DA-LES2 (+), BL-LES (★) et DNS (−). L'adimensionnement est réalisée en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l'AD.	121
6.9	Profils de vitesse sur le demi-canal pour DA-LES1 (▶), DA-LES2 (+), BL-LES (★) et DNS (−). La zone grise représente la région où la localisation physique est appliquée pour l'AD.	122
6.10	Composantes du tenseur de contrainte de Reynolds pour DA-LES1 (▶), DA-LES2 (+), BL-LES (★) et DNS (−). La zone grise représente la région où la localisation physique est appliquée pour l'AD.	122
6.11	Distribution du rapport \bar{v}_{SGS}/ν dans la direction verticale pour BL-LES (★), VD-LES (●), DA-LES2 (+) and DA-LES3 (×)	123
6.12	Profils de vitesse de demi-canal pour DA-LES2 (+), DA-LES3 (×), VD-LES (●) et DNS (−). L'adimensionnement est réalisé en utilisant la vitesse de frottement u_τ obtenue dans la simulation DNS. La zone grise représente la région où la localisation physique (clipping) est appliquée à des fins de DA.	124
6.13	Composantes du tenseur de contrainte de Reynolds pour DA-LES2 (+), DA-LES3 (×), VD-LES (●) et DNS (−). L'adimensionnement est réalisé en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l'AD.	125
6.14	Spectres de puissance du champ de vitesse obtenus par transformée en ondelettes de Morlet. Les résultats sont présentés pour la DA-LES2 (−), BL-LES (−) et la DNS (−). Pour cette dernière, des résultats supplémentaires obtenus par FFT sont présentés en gris.	126
6.15	Synchronisation du champ de vitesse. Les données en bleu représentent l'observation, tandis que les lignes noires indiquent la prédiction de l'ensemble. Les résultats sont présentés pour (colonne de gauche) le capteur 1, (colonne centrale) le capteur 2 et (colonne de droite) le capteur 3. La valeur de l'inflation d'état est égale à (première ligne) 0%, (deuxième ligne) 5%, (troisième ligne) 15% et (quatrième ligne) 25%.	130
6.16	Courbes de Φ : Du vert au violet - 0%, 5%, 15% et 25% d'inflation. La courbe rouge montre la limite moyenne obtenue en comparant les valeurs des simulations DA-LES à 1000 temps d'observations.	131

7.1	Schéma de la géométrie simplifiée de moteur étudiée comportant une seule soupape.	136
7.2	Zoom du maillage structuré utilisé pour la simulation haute-fidélité de référence. Le maillage est affiché en bleu clair sur la soupape et bleu foncé sur un plan en $z = 0$ pour le domaine d'étude.	137
7.3	Profils radiaux de vitesses axiales pour deux section $x = 0.124$ mm à gauche et $x = 0.126$ mm à droite. En orange : données expérimentales de Thobois et al. [130]. En bleu : données issues de la simulation LES haute-fidélité	138
7.4	Placement des capteurs utilisés pour mesurer l'évolution de la condition d'entrée dans la conduite d'admission. La section 2D $x = 0.124$ m régulièrement utilisée dans ce travail est indiquée en vert.	138
7.5	Évolution de la vitesse axiale u_x adimensionnée par la vitesse débitante u_b et de la pression p adimensionnée par la pression atmosphérique p_{atm} en fonction du temps. Les valeurs sont données pour 3 capteurs disposés le long de la conduite d'admission de la LES-HF en $x = 0$ bleu clair, 0.046 bleu foncé, 0.092 violet, au centre de la conduite d'entrée en $z = 0$ et $y = 0.0125$.	139
7.6	Champ de vitesse de la simulation haute fidélité de référence en $t = 9t_p$ sur une section telle que $z = 0$. La pression est représentée sur la surface de la soupape.	140
7.7	Les points oranges représentent la position des capteurs pour la mesure des données. Ils sont répartis sur 5 plans : $x = 0.124$, $x = 0.144$, $x = 0.164$, $y = 0$ et $z = 0$	140
7.8	Champs de viscosités de sous-maille moyens adimensionnés par la viscosité cinématique initiale ν_{SGS}/ν_{init} sur un plan 2D en $z = 0$ pour les configurations haute et basse fidélité.	141
7.9	Évolution de la vitesse axiale u_x adimensionnée par la vitesse débitante u_b en fonction du temps pour les configurations HF à gauche et LF à droite. Les valeurs sont données pour 3 capteurs disposés le long de la conduite d'admission de la LES-LF en $x = 0$ bleu clair, 0.046 bleu foncé, 0.092 violet, au centre de la conduite d'entrée en $z = 0$ et $y = 0.0125$.	142
7.10	Champ de vitesse de la simulation haute fidélité de référence en $t = 9t_p$ pour deux sections tel que $y = 0$ et $z = 0$ délimitées en vert. Les points rouges correspondent au capteurs d'observations sélectionnées pour la calibration de la condition d'entrée. Ces capteurs se situent sur des plans 2D $y = 0$ et $z = 0$ sur le chemin emprunté par le jet de soupape.	144
7.11	Schéma illustrant le temps advectif de la transmission de l'information d'entrée mise à jour aux zones d'observations.	145
7.12	Évolution de la condition d'entrée de vitesse axiale pour la distribution a priori 1 ligne du haut et a priori 2 ligne du bas. La courbe rouge correspond à la condition d'entrée de référence utilisée pour la LES-HF. La vitesse est adimensionnée par la vitesse débitante de référence u_b , le temps par la période de l'onde sinusoïdale de référence. Les courbes grises (gauche) correspondent aux 35 membres d'ensemble. La courbe bleue (droite) correspond à la moyenne des conditions des membres d'ensemble, la zone grisée indique l'écart type par rapport à la moyenne.	147

7.13	Évolution de la condition d'entrée de vitesse axiale pour la distribution <i>a priori</i> 1. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleu : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.	149
7.14	Évolution de la condition d'entrée de vitesse axiale pour la distribution <i>a priori</i> 2. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleu : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.	150
7.15	Évolution des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution <i>a priori</i> 1. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4. .	150
7.16	Évolution des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution <i>a priori</i> 2. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4. .	151
7.17	Courbe bleu : onde sinusoïdale résultante de la condition d'entrée pour des jeux de paramètres isolés. Courbe rouge : onde de référence. Ligne du haut : <i>a priori</i> 1. Ligne du bas : <i>a priori</i> 2. De gauche à droite : jeux de paramètres aux temps $t = 6t_p/9t_p/12t_p$.	151
7.18	Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les positions des capteurs du jeu SE-set1 d'observations sont indiquées par les points rouges. . . .	154
7.19	Erreur quadratique moyenne Φ en fonction de la période inférée. Courbe noire : LES-PE2. Courbe rouge : LES-OW250 avec 5% de confiance sur les observations sans inflation. Courbe Orange : LES-OW50 avec 5% de confiance sur les observations sans inflation. Courbes vertes : LES-OW10 avec un niveau de confiance sur les observations de 20% pour deux niveaux d'inflations de 0% pour la couleur claire et 5% pour la couleur foncée. Courbes bleues : LES-OW10 avec un niveau de confiance sur les observations de 5% pour deux niveaux d'inflations de 0% pour la couleur claire et 5% pour la couleur foncée.	156
7.20	Synchronisation de la composante axiale de la vitesse u_x adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (<i>a priori</i>) et les vertes dans le cas associé à chaque configuration (<i>a posteriori</i>).	158
7.21	Vitesses axiales de la simulation de référence LES-HF mesurées par les capteurs du plan $x = 0.124$. Gauche : mesures des capteurs. Droite : interpolation des valeurs sur une grille structurée en fonction de r/R et θ	159

7.22	Cartographie des modes principaux de la transformée de Fourier azimutale de la vitesse axiale $u_x - \overline{u_x}$ au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seul la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.	160
7.23	Transformées de Fourier inverses du mode 10 de la composante $u_x - \overline{u_x}$ au temps $t = 13t_p$. Les vitesses sont normalisées par la vitesse débitante de référence u_b	162
7.24	Cartographie des modes principaux de la transformée de Fourier azimutale de la vitesse u_y au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seul la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.	163
7.25	Champs de vitesses des différentes configurations pour une section $z = 0$ de la géométrie. Les vitesses sont normalisées par la vitesse débitante de référence u_b	164
A.1	Profils de vitesse sur le demi-canal pour DA-LESA (▶), DA-LES1 (+), BL-LES (*) et DNS (×)	174
A.2	Test de sensibilité pour la fonction de coût Φ pour les 408 capteurs utilisés dans l'AD. (a) Différents niveaux d'inflation des paramètres sont utilisés, de la couleur claire à la couleur foncée : 0%, 2% et 5% d'inflation. (b) Différentes valeurs pour la confiance imposée dans l'observation sont présentées, du vert au violet : 0.5%, 1%, 5% et 10% de confiance	175
B.1	Évolution du coefficient C_k du modèle Smagorinsky au cours des 100 premières phases d'assimilations ($12t_A$) sans aucune inflation. La courbe rouge correspond à l'algorithme du EnKF et la courbe verte au HLEnKF. La zone cible grisée correspond aux valeurs convergées établies avec l'EnKF et de l'inflation dans la Sec. 6.3.	178
B.2	Test de sensibilité pour la fonction de coût Φ pour les 408 capteurs utilisés dans l'AD à gauche (a) et non utilisés à droite (b). Deux niveaux d'inflation des paramètres sont utilisés, de la couleur claire à la couleur foncée : 0% et 5% d'inflation. Les couleurs bleues correspondent aux résultats de l'EnKF et les couleurs vertes à ceux du HLEnKF.	179
B.3	Synchronisation du champ de vitesse pour le HLEnKF sans inflation avec un niveau de localisation de covariance égal entre les procédures. Les données en bleu représentent l'observation, tandis que les lignes noires indiquent la prédiction de l'ensemble. Les résultats sont présentés pour (colonne de gauche) le capteur 1, (colonne centrale) le capteur 2 et (colonne de droite) le capteur 3. Aucune correction n'est effectuée sur la première ligne. La deuxième ligne correspond au EnKF sans localisation physique. La dernière ligne correspond au HLEnKF.	180
C.1	Évolution de l'écart type des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution <i>a priori</i> 1. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4.	183

C.2	Évolution de l'écart type des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution <i>a priori</i> 2. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4.	184
C.3	Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les points rouges correspondent au second jeu de capteurs d'observations sélectionnées pour la calibration de la condition d'entrée.	184
C.4	Évolution de la condition d'entrée de vitesse axiale pour la distribution <i>a priori</i> 2 avec un jeu d'observations différent. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleue : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.	185
C.5	Synchronisation de la composante de la vitesse u_y adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (<i>a priori</i>) et les vertes dans le cas associé à chaque configuration (<i>a posteriori</i>).	186
C.6	Synchronisation de la composante de la vitesse u_y adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (<i>a priori</i>) et les vertes dans le cas associé à chaque configuration (<i>a posteriori</i>).	187
C.7	Synchronisation de la pression p adimensionnée par p_{atm} en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (<i>a priori</i>) et les vertes dans le cas associé à chaque configuration (<i>a posteriori</i>).	188
C.8	Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les positions des capteurs du jeu SE-set2 d'observations sont indiquées par les points rouges.	189
C.9	Synchronisation de la composante axiale de vitesse u_x adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (<i>a priori</i>) et les vertes dans le cas associé à chaque configuration (<i>a posteriori</i>).	190
C.10	Cartographie des modes principaux de la transformée de Fourier azimuthale de la vitesse axiale $u_x - \bar{u}_x$ au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seule la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.	191

Liste des tableaux

4.1	Contenu du fichier <i>obs_field.txt</i> . N_{o_u} le nombre de capteurs de vitesse et N_{o_p} le nombre de capteurs de pression dans l'observation.	77
4.2	Contenu du fichier <i>obs_coordinates.txt</i> . Coordonnées des capteurs du fichier <i>obs_field.txt</i>	78
4.3	Configuration hyper-paramétrique des tests effectués sur la Cavité	81
4.4	Complexité du gain et temps d'exécution d'une phase d' <i>analyse</i> complète pour le cas du canal turbulent chapitre 6.	86
5.1	Coefficients globaux des modèles de turbulence RANS à optimiser via l'AD. Colonne de gauche à droite : valeurs classiques dans la littérature, caractéristiques de la distribution gaussienne bornée utilisée pour générer les <i>a priori</i> , valeurs optimisées.	94
5.2	Comparaison entre les constantes conventionnelles du modèle RANS $\mathcal{K} - \varepsilon$ et les paramètres initiaux employés pour l'EnKF ($N_e = 40$).	95
5.3	Coefficients de modèle optimisés $\mathcal{K} - \varepsilon$ obtenus avec les différentes stratégies basées sur l'EnKF.	99
6.1	Détail des maillages utilisés pour les simulations LES ainsi que pour la DNS de référence.	113
6.2	Détails des capteurs utilisés pour étudier la synchronisation.	128
7.1	Distributions <i>a priori</i> utilisées pour l'estimation paramétrique de la procédure d'assimilation de données.	145
7.2	Résumé des configurations étudiés avec l'estimation d'état pour la synchronisation des champs comparés à la distribution <i>a priori</i> LES-PE2 utilisée pour ces configurations.	154
7.3	Nombre de phases d'assimilation subies par une particule fluide dans les zones d'observations en fonction de la vitesse du jet et des configurations étudiées.	156
7.4	Capteurs utilisés pour le contrôle de la synchronisation des vitesses inférées avec l'observation	157

Liste des Algorithmes

1	Algorithme PIMPLE incompressible utilisé dans OpenFOAM	34
2	Algorithme du Filtre de Kalman d'Ensemble (EnKF)	60
3	Algorithme du Filtre de Kalman d'Ensemble hyper-localisé (HLEnKF)	67

Introduction

Sommaire

1.1	Contexte	19
1.2	Les limites des outils expérimentaux et numériques	20
1.3	Le paradigme du jumeau numérique et le pilotage par les données	23
1.4	Annonce de plan	25

1.1 Contexte

Les émissions des transports

En France, la part des émissions de gaz à effet de serre attribuée aux transports tous secteurs confondus représente 30%. C'est la première source d'émissions, devant l'industrie manufacturière et de construction ou encore devant l'agriculture [20]. Bien que la crise de la COVID-19 ait freiné les émissions mondiales, celles dues au secteur des transports continuent leur progression constante s'éloignant des objectifs fixés pour atteindre le *Net Zero Emission* d'ici à 2050 [54]. La réduction des gaz à effets de serres et particules fines est devenue un enjeu de premier plan pour ces secteurs. Dans "Perspectives des Transports 2023", le Forum International des Transports (FIT, organisme de l'OCDE) [40] précise qu'un scénario de réduction des émissions du secteur des transports repose principalement sur le recours aux nouvelles technologies de propulsion de véhicules et aux carburants alternatifs bas-carbone. L'emploi de carburants bas-carbone est primordial dans les secteurs aéronautiques, maritimes et automobiles. Le secteur aérien se tourne par exemple vers ce type de carburants, produit à partir de biomasse ou synthétiquement à partir d'hydrogène. Le secteur de l'automobile voit l'émergence de nouvelles technologies telles que l'électrification des véhicules, l'hybridation ou encore l'arrivée de carburants alternatifs.

Bien que la part des véhicules électriques ait atteint 14% des ventes mondiales en 2022 [54], les véhicules équipés de moteurs à combustion internes représentent encore plus de 90% du parc de véhicules mondial [54]. L'hydrogène apparaît comme carburant alternatif viable pour ce secteur. Le moteur à combustion hydrogène est une technologie prometteuse pour laquelle il faudra surmonter de nouveaux défis de modélisation et de conception dus à la complexité d'une telle installation.

Lien avec la mécanique des fluides

Dans ce contexte, le domaine de la mécanique des fluides présente encore beaucoup de défis à relever. La représentation d'évènements souvent complexes est encore difficile à caractériser et à représenter. La calibration des moteurs à combustion interne est trop souvent erronée du fait de la tolérance exigée par la prédiction difficile de la combustion en régime turbulent. L'apparition de technologies modernes de moteurs à combustion et de nouveaux carburants favorisent la sensibilité de la combustion aux caractéristiques de l'écoulement. Cela contribue à l'apparition de variations cycle à cycle de la combustion dans le moteur, responsable *in fine* d'une baisse de performance pour ce dernier. Par ailleurs, les cycles les plus extrêmes peuvent entraîner une baisse de l'efficacité de la combustion se traduisant en rejets de polluants [94, 1]. L'origine multifactorielle de ces variations est un sujet d'actualité dans la recherche [27, 28]. Elle établit un lien entre vitesse de combustion et aérodynamique interne, en particulier du fait des interactions complexes entre les jets de soupape durant l'admission, de la formation du mouvement de rotation appelé tumble et de son interaction complexe avec les parois.

La représentation paramétrique de tels phénomènes reste difficile à étudier via les outils expérimentaux malgré leur faculté à identifier les facteurs principaux mis en jeu. Les outils numériques traditionnels atteignent la limite fixée par leurs modèles et le coût de simulations complexes. De nouveaux outils numériques sont nécessaires, leur objectif est double :

- Augmenter la précision des résultats pour un coût équivalent aux modèles existants.
- Être capable de reproduire des phénomènes locaux issus de systèmes réels.

Ce deuxième point est essentiel pour la prédiction d'évènements extrêmes pouvant par exemple conduire à de fortes variations cycle à cycle dans un moteur. Pour comprendre quels outils peuvent permettre d'atteindre ces objectifs, il est important de comprendre quelles sont les limites des outils actuels.

1.2 Les limites des outils expérimentaux et numériques

Les sources d'erreurs

Dans "Turbulent flows", S.B.Pope [101] classe les sources d'erreur existantes entre les mesures expérimentales d'un système physique et son calcul via la simulation numérique. Si le propos est à la validation des modèles numériques par comparaisons avec les données expérimentales, cette classification est particulièrement pertinente pour identifier les sources qu'il est possible de minimiser via de nouveaux outils. Ces erreurs sont résumées sur la Fig. 1.1.

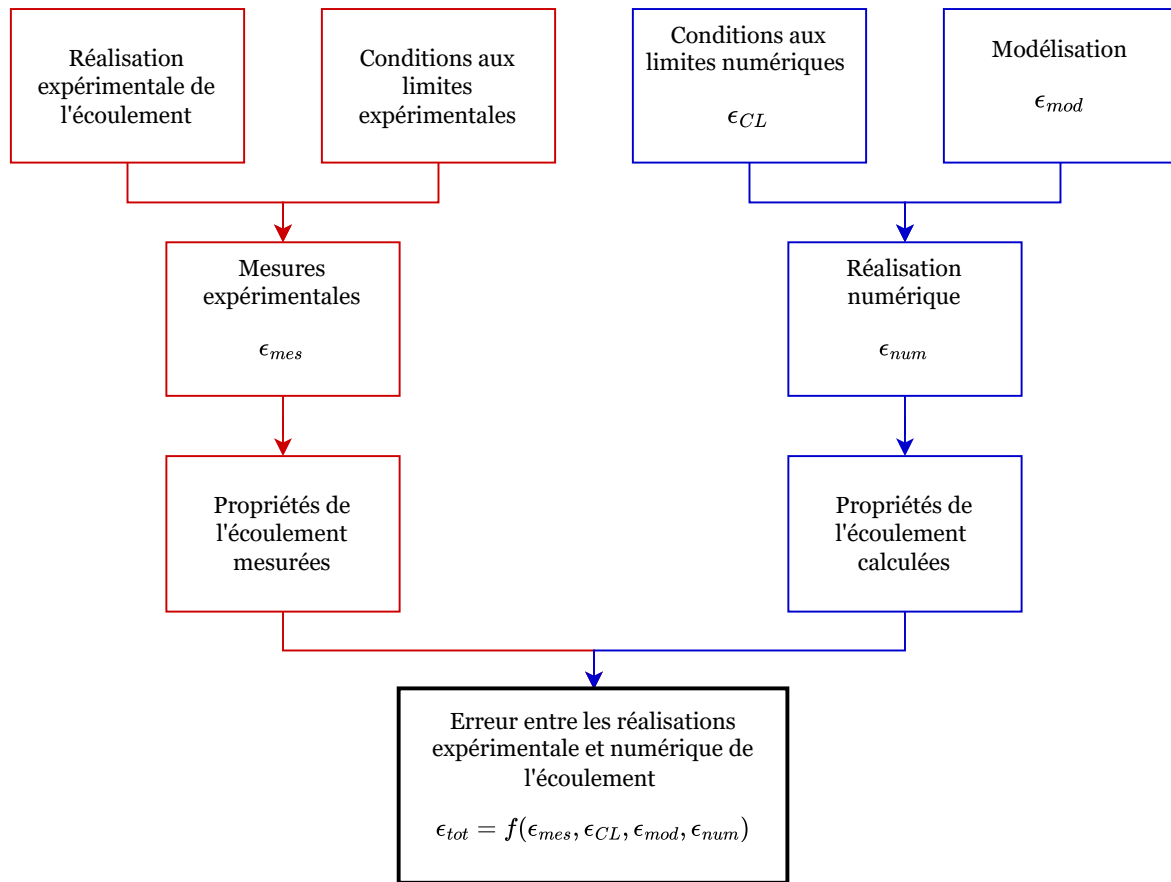


FIGURE 1.1 : Schéma adapté et traduit de [101]. Celui-ci illustre les différentes sources d'erreur menant aux écarts rencontrés entre une réalisation expérimentale et numérique d'un même écoulement. ϵ_{mes} : erreur due à la précision des instruments de mesure / ϵ_{CL} : erreur due au choix des conditions aux limites numériques / ϵ_{mod} : erreur due aux hypothèses de modèle / ϵ_{num} : erreur due aux méthodes numériques employées (discrétisations, schémas...)

- Erreur liée aux mesures
- Erreur liée aux imprécisions du modèle
- Erreur liée aux conditions aux limites numériques
- Erreur liée aux schémas numériques

Approches expérimentales

Les approches expérimentales, qui reposent sur des mesures pouvant être obtenues par diverses techniques, telles que les capteurs de pression et les fils chauds, permettent de fournir une représentation presque exacte des caractéristiques de l'écoulement sous la forme de mesures de pression et de vitesses. Toutefois, ces données sont généralement locales dans l'espace et le temps, et une représentation complète des écoulements est trop coûteuse. L'exactitude des

mesures est de plus soumise à la tolérance dans la précision des instruments, qui est une source d'erreur non négligeable.

D'autres méthodes de mesures comme la mesure par PIV (*Particle Image Velocimetry*) permettent l'obtention de champs de vitesse instantanés sur un plan 2D. Leur calibration est complexe et englobe différentes problématiques (réglage optique, ensemencement...). Elle peut également s'avérer coûteuse du fait des ressources nécessaires au traitement des données, qui peut également être une source d'erreur due au traitement statistique appliqué. Par ailleurs, la résolution à la fois temporelle et spatiale de la méthode peut également être limitante pour l'étude d'évènements locaux en temps et en espace. Sa mise en place dans des géométries complexes telles qu'un moteur à combustion interne nécessite des ressources non négligeables pouvant peser sur le coût d'une étude. Il faut donc se tourner vers l'approche numérique pour l'étude de champs complets.

Approches numériques

Les ressources de calcul nécessaires pour représenter complètement les écoulements turbulents par simulation numérique directe (DNS) sont prohibitives pour les nombres de Reynolds observés dans des applications réalistes. La mécanique des fluides numérique (CFD) d'ordre réduit, de par la modélisation de la turbulence basée sur les équations de Navier-Stokes moyennées (RANS) [148, 9] ou filtrées (LES) [119, 101], peut fournir une description statistique et/ou instationnaire des écoulements autour de systèmes complexes avec des ressources plus abordables. La précision de cette prédiction est cependant fortement affectée par les caractéristiques du modèle de turbulence nécessaire pour fermer de telles équations dynamiques. Les modèles, pilotés par un certain nombre de coefficients classiquement déterminés par des approches empiriques, échouent généralement à représenter les interactions de différents phénomènes physiques déclenchés par la turbulence. Ils sont la source d'écarts, notés ϵ_{mod} sur la Fig. 1.1, par rapport aux mesures du système réel.

Un système réel est également soumis à une forte variabilité des réalisations d'une même étude due aux variations infinitésimales des conditions initiales et des conditions aux limites amplifiées par la turbulence [101]. Les approches numériques, qui permettent la simulation d'effets instationnaires (LES, DNS), n'autorisent pas la représentation de cette variabilité, à conditions initiales et limites fixées. Le choix de ces conditions est de plus soumis à deux autres problèmes : (1) les conditions aux limites de systèmes réels ne sont pas toujours connues et doivent être estimées, (2) les contraintes de modélisations ne permettent pas toujours la simulation du système réel complet. Ces incertitudes se traduisent par des sources d'erreurs potentielles pouvant affecter la précision des simulations numériques (notées ϵ_{CL} sur la Fig. 1.1).

Enfin, la simulation numérique contient également des erreurs intrinsèques à la numérisation du problème de l'étude d'un système réel : l'erreur d'arrondi due à la précision de la machine, l'erreur de troncature des schémas numériques ou encore l'erreur de discrétisation de la résolution de maillage utilisée. Ces erreurs sont regroupées dans l'erreur numérique ϵ_{num} sur la Fig. 1.1. La réduction de cette erreur passe par l'utilisation de schémas numériques précis ou encore d'une résolution accrue. Elle ne sera donc pas le sujet d'optimisation.

Ainsi, la prédiction d'effets instationnaires réalistes et d'évènements extrêmes locaux est

souvent dépendante de modèles aux coûts inabordables. Elle est de plus soumise à la caractérisation exacte de différents paramètres tels que les conditions aux limites et initiales de la simulation. Il est par ailleurs difficile de reproduire une réalisation unique de l'écoulement observé sur le système réel, sans recourir à la reconstruction exacte des structures observées. De nouveaux outils sont donc nécessaires à la reproduction et la résolution des phénomènes complexes et multifactoriels mis en jeu.

1.3 Le paradigme du jumeau numérique et le pilotage par les données

Le paradigme du jumeau numérique

Après la mécanisation et l'invention de machines à vapeur, de l'électricité puis de l'informatique, l'Industrie 4.0 représente la quatrième révolution industrielle en cours et à venir. Elle repose sur l'arrivée de technologies intelligentes basées sur l'information [5]. La simulation numérique y est vue comme un outil pouvant avoir un impact durable sur l'économie, l'environnement et les progrès sociaux, aux côtés du big data ou des technologies de cloud. Dans le cadre de cette révolution numérique, les applications envisagées prédisent et contrôlent des configurations réelles, généralement appelées jumeaux physiques, à l'aide d'un homologue numérique, le jumeau numérique [104, 118]. Semeraro et al. donnent la définition suivante du paradigme du jumeau numérique illustré sur la Fig. 1.2 :

« A set of adaptive models that emulate the behaviour of a physical system in a virtual system getting real time data to update itself along its life cycle. The digital twin replicates the physical system to predict failures and opportunities for changing, to prescribe real time actions for optimizing and/or mitigating unexpected events observing and evaluating the operating profile system »

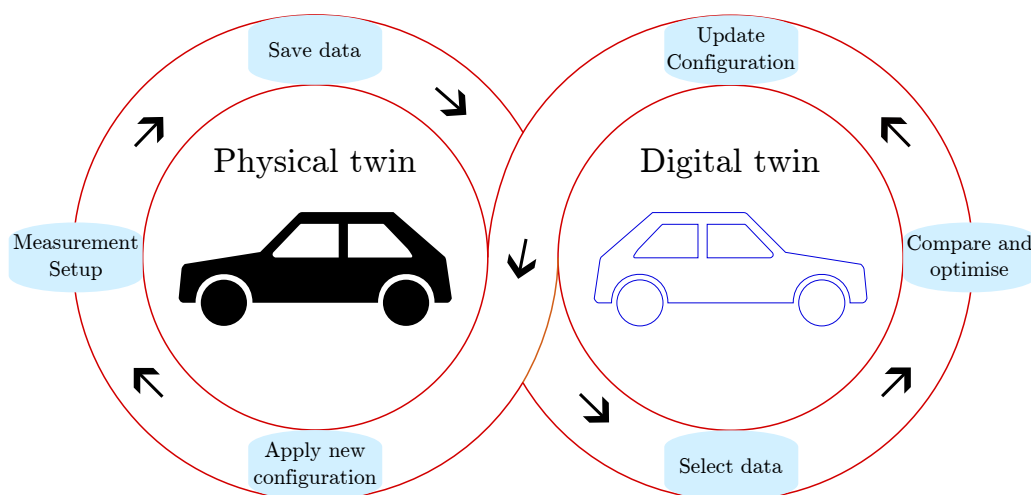


FIGURE 1.2 : Schéma illustrant le paradigme du jumeau numérique.

En mécanique des fluides, la représentation des caractéristiques instantanées de l'écoulement présente également un grand potentiel pour le développement durable d'applications et la diminution de leurs coûts. L'amélioration de la prédiction de la turbulence est ici un enjeu de premier plan qui trouve son potentiel dans la diminution des erreurs explicitées en Sec. 1.2. L'utilisation de données issues de systèmes réels est considérée pour l'amélioration de la prédiction de l'équivalent numérique. L'inférence du jumeau numérique, i.e. induire les caractéristiques générales d'un système à partir d'échantillons, permettrait la prévention d'évènements inattendus pouvant nuire à l'efficacité ou l'intégrité du jumeau physique. En pratique, l'utilisation de données mesurées sur le jumeau physique est envisagée pour l'amélioration de la prédiction de la simulation numérique. Cette amélioration ciblée sur des comportements instantanés et marginaux du système physique pourrait permettre l'étude et la prévention de ces derniers grâce à la simulation. La prise en compte de ces données peut être effectuée via différentes méthodes relevant de méthodes statistiques ou encore du *machine learning*. Leur but est d'optimiser le jumeau numérique.

Le pilotage par les données

Plusieurs techniques de réduction des erreurs de modélisation et des conditions aux limites (Fig. 1.1) existent dans la littérature. L'optimisation de ces paramètres a été étudiée pour améliorer les capacités prédictives des modèles d'ordre réduit pour la CFD. Par exemple, la quantification d'incertitudes a été largement utilisée pour les modèles RANS [149, 45, 69] ou LES [74, 76, 60, 113]. Elle permet le calibrage de la description paramétrique des modèles numériques. Le suivi de phénomènes instationnaires reste cependant difficile. La réduction de modèle, par exemple via l'utilisation de décompositions en modes propres orthogonaux (POD) est utilisée pour fournir un modèle capable d'identifier les caractéristiques principales d'un système pour un coût plus abordable [105, 110]. Toutefois, l'objectif n'est pas d'augmenter la précision de la prédiction, mais de diminuer le coût de la prédiction. Le *machine learning* est également largement utilisé dans le domaine de la mécanique des fluides numérique, les exemples sont nombreux [32, 117, 116, 11]. Il est néanmoins difficile d'extrapoler des informations en dehors de la base de données utilisée pour l'apprentissage. Ces techniques ne parviennent donc que partiellement à remplir les objectifs donnés dans la Sec. 1.1. Aucune de ces méthodes n'est capable du suivi temporel physique d'évènements particuliers exigé par le paradigme du jumeau numérique.

Parmi les méthodes disponibles dans la littérature, l'assimilation de données (AD) [37, 4] est une vaste famille d'outils conçus pour combiner des données expérimentales et numériques afin d'améliorer la précision de la prédiction numérique. L'AD utilise des outils de probabilités et de statistiques pour cibler un état physique présentant un écart minimal par rapport aux données assimilées, les observations. La figure 1.3 illustre le principe de réduction des erreurs des approches numériques. Une des forces de l'assimilation de données est de pouvoir combiner des observations avec un modèle numérique en prenant en compte leurs erreurs respectives. Les écarts de la modélisation et des conditions aux limites numériques sont compensés par les corrections de l'état et des paramètres de la simulation numérique. Pour une procédure optimale, l'erreur résultante de l'algorithme ϵ_{AD} tend vers 0. L'assimilation de données est largement utilisée pour améliorer la précision de la prédiction de modèles RANS [150, 153] ou encore LES [14, 87, 83]. Ces travaux montrent qu'il est possible d'inférer les capacités prédictives de

modèles numériques à partir d'observations du jumeau physique. Parmi les différentes familles d'assimilation de données présentées dans ce manuscrit, l'AD séquentielle semble être la plus prometteuse pour la reconstruction de structures fortement instationnaires, locales en temps et en espace. Le filtre de Kalman (KF) [56] et sa version ensembliste, le filtre de Kalman d'ensemble (EnKF), [34] sont les plus célèbres représentants de l'AD séquentielle. Grâce à cette caractéristique permettant d'effectuer une mise à jour instantanée et à la volée du champ physique, l'EnKF montre un potentiel d'intégration certain dans le paradigme du jumeau numérique.

Les objectifs de ces travaux de thèse sont multiples :

- Créer un outil pour le couplage d'un algorithme d'assimilation de données avec un logiciel de mécanique des fluides numérique.
- Augmenter la précision des résultats de simulations pour un coût équivalent via l'inférence des paramètres de simulations. En particulier :
 - La calibration des paramètres de fermetures de modèles de turbulence.
 - La calibration de la description paramétrique de conditions aux limites.
- La reproduction de phénomènes locaux issus de systèmes réels via l'inférence du champ de vitesse des simulations.

Ce travail est mené dans le cadre du projet ANR ALEKCIA (Augmented prediction and anaLysis of massive data for the idEntification of Key parameters Controlling Internal Aerodynamics). Pour répondre aux objectifs énoncés, l'EnKF est appliqué à trois cas d'étude. Chacun de ces cas permet de répondre à un des objectifs tel que l'inférence de fermetures de modèles pour les chapitres 5 et 6, de conditions aux limites pour le chapitre 7 ou encore la synchronisation de champs de vitesses avec les données observées pour les chapitres 6 et 7. L'intégration de l'algorithme d'AD séquentielle dans un solveur CFD est effectuée via le développement d'un outil C++ appelé C.O.N.ES (Coupling OpenFOAM with Numerical EnvironmentS). Il permet le couplage de l'outil de mécanique des fluides numérique OpenFOAM avec un programme d'assimilation de données utilisant le filtre de Kalman d'ensemble. Le calcul couplé ne nécessite aucun arrêt. Les données sont transférées en mémoire pour l'assimilation de données à tous les pas de temps souhaités par l'utilisateur, l'état augmenté est ensuite récupéré et le calcul du modèle numérique peut reprendre. Les données d'observations peuvent être composées de mesures de vitesses et/ou de pressions. L'outil intègre également certaines techniques avancées d'assimilation de données telles que l'inflation ou la localisation. Une nouvelle technique de localisation appelée *hyper-localisation* a également été développée et implémentée dans l'outil. Cette dernière permet de baisser considérablement le temps d'exécution de chaque phase d'assimilation de données et améliore localement la précision de la prédiction.

1.4 Annonce de plan

Les travaux effectués dans cette thèse présentent une difficulté croissante. Le cas étudié dans le [chapitre 7](#) est donc le plus complexe, coûteux et difficile à manipuler. Le manuscrit est organisé

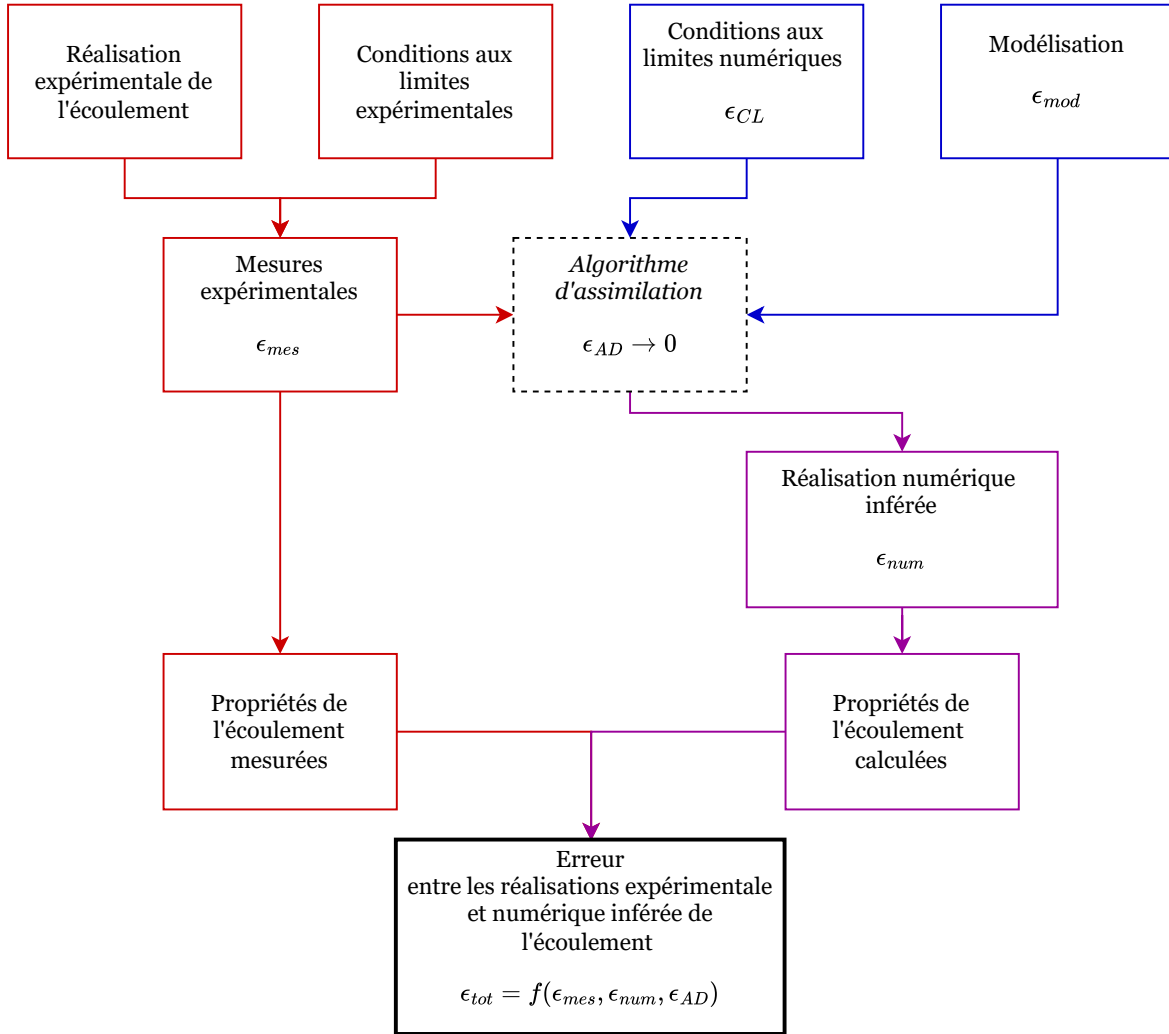


FIGURE 1.3 : Schéma illustrant les différentes sources d'erreur menant aux écarts rencontrés entre une réalisation expérimentale et numérique d'un même écoulement avec assimilation de données. Pour une inférence optimale de l'écoulement, ϵ_{CL} et ϵ_{mod} tendent vers 0.

de la manière suivante :

- Chapitre 2 : les méthodes numériques liées aux calculs de mécanique des fluides utilisées dans les différents cas d'études de ce manuscrit sont présentées.
- Chapitre 3 : plusieurs méthodes d'assimilation de données sont présentées, en particulier le filtre de Kalman d'ensemble utilisé pour les cas d'études de ce manuscrit. Une brève revue de littérature est également menée sur l'utilisation des méthodes d'assimilation de données en mécanique des fluides.
- Chapitre 4 : le fonctionnement de CONES est présenté en détails accompagnés d'un exemple d'optimisation paramétrique effectué sur un cas simple de cavité entraînée 2D laminaire.
- Chapitre 5 : le premier cas d'étude de ce manuscrit est traité. Il concerne l'optimisation de fermetures de modèles RANS dans une simulation 3D incompressible pour une configuration urbaine à partir de données expérimentales mêlant mesures de pression et vitesses.
- Chapitre 6 : le deuxième cas d'étude abordé dans ce chapitre concerne l'inférence d'une simulation LES incompressible dans un canal plan turbulent. La fermeture du modèle de sous-maille est dans un premier temps optimisée avant de se concentrer sur la synchronisation du champ de vitesse avec les données observées.
- Chapitre 7 : le troisième et dernier cas d'étude abordé est le plus complexe. Il concerne l'optimisation d'une condition aux limites d'entrée instationnaire de la simulation LES compressible d'une géométrie simplifiée de moteur. Le champ de vitesse est également corrigé pour fournir une synchronisation locale du champ de vitesse avec l'observation menant à une reconstruction locale des répartitions énergétiques.
- Chapitre 8 : une synthèse des travaux menés dans ce manuscrit est donnée avant une conclusion finale ainsi que quelques perspectives.

Approches numériques pour la turbulence

Sommaire

2.1	Les équations de Navier–Stokes	29
2.2	Algorithme de résolution et discrétisation	32
2.3	Modélisation de la turbulence	34
2.3.1	La turbulence moyennée	36
2.3.2	Simulation des grandes échelles	38

Ce chapitre vise à rappeler les équations qui régissent les problèmes de mécanique des fluides, i.e. les équations de Navier-Stokes. Dans un premier temps, le système d'équations général pour un écoulement compressible sera donné, puis les simplifications menant au cas incompressible seront montrées. Par la suite, une section présentera la description des algorithmes de résolution et schéma numériques utilisés dans OpenFOAM pour les trois cas d'études de ce manuscrit. Enfin, les différents modèles de turbulence utilisés seront explicités, notamment les modèles RANS pour *Reynolds Averaged Navier–Stokes*, modélisant seulement la turbulence en moyenne, ainsi que les modèles LES *Large Eddy Simulation* ou simulation des grandes échelles.

2.1 Les équations de Navier–Stokes

Plusieurs relations peuvent être déduites des lois fondamentales de la mécanique et de la thermodynamique. Elles sont données en notation indicielle dans un repère orthonormé avec $i = 1, 2, 3$ et $j = 1, 2, 3$ et suivant la sommation d'Einstein. Les composantes d'un vecteur sont notées a_i et les composantes d'un tenseur d'ordre deux sont notées a_{ij} où i désigne les lignes et j les colonnes. La sommation sur l'indice muet respecte cet ordre. [15] :

$$\text{Continuité : } \frac{D\rho}{Dt} + \rho \frac{\partial u_i}{\partial x_i} = 0 \quad (2.1)$$

$$\text{Dynamique : } \rho \frac{Du_i}{Dt} = \rho f_i + \frac{\partial \sigma_{ij}}{\partial x_j} \quad (2.2)$$

$$\text{Énergie : } \rho \frac{D}{Dt} \left(e + \frac{u_i u_i}{2} \right) = \rho u_i f_i + \frac{\partial (\sigma_{ij} u_i)}{\partial x_j} - \frac{\partial q_i}{\partial x_i} \quad (2.3)$$

où u_i sont les composantes de vitesse, ρ désigne la masse volumique, f_i les composantes de forces de volumes, σ_{ij} les composantes du tenseur des contraintes et q_i celles du vecteur de densité de flux de chaleur. D/Dt désigne la dérivée particulaire telle que pour une quantité scalaire (ou vectorielle) quelconque :

$$\frac{Da}{Dt} = \frac{\partial a}{\partial t} + u_j \frac{\partial a}{\partial x_j} \quad (2.4)$$

L'énergie interne e dépend de ρ et de la température T telle que $e = f_e(\rho, T)$. Différentes hypothèses sont faites à partir de ces relations pour déduire les équations de Navier–Stokes, soit les équations fondamentales régissant la mécanique en milieu fluide. Le tenseur des contraintes σ_{ij} peut être exprimé à partir des contraintes de pression et des contraintes de viscosité tel que :

$$\sigma_{ij} = -p\delta_{ij} + \tau_{ij} \quad (2.5)$$

ou p désigne la pression, τ_{ij} désigne le tenseur des contraintes visqueuses et δ_{ij} désigne le symbole de *Kronecker*, pour $i, j = 1, 2, 3$, $\delta_{ij} = 1$ si $i = j$ et 0 sinon. Pour un fluide *newtonien*, le tenseur des contraintes visqueuses dépend linéairement du tenseur des vitesses de déformation $S_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$.

$$\tau_{ij} = \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \mu' \frac{\partial u_k}{\partial x_k} \delta_{ij} \quad (2.6)$$

ou μ désigne la viscosité dynamique du fluide et μ' le second coefficient de viscosité. La condition de Stokes donne la relation suivante pour ces coefficients :

$$2\mu + 3\mu' = 0 \quad (2.7)$$

Le tenseur des contraintes visqueuses dépend alors uniquement de la viscosité dynamique tel que :

$$\tau_{ij} = \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{2}{3} \mu \frac{\partial u_k}{\partial x_k} \delta_{ij} \quad (2.8)$$

La loi de *Fourier* relie la densité de flux de chaleur au gradient de température tel que :

$$q_i = -\lambda \frac{\partial T}{\partial x_i} \quad (2.9)$$

avec λ la conductivité thermique du fluide. Le fluide est considéré comme un gaz parfait, donnant également la relation :

$$\frac{p}{\rho} = RT \quad (2.10)$$

ou R désigne la constante du gaz considéré. Enfin, l'énergie interne du système peut s'exprimer à partir des lois de *Joule* :

$$de = C_v dT \quad (2.11)$$

où C_v désigne le coefficient de chaleur spécifique à volume constante. Ces hypothèses mènent aux équations de bilan pour un fluide *newtonien*, soit les fameuses équations de *Navier–Stokes* :

$$\text{Conservation de la masse : } \frac{D\rho}{Dt} + \rho \frac{\partial u_i}{\partial x_i} = 0 \quad (2.12)$$

$$\text{Conservation de la quantité de mouvement : } \rho \frac{Du_i}{Dt} = \rho f_i - \frac{\partial p}{\partial x_i} + 2 \frac{\partial(\mu S_{ij})}{\partial x_j} - \frac{2}{3} \frac{\partial(\mu \vartheta)}{\partial x_i} \quad (2.13)$$

$$\begin{aligned} \text{Conservation de l'énergie : } \rho \frac{D}{Dt} \left(C_v T + \frac{u_i u_i}{2} \right) &= \rho u_i F_i - \frac{\partial(p u_i)}{\partial x_i} + \\ & 2 \frac{\partial(\mu S_{ij} u_i)}{\partial x_j} - \frac{2}{3} \frac{\partial(\mu \vartheta u_i)}{\partial x_i} + \frac{\partial}{\partial x_i} \left(\lambda \frac{\partial T}{\partial x_i} \right) \end{aligned} \quad (2.14)$$

avec $\vartheta = \partial u_k / \partial x_k$

Dans le cas des écoulements incompressibles, $\rho = \text{cste}$ et le problème thermique est découplé du problème vitesse-pression, les équations de Navier–Stokes se simplifient alors tel que :

$$\frac{\partial u_i}{\partial x_i} = 0 \quad (2.15)$$

$$\frac{\partial u_i}{\partial t} + \frac{\partial u_i u_j}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} + f_i \quad (2.16)$$

où $\nu = \mu/\rho$ désigne la viscosité cinématique du fluide.

L'équation de Poisson, régissant l'évolution de la pression, peut être obtenue en prenant la divergence de l'équation de quantité de mouvement 2.16 tel que :

$$\frac{1}{\rho} \frac{\partial^2 p}{\partial x_i \partial x_i} = \frac{\partial f_i}{\partial x_i} - \frac{\partial u_i}{\partial x_j} \frac{\partial u_j}{\partial x_i} = \frac{\partial f_i}{\partial x_i} - \frac{\partial^2 u_i u_j}{\partial x_i \partial x_j} \quad (2.17)$$

2.2 Algorithme de résolution et discrétisation

Les simulations numériques présentées dans ce manuscrit sont réalisées à l'aide de la plateforme open source C++ *OpenFOAM* [99]. Cette bibliothèque comprend un certain nombre de solveurs basés sur la méthode des volumes finis (FVM) [39] appliquée aux équations de Navier-Stokes, ainsi qu'un certain nombre d'outils pour le pré-traitement, le post-traitement et la manipulation des données. En raison de la licence libre et du très grand nombre de modules disponibles, permettant des analyses multi-physiques étendues, cette librairie est largement utilisée dans la littérature pour la recherche en mécanique des fluides [76, 21, 8]. Ses solveurs sont généralement ségrégués. Pour tous les cas étudiés, les équations sont résolues à l'aide d'une boucle PIMPLE, fusion des termes SIMPLE (*Semi-Implicit Method for Pressure-Linked Equations*) et PISO (*Pressure Implicit with Splitting of Operators*). [47, 39]. Les algorithmes SIMPLE et PISO sont des algorithmes de résolution des équations de Navier-Stokes discrétisées. Ces solutions utilisent le couplage vitesse-pression dans différentes boucles pour obtenir la solution finale à chaque pas de temps. Ceux-ci présentent de grandes similarités (notamment dans le cadre de l'implémentation faite par OpenFOAM) bien que le premier soit dédié à la résolution d'écoulements stationnaires et le deuxième aux écoulements instationnaires.

Le couplage vitesse-pression est à présent détaillé dans le cadre d'OpenFOAM pour le cas incompressible [47]. Dans ce cas, la pression considérée est la pression normalisée $p' = p/\rho$, qui sera notée p par souci de concision. Plusieurs ingrédients sont utilisés dans les algorithmes cités, ceux-ci se présentent sous la forme donnée ci-dessous. La notation est importante pour tout utilisateur d'OpenFOAM. Elle est utilisée dans le code de la librairie C++ tel quel. La première étape consiste à établir une équation matricielle pour la vitesse u_i à partir des termes de l'équation de quantité de mouvement. Les termes de l'équation matricielle sont divisés ici en termes diagonaux $a_p u_i$ et hors diagonale $H(u_i) = \sum_N a_n(u_i)_N$. a_p et a_N sont des fonctions de u_i désignant les coefficients associés aux centres des cellules pour a_p et aux termes advectifs des cellules voisines pour $a_N(u_i)_N$. Les termes transitoires sont omis par souci de concision. L'équation de quantité de mouvement, sans gradient de pression, devient alors :

$$a_p u_i - H(u_i) \equiv \frac{\partial u_i}{\partial t} + \frac{\partial u_i u_j}{\partial x_j} - \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} - f_i \quad (2.18)$$

L'équation de Poisson permet d'obtenir une expression discrétisée matricielle pour la pression. C'est la première étape de la boucle PISO.

$$\frac{\partial}{\partial x_i} \left(\frac{1}{a_p} \frac{\partial p}{\partial x_i} \right) = \frac{\partial}{\partial x_i} \left[\frac{H(u_i)}{a_p} \right] \quad (2.19)$$

Le flux de vitesse $\phi_i^f = S_i^f u_i^f$ avec S_i^f la surface d'une face, est ensuite interpolé aux faces des mailles du domaine grâce à la forme matricielle de l'équation de quantité de mouvement précédemment exposée. Ceci permet d'obtenir l'expression du correcteur de flux.

$$\phi_i^f := S_i^f \left[\left(\frac{H(u_i)}{a_p} \right)^f - \left(\frac{1}{a_p} \right)^f \frac{\partial p^f}{\partial x_i} \right] \quad (2.20)$$

Le nouveau champ de vitesse est enfin obtenu à partir du champ de vitesse précédent ainsi qu'à partir du champ de pression nouvellement calculé grâce à la forme matricielle de l'équation de quantité de mouvement. C'est la dernière étape de l'algorithme.

$$u_i := \frac{H(u_i)}{a_p} - \frac{1}{a_p} \frac{\partial p}{\partial x_i} \quad (2.21)$$

La température peut être optionnellement obtenue à partir de l'équation de l'énergie en considérant que μ , λ et C_v sont constants, avec r_i le taux de dissipation par viscosité :

$$\frac{\partial T}{\partial t} + u_j \frac{\partial T}{\partial x_j} - \frac{\lambda}{\rho C_v} \frac{\partial^2 T}{\partial x_j \partial x_j} = r_i \quad (2.22)$$

L'algorithme SIMPLE effectue toutes ces étapes une seule fois afin d'obtenir une solution physique des champs mis en jeu. Cependant, les mises à jour des variables physiques sont généralement relaxées pour améliorer la convergence du système. La valeur typique utilisée par l'algorithme d'OpenFoam est un coefficient de relaxation de 0.7. Celui-ci peut être réduit pour améliorer mais ralentir la convergence du système. Dans le cas d'une solution instationnaire ou en régime transitoire, la solution avance dans le temps avec un pas de temps assez faible pour maintenir une résolution temporelle suffisante pour le système. L'algorithme PISO intervient ici dans la possibilité de résoudre tour à tour la pression et la vitesse dans une boucle faisant intervenir l'équation de pression, le correcteur de flux ainsi que le correcteur de quantité de mouvement. Ceci est nécessaire afin d'améliorer la précision des champs p , u_i et ϕ_i^f pour chaque pas de temps. Une étape de prédiction de la quantité de mouvement peut être ici ajoutée pour améliorer la convergence du système dans le cas d'écoulements rapides où l'échange de quantité de mouvement est prépondérant.

L'algorithme PIMPLE, fusion des deux algorithmes précédents, permet la résolution de problèmes transitoires ou pseudo-transitoires avec une convergence améliorée en combinant les contrôles du SIMPLE et du PISO, sa version incompressible est décrite par l'Alg. 1. Il permet l'amélioration de la convergence du système par l'ajout de boucles externes à chaque pas de temps. Ce nombre de boucles est contrôlé par la variable N_{PIMPLE} , correspondant à la variable $nOuterCorrectors$ dans le fichier *fvSolution* d'OpenFOAM. Le fonctionnement itératif appliqué pour chaque pas de temps considère donc l'écoulement stationnaire à un temps donné, le comportement est alors celui de l'algorithme SIMPLE. Des facteurs de relaxations peuvent être utilisés pour améliorer la convergence du système. Le nombre de boucles internes est lui décidé par la variable notée N_{PISO} ici, correspondant à la variable $nCorrectors$ dans le fichier *fvSolution* d'OpenFOAM. Il est donc à noter que pour $N_{PIMPLE} = 1$, l'algorithme est purement PISO. Si $N_{PISO} = 1$ également, l'algorithme est purement SIMPLE.

Dans le contexte des calculs augmentés par assimilation de données étudiés dans ce manuscrit, seul le champ de vitesse est mis à jour pendant les phases d'AD, étant donné le caractère explicite de cette variable. Les autres variables physiques sont recalculées par l'algorithme PIMPLE à partir de cette dernière et ne nécessitent donc pas d'être mises à jour directement par l'AD. En revanche, une fois le champ de vitesse mis à jour après une phase d'assimilation de données, les autres grandeurs physiques doivent être mises à jour à partir de ce dernier avant de reprendre la boucle de calcul suivante. Dans le cas des travaux effectués et présentés dans

les [chapitre 5](#) et [chapitre 6](#) portant sur l'étude d'écoulements incompressibles, seule la pression a besoin d'être mise à jour après une phase d'AD pour respecter la conservation du système physique. Cette mise à jour est alors effectuée via une simple équation de Poisson. En revanche, dans le cas du [chapitre 7](#), l'écoulement étudié est compressible. Une boucle PISO complète est alors effectuée après la phase d'AD pour mettre à jour les champs de pression et de température. Il est à noter que l'étape de prédiction de quantité de mouvement n'est pas utilisée lors de cette mise à jour afin d'éviter de remplacer les données de vitesses issues de l'AD.

Algorithme 1 : Algorithme PIMPLE incompressible utilisé dans OpenFOAM

Départ : Initialisation des champs, du temps initial $t = t_{init}$ et du pas de temps Δ_t

while $t < t_{final}$ **do**

for $k = 0$ to N_{PIMPLE} **do**

1 Résolution de l'équation d'énergie pour T (optionnel) :

$$\frac{\partial T}{\partial t} + u_j \frac{\partial T}{\partial x_j} - \frac{\lambda}{\rho C_p} \frac{\partial^2 T}{\partial x_j \partial x_j} = r_i$$

2 Déclaration de la matrice de quantité de mouvement :

$$a_p u_i - H(u_i)$$

3 Prédiction de quantité de mouvement (optionnel) :

$$a_p u_i - H(u_i) = -\frac{\partial p}{\partial x_i}$$

for $m = 0$ to N_{PISO} **do**

4 Équation de pression :

$$\frac{\partial}{\partial x_i} \left(\frac{1}{a_p} \frac{\partial p}{\partial x_i} \right) = \frac{\partial}{\partial x_i} \left[\frac{H(u_i)}{a_p} \right]$$

5 Correcteur de flux :

$$\phi_i^f := S_i^f \left[\left(\frac{H(u_i)}{a_p} \right)^f - \left(\frac{1}{a_p} \right)^f \frac{\partial p}{\partial x_i} \right]$$

6 Correcteur de quantité de mouvement :

$$u_i := \frac{H(u_i)}{a_p} - \frac{1}{a_p} \frac{\partial p}{\partial x_i}$$

7 Mise à jour du temps :

$$t = t + \Delta_t$$

La résolution des systèmes linéaires est effectuée grâce aux méthodes PCG (*preconditioned (bi-) conjugate gradient*), GAMG (*generalised geometric-algebraic multi-grid*) ou encore Gauss-Seidel symétrique. Toutes les méthodes de discrétisation ou de résolution de systèmes linéaires sont natives à OpenFOAM version 8 ([chapitre 5](#)) ou version 9 ([chapitre 6](#) et [chapitre 7](#)).

2.3 Modélisation de la turbulence

Les travaux menés dans ce manuscrit concernent exclusivement des écoulements turbulents. À la différence du laminaire, l'écoulement turbulent est obligatoirement instationnaire et présente des

phénomènes irréguliers. La turbulence se compose de structures tourbillonnaires aux multiples échelles. Le mouvement de ces structures est difficile à prédire pour un système réel. Les grandes structures de la turbulence sont généralement de l'ordre de grandeur de l'échelle caractéristique du cas étudié et sont donc souvent affectées par les considérations uniques de la géométrie. Au contraire, les petites échelles de la turbulence sont indépendantes de la géométrie étudiée et présentent donc un caractère commun à toutes les topologies d'écoulements. L'augmentation du nombre de Reynolds, traduisant la compétition entre les effets inertiels et visqueux, est un marqueur de la turbulence. Un nombre élevé est synonyme de phénomènes fortement non-linéaires aux interactions variées. Un grand nombre de Reynolds peut par exemple favoriser la séparation entre les petites et les grandes échelles de la turbulence. La région centrale, dite inertielle, est souvent responsable du développement d'interactions triadiques pour des gammes d'échelles qui augmentent avec le nombre de Reynolds [115]. Les grandes échelles de la turbulence sont également les plus énergétiques. L'énergie cinétique de ces échelles est transférée aux plus petites structures par lesquelles elle est dissipée.

Les équations de Navier-Stokes sont capables de décrire la complexité instationnaire de la turbulence. Cependant, l'évolution du champ de vitesse pour un écoulement turbulent est particulièrement dépendante des conditions aux limites et initiales de l'étude. Des variations infinitésimales sont à l'origine de réalisations différentes d'un même cas. Pour des conditions quasi identiques, le champ de vitesse n'a donc pas la même valeur. La mécanique des fluides numérique résout les équations de Navier-Stokes et peut faire usage des différents modèles de turbulence pour la simulation d'écoulements fluides. Trois grandes familles se distinguent [101, 38], celles-ci sont illustrées sur la Fig. 2.1 :

- DNS (*Direct Numerical Simulation*) ou simulation numérique directe : Les équations de Navier-Stokes sont entièrement résolues, des plus grandes échelles de la turbulence aux plus petites structures. Les résolutions spatiale et temporelle nécessaires pour la résolution directe sont cependant très élevées, ce faisant le coût d'une telle approche est prohibitif.
- LES (*Large Eddy Simulation*) ou simulation des grandes échelles (voir Sec. 2.3.2) : Le filtrage des équations de Navier-Stokes entraîne l'apparition de tensions de sous-mailles notées $\tilde{\tau}_{ij}$. Les grandes échelles de la turbulence sont résolues et les échelles de sous-maille sont généralement modélisées via l'ajout d'un terme de sous-maille, gouverné par la viscosité de sous-maille ν_{SGS} (*subgrid scale*) tel qu'établi par Smagorinsky en 1963 [119]. Ceci permet de représenter l'instationnarité de la turbulence sans engendrer un coût comparable à la DNS, bien que la recherche de précision puisse entraîner une résolution proche des résolutions utilisées pour la DNS.
- RANS (*Reynolds Averaged Navier Stokes*) (voir Sec. 2.3.1) : La vitesse peut être décomposée par sa valeur moyenne et sa fluctuation $u_i = \bar{u}_i + u'_i$, où $\bar{\cdot}$ désigne la moyenne et u'_i le terme de fluctuation. Le traitement moyen des équations de Navier-Stokes laisse alors apparaître les termes de covariances de vitesses $\overline{u'_i u'_j}$, les tensions de Reynolds. La fermeture de ces équations nécessite de nouvelles approximations, ou modèles de turbulence, à cause des inconnues supplémentaires engendrées par les tensions de Reynolds. Elles sont souvent modélisées via l'ajout d'un terme gouverné par la viscosité turbulente ν_t ,

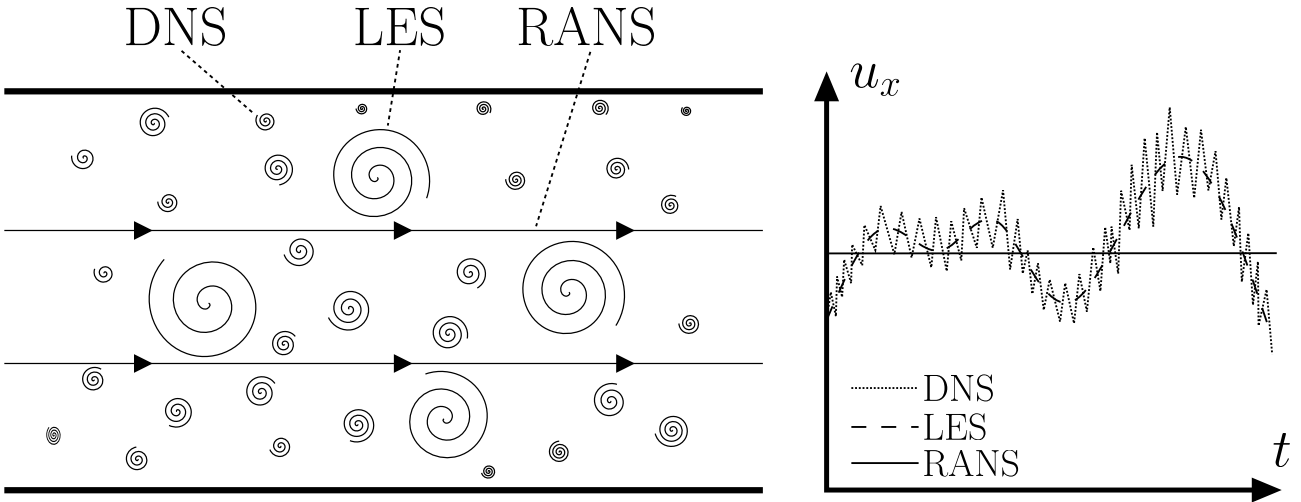


FIGURE 2.1 : Schéma adapté et traduit de [38] illustrant le mouvement des structures turbulentes à gauche et l'évolution de la première composante de vitesse en un point en fonction du temps à droite pour différents modèles de turbulence.

selon l'hypothèse de Boussinesq. La représentation de l'écoulement est une représentation moyenne et stationnaire.

La LES est capable de représenter les grandes échelles de la turbulence comme le montre le schéma Fig. 2.1. Cette approche est privilégiée pour les cas aux géométries complexes et haut nombre de Reynolds pour lesquels la DNS serait trop coûteuse. Elle est également préférée pour la simulation de phénomènes fortement instationnaires que les modèles RANS ne sont pas capables de représenter [38]. Le choix du modèle optimal est cependant fortement dépendant du cas d'étude et de la précision des résultats attendus. C'est un compromis entre coût et précision. Différents modèles de turbulences, décrits dans les sections suivantes, sont utilisés à travers le manuscrit, tel que certains modèles RANS dans le chapitre 5 ou LES dans le chapitre 6 et le chapitre 7.

2.3.1 La turbulence moyennée

L'étude d'un écoulement turbulent est établie d'un point de vue statistique. Les équations de Navier–Stokes moyennées pour un fluide newtonien et un écoulement stationnaire et incompressible sont données ci-après [101] :

$$\bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial \bar{\tau}_{ij}}{\partial x_j} - \frac{\partial \tau_{ij}^T}{\partial x_j} \quad (2.23)$$

$$\nabla^2 \bar{p} = -\frac{\partial \bar{u}_j}{\partial x_i} \frac{\partial \bar{u}_j}{\partial x_i} - \frac{\partial}{\partial x_i} \left(\frac{\partial \tau_{ij}^T}{\partial x_j} \right) \quad (2.24)$$

où Eq. 2.23 est l'équation de quantité de mouvement et Eq. 2.24 est l'équation de Poisson. p est ici la pression normalisée. τ_{ij} est le tenseur des contraintes visqueuses, comme déjà énoncé.

τ_{ij}^T est le tenseur des contraintes de Reynolds résultant de l'étude statistique de l'état turbulent du système. La barre supérieure indique l'opération de moyenne effectuée pour obtenir les équations 2.23 - 2.24.

La modélisation de type RANS affecte la fermeture des équations, notamment pour le terme τ_{ij}^T . Le modèle $\mathcal{K} - \varepsilon$ [66, 148] utilise le concept de viscosité turbulente pour créer un lien entre τ_{ij}^T et le gradient de la vitesse moyenne $\bar{\mathbf{u}}$:

$$-\tau_{ij}^T = 2\nu_T \bar{S}_{ij} - \frac{2}{3}\mathcal{K}\delta_{ij} \quad (2.25)$$

où ν_T est la viscosité turbulente, \mathcal{K} est l'énergie cinétique turbulente et \bar{S}_{ij} est le tenseur des vitesses de déformation moyen :

$$\bar{S}_{ij} = \frac{1}{2} \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \quad (2.26)$$

Dans le modèle $\mathcal{K} - \varepsilon$, ν_T est exprimé comme une fonction algébrique de \mathcal{K} et du taux de dissipation d'énergie cinétique turbulente ε :

$$\nu_T = C_\mu \frac{\mathcal{K}^2}{\varepsilon} \quad (2.27)$$

où C_μ est une constante de modèle à étalonner. Pour fermer le problème, deux équations de modèle pour \mathcal{K} et ε doivent être incluses :

$$\frac{\partial \mathcal{K}}{\partial t} + \bar{u}_j \frac{\partial \mathcal{K}}{\partial x_j} = \frac{\partial}{\partial x_j} \left[\left(\nu + \frac{\nu_T}{\sigma_{\mathcal{K}}} \right) \frac{\partial \mathcal{K}}{\partial x_j} \right] + \mathcal{P} - \varepsilon \quad (2.28)$$

$$\frac{\partial \varepsilon}{\partial t} + \bar{u}_j \frac{\partial \varepsilon}{\partial x_j} = \frac{\partial}{\partial x_j} \left[\left(\nu + \frac{\nu_T}{\sigma_\varepsilon} \right) \frac{\partial \varepsilon}{\partial x_j} \right] + C_{\varepsilon 1} \frac{\varepsilon}{\mathcal{K}} \mathcal{P} - C_{\varepsilon 2} \frac{\varepsilon^2}{\mathcal{K}} \quad (2.29)$$

où le terme de production $\mathcal{P} = \nu_T \bar{S}^2$, $\bar{S} = \sqrt{2\bar{S}_{ij}\bar{S}_{ij}}$. Le modèle est complet une fois que les cinq constantes C_μ , $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, $\sigma_{\mathcal{K}}$ et σ_ε sont déterminées. Launder et Sharma [66] fournissent des valeurs étalonnées via l'analyse de cas tests académiques, tels que la dissipation libre de la turbulence isotrope homogène ou le canal plan turbulent.

Un deuxième modèle de turbulence très populaire implémenté dans de nombreux solveurs CFD est le modèle $\mathcal{K} - \omega$ SST [148]. Sa formulation repose sur des fonctions de mélange qui sont régies par un certain nombre de paramètres, notamment la distance à la paroi la plus proche δ . Les fonctions de mélange fournissent une modélisation hybride combinant les caractéristiques des deux modèles classiques $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ [148]. Plus précisément, le modèle $\mathcal{K} - \omega$ SST se comporte comme le modèle $\mathcal{K} - \omega$ près de la surface des corps, et comme le modèle $\mathcal{K} - \varepsilon$ à mesure que la distance augmente. La formulation de ce modèle par OpenFOAM est basée sur le travail de Menter et al. [77], où deux équations de modèle pour \mathcal{K} et ω doivent être incluses. Un limiteur pour le terme de production dans l'équation d'énergie cinétique turbulente est introduit pour exclure des niveaux de turbulence très élevés dans les régions à forte déformation normale. Par conséquent, le terme de production est $\tilde{\mathcal{P}} = \min(\mathcal{P}, 10 \beta^* \mathcal{K} \omega)$. Le taux de dissipation d'énergie

cinétique turbulence $\varepsilon = \beta^* \mathcal{K} \omega$ est défini par l'ajout du paramètre restrictif β^* pour la même raison.

$$\frac{\partial \mathcal{K}}{\partial t} + \bar{u}_j \frac{\partial \mathcal{K}}{\partial x_j} = \frac{\partial}{\partial x_j} \left[(\nu + \sigma_{\mathcal{K}} \nu_T) \frac{\partial \mathcal{K}}{\partial x_j} \right] + \tilde{\mathcal{P}} - \beta^* \mathcal{K} \omega \quad (2.30)$$

$$\frac{\partial \omega}{\partial t} + \bar{u}_j \frac{\partial \omega}{\partial x_j} = \frac{\partial}{\partial x_j} \left[(\nu + \sigma_{\omega} \nu_T) \frac{\partial \omega}{\partial x_j} \right] + \alpha \bar{S}^2 - \beta \omega^2 + 2(1 - F_1) \sigma_{\omega 2} \frac{1}{\omega} \frac{\partial \mathcal{K}}{\partial x_i} \frac{\partial \omega}{\partial x_i} \quad (2.31)$$

F_1 est une fonction de mélange caractérisée par :

$$F_1 = \tanh \left(\left\{ \min \left[\max \left(\frac{\sqrt{\mathcal{K}}}{\beta^* \omega \delta}, \frac{500\nu}{\delta^2 \omega} \right), \frac{4\sigma_{\omega 2} \mathcal{K}}{CD_{\mathcal{K}\omega} \delta^2} \right] \right\}^4 \right) \quad (2.32)$$

avec $CD_{\mathcal{K}\omega} = \max \left(2\sigma_{\omega 2} \frac{1}{\omega} \frac{\partial \mathcal{K}}{\partial x_i} \frac{\partial \omega}{\partial x_i}, 10^{-10} \right)$. F_1 contrôle le mélange entre les modèles $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$. Plus précisément, $F_1 \approx 0$ correspond à un modèle $\mathcal{K} - \varepsilon$ loin de la surface du corps, et $F_1 \approx 1$ correspond au modèle $\mathcal{K} - \omega$ à proximité du corps. Pour fermer le problème, la viscosité turbulente ν_T est définie comme :

$$\nu_T = \frac{a_1 \mathcal{K}}{\max(a_1 \omega, \bar{S} F_2)} \quad (2.33)$$

où a_1 est une constante ($a_1 = 0,31$) et F_2 est une autre fonction de mélange :

$$F_2 = \tanh \left(\left[\max \left(\frac{2\sqrt{\mathcal{K}}}{\beta^* \omega \delta}, \frac{500\nu}{\delta^2 \omega} \right) \right]^2 \right) \quad (2.34)$$

Les neuf constantes globales qui déterminent le modèle sont $\sigma_{\mathcal{K}1}$, $\sigma_{\mathcal{K}2}$, $\sigma_{\omega 1}$, $\sigma_{\omega 2}$, α_1 , α_2 , β_1 , β_2 et β^* . À l'exception de cette dernière, tous les autres coefficients sont définis par deux indices 1, 2. Les coefficients inclus dans les équations 2.30 - 2.31 sont déterminés via une interpolation linéaire entre les deux valeurs fournies, qui est pilotée par la fonction F_1 . Par exemple, $\alpha = \alpha_1 F_1 + \alpha_2 (1 - F_1)$. De cette dernière relation peut être déduit le fait que l'indice 1 fait référence aux valeurs pour le modèle $\mathcal{K} - \omega$ alors que l'indice 2 renvoie des valeurs pour le modèle $\mathcal{K} - \varepsilon$. Menter a déterminé des valeurs pour les neuf paramètres [78]. En pratique, de manière similaire au modèle $\mathcal{K} - \varepsilon$, ces coefficients ne sont pas des constantes mais ils sont fonctions de la dynamique locale de l'écoulement et de leur interaction avec les caractéristiques globales (voir discussions dans les réf. [70, 149, 32]).

2.3.2 Simulation des grandes échelles

Dans le formalisme LES, les équations de Navier-Stokes sont filtrées pour obtenir une réduction globale des degrés de liberté du système physique. Par souci de concision, seul le filtrage des équations incompressibles 2.15 et 2.16 est explicité :

$$\frac{\partial \tilde{u}_j}{\partial x_j} = 0 \quad (2.35)$$

$$\frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial \tilde{u}_i \tilde{u}_j}{\partial x_j} = -\frac{1}{\rho} \frac{\partial \tilde{p}}{\partial x_i} + \nu \frac{\partial^2 \tilde{u}_i}{\partial x_j \partial x_j} - \frac{\partial \tilde{\tau}_{ij}}{\partial x_j} + \tilde{f}_i \quad (2.36)$$

Le symbole tilde représente les variables filtrées et $\tilde{\tau}_{ij} = \overline{u_i u_j} - \tilde{u}_i \tilde{u}_j$ est le tenseur des contraintes à l'échelle de sous-maille.

Smagorinsky

Dans le modèle de Smagorinsky [119], la partie déviatorique de $\tilde{\tau}_{ij}$ est modélisée par un modèle de viscosité turbulente, comme pour les modèles RANS :

$$\tilde{\tau}_{ij} - \frac{1}{3} \tilde{\tau}_{kk} \delta_{ij} = -2\nu_{SGS} \tilde{S}_{ij}, \quad \nu_{SGS} = (C_S \Delta)^2 \sqrt{2\tilde{S}_{ij} \tilde{S}_{ij}} \quad (2.37)$$

où $\tilde{S}_{ij} = \frac{1}{2} \left(\frac{\partial \tilde{u}_i}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_i} \right)$ est le tenseur de vitesse de déformation du champ de vitesse résolu, Δ est la largeur du filtre et C_S est un coefficient de modèle qui peut être sélectionné par l'utilisateur. Les valeurs classiques trouvées dans la littérature sont $C_S \in [0.1, 0.2]$. Cette formulation, qui est dérivée de la théorie de la turbulence asymptotique de Kolmogorov, ne parvient pas à fournir une prédiction précise des interactions entre les variables physiques résolues et filtrées. La raison est que le tenseur de contrainte de sous-maille dans l'équation 2.37 est intrinsèquement dissipatif et affecte toutes les échelles simulées de l'écoulement [114]. Malgré ces caractéristiques négatives, la mise en œuvre directe et simple d'un tel modèle en a fait un choix populaire pour la plupart des solveurs. Le modèle de Smagorinsky utilisant une valeur constante de C_S est connu pour obtenir un comportement erroné de la viscosité de sous-maille à proximité de la paroi. Ce problème affecte la prédiction globale des statistiques de l'écoulement. La précision peut être partiellement améliorée en combinant le modèle avec des fonctions de paroi, pour des écoulements délimités par une paroi relativement simple, présentant peu d'irrégularités de surfaces. Cependant, le modèle de Smagorinsky est inadéquat pour capturer avec précision la dynamique de l'écoulement dans des géométries complexes, caractérisées par des nombres de Reynolds locaux à proximité de la paroi présentant une forte variance dans le temps et l'espace. À l'inverse, d'autres approches comme le Smagorinsky dynamique [80], le WALE [95] et le modèle Sigma [96] offrent des performances améliorées lorsqu'elles sont appliquées à un maillage raffiné en paroi et dans des conditions de non-glissement. Ces dernières présentent également des inconvénients comme la stabilisation de la constante C_S pour le Smagorinsky dynamique. Dans ce cas, des valeurs négatives pour C_S , précédemment interprétées comme un transfert d'énergie des petites aux grandes échelles (*backscatter*), peuvent perturber considérablement la simulation. Des méthodes de stabilisation avancées ont été développées [133, 132], mais aucune d'entre elles n'est pleinement satisfaisante dans des applications industrielles.

L'implémentation du modèle Smagorinsky dans OpenFOAM repose sur deux constantes de modèle, le paramètre C_k et le paramètre de dissipation normalisé C_ε . Ces derniers présentent généralement une sensibilité élevée aux effets de production de turbulence et au manque d'homogénéité de l'écoulement [138]. Dans le cas d'un équilibre de la turbulence, comme dans la

théorie de Kolmogorov, $C_\varepsilon = \text{const}$, et sa valeur peuvent être définies par l'utilisateur. OpenFOAM suggère une valeur par défaut de $C_\varepsilon = 1.048$, qui se situe dans la gamme des résultats expérimentaux et numériques. Dans ce cadre, la relation entre la constante C_S du modèle de Smagorinsky et C_k est :

$$C_S^2 = C_k \sqrt{\frac{C_k}{C_\varepsilon}} \quad (2.38)$$

Le filtrage LES est réalisé implicitement en utilisant la résolution du maillage. La largeur du filtre Δ est ainsi associée à la largeur du filtre géométrique Δ_g , qui est localement proportionnelle au volume de chaque cellule V_c (option *cube-root volume* dans OpenFOAM) et plus précisément $\Delta = \Delta_g = \sqrt[3]{V_c}$. Le modèle de Smagorinsky conduit à une estimation incorrecte de la viscosité turbulente ν_{SGS} [96] et à une contrainte de cisaillement turbulente non nulle à l'approche de la paroi. Ainsi, pour surmonter ce problème, l'utilisation de la fonction d'amortissement de van Driest a été proposée [81] de manière à appliquer une mise à l'échelle près de la paroi. L'implémentation dans OpenFOAM de la fonction d'amortissement est décrite comme :

$$\Delta = \min\left(\frac{\zeta y}{C_S} D, \Delta_g\right) \quad (2.39)$$

$$D = 1 - e^{-\frac{y^+}{A^+}} \quad (2.40)$$

avec $y^+ = u_\tau y / \nu$, u_τ la vitesse de frottement, $\zeta = 0,41$ et $A^+ = 26$

Les simulations numériques menées notamment dans le chapitre 6 sont effectuées en utilisant le Smagorinsky avec l'estimation de Δ basée sur le *cube-root volume* ainsi qu'en utilisant la fonction d'amortissement de van Driest. Des augmentations pilotées par les données seront effectuées, optimisant les coefficients libres des modèles.

WALE

Comme précisé plus haut, le modèle LES-WALE pour *Wall Adapting Local Eddy-viscosity* a été développé pour dépasser les limitations du modèle Smagorinsky notamment proche de la paroi. Nicoud et Ducros [95] mentionne deux limites principales au modèle Smagorinsky, basé sur le second invariant de la partie symétrique de \tilde{S}_{ij} (1) : sa formulation est uniquement reliée au tenseur des vitesses de déformation et non au taux de rotation (2) : l'ordre de grandeur de l'invariant proche de la paroi est $\mathcal{O}(1)$. Le but du modèle WALE est de reproduire le comportement attendu de la viscosité de sous-maille à l'approche de la paroi $\nu_{SGS} = \mathcal{O}(y^3)$ en utilisant à la fois le tenseur des vitesses de déformation et le taux de rotation. Pour cela, la partie symétrique sans trace du tenseur du carré du gradient de vitesse est utilisée :

$$S_{ij}^d = \frac{1}{2} \left(\frac{\partial \tilde{u}_k}{\partial x_i} \frac{\partial \tilde{u}_j}{\partial x_k} + \frac{\partial \tilde{u}_k}{\partial x_j} \frac{\partial \tilde{u}_i}{\partial x_k} \right) - \frac{1}{3} \delta_{ij} \frac{\partial \tilde{u}_k}{\partial x_m} \frac{\partial \tilde{u}_m}{\partial x_k} \quad (2.41)$$

En considérant $\tilde{\Omega}_{ij} = \frac{1}{2} \left(\frac{\partial \tilde{u}_i}{\partial x_j} - \frac{\partial \tilde{u}_j}{\partial x_i} \right)$ la partie anti-symétrique du gradient de vitesse, correspondant au taux de rotation, la relation précédente peut s'exprimer par :

$$S_{ij}^d = \tilde{S}_{ik}\tilde{S}_{kj} + \tilde{\Omega}_{ik}\tilde{\Omega}_{kj} - \frac{1}{3}\delta_{ij}[\tilde{S}_{mn}\tilde{S}_{mn} - \tilde{\Omega}_{mn}\tilde{\Omega}_{mn}] \quad (2.42)$$

L'implémentation de la méthode dans OpenFOAM se base sur la formulation suivante de la viscosité de sous-maille :

$$\nu_{SGS} = C_k \Delta \sqrt{\mathcal{K}_{SGS}} \quad (2.43)$$

où \mathcal{K}_{SGS} est l'énergie cinétique de sous-maille, elle s'exprime par :

$$\mathcal{K}_{SGS} = \left(\frac{C_w^2 \Delta}{C_k} \right)^2 \frac{(S_{ij}^d S_{ij}^d)^3}{\left((\tilde{S}_{ij} \tilde{S}_{ij})^{5/2} + (S_{ij}^d S_{ij}^d)^{5/4} \right)^2} \quad (2.44)$$

La largeur du filtre Δ est également associée à la largeur du filtre géométrique Δ_g , comme pour le Smagorinsky. Les valeurs des constantes peuvent être modifiées par l'utilisateur, par défaut $C_k = 0.094$ et $C_w = 0.325$. Le remplacement de l'expression 2.44 dans l'expression de la viscosité de sous-maille 2.43 donne :

$$\nu_{SGS} = \left(C_w^2 \Delta \right)^2 \frac{(S_{ij}^d S_{ij}^d)^{3/2}}{(\tilde{S}_{ij} \tilde{S}_{ij})^{5/2} + (S_{ij}^d S_{ij}^d)^{5/4}} \quad (2.45)$$

Cette dernière relation correspond à l'équation (13) donnée par Nicoud et Ducros [95]. Elle permet l'identification de structures turbulentes dont les vitesses de déformations et/ou le taux de rotation sont élevés. La viscosité de sous-maille est également correctement évaluée à l'approche de la paroi grâce au terme de normalisation utilisé dans l'expression 2.45. Le modèle décroît naturellement dans les zones de cisaillement. Cependant, certains travaux tel que ceux de Bricteux et al. [10] ont montré les limitations du WALE notamment pour les écoulements purement rotationnels. Le modèle Sigma, développé plus tard par Nicoud et al. [96], corrige ces limitations. Ce dernier n'étant pas disponible dans OpenFOAM, le modèle WALE est utilisé dans le chapitre 7 où l'objectif n'est plus d'optimiser les coefficients libres du modèle mais d'optimiser une condition d'entrée instationnaire. L'étude ne se consacrant pas à l'optimisation du modèle, le WALE, équivalent plus élaboré et offrant une plus grande précision en particulier proche paroi que le Smagorinsky, a été choisi.

Assimilation de données : méthodes conventionnelles et développements

Sommaire

3.1	Inférence bayésienne	44
3.2	Méthodes variationnelles	48
3.2.1	3DVar	49
3.2.2	4DVar	50
3.2.3	Application des méthodes variationnelles en mécanique des fluides	51
3.3	Le filtre de Kalman appliqué à la mécanique des fluides	54
3.3.1	Le filtre de Kalman et les contraintes de linéarité	55
3.3.2	Le filtre de Kalman d'ensemble	57
3.3.3	Corrections des défauts du EnKF	59
3.3.4	HLEnKF : <i>Hyper-Localized Ensemble Kalman Filter</i>	63
3.3.5	Application des méthodes séquentielles en mécanique des fluides	66

L'assimilation de données (AD) [24, 4] inclut un large spectre de techniques basées sur les données dont l'objectif principal est d'obtenir une prédiction *augmentée* d'un processus aléatoire étudié, combinant différentes sources d'information. Généralement, un modèle représentatif d'un phénomène physique est amélioré par l'assimilation d'observations de ce phénomène physique. Les observations sont donc considérées comme des données haute-fidélité alors que le modèle est, la plupart du temps, une représentation d'ordre réduit du phénomène physique, contenant des hypothèses simplificatrices et autres biais diminuant la qualité de la prédiction. La prédiction du modèle d'ordre réduit est dite *inférée*, c'est-à-dire induite, par l'assimilation de

données haute-fidélité. Les outils sont généralement regroupés en deux catégories principales. Les approches *variationnelles* exécutent la stratégie d'assimilation de données via un problème d'optimisation. Les approches *séquentielles* s'appuient sur des approches probabilistes basées sur le théorème de Bayes. La première section de ce chapitre revient sur les généralités conduisant à la définition des deux grandes familles d'algorithmes d'assimilation de données, qui seront abordées dans les deux sections suivantes. Chacune de ces deux sections contient une brève revue de littérature des utilisations de ces algorithmes pour la mécanique des fluides numérique.

3.1 Inférence bayésienne

L'inférence bayésienne constitue un socle théorique pour une multitude d'algorithmes tel que les algorithmes d'assimilation de données. Il peut se résumer ainsi : la probabilité (notée $Prob(\cdot)$) d'un évènement A peut être inférée à partir de l'observation d'un évènement B interdépendant de l'évènement A . La prédiction de A est alors améliorée sachant B . Les termes de probabilités *a priori* ainsi qu'*a posteriori* sont régulièrement utilisés pour décrire la prévision de l'évènement ainsi que son état inféré. Le théorème de Bayes est le suivant :

$$Prob(A|B) = \frac{Prob(B|A)Prob(A)}{Prob(B)} \quad (3.1)$$

La prédiction *a posteriori* $Prob(A|B)$ de l'évènement A sachant B est égale au produit de la probabilité de l'évènement B sachant A (aussi appelée vraisemblance) et de la prévision de la probabilité de l'évènement A (*a priori*) $Prob(B|A)P(A)$, normalisé par l'observation de l'évènement B , $Prob(B)$.

Notons à présent $f(\cdot)$ la fonction de densité de probabilité, u est la description imparfaite d'un système physique et α les quantités mesurées issues de l'état réel. Le terme vecteur d'état est utilisé pour désigner u alors que α est le vecteur d'observations. Pour les variables évoquées, le théorème de Bayes devient $f(u|\alpha) = f(\alpha|u)f(u)/f(\alpha)$. Attardons-nous sur ces quatre termes :

- Distribution *a priori* $f(u)$: correspond à l'état *a priori* du système. Fais typiquement référence aux conditions initiales.
- Vraisemblance $f(\alpha|u)$: correspond à la distribution des mesures par rapport au système u , autrement dit la probabilité que l'état u corresponde aux mesures α . Dans la mesure où α représente l'observation de l'état réel inconnu du système, $f(\alpha|u)$ quantifie la distribution des erreurs sur la mesure.
- Distribution *a posteriori* $f(u|\alpha)$: correspond à l'état *a posteriori* du système en prenant en compte les données *a priori* ainsi que les observations.
- Distribution marginale $f(\alpha) = \int f(\alpha|u^r)f(u^r)du^r$: agit comme un facteur de normalisation du numérateur. $f(\alpha)$ n'est pas égal à 1 puisque le calcul se réfère à l'état réel et inconnu du système.

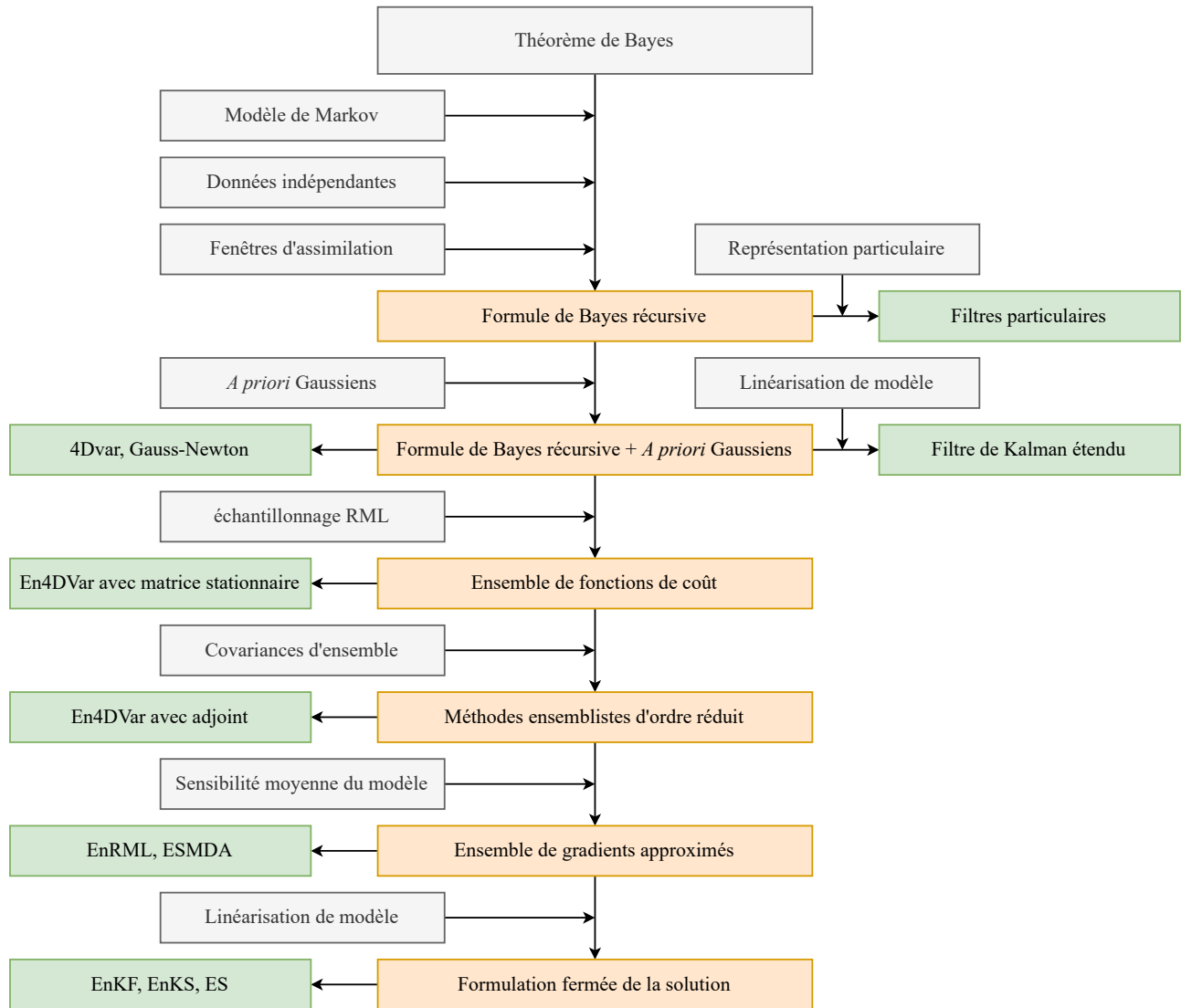


FIGURE 3.1 : Schéma adapté et traduit de [37]. Celui-ci montre de manière unifiée les hypothèses conduisant aux algorithmes principaux d'assimilation de données.

Ceci décrit parfaitement l'ambition donnée à l'assimilation de données : améliorer l'état d'un modèle u grâce à l'observation de cet état α à partir de son état réel u^r . Dans ses "Data Assimilation Fundamentals" [37], Evensen décrit les algorithmes principaux d'assimilation de manière unifiée, ayant Bayes pour base commune, comme le montre la figure 3.1. Les différentes hypothèses énoncées par les cases grises permettent successivement l'obtention de systèmes plus complexes en orange, donnant leurs noms aux algorithmes principaux d'assimilation de données en vert.

L'inférence bayésienne est difficilement utilisable tel quel pour un système continu. Plusieurs approximations permettent l'assimilation de données grâce au théorème de Bayes. La première est de présumer que le modèle du système étudié suit un processus de Markov, i.e. l'état futur du modèle ne dépend que de l'état présent et ne dépend pas des fenêtres d'assimilation

précédentes. Considérons un calcul séquentiel à travers les fenêtres d'assimilations $\iota \in (1, \dots, I)$, cette propriété peut s'exprimer par [37] :

$$f(u_\iota | u_{\iota-1}, u_{\iota-2}, \dots, u_0) = f(u_\iota | u_{\iota-1}) \quad (3.2)$$

Le terme de fenêtre d'assimilation temporelle désigne les intervalles de temps (fixes ou non selon les algorithmes) entre chaque jeu de mesures/observations disponibles. Les algorithmes d'assimilation ont des comportements différents vis-à-vis de la propagation de l'information de fenêtre à fenêtre ou encore des temps de mises à jour.

Une deuxième approximation, souvent erronée, est de considérer que les erreurs de mesures d'observations sont indépendantes entre les fenêtres d'assimilation de données [37]. Or, dans la plupart des systèmes physiques, ce genre de corrélations existe généralement lorsque l'on considère des fenêtres d'assimilations très proches dans le temps. Enfin, l'état final de la fenêtre I est donné par la prise en compte de l'information de l'ensemble des fenêtres d'assimilation dans la formule de Bayes. Si l'on considère l'indépendance des fenêtres d'assimilation grâce au processus de Markov, l'état final peut être approximé à partir de la fenêtre précédente, qui est elle-même approximé à partir de la fenêtre précédente etc... La formule de Bayes prend alors sa forme récursive :

$$f(u_\iota | \alpha_\iota) = \frac{f(\alpha_\iota | u_\iota) f(u_{\iota-1} | \alpha_{\iota-1})}{f(\alpha_\iota)} \quad (3.3)$$

Une autre approximation menant à la formulation des algorithmes principaux d'assimilation de données et à la simplification du problème est l'hypothèse gaussienne. La distribution *a priori* du vecteur d'état u ainsi que l'erreur sur les observations attribuées aux mesures ϵ sont considérées comme des distributions gaussiennes.

$$f(u) = \mathcal{N}(u_0, Q) \quad (3.4)$$

$$f(\alpha | u) = f(\bar{\alpha} | u) + f(\epsilon) \text{ avec } f(\epsilon) = \mathcal{N}(0, R) \quad (3.5)$$

u_0 représente l'état initial *a priori* de son affectation par le modèle. Il constitue la moyenne affectée par Q sa matrice de covariance d'erreur. R est la matrice de covariance d'erreur des observations.

Processus physique scalaire

Par exemple, prenons l'évolution de l'état u , qui ici n'est plus constitué que d'un seul scalaire [4, 147]. u est normalement distribué suivant une loi normale de moyenne μ et de variance τ^2 telle que $u \sim \mathcal{N}(\mu, \tau^2)$. Considérons un vecteur d'observations indépendantes $\alpha = (\alpha_1, \dots, \alpha_n)$. Ce dernier est bruité et conditionné par la valeur de u . Il suit donc également une loi normale telle que $\alpha | u \sim \mathcal{N}(u, \sigma^2)$. Sa densité de probabilité est donnée par :

$$f(\alpha | u) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-0.5(\alpha_i - u)^2/\sigma^2\right) \propto \exp\left(-0.5 \sum_{i=1}^n (\alpha_i - u)^2/\sigma^2\right) \quad (3.6)$$

En reprenant la formule de Bayes $f(u | \alpha) \propto f(\alpha | u) f(u)$:

$$\begin{aligned}
 f(u|\alpha) &\propto \exp\left(-0.5\left[\sum_{i=1}^n(\alpha_i - u)^2/\sigma^2 + (u - \mu)^2/\tau^2\right]\right) \\
 &\propto \exp\left(-0.5\left[u^2(n/\sigma^2 + 1/\tau^2) - 2\left(\sum_{i=1}^n \alpha_i/\sigma^2 + \mu/\tau^2\right)u\right]\right)
 \end{aligned} \tag{3.7}$$

La distribution *a posteriori* devient alors :

$$u|\alpha \sim \mathcal{N}(\mu_{u|\alpha}, \sigma_{u|\alpha}^2) \tag{3.8}$$

avec

$$\mu_{u|\alpha} = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \left(\sum_{i=1}^n \alpha_i/\sigma^2 + \mu/\tau^2\right) \tag{3.9}$$

$$\sigma_{u|\alpha}^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \tag{3.10}$$

L'espérance du système *a posteriori* s'exprime par :

$$E(u|\alpha) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2} (n\bar{\alpha}/\sigma^2 + \mu/\tau^2) = \omega_\alpha \bar{\alpha} + \omega_\mu \mu \tag{3.11}$$

Avec $\bar{\alpha} = \sum_{i=1}^n \alpha_i/n$, $\omega_\alpha = n\tau^2/(n\tau^2 + \sigma^2)$ et $\omega_\mu = \sigma^2/(n\tau^2 + \sigma^2)$. L'état final est donc une moyenne pondérée de la moyenne *a priori* du système et des mesures d'observations de l'état réel inconnu. Le calcul de l'espérance du système *a posteriori* peut également s'exprimer sous la forme :

$$E(u|\alpha) = \mu + \left(\frac{n\tau^2}{\sigma^2 + n\tau^2}\right)(\bar{\alpha} - \mu) = \mu + K(\bar{\alpha} - \mu) \tag{3.12}$$

Et la variance *a posteriori* par :

$$var(u|\alpha) = (1 - K)\tau^2 \tag{3.13}$$

Le terme K est considéré comme un gain. Il est fondamental dans les algorithmes principaux d'assimilations de données. Comme indiqué par les équations précédentes, le gain agit comme un facteur correctif à la différence $(\bar{\alpha} - \mu)$, appelée innovation, ainsi qu'à τ dans le calcul de la variance finale.

Processus physique vectoriel

Généralisons à présent les expressions précédentes dans le cas où le processus physique est décrit par un vecteur $u \sim \mathcal{N}(\mu, P)$ avec μ la moyenne et P la matrice de covariance de u . L'observation est décrite par un vecteur de mesures de l'état réel du système u tel que $\alpha|u \sim \mathcal{N}(Hu, R)$ avec H la matrice de projection des valeurs de l'état sur l'espace des mesures et R la matrice de

covariance d'erreur des mesures (dont le contenu est explicité dans les sections suivantes). Les expressions dans le cas vectoriel se généralisent tel que :

$$u|\alpha \sim \mathcal{N}(\mu_{u|\alpha}, \Sigma_{u|\alpha}) \quad (3.14)$$

avec

$$\mu_{u|\alpha} = (H^T R^{-1} H + P^{-1})^{-1} (H^T R^{-1} \alpha + P^{-1} \mu) \quad (3.15)$$

$$\Sigma_{u|\alpha} = (H^T R^{-1} H + P^{-1})^{-1} \quad (3.16)$$

L'espérance conditionnelle et la variance *a posteriori* deviennent alors :

$$\begin{aligned} E(u|\alpha) &= (H^T R^{-1} H + P^{-1})^{-1} H^T R^{-1} \alpha + (H^T R^{-1} H + P^{-1})^{-1} P^{-1} \mu \\ &= \mu + K(\alpha - H\mu) \end{aligned} \quad (3.17)$$

Et

$$\Sigma(u|\alpha) = (H^T R^{-1} H + P^{-1})^{-1} = (I - KH)P \quad (3.18)$$

Avec

$$K = PH^T(HPH^T + R)^{-1} \quad (3.19)$$

Ces expressions sont similaires au cas scalaire. Elles montrent l'importance du gain K ainsi que l'importance des données *a priori* et des observations dans l'espérance conditionnelle *a posteriori* des algorithmes basés sur l'inférence Bayésienne.

3.2 Méthodes variationnelles

La définition d'*a priori* Gaussiens permet de simplifier l'estimation de la fonction de densité de probabilité postérieure. Ceci simplifie l'estimation de la solution *Maximum a Posteriori* ou MAP [37], qui est à la base des algorithmes variationnels comme le 3DVar ou 4DVar. Cette solution peut être vue comme la plus probable pour u étant donné les mesures α .

$$u_{\text{MAP}} = \operatorname{argmax}(u|\alpha) \quad (3.20)$$

Ceci revient à minimiser une fonction de coût $\mathcal{J}(u)$ tel que :

$$u_{\text{MAP}} = \operatorname{argmin} \mathcal{J}(u) \quad (3.21)$$

Cette fonction de coût est souvent écrite dans un contexte discrétisé, facilitant sa manipulation et son application dans des cas concrets. Dans le contexte décrit dans ce manuscrit, \mathcal{J} peut se traduire par :

$$\mathbf{J}(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{u}^p)^T \mathbf{P}^{-1}(\mathbf{u} - \mathbf{u}^p) + \frac{1}{2}(\mathbf{H}\mathbf{u} - \boldsymbol{\alpha})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{u} - \boldsymbol{\alpha}) \quad (3.22)$$

Où \mathbf{u}^p représente l'état précédent *a priori* de l'analyse, on parle aussi d'état prévisionnel, et $\boldsymbol{\alpha}$ les mesures faites à partir de l'état réel du système. \mathbf{H} est la matrice d'observation, elle permet de projeter les valeurs de l'état \mathbf{u} sur l'espace des mesures d'observations $\boldsymbol{\alpha}$. \mathbf{R} et \mathbf{P} sont les matrices de covariance d'erreurs des mesures ainsi que celles associées à l'évolution du système par rapport à son état précédent.

3.2.1 3DVar

Prenons le cas d'un système stationnaire. Un état prévisionnel $\mathbf{u}^p = \mathbf{u}^p + \boldsymbol{\eta}^p$ ainsi que des mesures de l'état réel du système sont disponibles. Ces dernières se présentent sous la forme $\boldsymbol{\alpha} = \mathbf{H}\mathbf{u}^r + \boldsymbol{\epsilon}$. L'analyse effectuée dans ce cas se nomme analyse variationnelle en trois dimensions (*three-dimensional variational*) ou 3DVar. Son objectif est la minimisation de la fonction de coût \mathbf{J} précédemment évoquée. La recherche de ce minimum équivaut à la recherche de l'état pour lequel le gradient de \mathbf{J} est nul :

$$\nabla \mathbf{J}(\mathbf{u}_{\text{MAP}}) = 0 \quad (3.23)$$

C'est-à-dire

$$\nabla \mathbf{J}(\mathbf{u}^a) = \mathbf{P}^{-1}(\mathbf{u}^a - \mathbf{u}^p) - \mathbf{H}^T \mathbf{R}^{-1}(\boldsymbol{\alpha} - \mathbf{H}\mathbf{u}^a) = 0 \quad (3.24)$$

Où $\mathbf{u}^a = \mathbf{u}_{\text{MAP}}$ représente l'état *a posteriori* analysé du système. La résolution de cette équation entraîne :

$$\begin{aligned} \mathbf{P}^{-1}(\mathbf{u}^a - \mathbf{u}^p) &= \mathbf{H}^T \mathbf{R}^{-1}(\boldsymbol{\alpha} - \mathbf{H}\mathbf{u}^a) \\ (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})\mathbf{u}^a &= \mathbf{H}^T \mathbf{R}^{-1} \boldsymbol{\alpha} + \mathbf{P}^{-1} \mathbf{u}^p \\ \mathbf{u}^a &= (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{R}^{-1} \boldsymbol{\alpha} + \mathbf{P}^{-1} \mathbf{u}^p) \\ \mathbf{u}^a &= (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{R}^{-1} \boldsymbol{\alpha} + (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})\mathbf{u}^p - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{u}^p) \\ \mathbf{u}^a &= \mathbf{u}^p + (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\boldsymbol{\alpha} - \mathbf{H}\mathbf{u}^p) \\ \mathbf{u}^a &= \mathbf{u}^p + \mathbf{K}(\boldsymbol{\alpha} - \mathbf{H}\mathbf{u}^p) \end{aligned}$$

Ceci fait apparaître le terme appelé *innovation* $\boldsymbol{\alpha} - \mathbf{H}\mathbf{u}^p$ traduisant la différence entre l'observation de l'état réel et la prévision de l'observation par l'état prévisionnel. Le gain $\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}$ apparaît également à nouveau ici tel qu'établi lors du calcul de l'espérance conditionnelle pour un processus Gaussien. $\mathbf{H}\mathbf{P}\mathbf{H}^T$ correspond à la covariance des erreurs de l'état prévisionnel du système projeté sur l'espace des observations, \mathbf{R} traduit la covariance des erreurs de mesures des observations. Le numérateur $\mathbf{P}\mathbf{H}^T$ projette à nouveau la matrice de covariance \mathbf{P} dans l'espace de l'état \mathbf{u} . Ce rapport de matrices de covariances est identique au rapport des variances identifié lors de l'estimation de l'espérance *a posteriori* pour un processus Gaussien scalaire $n\tau^2/(\sigma^2 + n\tau^2)$. Ceci démontre que la meilleure estimation de l'état *a posteriori* est une moyenne pondérée des observations de l'état réel et des observations prévisionnelles. Cette pondération s'effectue par le calcul du rapport des variances. On montrera Sec. 3.3 la très grande similarité du 3DVar avec le filtre de Kalman, qui peut être assimilé à un 3DVar séquentiel.

3.2.2 4DVar

Dans le cas instationnaire, le 3DVar ne s'applique plus. La prise en compte de la dimension temporelle conduit à l'utilisation de l'analyse variationnelle en quatre dimensions (*four-dimensional variational*) ou 4DVar. Supposons un système dynamique évoluant dans le temps tel qu'un écoulement instationnaire. Les équations d'état du système dynamique sont décrites via la matrice \mathbf{M}_{k+1} traduisant l'évolution de l'état du système du temps k à $k+1$ pour $k = 0, 1, \dots, k_f$:

$$\mathbf{u}_{k+1} = \mathbf{M}_{k+1} \mathbf{u}_k \quad (3.25)$$

Ici, le modèle est considéré comme déterministe, c'est-à-dire qu'il ne produit aucune erreur. Les observations mesurées sur l'état réel du système sont également disponibles à chaque pas de temps k tel que :

$$\boldsymbol{\alpha}_k = \mathbf{H}_k \mathbf{u}_k^r + \boldsymbol{\epsilon}_k \quad (3.26)$$

Les erreurs d'observations $\boldsymbol{\epsilon}$ se basent sur la matrice de covariance d'erreur \mathbf{R}_k et comme évoqué Sec. 3.1 celles-ci ne sont pas corrélées dans le temps. Il est à noter que \mathbf{H}_k ne varie généralement pas dans le temps. Dans le cas où l'état initial \mathbf{u}_0^p ainsi que sa matrice de covariance d'erreur \mathbf{P}_0^p sont connus, alors la solution de l'état \mathbf{u}_{k+1} est définie par la condition initiale correspondante \mathbf{u}_0^p . La fonction de coût devient :

$$\mathbf{J}(\mathbf{u}_0) = \frac{1}{2}(\mathbf{u}_0 - \mathbf{u}_0^p)^T (\mathbf{P}_0^p)^{-1} (\mathbf{u}_0 - \mathbf{u}_0^p) + \frac{1}{2} \sum_{k=0}^{k_f} (\mathbf{H}_k \mathbf{u}_k - \boldsymbol{\alpha}_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{u}_k - \boldsymbol{\alpha}_k) \quad (3.27)$$

Cette dernière dépend maintenant de toutes les innovations disponibles dans la fenêtre d'assimilation et non pas d'un temps unique comme pour le 3DVar. Cette version est appelée *strong constraint 4DVar*, que l'on pourrait traduire par 4DVar à forte contrainte. Ici la minimisation de la fonction de coût fourni l'état initial du modèle qui conviendra le mieux. Que se passe-t-il maintenant lorsque le modèle n'est plus purement déterministe ?

$$\mathbf{u}_{k+1} = \mathbf{M}_{k+1} \mathbf{u}_k + \boldsymbol{\eta}_{k+1} \quad (3.28)$$

Le bruit de modèle noté $\boldsymbol{\eta}_k$ a une matrice de covariance \mathbf{Q}_k . Ce dernier n'est pas corrélé temporellement avec les bruits de mesure ou avec l'erreur initiale de l'état. \mathbf{J} devient alors :

$$\begin{aligned} \mathbf{J}(\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_K) &= \frac{1}{2}(\mathbf{u}_0 - \mathbf{u}_0^p)^T (\mathbf{P}_0^p)^{-1} (\mathbf{u}_0 - \mathbf{u}_0^p) \\ &\quad + \frac{1}{2} \sum_{k=0}^{k_f} (\mathbf{H}_k \mathbf{u}_k - \boldsymbol{\alpha}_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{u}_k - \boldsymbol{\alpha}_k) \\ &\quad + \frac{1}{2} \sum_{k=0}^{k_f-1} (\mathbf{u}_{k+1} - \mathbf{M}_{k+1} \mathbf{u}_k)^T \mathbf{Q}_{k+1}^{-1} (\mathbf{u}_{k+1} - \mathbf{M}_{k+1} \mathbf{u}_k) \end{aligned} \quad (3.29)$$

La fonction de coût à minimiser dépend maintenant de tous les états consécutifs et non plus uniquement de l'état initial. Ce modèle n'est plus purement déterministe. Cette méthode est appelée *weak constraint 4DVar*, que l'on pourrait traduire par 4DVar à faible contrainte.

Le calcul du gradient $\nabla \mathbf{J}$ implique le stockage de larges matrices, ce qui rend l'application directe des méthodes 4DVar très coûteuse. Pour contourner le problème, l'algorithme variationnel est généralement résolu grâce à l'approche de l'adjoint qui ne sera pas montrée en détails dans ce manuscrit (voir [4, 37]). Dans le cas du *strong constraint 4DVar*, le gradient peut s'exprimer tel que :

$$\delta \mathbf{J} = (\nabla_{\mathbf{u}_0} \mathbf{J})^T \delta \mathbf{u}_0 \quad (3.30)$$

L'adjoint permet d'exprimer cette équation tel que :

$$\delta \mathbf{J} = \mathbf{A}_0^T \delta \mathbf{u}_0 \quad (3.31)$$

Les termes \mathbf{A}_k sont les états adjoints, qui pour permettre la résolution de la minimisation de la fonction de coût s'expriment par :

$$\mathbf{A}_{k_f} = H_{k_f}^T \mathbf{R}_{k_f}^{-1} (\mathbf{H}_{k_f} \mathbf{u}_{k_f} - \boldsymbol{\alpha}_{k_f}) \quad (3.32)$$

$$\mathbf{A}_k = H_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{u}_k - \boldsymbol{\alpha}_k) + M_k^T \mathbf{A}_{k+1} \quad \text{avec } k = k_f - 1, \dots, 1 \quad (3.33)$$

$$\mathbf{A}_0 = (\mathbf{P}_0^p)^{-1} (\mathbf{u}_0 - \mathbf{u}_0^p) + H_0^T \mathbf{R}_0^{-1} (\mathbf{H}_0 \mathbf{u}_0 - \boldsymbol{\alpha}_0) + M_0^T \mathbf{p}_1 \quad (3.34)$$

Où H_k et M_k sont les matrices jacobiennes de \mathbf{H}_k et \mathbf{M}_k . L'algorithme du *strong constraint 4DVar* se dessine alors. Il nécessite le calcul et le stockage de la solution jusqu'au temps k , puis le calcul inverse de l'adjoint à partir de la solution finale jusqu'au temps initial et enfin le calcul du gradient pour obtenir le minimum de la fonction.

3.2.3 Application des méthodes variationnelles en mécanique des fluides

Les méthodes variationnelles sont maintenant largement étudiées en mécanique des fluides numériques. Le début de cette popularisation a été effectué dans les années 2000, notamment avec les travaux de Ruhnau et al. [111] à partir de 2005. Ces travaux se concentrent cependant uniquement sur la reconstruction de champs complets sur les plans des données PIV (*particle image velocimetry*) associés, notamment en régularisant les champs reconstruits grâce à l'utilisation de l'équation de transport de la vorticit e pour un  coulement incompressible [112]. N anmoins, cette section se concentrera plus sp ecifiquement sur les travaux men es ces dix derni eres ann ees, comprenant notamment l'utilisation de m ethodes 4D-Var pour la reconstruction de champs de vitesse sur des simulations 2D, 3D, RANS, LES et DNS, l'optimisation de fermetures de mod eles de turbulence ou encore la calibration de champs initiaux ou sources scalaires. Enfin, certains travaux [86, 88] comprennent  galement l' tude du placement des capteurs d'observations et sont donc valables pour n'importe quelle m ethodes d'assimilation de donn ees.

- **Reconstruction de champs physiques**

La reconstruction de champs de vitesses   partir de donn ees hautes fid elit es est un des axes privil egi es d' tude avec les m ethodes variationnelles d'assimilation de donn ees. Le travail effectu e par Foures et al. [41] a notamment ouvert la voie en inf erant le champ de vitesse dans un  coulement RANS autour d'un cylindre en 2D   partir de donn ees DNS

moyennées. L’auteur note une reconstruction satisfaisante du champ ainsi que le début d’une optimisation des termes de second ordre malgré l’utilisation de champs moyens uniquement. Par la suite, l’inférence de champs issus de simulations RANS passe surtout par l’optimisation de la fermeture du modèle ainsi que par l’ajout de termes sources dans les équations de bilan (paragraphe suivant). La reconstruction pure du champ de vitesse continue cependant à être étudiée pour des simulations plus haute-fidélité où les équations de Navier–Stokes sont tout ou partie résolues. Bauweraerts et al. [6] ont par exemple étudié cette possibilité à partir de données synthétiques reproduisant des données LIDAR sur une couche limite atmosphérique modélisé par LES. La zone reconstruite dans le cône d’influence des données entraîne une erreur sur la variance de 15 à 25%, puis 100% à l’extérieur de cette région. Li et al. [67] utilisent le 4D-Var pour reconstruire les petites échelles de la turbulence à partir de données DNS haute-fidélité sur un maillage plus grossier. Les champs obtenus sont corrélés à 95% notamment pour l’enstrophie. Wang et Zaki étudient de près la reconstruction des champs via une série d’articles dédiés à l’utilisation de méthodes variationnelles, en particulier, la méthode développée dans [143] est utilisée par la suite [144] pour la reconstruction de champs DNS initiaux dans un canal plan turbulent. La reconstruction du champ initial de vitesse à partir d’une grille d’observations grossière en utilisant la méthode variationnelle entraîne une réduction de 80% de l’erreur comparée à une simple interpolation de la grille grossière vers le maillage DNS. Enfin, Chandramouli et al. [14] étudient la reconstruction du sillage d’un cylindre à partir de données DNS, émulant des mesures PIV. La reconstruction est effectuée sur une simulation LES plus grossière où cette fois l’obstacle est modélisé par l’utilisation de la méthode des frontières immergées (IBM).

- **Optimisation de fermetures de modèles**

L’optimisation de fermetures de modèles est également un axe privilégié par l’inférence d’écoulements à partir de méthodes variationnelles, et particulièrement 4D-Var. Symon et al. [126] assimile des données PIV pour la reconstruction d’un champ 2D RANS autour d’un profil d’aile. Un terme de forçage est utilisé dans l’équation de quantité de mouvement. Ils mettent entre autres en évidence la résolution de grille de PIV minimale pour une bonne reconstruction du champ de vitesse. Ghosh et al. [44] étudient également l’inférence de champs 2D RANS avec forçage. L’originalité de l’article réside dans l’utilisation d’observations très éloignées de la moyenne (*outliers*) et d’une fonction de coût *Huber*. Concernant la modélisation RANS 3D, Li et al. ont d’abord étudié l’optimisation des constantes de modèles RANS 3D sur différents cas et ont montré une amélioration de la prédiction en accord avec les données d’observations. Ben Ali et al. [7, 8] ont mené une étude similaire sur un corps non profilé (comme étudié dans le chapitre 5). Ils ont d’abord inféré les constantes du modèle $\mathcal{K} - \varepsilon$ à partir de données expérimentales issues de sondes de pression, puis des termes de forçages ont été ajoutés à l’équation de quantité de mouvement, améliorant la prédiction du sillage. Franceschini et al. [42] ont également ajouté différents termes sources pour l’inférence d’une marche descendante en utilisant le modèle RANS-SA 3D (*Spalart-Allmaras*). Ils ont spécifiquement comparé l’utilisation d’un terme source dans l’équation de quantité de mouvement et dans l’équation du mo-

dèle de turbulence. Ils mettent notamment en lumière la rigidité du modèle de turbulence aux changements par terme source. Le premier cas est alors meilleur lorsque beaucoup d'observations sont disponibles et le deuxième lorsque peu sont disponibles. Enfin, Cato et al. [13] se sont appliqués à comparer différentes approches pour l'inférence de modèle RANS-SA 3D i.e. l'addition de termes de forçage dans l'équation de quantité de mouvement, dans l'équation de modèle *Spalart-Allmaras* ou encore la correction du champ de viscosité turbulente. Les deux premières approches se sont montrées plus efficaces que la dernière. Mons et al. [87] ont également étudié l'inférence des paramètres de modèle LES avec ajout d'un terme dans l'équation de quantité de mouvement sur le canal plan turbulent. Ceci est largement abordé dans le chapitre 6.

- ***Calibration de conditions/champs initiaux ou autres sources scalaires***

L'assimilation de données variationnelle est également régulièrement utilisée pour le calibrage ou la reconstruction de champs initiaux et autres sources scalaires. Par exemple, Gronsksis et al. [48] infèrent des conditions aux limites et initiales d'une configuration 2D DNS avec le 4DVar. Mons et al. [91] infèrent l'écoulement autour d'un cylindre pour un écoulement laminaire 2D compressible à partir d'observations synthétiques. En particulier, l'optimisation se porte sur les conditions d'entrée et initiales de l'écoulement. Trois méthodes différentes sont comparées, le 4DVar classique, le 4DEnVar et enfin le EnKS (Ensemble Kalman Smoother). Cette dernière est une variante du EnKF permettant l'utilisation d'observations successives pour l'estimation de l'état en un temps donné. Le 4DEnVar contourne l'utilisation d'un adjoint du gradient de la fonction de coût en utilisant un ensemble de simulations comme dans les techniques ensemblistes séquentielles abordées dans la section suivante. Cette méthode a notamment été utilisée par Yang et al. [151] pour l'inférence d'un modèle "shallow water". L'EnKS se montre plus sensible aux conditions initiales et aux observations tandis que le 4DEnVar est moins sensible aux variations de l'ensemble. Globalement, le 4DVar montre de meilleurs résultats que les autres méthodes, cependant son utilisation est plus coûteuse pour un bénéfice relatif. Par la suite, Mons et al. se sont consacrés à l'utilisation de la méthode 4DEnVar dans différents cas. La méthode est augmentée grâce à l'utilisation de POD pour l'identification des sources de pollution et caractéristiques de vent dans des zones urbaines avec des simulations LES [90]. Elle est également couplée à une méthode de Kriging pour la reconstruction spatiale d'une source scalaire dans un canal plan turbulent en DNS [89].

- ***Aparté sur le placement optimal de capteurs d'observations***

Mons et al. ont également étudié le placement optimal des capteurs d'observations avec des méthodes variationnelles. Ces études sont cependant valables pour toute méthode d'assimilation de données puisqu'elles optimisent le placement des capteurs *a priori* de l'assimilation. Dans un premier travail [86], le placement des capteurs d'observations est optimisé via l'étude de la sensibilité de l'écoulement aux changements de conditions aux limites, conditions initiales ou encore aux paramètres de modèle. Les lieux où la variabilité est la plus forte sont alors utilisés pour des observations de vitesses. La méthode est validée sur le cas 2D laminaire d'un cylindre en rotation. Un cadre plus complet est présenté dans un travail distinct [88] où, outre les approches de placement linéaire des

capteurs, des méthodes de second ordre basées sur l'utilisation d'adjoints sont explorées. Ils montrent une reconstruction du champ à partir de très peu de mesures de vitesse placées à des endroits optimisés. Un autre avantage de cette approche est sa capacité à estimer le nombre minimum de capteurs requis pour atteindre un niveau souhaité de précision de reconstruction. Cependant, il est essentiel de noter que ces travaux s'appuient sur des équations RANS et/ou laminaire et un nombre de Reynolds relativement faible. L'optimisation d'un modèle de turbulence réel est identifiée comme perspective pour une future étude. En outre, les difficultés liées à la résolution du problème de placement des capteurs découlent des limites du processus de mesure lui-même. Par exemple, obtenir des mesures denses de l'ensemble de l'écoulement moyen peut être problématique, et de telles données complètes peuvent ne pas toujours être disponibles, en particulier pour des mesures expérimentales.

Les méthodes variationnelles sont donc largement utilisées pour la reconstruction de champs et pour la calibration de fermetures de modèles en mécanique des fluides numérique. En particulier, l'ajout de termes sources dans les équations de bilan est une méthode privilégiée pour la correction de modèles basés sur la viscosité turbulente. Les champs corrigés à partir de la méthode 4D-Var montrent également d'excellents résultats. L'application de cette méthode aux modèles non-linéaires de mécanique des fluides est rendue possible grâce à l'utilisation d'un adjoint pour le gradient de la fonction de coût. Cependant, l'implémentation de l'approche de l'adjoint peut s'avérer complexe pour ce type d'applications [91]. Son coût de calcul assez élevé peut également rendre prohibitif son utilisation. Certains préféreront donc l'utilisation de méthodes ensemblistes telles que le 4DVar pour contourner l'utilisation de l'adjoint grâce à un ensemble d'états, comme peut le faire le EnKF (Sec. 3.3.2). Cependant, toutes les applications citées impliquent l'inférence d'un écoulement statistiquement stationnaire, l'inférence d'un seul champ instantané tel qu'un champ initial ou encore l'optimisation de variables globales. Bien que Gronska et al. [48] et Mons et al. [91] approchent l'inférence d'un comportement instationnaire, aucun de ces cas d'études ne se consacre à l'inférence d'un comportement dynamique nécessaire pour la prédiction augmentée d'événements extrêmes.

3.3 Le filtre de Kalman appliqué à la mécanique des fluides

Le filtre de Kalman (KF) est un outil d'AD bien connu introduit pour la première fois en 1960 par R.E. Kalman [56] pour estimer un état de système augmenté combinant la prédiction d'un modèle avec des observations de données externes haute-fidélité mais présentes en faible nombre. Les deux sources d'information sont affectées par des incertitudes, qui sont assimilées à des variables aléatoires Gaussiennes. Cette section reviendra dans un premier temps sur le filtre de Kalman classique tel qu'énoncé en 1960 puis sur son cadet ensembliste. Les méthodes d'inflation, de localisation et d'hyper-localisation seront ensuite abordées avant de terminer ce chapitre par une revue de littérature concernant les méthodes séquentielles en mécanique des fluides.

3.3.1 Le filtre de Kalman et les contraintes de linéarité

Notons à nouveau le vecteur d'état \mathbf{u} et le vecteur d'observations $\boldsymbol{\alpha}$. Les valeurs prédites par le modèle aux coordonnées des observations sont indiquées par $\mathbf{s} = \mathbf{H}\mathbf{u}$. \mathbf{H} est donc une matrice de projection qui permet d'interpoler les valeurs de l'état du modèle sur les coordonnées de l'espace d'observation. La matrice d'état \mathbf{u} au temps $k + 1$ est calculée par le modèle à partir de l'état précédent tel que :

$$\mathbf{u}_{k+1} = \mathbf{M}_{k+1}\mathbf{u}_k + \boldsymbol{\eta}_k \quad (3.35)$$

Les observations sont mesurées à partir de l'état considéré comme l'état réel ou vrai du système tel que :

$$\boldsymbol{\alpha}_k = \mathbf{H}_k\mathbf{u}_k^r + \boldsymbol{\epsilon}_k \quad (3.36)$$

Comme vu précédemment en sec. 3.1, une des hypothèses fondamentale est de considérer que les incertitudes des variables aléatoires mises en jeu suivent des distributions gaussiennes. Les erreurs $\boldsymbol{\eta}$ ainsi que $\boldsymbol{\epsilon}$ sont donc définies tel que :

$$\boldsymbol{\eta}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \quad (3.37)$$

$$\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \mathbf{R}_k) \quad (3.38)$$

Les matrices \mathbf{Q}_k et \mathbf{R}_k sont les matrices de covariance d'erreur de modèle et de mesures respectivement. Ces erreurs sont des bruits blancs gaussiens. Le filtre de Kalman est un algorithme séquentiel qui nécessite la mise à jour de sa matrice de covariance d'erreur pour les mises à jour consécutives de l'état. La matrice de covariance d'erreur \mathbf{P}_{k+1}^p peut être définie tel que :

$$\mathbf{P}_k^{p/a} = \text{cov}(\mathbf{e}_k^{p/a}) = E \left[\mathbf{e}_k^{p/a} (\mathbf{e}_k^{p/a})^T \right] \quad (3.39)$$

avec

$$\mathbf{e}_k^{p/a} = \mathbf{u}_k^{p/a} - \mathbf{u}_k^r \quad (3.40)$$

où l'exposant r désigne l'état réel du système. L'exposant p (*prévision*) fait référence à l'état prévisionnel de la matrice d'état suite à l'avancement temporel des grandeurs physiques par le modèle. L'exposant a (*analyse*) représente l'état augmenté final de l'algorithme à la suite de la phase d'*analyse*. Si la dérivation de la matrice de covariance d'erreur est possible à partir de substitutions dans sa formulation 3.39, il est aussi possible d'identifier sa formulation à partir de la fonction de coût établie pour le 3DVar sec. 3.1. Prenons l'hessienne de la fonction de coût (matrice des dérivées partielles secondes), pour rappel (l'indexe de temps est omis par simplicité) :

$$\mathbf{J}(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{u}^p)^T (\mathbf{P}^p)^{-1} (\mathbf{u} - \mathbf{u}^p) + \frac{1}{2}(\mathbf{H}\mathbf{u} - \boldsymbol{\alpha})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{u} - \boldsymbol{\alpha}) \quad (3.41)$$

Donc

$$\nabla\nabla\mathbf{J}(\mathbf{u}^p) = (\mathbf{P}^p)^{-1} + \mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T \quad (3.42)$$

Cette dernière ne dépend pas de l'état mais seulement de la matrice de covariance d'erreur prévisionnelle ainsi que de la matrice de covariance des erreurs d'observations. La distribution analysée étant également gaussienne, il vient :

$$\mathbf{J}(\mathbf{u}) = \frac{1}{2}(\mathbf{u} - \mathbf{u}^a)^T(\mathbf{P}^a)^{-1}(\mathbf{u} - \mathbf{u}^a) + \text{constante} \quad (3.43)$$

Et donc

$$\nabla\nabla\mathbf{J}(\mathbf{u}^a) = (\mathbf{P}^a)^{-1} \quad (3.44)$$

Les deux matrices hessiennes devant être égales, le résultat sous-entend que l'inverse de la matrice hessienne de la fonction de coût pour l'état prévisionnel est égal à la matrice de covariance d'erreur analysée, conduisant à la formulation conventionnelle de la mise à jour de la matrice de covariance d'erreur dans le filtre de Kalman :

$$\mathbf{P}^a = (\mathbf{P}^p + \mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T)^{-1} = \mathbf{P}^p - \mathbf{P}^p\mathbf{H}^T(\mathbf{H}\mathbf{P}^p\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{R} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^p \quad (3.45)$$

Le gain de Kalman \mathbf{K}_{k+1} est obtenu à partir de la manipulation de cette matrice, sa formulation a déjà été identifiée dans les algorithmes précédents :

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^p\mathbf{H}_{k+1}^T(\mathbf{H}_{k+1}\mathbf{P}_{k+1}^p\mathbf{H}_{k+1}^T + \mathbf{R}_{k+1})^{-1} \quad (3.46)$$

La matrice \mathbf{K} intègre respectivement le niveau de confiance dans le modèle et dans l'observation, qui sont mesurés par la variance des incertitudes affectant les deux sources d'information. Considérons l'avancement temporel du modèle depuis le pas de temps k jusqu'à $k + 1$ dans le cas où l'observation est disponible pour ce dernier temps. L'état augmenté est obtenu comme suit :

$$\mathbf{u}_{k+1}^a = \mathbf{u}_{k+1}^p + \mathbf{K}_{k+1}(\boldsymbol{\alpha}_{k+1} - \mathbf{s}_{k+1}) \quad (3.47)$$

Cette équation dérive donc de l'inférence Bayésienne. C'est également ce qui a permis d'établir la formulation de l'algorithme 3DVar sec. 3.2.1, présentant une formulation de la mise à jour de l'état identique au filtre de Kalman.

Résumons la présentation du filtre de Kalman conventionnel par son algorithme. L'avancement temporel du système est noté $k = 0, 1, \dots, k_f$ tel que précédemment. Les fenêtres d'assimilations sont définies par l'intervalle entre chaque mesure d'observation disponible. Pour simplifier les équations, supposons qu'un vecteur d'observation est disponible à chaque pas de temps, ce qui n'est généralement pas le cas. L'algorithme se décompose donc en deux étapes, dans un premier temps le vecteur d'état ainsi que la matrice de covariance d'erreurs sont prédits par l'avancement temporel du modèle :

$$\mathbf{u}_{k+1} = \mathbf{M}_{k+1}\mathbf{u}_k + \eta_k \quad (3.48)$$

$$\mathbf{P}_{k+1}^p = \mathbf{M}_{k+1}\mathbf{P}_k^p\mathbf{M}_{k+1}^T + \mathbf{Q}_k \quad (3.49)$$

La matrice de covariance d'erreur de modèle étant difficile à obtenir dans un cas concret, elle est souvent négligée, considérant ainsi le modèle de CFD comme parfait. La deuxième étape, communément appelée phase d'analyse, met à jour le vecteur d'état grâce aux observations tel que :

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^p \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{P}_{k+1}^p \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1})^{-1} \quad (3.50)$$

$$\mathbf{u}_{k+1}^a = \mathbf{u}_{k+1}^p + \mathbf{K}_{k+1}(\boldsymbol{\alpha}_{k+1} - \mathbf{s}_{k+1}) \quad (3.51)$$

$$\mathbf{P}_{k+1}^a = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}_{k+1}^p \quad (3.52)$$

La non-linéarité des modèles de mécanique des fluides empêche la mise à jour prévisionnelle de \mathbf{P} . Le *filtre de Kalman étendu* cherche à corriger ce problème en linéarisant les équations. Cependant, le principal inconvénient du filtre de Kalman classique réside dans les manipulations coûteuses de la matrice \mathbf{P} . En effet, dans le cas d'un problème de mécanique des fluides numérique, le champ de vitesse peut facilement être composé de millions de cellules. Or, par exemple, pour une matrice d'état à un million de degrés de liberté N , la matrice de covariance d'erreur est composée de $N \times N$ degrés de liberté. Le coût de stockage d'une telle matrice dans la mémoire vive s'élèverait à 7450 Go, ce qui est impossible. L'application du filtre de Kalman passe alors par l'estimation des erreurs à partir d'un ensemble de vecteurs d'états, ce qui permet d'estimer le gain de Kalman \mathbf{K} sans avoir à stocker et à mettre à jour la matrice \mathbf{P} .

3.3.2 Le filtre de Kalman d'ensemble

Le filtre de Kalman d'Ensemble (EnKF) [35] se base sur le filtre de Kalman classique. Il est de nos jours largement plébiscité pour sa robustesse et sa facilité d'utilisation. L'EnKF a notamment fait ses preuves dans le domaine des sciences météorologiques [4]. Les inconvénients du filtre classique sont contournés grâce à la méthode de Monte Carlo pour estimer la matrice de covariance d'erreur \mathbf{P} en utilisant un ensemble de réalisations pseudo-aléatoires. La procédure utilisée dans cette étude est présentée qualitativement sur la Fig. 3.2 et un algorithme détaillé de l'EnKF (y compris ses améliorations) est fourni dans l'Alg. 2. Un ensemble de N_e états physiques \mathbf{u} , chacun d'eux décrit par N degrés de liberté, est avancé dans le temps à l'aide d'un modèle \mathcal{M} (lignes grises sur la Fig. 3.2), qui peut dans ce cas être non linéaire. Une matrice d'état \mathbf{U} de taille $[N, N_e]$ est assemblée et mise à jour à chaque phase d'*analyse*, ce qui est indiqué par la flèche rouge sur la Fig. 3.2. Chaque colonne $i = 1, \dots, N_e$ de la matrice d'état représente un état physique \mathbf{u}_i obtenu par le i^{th} membre de l'ensemble. Compte tenu de l'avancement temporel de la solution depuis l'instant k jusqu'à $k + 1$ comme dans l'équation 3.47 pour le KF, l'EnKF fournit une estimation d'ensemble de la matrice de covariance d'erreur \mathbf{P} en utilisant l'hypothèse d'indépendance statistique des membres. La méthode de Monte Carlo permet de fournir une estimation de la covariance du système à l'aide de l'estimateur non biaisé suivant :

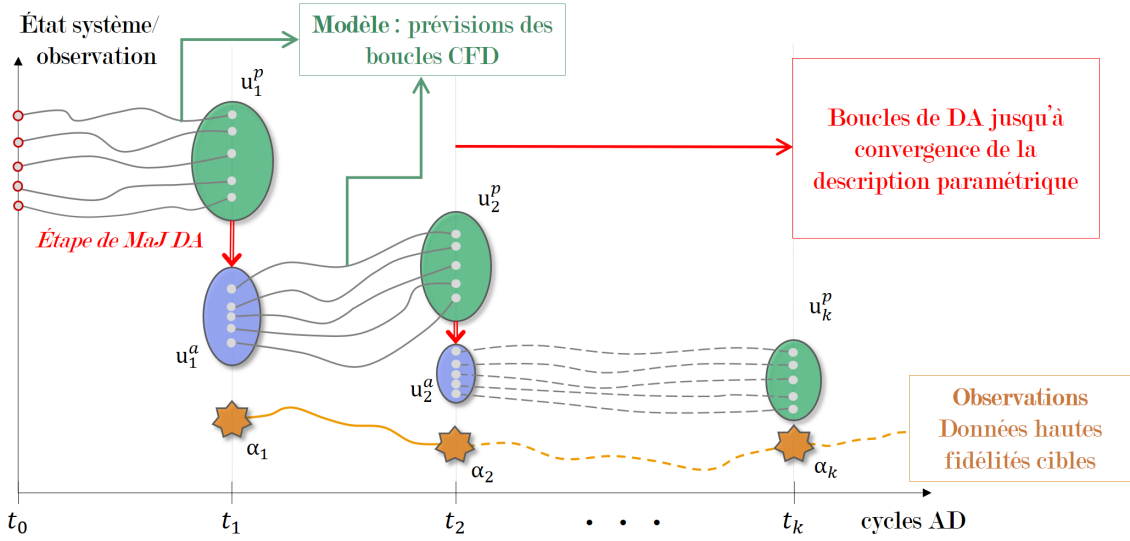


FIGURE 3.2 : Schéma représentant le filtre de Kalman d'ensemble. L'état du modèle u_k^f est mis à jour via la phase d'analyse de l'algorithme EnKF en prenant en compte les observations α_k . Chaque ligne grise représente la trajectoire d'un membre de l'ensemble.

$$\mathbf{P}^p = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{u}_i^p - \langle \mathbf{u}^p \rangle) (\mathbf{u}_i^p - \langle \mathbf{u}^p \rangle)^T = \mathbf{\Gamma} \mathbf{\Gamma}^T \quad (3.53)$$

où $\mathbf{\Gamma}$ est la matrice d'anomalies, qui est dérivée de la matrice d'état \mathbf{U} des membres de l'ensemble. Elle quantifie l'écart des vecteurs d'état par rapport à leur moyenne d'ensemble :

$$[\mathbf{\Gamma}_{k+1}]_i = \frac{\mathbf{u}_{i,k+1}^p - \langle \mathbf{u} \rangle_{k+1}^p}{\sqrt{N_e - 1}}, \quad \langle \mathbf{u} \rangle_{k+1}^p = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{u}_{i,k+1}^p \quad (3.54)$$

Afin d'obtenir une matrice d'observations cohérente dans le cadre d'un calcul ensembliste, le vecteur de N_o observations disponibles est artificiellement perturbé pour obtenir N_e ensembles de valeurs. Pour ce faire, un bruit gaussien borné basé sur la matrice de covariance de l'erreur de mesure \mathbf{R}_{k+1} est ajouté au vecteur d'observation :

$$\boldsymbol{\alpha}_{i,k+1} = \boldsymbol{\alpha}_{k+1} + \epsilon_{i,k+1}, \quad \text{with } \epsilon_{i,k+1} \sim \mathcal{N}(0, \mathbf{R}_{k+1}) \quad (3.55)$$

Comme pour le filtre de Kalman classique, les réalisations du modèle et les observations sont combinées sur l'espace d'observation à l'aide de la matrice de projection \mathbf{H} :

$$\mathbf{s}_{i,k+1} = \mathbf{H} \mathbf{u}_{i,k+1}^p \quad (3.56)$$

L'avantage du filtre de Kalman d'ensemble réside dans le fait que le gain de Kalman est directement estimé via les membres d'ensemble, sans devoir calculer la matrice de covariance d'erreur \mathbf{P} . La covariance des erreurs de l'état prévisionnel du système projeté sur l'espace des observations $\mathbf{H} \mathbf{P}^p \mathbf{H}^T$ est estimé via l'ensemble des vecteurs \mathbf{s}_i tandis que le terme $\mathbf{P}^p \mathbf{H}^T$

est estimé via le produit des matrices d'anomalies citées $\mathbf{\Gamma}(\mathbf{S})^T$. Ces éléments fournissent une estimation du gain de Kalman telle que :

$$\mathbf{K}_{k+1} = \mathbf{\Gamma}_{k+1}(\mathbf{S}_{k+1})^T \left[\mathbf{S}_{k+1}(\mathbf{S}_{k+1})^T + \mathbf{E}_{k+1}(\mathbf{E}_{k+1})^T \right]^{-1} \quad (3.57)$$

avec

$$[\mathbf{S}_{k+1}]_i = \frac{\mathbf{s}_{i,k+1} - \langle \mathbf{s} \rangle_{k+1}}{\sqrt{N_e - 1}}, \quad \langle \mathbf{s} \rangle_{k+1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{s}_{i,k+1} \quad (3.58)$$

$$[\mathbf{E}_{k+1}]_i = \frac{\epsilon_{i,k+1} - \langle \epsilon \rangle_{k+1}}{\sqrt{N_e - 1}}, \quad \langle \epsilon \rangle_{k+1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \epsilon_{i,k+1} \quad (3.59)$$

Pour un ensemble de taille limitée, \mathbf{R}_{k+1} est préféré au produit des matrices d'anomalies des erreurs $\mathbf{E}_{k+1}(\mathbf{E}_{k+1})^T$ dans l'équation 3.57. Cela permet la simplification de l'algorithme ainsi qu'une diminution du coût de calcul [12, 52]. La matrice \mathbf{R}_{k+1} est généralement choisie diagonale de sorte que les erreurs de mesures soient indépendantes. La quantité d'erreur attribuée à chaque mesure est un hyper-paramètre sujet à optimisation. Pour finir, l'état physique prédit par chaque membre de l'ensemble est mis à jour à l'aide du gain de Kalman :

$$\mathbf{u}_{i,k+1}^a = \mathbf{u}_{i,k+1}^p + \mathbf{K}_{k+1}(\boldsymbol{\alpha}_{i,k+1} - \mathbf{s}_{i,k+1}) \quad (3.60)$$

Les approches basées sur l'EnKF peuvent également optimiser simultanément les paramètres libres du modèle pour minimiser l'écart entre le modèle et l'observation lors de la phase d'*analyse*. Ces paramètres sont généralement assemblés dans un tableau appelé $\boldsymbol{\theta}$. Une stratégie simple pour effectuer une telle optimisation est ce que l'on appelle l'*état étendu* [4]. Ici, le problème du EnKF est résolu pour un vecteur d'état \mathbf{u}_{ext} défini comme :

$$\mathbf{u}_{\text{ext}} = \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\theta} \end{bmatrix} \quad (3.61)$$

La taille de l'état étendu est désormais égale à $N_{\text{ext}} = N + N_{\boldsymbol{\theta}}$, où $N_{\boldsymbol{\theta}}$ est le nombre de paramètres à optimiser. Cette modification entraîne une augmentation négligeable du coût de calcul si $N_{\boldsymbol{\theta}} \ll N$ et fournit simultanément une estimation d'état mise à jour et une description paramétrique optimisée pour le modèle à la fin de la phase d'*analyse*. Dans certains cas, il sera préférable d'effectuer l'optimisation paramétrique en amont de l'estimation d'état, cette dernière intervenant après convergence des paramètres.

3.3.3 Corrections des défauts du EnKF

Attardons-nous à présent sur les méthodes de localisation et d'inflation. Celles-ci sont très régulièrement utilisées pour corriger les défauts inhérents à l'utilisation d'un nombre limité de membres d'ensemble dans l'EnKF.

3.3.3.1 L'inflation

L'un des inconvénients majeurs du filtre de Kalman d'Ensemble est l'effondrement rapide de la variabilité de la matrice d'état. La conséquence de la réduction indésirable de la variabilité est la

Algorithme 2 : Algorithme du Filtre de Kalman d'Ensemble (EnKF)

Input : \mathcal{M} , \mathcal{H} , \mathbf{R}_{k+1} , et distributions initiales des vecteurs d'états $\mathbf{u}_{i,0}^a$, où généralement

$\mathbf{u}_{i,0}^a \sim \mathcal{N}(\mu_N, \sigma_N^2)$, k_f le temps final et N_e le nombre de membres

for $k = 0$ to $k_f - 1$ **do**

for $i = 1$ to N_e **do**

1 Avancement temporel des vecteurs d'état :

$$\mathbf{u}_{i,k+1}^p = \mathcal{M}\mathbf{u}_{i,k}^a$$

2 Création de la matrice d'observation à partir des données d'observations en introduisant des erreurs :

$$\boldsymbol{\alpha}_{i,k+1} = \boldsymbol{\alpha}_{k+1} + \boldsymbol{\epsilon}_{i,k+1}, \text{ avec } \boldsymbol{\epsilon}_{i,k+1} \sim \mathcal{N}(0, \mathbf{R}_{k+1})$$

3 Projection de la solution de modèle sur l'espace des observations :

$$\mathbf{s}_{i,k+1} = \mathcal{H}\mathbf{u}_{i,k+1}^p$$

4 Calcul des moyennes d'ensemble :

$$\langle \mathbf{u} \rangle_{k+1}^p = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{u}_{i,k+1}^p, \quad \langle \mathbf{s} \rangle_{k+1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{s}_{i,k+1},$$

for $i = 1$ to N_e **do**

5 Calcul des matrices d'anomalies :

$$[\boldsymbol{\Gamma}_{k+1}]_i = \frac{\mathbf{u}_{i,k+1}^p - \langle \mathbf{u} \rangle_{k+1}^p}{\sqrt{N_e - 1}}, \quad [\mathbf{S}_{k+1}]_i = \frac{\mathbf{s}_{i,k+1} - \langle \mathbf{s} \rangle_{k+1}}{\sqrt{N_e - 1}},$$

6 Calcul du gain de Kalman :

$$\mathbf{K}_{k+1} = \boldsymbol{\Gamma}_{k+1}(\mathbf{S}_{k+1})^T \left[\mathbf{S}_{k+1}(\mathbf{S}_{k+1})^T + \mathbf{R}_{k+1} \right]^{-1}$$

7 Localisation du gain de Kalman :

$$\mathbf{K}_{k+1}^{loc} = [\mathbf{L}]_{m,n} [\mathbf{K}_{k+1}]_{m,n}$$

for $i = 1$ to N_e **do**

8 Mise à jour de la matrice d'état :

$$\mathbf{u}_{i,k+1}^a = \mathbf{u}_{i,k+1}^p + \mathbf{K}_{k+1}^{loc}(\boldsymbol{\alpha}_{i,k+1} - \mathbf{s}_{i,k+1})$$

9 Inflation de la matrice d'état :

$$\mathbf{u}_{i,k+1}^a = (1 + \lambda_i)\mathbf{u}_{i,k+1}^a$$

convergence de la matrice d'état vers un optimum localisé, fortement lié à l'état antérieur fourni. Si ce dernier n'est pas précis, la précision de l'optimisation via l'EnKF peut être gravement affectée. Il est possible d'augmenter la variabilité globale du système et de diminuer les erreurs d'échantillonnage en utilisant un nombre plus élevé de membres dans l'ensemble, gagnant ainsi en précision dans la prédiction de l'EnKF. Cependant, cette stratégie n'est pas concevable pour des applications de dynamique des fluides où les coûts de calcul excluent l'utilisation de grands ensembles. En fait, le nombre de membres généralement utilisé pour les analyses tridimensionnelles est d'environ $N_e \in [40, 100]$ [87, 83], ce qui est assez loin de la limite de convergence classique de Monte-Carlo.

Ce problème est généralement atténué en gonflant la variance de l'ensemble *après* la phase d'*analyse* comme illustré sur la Fig. 3.3. Ceci peut être facilement obtenu en augmentant l'écart entre chaque vecteur d'état $\mathbf{u}_{i,k+1}^a$ et la moyenne d'ensemble $\langle \mathbf{u}^a \rangle$ par des opérations algébriques pilotées via un coefficient λ . Cette procédure est appelée *inflation multiplicative*. La manière dont cette procédure est effectuée peut être *déterministe* ou *stochastique* :

$$\text{deterministe} \quad \mathbf{u}_i^a \longrightarrow \langle \mathbf{u}^a \rangle + \lambda_i (\mathbf{u}_i^a - \langle \mathbf{u}^a \rangle) \quad \text{avec } \lambda_i > 1 \quad (3.62)$$

$$\text{stochastique} \quad \mathbf{u}_i^a \longrightarrow (1 + \lambda_i) \mathbf{u}_i^a \quad \text{avec } \lambda_i \sim \mathcal{N}(0, \sigma) \quad (3.63)$$

L'implémentation déterministe peut être très efficace lors des phases initiales d'*analyse* du calcul. Considérant qu'il est appliqué à l'écart par rapport aux valeurs moyennes de l'ensemble, le processus est assez stable et des valeurs plus élevées de λ peuvent être utilisées. Néanmoins, il est moins efficace lorsque l'ensemble présente un fort effondrement de la variabilité de la solution physique ($\mathbf{u}_i^a - \langle \mathbf{u}^a \rangle \approx 0$). Au contraire, l'inflation stochastique est très utile pour atténuer un effondrement rapide de la variabilité, lui permettant ainsi de cibler une solution optimale globale. La distribution gaussienne utilisée pour déterminer λ_i dans l'Eq. 3.63 est généralement tronqué pour éviter la génération de valeurs aberrantes qui pourraient conduire à la divergence de l'EnKF.

3.3.3.2 La localisation

Les composantes de la matrice d'état correspondent aux valeurs des variables de l'écoulement (à savoir le champ de vitesse) en des points spécifiques du domaine physique, généralement le centre des éléments du maillage. Comme discuté dans la Sec. 3.3.2 et montré dans l'Eq. 3.57, le gain de Kalman établit une corrélation entre ces valeurs et les valeurs de la matrice d'état projetées dans l'espace d'observation, c'est-à-dire les capteurs auxquels des données haute-fidélité sont disponibles. Considérant que la corrélation physique décroît naturellement avec la distance dans les systèmes continus, les approximations utilisées pour déterminer le gain de Kalman dans le cadre de l'EnKF peuvent conduire à des corrélations non physiques sur la matrice d'état analysée, en particulier pour les grands domaines. Ces effets peuvent être responsables de problèmes critiques tels que des solutions non physiques, pouvant conduire à des divergences dans les calculs. Encore une fois, ces problèmes peuvent être réduits en augmentant le nombre de membres d'ensemble, ce qui ne constitue pas une solution viable pour les applications de mécanique des fluides numérique. Par conséquent, différentes stratégies

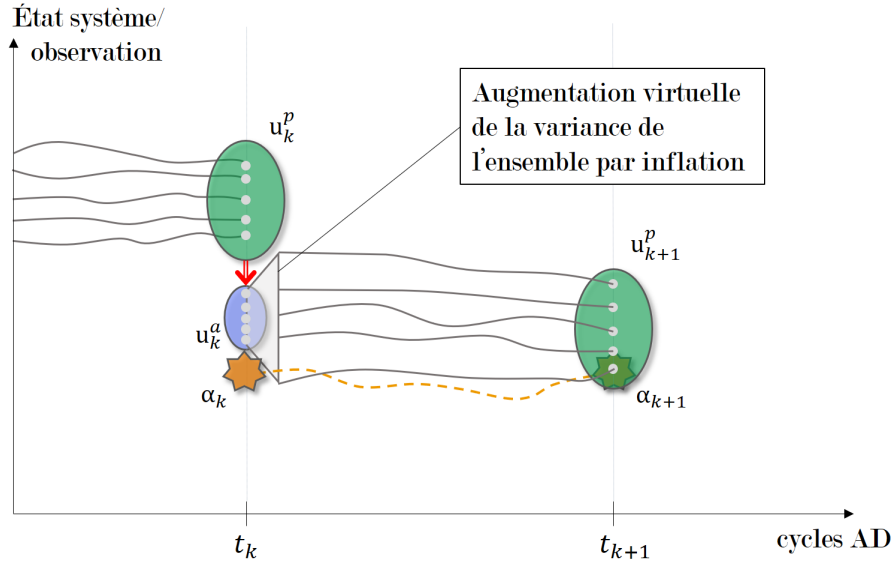


FIGURE 3.3 : Schéma représentant l'inflation utilisée pour le filtre de Kalman d'ensemble. Chaque ligne grise représente la trajectoire d'un membre de l'ensemble. L'augmentation de la variance de l'ensemble est représentée par le trapèze gris écartant les 5 membres de la moyenne.

doivent être employées pour atténuer les effets des corrélations parasites. La stratégie la plus utilisée pour les réduire est d'opérer sur les coefficients corrélant les variables de l'EnKF qui sont calculés en des points éloignés les uns des autres. Dans ce cas, on pourrait s'attendre à ce que les phénomènes physiques soient pour la plupart décorrélés. Deux stratégies possibles peuvent être adoptées pour obtenir ce résultat [4]. La *localisation de covariance* opère directement sur les coefficients de la matrice de covariance d'erreur \mathbf{P}_{k+1} , en les pré-multipliant par un terme qui tend vers zéro à mesure que la distance physique entre les capteurs d'observation et les éléments de l'état augmente. Ce processus est mathématiquement effectué par une multiplication entre la matrice de covariance et une matrice de correction appelée \mathbf{L} . Cette expression peut être directement ajoutée à l'algorithme sans aucune modification structurelle. Le gain de Kalman localisé devient :

$$[\mathbf{P}_{k+1}]_{m,n} [\mathbf{L}]_{m,n} \longrightarrow \mathbf{K}_{k+1}^{loc} = [\mathbf{L}]_{m,n} [\mathbf{K}_{k+1}]_{m,n} \quad (3.64)$$

La structure de la matrice \mathbf{L} doit être définie par l'utilisateur. Dans les systèmes fluides, et en particulier pour la turbulence, la corrélation diminue rapidement dans l'espace. Par conséquent, une structure généralement utilisée pour la matrice de localisation est une forme de décroissance exponentielle :

$$\mathbf{L}(m, n) = e^{-\Delta_{m,n}^2/l} \quad (3.65)$$

ou encore

$$\mathbf{L}(m, n) = e^{-\frac{1}{2}(\Delta_{m,n}/l)^2} \quad (3.66)$$

où $\Delta_{m,n}$ est la distance entre le capteur d'observation donné et le point d'évaluation du modèle. l est une échelle de longueur de corrélation qui peut être ajustée en fonction des caractéristiques locales du cas test.

Une autre façon de localiser le gain de Kalman est d'utiliser la *localisation physique*, on parle aussi d'*analyse locale* [37]. Le principe est assez simple. Au lieu d'exécuter l'EnKF sur l'ensemble du domaine physique, le calcul est effectué sur un domaine *réduit*. L'espace réduit contient l'ensemble des capteurs utilisés pour les mesures d'observation. Cette stratégie présente également l'avantage de réduire le nombre de degrés de liberté opérant dans la procédure d'assimilation, ce qui peut produire un gain significatif en termes de ressources de calcul nécessaires. Le *localisation de covariance* est couramment utilisée avec la *localisation physique* pour éviter les discontinuités de l'état physique mis à jour, en particulier à l'interface du domaine réduit. Cette opération évite une divergence potentielle des exécutions du modèle. Cette méthode est très efficace pour accélérer le calcul et améliorer simultanément la stabilité du calcul et la précision de la prédiction pour une taille d'ensemble réduite telle que celles actuellement utilisables pour les études basées sur la CFD [141].

Une autre manière de localiser le calcul est utilisée dans le *Local Ensemble Transform Kalman Filter* (LETKF) [53]. La méthode utilisée consiste à effectuer un *Ensemble Transform Kalman Filter* par cellule dans le domaine étudié. Cette variante consiste à calculer le gain de Kalman dans l'espace de l'ensemble. La mise à jour de l'état *a posteriori* est donc modifiée en conséquence pour correspondre à cette nouvelle formulation. Le nombre d'observations pour chaque ETKF est réduit, car seules les observations placées à une certaine distance de chaque cellule sont choisies. Cette méthode a pour avantage de réduire les coûts de calcul à cause de la diminution du nombre d'observations dans chaque analyse, cependant les cellules pour lesquels la procédure est réalisée ainsi que le nombre d'observations choisies peut impacter le résultat final et sont donc sujets à hyper-paramétrisation. Toutes les méthodes de localisation sont qualitativement illustrées sur la Fig. 3.4.

3.3.4 HLEnKF : *Hyper-Localized Ensemble Kalman Filter*

La décomposition de la matrice de gain de Kalman montre que la localisation de covariance, appliquée telle quelle au gain de Kalman, est biaisée. Par exemple, pour un système à deux membres, un degré de liberté et deux observations, le gain peut se décomposer de cette manière :

$$\mathbf{K} = (\mathbf{\Gamma}_{11}\mathbf{\Gamma}_{12}) \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{21} \\ \mathbf{S}_{12} & \mathbf{S}_{22} \end{pmatrix} \left[\begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{21} \\ \mathbf{S}_{12} & \mathbf{S}_{22} \end{pmatrix} + \begin{pmatrix} \mathbf{R}_{11} & 0 \\ 0 & \mathbf{R}_{22} \end{pmatrix} \right]^{-1} \quad (3.67)$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{\Gamma}_{11}\mathbf{S}_{11} + \mathbf{\Gamma}_{12}\mathbf{S}_{12} & \mathbf{\Gamma}_{11}\mathbf{S}_{21} + \mathbf{\Gamma}_{12}\mathbf{S}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{11}^2 + \mathbf{S}_{12}^2 + \mathbf{R}_{11} & \mathbf{S}_{11}\mathbf{S}_{21} + \mathbf{S}_{12}\mathbf{S}_{22} \\ \mathbf{S}_{11}\mathbf{S}_{21} + \mathbf{S}_{12}\mathbf{S}_{22} & \mathbf{S}_{21}^2 + \mathbf{S}_{22}^2 + \mathbf{R}_{22} \end{pmatrix}^{-1} \quad (3.68)$$

Les valeurs du gain de Kalman correspondantes à chaque observation contiennent des corrélations entre toutes les observations présentes dans l'estimation des covariances d'erreur sur la mesure. La pondération appliquée par la localisation affecte donc des valeurs du gain de Kalman contenant des corrélations entre les différents capteurs d'observation sans effacer ces corrélations, quelle que soit la distance physique.

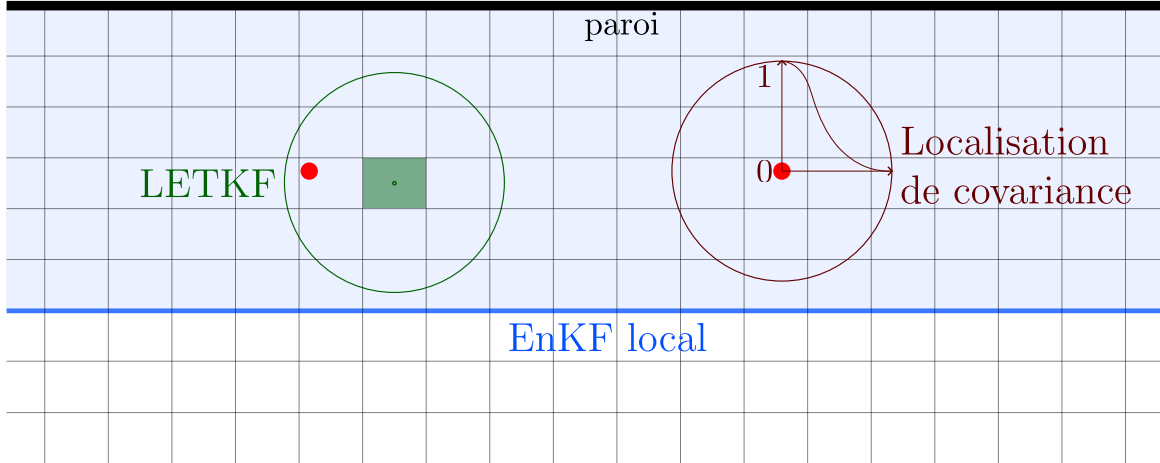


FIGURE 3.4 : Schéma représentant les différentes méthodes de localisation sur une grille cartésienne. Les points rouges représentent les observations. En bleu : l'EnKF est exécuté sur un domaine réduit. En rouge foncé : la localisation de covariance pondère la correction apportée par l'EnKF dans une zone donnée. En vert : Exemple du comportement du LETKF pour une des mailles du domaine.

La méthode de localisation utilisée dans le LETKF (Sec. 3.3.3.2), dont le fonctionnement est illustré sur la Fig. 3.4, exécute l'algorithme d'assimilation de manière séparée pour chaque cellule. Avec une gestion adéquate des observations considérées, elle permet donc également de réduire les corrélations entre les observations lorsqu'elles ne sont pas nécessaires. Partant de ce principe, une nouvelle méthode (à notre connaissance) a été développée en prenant le point de vue de l'observation pour le EnKF classique. L'*hyper-localisation* consiste à exécuter un filtre de Kalman d'ensemble *par capteur d'observations* dans le but de supprimer les corrélations d'observations lorsqu'elles ne sont pas nécessaires. En reprenant l'exemple :

$$\mathbf{K}_1 = (\mathbf{\Gamma}_{11}\mathbf{S}_{11} + \mathbf{\Gamma}_{12}\mathbf{S}_{12}) (\mathbf{S}_{11}^2 + \mathbf{S}_{12}^2 + \mathbf{R}_{11})^{-1} \quad (3.69)$$

$$\mathbf{K}_2 = (\mathbf{\Gamma}_{11}\mathbf{S}_{21} + \mathbf{\Gamma}_{12}\mathbf{S}_{22}) (\mathbf{S}_{21}^2 + \mathbf{S}_{22}^2 + \mathbf{R}_{22})^{-1} \quad (3.70)$$

Cette méthode permet d'appliquer une localisation sur un gain considérant *uniquement* les observations situées à une distance physique de chaque cellule dans le calcul de chaque gain de Kalman. Elle est particulièrement efficace lorsque ces corrélations ne correspondent à aucune réalité physique. Dans ce cas, la précision de la prédiction est améliorée. L'amélioration des résultats est cependant fortement dépendante de la pertinence physique des observations pour une cellule donnée, et donc du nombre d'observations considérées pour l'inférence de cette dernière. L'algorithme du LETKF est dans ce cas pertinent pour sa capacité à considérer différentes observations d'une cellule à l'autre, mais doit en contrepartie satisfaire une gestion complexe des discontinuités des corrections de la matrice d'état. Un des autres avantages du HLEnKF est la forte diminution du coût de calcul, qui dépend fortement du nombre d'observations utilisées dans le calcul du gain, une comparaison de coûts en fonction de plusieurs hyper-paramètres est donnée dans la Sec. 4.5.

Le choix des capteurs pour la mesure des observations n'est généralement pas sujet à modification. L'hyper-localisation est pertinente lorsque ces capteurs ne sont pas regroupés. C'est souvent le cas pour les applications de mécanique des fluides. La taille de la sous-région d'optimisation associée à chaque capteur est choisie en fonction de la fonction de localisation de covariance Eq. 3.65 de sorte que la correction apportée par le EnKF soit nulle à la frontière de chaque sous-région. La correction est donc maximale aux coordonnées du capteur d'observation et décroît de manière exponentielle jusqu'à une valeur proche de zéro à la frontière de la sous-région. La gestion des sous-régions pour l'hyper-localisation est qualitativement résumée sur la figure 3.5 montrant la différence entre une localisation physique classique par rapport à l'hyper-localisation. Toutes les cellules de chaque sous région ne sont donc affectées que par le capteur d'observation présent dans leur région d'attribution, d'où le terme d'hyper-localisation.

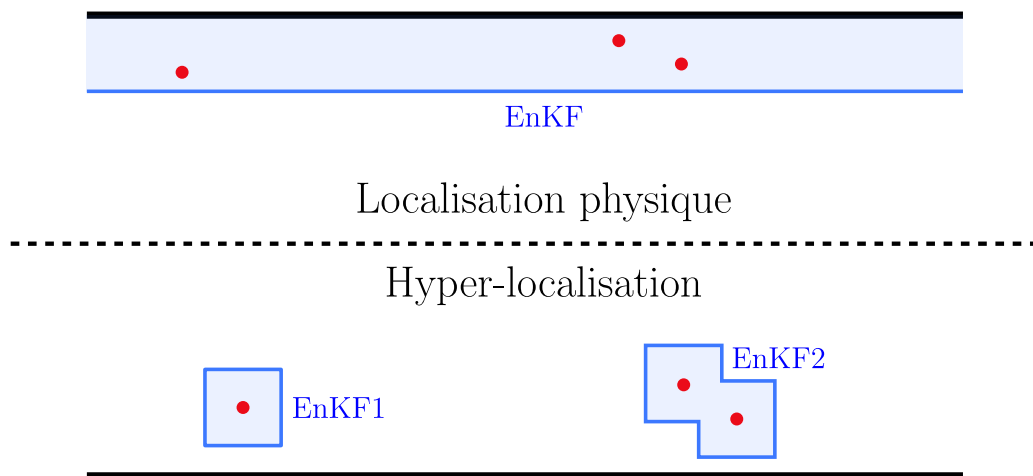


FIGURE 3.5 : Schéma représentant le principe de l'hyper-localisation. Les points rouges représentent les observations. L'EnKF local classique s'exécute sur un domaine réduit contenant les observations. Pour l'hyper-localisation, un EnKF est exécuté autour de chaque observation. Les régions sont fusionnées si elles s'intersectent.

L'inconvénient principal de cette méthode est la gestion des sous-régions coïncidentes lorsque deux ou plusieurs capteurs sont trop rapprochés. On préférera alors fusionner les sous-régions, créant ainsi une forme de sous-région arbitraire, de manière à créer une analyse locale compatible toutes les observations à région coïncidentes dans le voisinage (voir "EnKF2" sur la Fig. 3.5). Cette complexité dans la gestion algorithmique est cependant transparente une fois la solution proprement programmée. Ce dernier intègre une boucle *for* supplémentaire permettant de gérer l'exécution successive des EnKF hyper-localisés. L'algorithme complet du HLEnKF est donné par l'Alg. 3.

La gestion de la mise à jour de la description paramétrique est également différente. Cette dernière n'étant pas affectée par une quelconque distance physique, chaque EnKF_i apporte une correction différente des paramètres optimisés lors du calcul. Pour chaque membre, la valeur renvoyée vers le modèle est la moyenne calculée à partir des valeurs corrigées par chaque EnKF_i . Ceci a pour conséquence de diminuer la rapidité de la réponse du système dans la convergence de la description paramétrique mais d'augmenter à la fois la précision et la stabilité de la mise

à jour ainsi que d'atténuer la diminution de variance (voir la Sec. 4.4.3).

La précision de la mise à jour de la matrice d'état est accrue dans les régions hyper-localisées. Cette méthode se révèle particulièrement efficace dans le cas d'écoulements pilotés par la turbulence tel qu'exposé par le cas du chapitre 6 où le niveau de synchronisation du champ de vitesse avec l'observation est supérieur grâce à l'hyper-localisation. Elle est aussi particulièrement pertinente pour la réduction de coûts de calculs dans des cas plus exigeants comme pour le chapitre 7.

3.3.5 Application des méthodes séquentielles en mécanique des fluides

Les applications et les méthodes utilisées dans cette section sont plus variées que pour l'équivalent variationnel. Quelques applications se concentrent sur l'utilisation du filtre de Kalman ou du filtre de Kalman étendu, en linéarisant le système d'équations. Cependant, la plupart ont recours au filtre de Kalman d'ensemble, peu l'utilisent sans compléments comme la localisation ou l'inflation. Un certain nombre de ces articles utilisent également l'EnKF couplé à différentes méthodes telles que des décompositions de Karhunen-Loeve ou POD (*proper orthogonal decomposition*). Les premiers travaux montrant l'utilisation du filtre de Kalman pour la mécanique des fluides numérique semblent apparaître au début des années 2000. Certains de ces travaux, notamment Hoepffner, Chevalier et al. [51, 18] seront abordés dans le premier paragraphe aux côtés d'autres travaux utilisant le filtre de Kalman. Comme pour les méthodes variationnelles, les autres paragraphes se concentreront principalement sur les travaux effectués durant la dernière décennie et utilisant le filtre de Kalman d'ensemble.

- **Les quelques applications du filtre de Kalman classique**

Les applications utilisant le filtre de Kalman sont assez rares. Comme cités plus haut, Hoepffner et al. [51] puis Chevalier et al. [18] utilisent le filtre de Kalman pour le contrôle d'écoulements dans une série d'articles. Le problème de contrôle implique deux gros défis : le contrôle effectif de l'écoulement à partir d'actuateurs basé sur l'état du système et l'estimation précise de cet état. Les deux travaux se concentrent sur l'estimation précise de l'état du système d'un écoulement de paroi à partir de données de frottement bruitées et de mesures de pression. La solution est d'abord mise en œuvre sur un problème laminaire à faible nombre de Reynolds puis étendu à un système faiblement turbulent dans un canal plan. L'estimation de l'état est effectuée grâce au filtre de Kalman à partir des équations de Navier–Stokes linéarisées puis à partir du filtre de Kalman étendu. La partie finale de cette série d'articles par Colburn et al. [19] implique cette fois l'utilisation du filtre de Kalman d'ensemble, à partir duquel les résultats sont améliorés d'un ordre de grandeur. Suzuki et al. [125] ont également participé à approfondir cette étude avec l'utilisation du EnKF sur une DNS.

Avant cela, Suzuki et al. [124] ont utilisé un filtre de Kalman étendu pour la combinaison de données PTV (Particle Tracking Velocimetry) et d'une simulation d'écoulement de jet 2D. Inspirés par ce travail, le filtre de Kalman est utilisé par Meldi et al. [75] pour l'estimation augmentée d'écoulements incompressible 3D turbulents. Le filtre de Kalman est intégré à la boucle PISO. L'équation de prédiction de quantité de mouvement, naturellement linéarisée dans le PISO, est considérée comme le modèle utilisé par l'étape de *prévision* du

Algorithme 3 : Algorithme du Filtre de Kalman d'Ensemble hyper-localisé (HLEnKF)

Input : \mathcal{M} , \mathcal{H} , \mathbf{R}_{k+1} , et distributions initiales des vecteurs d'états $\mathbf{u}_{i,0}^a$, où généralement $\mathbf{u}_{i,0}^a \sim \mathcal{N}(\mu_N, \sigma_N^2)$, k_f le temps final, N_e le nombre de membres et N_r le nombre de régions hyper-localisées

for $k = 0$ to $k_f - 1$ **do**

for $j = 0$ to N_r **do**

for $i = 1$ to N_e **do**

1 Avancement temporel des vecteurs d'état :

$$\mathbf{u}_{i,j,k+1}^p = \mathcal{M}\mathbf{u}_{i,j,k}^a$$

2 Création de la matrice d'observation à partir des données d'observations en introduisant des erreurs :

$$\boldsymbol{\alpha}_{i,j,k+1} = \boldsymbol{\alpha}_{k+1} + \epsilon_{i,j,k+1}, \text{ avec } \epsilon_{i,j,k+1} \sim \mathcal{N}(0, \mathbf{R}_{k+1})$$

3 Projection de la solution de modèle sur l'espace des observations :

$$\mathbf{s}_{i,j,k+1} = \mathcal{H}\mathbf{u}_{i,j,k+1}^p$$

4 Calcul des moyennes d'ensemble :

$$\langle \mathbf{u} \rangle_{k+1}^p = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{u}_{i,j,k+1}^p, \quad \langle \mathbf{s} \rangle_{k+1} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{s}_{i,j,k+1},$$

for $i = 1$ to N_e **do**

5 Calcul des matrices d'anomalies :

$$[\boldsymbol{\Gamma}_{k+1}]_{i,j} = \frac{\mathbf{u}_{i,j,k+1}^p - \langle \mathbf{u} \rangle_{k+1}^p}{\sqrt{N_e - 1}}, \quad [\mathbf{S}_{k+1}]_{i,j} = \frac{\mathbf{s}_{i,j,k+1} - \langle \mathbf{s} \rangle_{k+1}}{\sqrt{N_e - 1}},$$

6 Calcul du gain de Kalman :

$$\mathbf{K}_{j,k+1} = \boldsymbol{\Gamma}_{j,k+1}(\mathbf{S}_{j,k+1})^T \left[\mathbf{S}_{j,k+1}(\mathbf{S}_{j,k+1})^T + \mathbf{R}_{j,k+1} \right]^{-1}$$

7 Localisation du gain de Kalman :

$$\mathbf{K}_{j,k+1}^{loc} = [\mathbf{L}_j]_{m,n} [\mathbf{K}_{j,k+1}]_{m,n}$$

for $i = 1$ to N_e **do**

8 Mise à jour de la matrice d'état :

$$\mathbf{u}_{i,j,k+1}^a = \mathbf{u}_{i,j,k+1}^p + \mathbf{K}_{j,k+1}^{loc}(\boldsymbol{\alpha}_{i,j,k+1} - \mathbf{s}_{i,j,k+1})$$

9 Inflation de la matrice d'état :

$$\mathbf{u}_{i,j,k+1}^a = (1 + \lambda_i)\mathbf{u}_{i,j,k+1}^a$$

for $i = 1$ to N_e **do**

10 Calcul de la moyenne des paramètres de chaque région :

$$\bar{\theta}_i = \frac{1}{N_r} \sum_{j=1}^{N_r} \theta_{i,j}$$

filtre de Kalman. La méthode du filtre de Kalman d'ordre réduit est appliquée au problème de couche de mélange, exhibant des instabilités de Kelvin-Helmoltz classiquement rencontrées dans les problèmes industriels. Les résultats montrent une synchronisation partielle du modèle LES avec les observations DNS fournies. M. Meldi a ensuite exploré la résolution temporelle des phases d'assimilation nécessaire pour atteindre la convergence optimale du processus d'assimilation [73]. Ce résultat, valable pour la plupart des algorithmes d'assimilation, montre qu'un temps de fenêtre d'observation inférieur au temps caractéristique de la physique de l'écoulement est nécessaire pour une synchronisation optimale des champs.

- ***Décompositions et modèles d'ordre réduit***

La reconstruction des champs physiques est une des applications phares du filtre de Kalman d'ensemble. Cependant, les applications de ce dernier semblent encore se concentrer majoritairement sur des systèmes stationnaires, malgré l'aspect séquentiel du EnKF. Certains travaux mettent en place différentes méthodes de décomposition pour l'inférence des données. Par exemple, Xiao et al. [150] infèrent le champ des tensions de Reynolds via une décomposition de Karhunen-Loeve à partir d'un nombre limité de données DNS pour une simulation RANS 2D, montrant une amélioration des résultats comparé à la référence. Cette décomposition est également utilisée par Zhang et al [153] pour la quantification d'incertitudes sur les tensions de Reynolds dans une simulation RANS 2D. Plusieurs variantes du EnKF sont comparées tel que le EnKF-MDA ou le EnRML, montrant les résultats les plus encourageants. Da Silva et al. [23] utilisent une décomposition de type POD pour la prise en compte de l'erreur de discrétisation due à l'utilisation de maillages grossiers. Malgré l'utilisation d'un faible nombre de mesures de pressions, la réduction d'erreur est de 80% pour une simulation de profil d'aile 2D. Très récemment, Wang et al. [146] étudient spécifiquement l'utilisation de modèles d'ordres réduits pour le modèle dynamique et pour l'opérateur d'observations. Le nombre de modes POD optimal est évalué via l'estimation de l'erreur entre deux estimations successives. La méthode du ROM-EnKF ainsi développée est cependant testée sur un modèle 1D de propagation d'ondes de tsunami. Les auteurs soulignent notamment les difficultés pouvant apparaître dans l'utilisation de modèles d'ordres supérieurs et non-linéaires.

- ***Reconstruction de champs physiques et variantes du EnKF***

Comme vu précédemment, l'utilisation de modèles d'ordre réduit pour le EnKF est encore très rare et les applications concernées sont relativement peu coûteuses, tel que des cas RANS 2D ou encore 1D. D'autres études se concentrent sur la reconstruction de champs physiques à partir du EnKF classique, souvent en ayant recours aux méthodes de localisation et d'inflation. Par exemple, Labahn et al. [62] utilisent des données TPIV (*tomographic particle image velocimetry*) et autres mesures d'espèce chimiques (OH-PLIF) pour l'inférence d'un jet de flamme dans une simulation LES. Ils montrent une amélioration de la prédiction des dynamiques d'extinction et de rallumage observées dans la flamme à partir de seulement 12 membres d'ensemble. Dans une deuxième partie [63], différents hyperparamètres du EnKF sont testés tels que la distance de localisation, pour laquelle une valeur ni trop faible ni trop élevée semble optimale, la fréquence d'assimilation, le nombre

de membres d'ensembles et le nombre d'observations disponibles pour lesquels des valeurs plus élevées donnent de meilleurs résultats.

Moldovan et al. [82] ont développé une méthode multi-grille pour l'inférence de champs CFD via l'utilisation du filtre de Kalman d'ensemble. Le MGenKF (*Multigrid Ensemble Kalman Filter*) vise à réduire le coût d'un EnKF classique en considérant différents niveaux de résolutions pour un même modèle. Les solutions basses résolutions sont alors considérées comme les membres de l'ensemble permettant de fournir une correction sur une simulation principale haute résolution. Les contributions des grandes échelles de la turbulence, associées aux caractéristiques uniques de l'écoulement, sont conservées dans le processus. L'algorithme, testé sur une couche de mélange 2D, semble convenir à l'étude de phénomènes instationnaires. Ce dernier est amélioré dans une seconde publication [84] puis mis à contribution pour la simulation LES d'un cas canonique de cylindre rectangulaire (BARC) [83]. La prédiction augmentée montre de bien meilleurs résultats pour les quantités physiques étudiées, malgré une asymétrie des observations utilisées.

Une autre variante du EnKF visant à réduire le coût de l'algorithme a récemment été développée par Zhang et al. [152]. Le PEnKI (*parallel ensemble Kalman inversion*) se concentre sur la parallélisation du EnKF via la décomposition de la simulation en sous-domaine et l'exécution de EnKFs séparément dans chacun des sous-domaines. Des méthodes de régularisation sont utilisées aux frontières des domaines décomposées pour compenser l'absence d'information sur les covariances entre sous-domaines. Hormis son exécution parallèle, la méthode semble similaire à la méthode de localisation développée par Hunt et al. (LETKF) [53]. La méthode est testée sur une simulation 2D où le coût du calcul est effectivement réduit.

- **Estimation paramétrique**

L'estimation de paramètres régissant la prédiction du modèle est un autre des aspects importants du EnKF. Dans la plupart des études, ces paramètres correspondent aux coefficients globaux pour la fermeture des modèles de turbulence ou encore aux paramètres de conditions aux limites, ou géométriques de la simulation. Kato et al. [58] calibrent par exemple les coefficients de fermeture du modèle Spalart-Allmaras sur une plaque plane 2D. Une seconde étude [57] se consacre à la calibration de l'angle d'attaque, du nombre de Mach et de la viscosité turbulente pour un profil d'aile 3D. Les paramètres corrigés compensent la prédiction erronée des modèles de turbulence RANS et permettent une prédiction améliorée vis-à-vis des données de référence expérimentales. Zhao et al. [154] infère également les coefficients de fermeture de modèles RANS pour un écoulement 3D RANS autour d'un bâtiment. Son cas est plus largement abordé dans le chapitre 5.

Quelques études se consacrent à l'inférence de conditions d'entrée. Par exemple, Sousa et al. [122] calibrent la distribution de vitesse d'entrée pour une simulation 3D-RANS de canopée urbaine à partir de mesures par anémomètres dans le milieu urbain. Par ailleurs, la solution inférée performe mieux que la solution basée sur les mesures de stations météo. Ströfer et al. [123] se consacrent plus particulièrement à la correction des matrices de covariances lors de la modification de champs latents tels que le champ de viscosité turbulente. L'objectif est notamment de s'assurer que les conditions aux limites sont respectées

lors de la modification de ce champ. Un test sur une simulation 2D montre que la qualité des résultats et la continuité des conditions aux limites sont assurées. Plus récemment, Moussie et al. [92] ont inféré la condition d'entrée d'une simulation 3D DES sur le cas du cylindre rectangulaire (BARC). La condition de turbulence synthétique décrite par 8 paramètres est optimisée via l'utilisation de 451 mesures de vitesses expérimentales par PIV. L'optimisation de l'entrée entraîne une amélioration de la prédiction de l'écoulement par le modèle et des quantités physiques tel que le champ de vitesse, de pression ou encore les contraintes de cisaillement des tensions de Reynolds. Enfin, Valero et Meldi [137] ont utilisé le filtre de Kalman d'ensemble pour inférer la méthode des frontières immergées intégrée aux équations dynamiques via la calibration de la description paramétrique du modèle. L'étude se concentre sur l'étude d'une telle configuration sur le cas canonique du canal turbulent à partir de capteurs de vitesses à la paroi. La méthode développée se montre particulièrement avantageuse en termes de coût par rapport à la précision de la prédiction, en particulier comparé aux données DNS de référence.

Comme vu précédemment, le filtre de Kalman n'est pas aussi répandu que son cousin variationnel. Il n'est pas présent sous une forme unique comme c'est le cas pour le 4DVar (bien que certains utilisent la version ensembliste du 4DVar). Cette brève revue de littérature montre que chaque étude fait état d'une version plus ou moins différente du filtre de Kalman d'ensemble. A fortiori, le filtre de Kalman dans sa version classique ou étendue est délaissé au profit de son cadet ensembliste, à cause du coût de calcul prohibitif de la matrice de covariances d'erreurs et de la nécessité de linéariser le modèle utilisé. Un nombre réduit d'études montrent cependant que l'utilisation de données expérimentales, souvent éparses et présentant des dynamiques invisibles pour un modèle d'ordre réduit, permet l'amélioration de ces mêmes modèles de manière locale (correction d'un champ) ou globale (correction d'une description paramétrique). En particulier, l'aspect séquentiel du filtre de Kalman d'ensemble est un avantage certain pour le suivi en temps réel de systèmes dynamiques tel qu'utilisé pour les problèmes de contrôle. Pour toutes ces raisons, le filtre de Kalman d'ensemble semble être un bon choix pour la reconstruction de champs physiques, la calibration de paramètres globaux ou encore la représentation d'évènements extrêmes aux dynamiques rapides à partir de données haute-fidélité. Il est utilisé pour les trois cas d'étude présentés dans ce manuscrit [chapitre 5](#), [chapitre 6](#) et [chapitre 7](#).

C.O.N.E.S : Coupling OpenFOAM with Numerical EnvironmentS

Sommaire

4.1	Besoins et objectifs	71
4.2	Fonctionnement du couplage	72
4.3	Formats et possibilités de calcul	75
4.4	Exemple : Cavity test case	77
4.4.1	Optimisation avec un EnKF classique	79
4.4.2	Effets des hyper-paramètres	80
4.4.3	(Hyper)-localisation	82
4.5	Consommation et temps d'exécutions	84

4.1 Besoins et objectifs

L'outil CONES pour "*Coupling OpenFOAM with Numerical EnvironmentS*" a été créé dans ce travail de thèse pour répondre au besoin de plate-forme numérique des activités de recherches, i.e. coupler la bibliothèque C++ OpenFOAM avec toutes sortes de codes maison, notamment pour l'assimilation de données. Pour cette raison, CONES permet actuellement le couplage de plusieurs solveurs OpenFOAM dédiés au calcul d'écoulements laminaires, turbulents, stationnaires ou instationnaires ainsi qu'incompressibles ou compressibles avec un code C++ maison pour l'assimilation de données avec un filtre de Kalman d'ensemble. L'outil utilise un coupleur existant, appelé CWIPI, développé par le CERFACS et l'ONERA [106]. Celui-ci sera détaillé dans la section suivante. L'intérêt d'une telle plate-forme est d'abord de répondre aux besoins

présents de l'équipe. Assimiler des données à l'aide d'un algorithme ensembliste et qui plus est, séquentiel, nécessite le calcul en temps réel de multiples simulations ainsi que l'assemblage et le traitement de ces données dans le processus du EnKF. Un tel fonctionnement pourrait être programmé directement dans la bibliothèque OpenFOAM. En dehors des considérations techniques avantageuses de CWIPI, son utilisation est intéressante pour plusieurs raisons :

- La possibilité de changer d'environnement numérique facilement sans affecter le couplage.
- Le transfert de champs physiques complets, de manière quasi instantanée entre deux maillages coïncidents ou non.
- L'interpolation automatique des champs physiques transférés dans le cas du couplage entre deux maillages non coïncidents, notamment utilisé pour les algorithmes de type *Multi-grid Ensemble Kalman Filter* (MGENKF) [82].

Ce dernier point illustre un des enjeux cruciaux dans le développement de CONES i.e. sa capacité à pouvoir accueillir dans le futur de nouveaux environnements numériques facilement et sans refonte en profondeur de l'outil. Bien que ce dernier soit calibré pour l'utilisation d'algorithmes ensemblistes, son utilisation en série et/ou statique est tout à fait possible. L'équipe de recherche a pour objectif d'améliorer l'outil en intégrant de nouvelles méthodes, de nouveaux algorithmes d'assimilation ou encore en améliorant l'expérience utilisateur pour faciliter la prise en main auprès de personnes étrangères aux algorithmes d'assimilation de données. Le dépôt public est accessible via le lien : <https://github.com/MiguelMValero/CONES.git>.

4.2 Fonctionnement du couplage

Le fonctionnement complet de CONES est abordé dans cette section. Comme énoncé dans la section précédente, l'outil s'appuie sur le coupleur CWIPI pour effectuer les échanges entre le solveur de mécanique des fluides et le code d'assimilations de données. CWIPI pour "*Coupling With Interpolation Parallel Interface*" est un coupleur développé pour répondre aux besoins grandissants de couplage de codes multi-physiques, notamment fluide/ thermique / structure, etc. conséquence de l'augmentation des moyens de calculs. Il est calibré pour l'échange de données hautement parallélisées pour des maillages de grandes tailles. CWIPI s'appuie sur l'environnement MPI pour "*Message Passing Interface*" pour effectuer ses échanges. Cette norme, conçue en 1993, s'est très largement répandue au sein des architectures de calcul et est devenue un standard pour le calcul en parallèle. CWIPI pourra donc être compilé avec une bibliothèque MPI telle que OpenMPI ou encore IntelMPI. La bibliothèque CONES ne se compose que d'une poignée de fichiers : la bibliothèque *Pstream* d'OpenFOAM modifié pour accueillir CWIPI, le solveur dédié au cas d'étude également modifié et enfin une petite bibliothèque de fichiers pour l'interface de CONES accueillant l'environnement numérique utilisé (ici l'EnKF).

La figure 4.1 résume le fonctionnement de CONES et notamment son couplage via CWIPI. L'environnement MPI doit dans un premier temps être initialisé avant l'environnement CWIPI. Vient ensuite la déclaration du couplage entre les deux programmes. Le prototype de la fonction de couplage de CWIPI est le suivant :

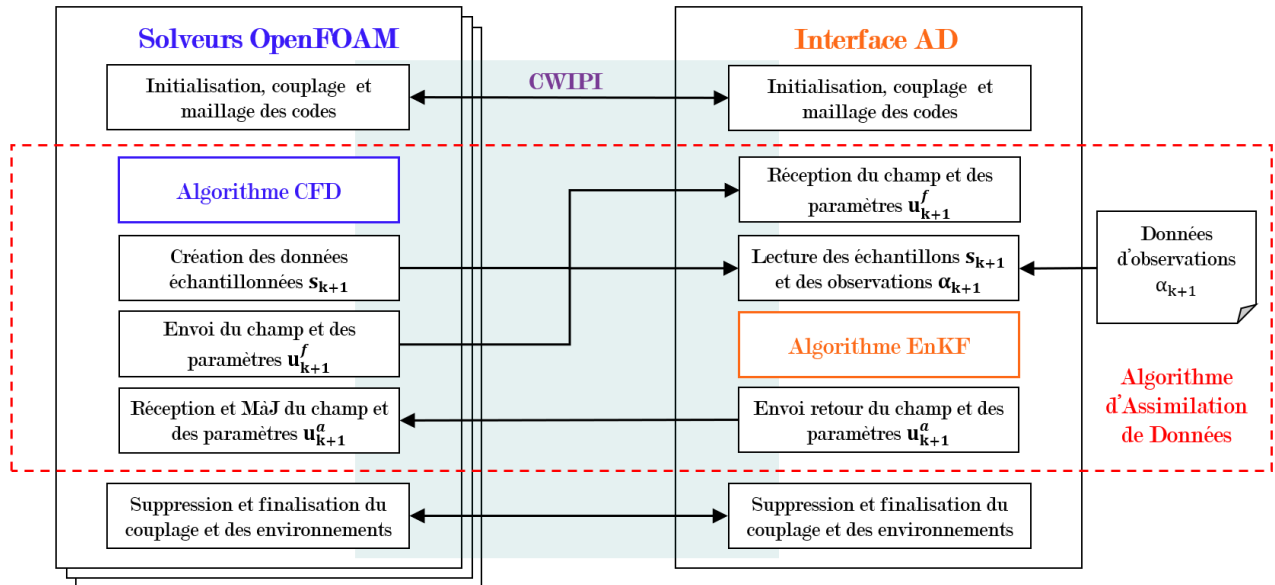


FIGURE 4.1 : Schéma de fonctionnement de CONES

`void cwipi_create_coupling`(nom-couplage, type-couplage, application-couplée, dimensions, tolérance géométrique, type-maillage, type-solveur ..) ;

L'utilisation de CONES avec l'algorithme du EnKF nécessite le couplage de N_e membres d'ensembles, c'est-à-dire solveurs OpenFOAM, avec une seule application gérant l'EnKF. Chaque membre d'ensemble possède donc un `cwipi_create_coupling`, alors que cette fonction est intégrée dans une boucle parcourant les N_e membres du côté du EnKF. Pour rappel, le coupleur réalise un échange de données entre deux maillages, coïncidents ou non. Dans le cas du EnKF, les maillages sont parfaitement coïncidents dans la mesure où tous les membres d'ensembles sont identiques. La déclaration de ces maillages est indépendante de la déclaration d'OpenFOAM et doit être effectuée indépendamment pour chaque application. Toutes les informations nécessaires sont récupérées via les fonctions d'OpenFOAM. Les deux fonctions de CWIPI permettant la déclaration des maillages sont les suivantes :

- `void cwipi_define_mesh(...)` ; permet la déclaration pour des maillages fait d'hexaèdres (ou polyèdres au nombre de faces inférieur). Convient parfaitement pour les maillages structurés (notamment réalisés avec l'outil *BlockMesh* d'OpenFOAM). L'utilisation de cette fonction est obligatoire.
- `void cwipi_add_polyhedra(...)` ; permet l'ajout de polyèdres d'ordre supérieur dans le cas de maillages non structurés (notamment réalisés avec l'outil *SnappyHexMesh* d'OpenFOAM).

L'utilisation de ces deux fonctions réunies permet la déclaration de maillages quelconques dans CONES. Elle est automatique et ne nécessite aucun *input* de la part de l'utilisateur. Chaque maillage est ensuite localisé par CWIPI grâce à la fonction `cwipi_locate` qui permet d'établir la correspondance des deux maillages au sein d'un même couplage. Le temps d'exécution de cette fonction dépend directement de la tolérance géométrique renseignée lors de la

création du couplage. Son coût sera détaillé dans la section 4.5. L'échange de données est réalisé de manière asynchrone dans CONES. Certaines fonctions, tel que `cwipi_locate` constituent un point d'arrêt et donc de synchronisation pour les membres d'ensembles et le code du EnKF. Les échanges principaux sont réalisés via les fonctions `cwipi_irecv` et `cwipi_issend` qui couplées aux fonctions `cwipi_wait_issend` et `cwipi_wait_irecv` constituent également des points de synchronisations.

La figure 4.1 reprend les étapes de l'algorithme du filtre de Kalman d'ensemble exposé dans la section 3.3.2 au sein de CONES :

- La première étape est l'étape de *prévision* effectuée par le solveur de mécanique des fluides utilisé.
- Vient ensuite l'étape d'interpolation des données des N_e membres aux coordonnées des capteurs utilisés pour les observations, notées s_{k+1} . Cette étape est responsable de la création d'un fichier par membre, ces derniers contenant les informations interpolées. Les fonctions d'interpolations utilisées sont les fonctions issues de la bibliothèque d'OpenFOAM (notamment `interpolationCellPoint : interpolate(...)`).
- Les N_e champs de vitesses sont ensuite envoyés au code du EnKF par les membres d'ensemble. Ces derniers sont assemblés pour former une seule grande matrice appelée matrice d'état, laquelle sera plus ou moins réduite selon la méthode de localisation du EnKF choisie.
- L'envoi de paramètres, formant la matrice d'état étendue, est assuré par les primitives `MPI_SEND` et `MPI_RECV`. Il existe plusieurs fonctions dans CONES traitant l'échange de paramètres différents (constantes de modèle, conditions aux limites...)
- Les fichiers contenant les interpolations ainsi que le fichier contenant les observations sont ensuite lus par l'outil. Le fichier contenant les observations peut contenir un seul vecteur dans le cas où le calcul est stationnaire, ou autant de vecteurs que disponibles dans le cas où le calcul est instationnaire.
- Toutes ces informations permettent l'exécution du filtre de Kalman d'ensemble (dont le coût par phase d'analyse est détaillé en section 4.5) et donc la mise à jour de la matrice d'état étendue.
- De la même manière que lors de l'envoi des données, les informations contenues dans la matrice d'état étendues sont renvoyées vers les membres d'ensemble pour reprendre les étapes de *prévision* ou pour mettre fin au calcul.

Lors de la finalisation du calcul, les couplages doivent être terminés et les environnements CWIPI puis MPI doivent être finalisés sans quoi le calcul se terminerait avec une erreur et la création de *core dumps* volumineux (sauvegarde des informations contenues dans la mémoire).

4.3 Formats et possibilités de calcul

CONES offre actuellement plusieurs possibilités pour l'assimilation de données via l'EnKF. Le choix des méthodes employées se fait grâce au fichier de configuration appelé *conesDict* dont le fonctionnement est calqué sur les fichiers de configuration d'OpenFOAM. Ce fichier permet de configurer les hyper-paramètres de l'outil, i.e. les paramètres utilisés pour configurer l'algorithme du EnKF. Le bruit ajouté aux observations, le nombre de membres dans l'ensemble, le pourcentage d'inflation ou encore l'échelle de corrélation de la localisation de covariance sont des hyper-paramètres. Ils ont un impact sur la précision des résultats mais ne font pas partie du système physique. L'ensemble des variables contenues dans le fichier *conesDict* et nécessaires au bon fonctionnement de l'algorithme est résumé ci-après :

- **paramEstSwitch** : Booléen pour l'activation de la mise à jour des paramètres lors des phases d'*analyse*
- **stateEstSwitch** : Booléen pour l'activation de la mise à jour de l'état lors des phases d'*analyse*
- **observationWindow** : Nombre de *prévisions* (boucles CFD) entre chaque phase d'*analyse* du EnKF
- **ensemble** : Nombre de membres dans l'ensemble
- **numberParameters** : Nombre de paramètres à optimiser dans l'EnKF
- Sous dictionnaire dédié aux variables liées à l'observation :
 - **obsCoordFile** : Nom du fichier .txt contenant les coordonnées des capteurs d'observations (par défaut *obs_coordinates.txt*)
 - **obsDataFile** : Nom du fichier .txt contenant les données des observations (par défaut *obs_field.txt*)
 - **numberObsProbesVelocity, numberObsProbesPressure** : Nombre de capteurs de vitesse et de pression dans l'observation
 - **obsVelocityComponents** : Booléens pour l'activation de l'assimilation des composantes u_x , u_y et u_z de vitesse
 - **velocityObsNoise, pressureObsNoise** : Correspond à l'écart type de la distribution utilisée dans la matrice \mathbf{R} du EnKF, appliqué aux observations de vitesse ou de pression
 - **obsTimeDependency** : Booléen contrôlant le type d'observation (false = stationnaire, true = instationnaire)
 - **obsTimeStep** : Pas de temps de l'observation dans le cas instationnaire
 - **obsStartTime** : Temps initial de l'observation dans le cas instationnaire
 - **obsType** : Commutateur pour le choix de la nature des observations (U = vitesses, p = pressions, Up = vitesses et pressions)

- `obsNoiseType` : Contrôle le type de bruit donné aux observations avec la matrice \mathbf{R} (absVal = valeurs absolues, relVal = pourcentage)
- Sous dictionnaire dédié aux variables pour l’inflation :
 - `stateInflation`, `parametersInflation` : Pourcentage d’inflation sur la matrice d’état et/ou sur les paramètres
 - `inflationType` : Type d’inflation (stochastique ou déterministe)
- Sous dictionnaire dédié aux variables pour la localisation :
 - `clippingSwitch` : Booléen pour l’activation de la localisation physique avec le fichier `clippingCells.txt`
 - `clippingFile` : Nom du fichier contenant les IDs de cellules contenues dans les clippings si `clippingSwitch` est vrai (par défaut `clippingCells.txt`)
 - `localisationSwitch` : Booléen pour l’activation de la localisation de covariance
 - `correlationScale` : Échelles de corrélation l_x , l_y et l_z utilisées pour la localisation de covariance (à indiquer dans un vecteur)
- `numberOutputs` : Nombre de fichiers de sortie créés (contenant la matrice d’état étendu) e.g. 2 outputs créera deux fichiers .txt contenant les matrices d’état étendues des deux dernières phases d’*analyses*.
- `verbosityLevel` : Contrôle la densité d’informations de sortie dans le fichier .log
- `geometricTolerance` : La tolérance géométrique du coupleur CWIPI (par défaut 0.1)
- `epsilon` : Valeur utilisée pour les comparaisons avec 0

Les formats et l’organisation des informations dans les fichiers ont leur importance. En premier lieu, le fichier `obs_field.txt` contenant les mesures d’observations est organisé comme indiqué Tab. 4.1. Chaque colonne correspond à un nouveau temps avec un pas de temps égal à l’échantillonnage choisi lors des mesures. Dans le cas où l’observation est stationnaire, ce fichier ne se compose donc que d’une seule colonne. Les coordonnées correspondant aux capteurs utilisés pour mesurer les valeurs Tab. 4.1 sont renseignées dans un autre fichier appelé `obs_coordinates.txt`, voir Tab. 4.2.

Un autre fichier peut être nécessaire en *input* de l’outil, le fichier appelé `clippingCells.txt`, qui contient les informations relatives aux méthodes de localisation physique ou d’hyper-localisation. Celui-ci contient, organisé également en colonne, le nombre de cellules du domaine localisé ainsi que les IDs de chaque cellule. Par exemple, pour un domaine contenant les cellules 1 à 4, il serait indiqué : 4 0 1 2 3. Dans le cas de l’hyper-localisation, le fichier contient les mêmes informations mais pour tous les domaines impliqués, e.g. pour deux domaines aléatoires de 3 cellules, le fichier contiendrait : 3 0 1 2 3 16 17 18. L’obtention de l’ID des cellules des domaines impliqués peut s’obtenir facilement à l’aide de la fonction `topoSet` d’OpenFOAM. Des scripts pour préparer ce fichier sont par ailleurs fournis avec CONES.

<i>obs_field.txt</i> :				
$u_{x_1}^1$,	$u_{x_1}^2$,	$u_{x_1}^k$
$u_{x_2}^1$,	$u_{x_2}^2$,	\dots
				$u_{x_2}^k$
		\vdots		
$u_{x_{N_{ou}}}^1$,	$u_{x_{N_{ou}}}^2$,	$u_{x_{N_{ou}}}^k$

$u_{y_1}^1$,	$u_{y_1}^2$,	$u_{y_1}^k$
$u_{y_2}^1$,	$u_{y_2}^2$,	\dots
				$u_{y_2}^k$
		\vdots		
$u_{y_{N_{ou}}}^1$,	$u_{y_{N_{ou}}}^2$,	$u_{y_{N_{ou}}}^k$

$u_{z_1}^1$,	$u_{z_1}^2$,	$u_{z_1}^k$
$u_{z_2}^1$,	$u_{z_2}^2$,	\dots
				$u_{z_2}^k$
		\vdots		
$u_{z_{N_{ou}}}^1$,	$u_{z_{N_{ou}}}^2$,	$u_{z_{N_{ou}}}^k$

p_1^1	,	p_1^2	,	p_1^k
p_2^1	,	p_2^2	,	\dots
				p_2^k
		\vdots		
$p_{N_{op}}^1$,	$p_{N_{op}}^2$,	$p_{N_{op}}^k$

TABLE 4.1 : Contenu du fichier *obs_field.txt*. N_{ou} le nombre de capteurs de vitesse et N_{op} le nombre de capteurs de pression dans l'observation.

Aucun autre fichier d'entrée n'est nécessaire au fonctionnement de CONES. Ce dernier produit en revanche plusieurs fichiers de sortie. Tous les fichiers sont contenus dans un dossier *results* à créer au préalable. Un script dédié permet l'initialisation des paramètres, la création du dossier *results*, d'un dossier correspondant à chaque membre et enfin le lancement du calcul. CONES crée au cours du calcul un fichier appelé *UpdatedCoefficients* qui contient les valeurs des paramètres mis à jour ainsi que leur écart type pour chaque phase d'*analyse*. Le fichier *NRSMD* contient les valeurs de la racine de l'erreur quadratique moyenne normalisée des observations (voir [chapitre 6](#) et [chapitre 7](#)). Enfin, l'outil peut également éditer des fichiers contenant les matrices d'état, comme indiqué par la variable `numberOutputs`.

4.4 Exemple : Cavity test case

La validation de CONES, des méthodes et des améliorations développées pour le filtre de Kalman d'ensemble a été principalement menée sur un cas test simple appelé *lid-driven cavity flow*. Ce cas est un des premiers tutoriels proposé par OpenFOAM dans son guide. Il a été choisi en particulier pour la simplicité de sa modélisation. La cavité est un modèle carré 2D, de côté $a = 0.1m$ constitué de 20 mailles, pour un total de $20 \times 20 = 400$ mailles. La vitesse imposée sur

<i>obs_coordinates.txt</i> :		
X_{u_1}	, Y_{u_1}	, Z_{u_1}
X_{u_2}	, Y_{u_2}	, Z_{u_2}
	\vdots	
$X_{N_{ou}}$, $Y_{N_{ou}}$, $Z_{N_{ou}}$

X_{u_1}	, Y_{u_1}	, Z_{u_1}
X_{p_2}	, Y_{p_2}	, Z_{p_2}
	\vdots	
$X_{N_{op}}$, $Y_{N_{op}}$, $Z_{N_{op}}$

TABLE 4.2 : Contenu du fichier *obs_coordinates.txt*. Coordonnées des capteurs du fichier *obs_field.txt*.

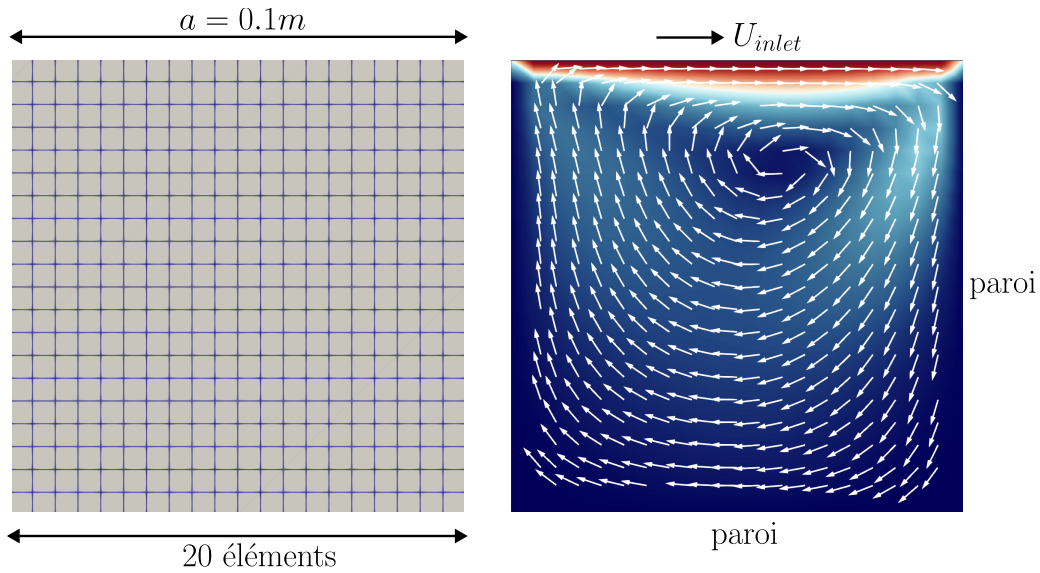


FIGURE 4.2 : Géométrie du cas d'étude de la cavité 2D. L'écoulement est entraîné par la vitesse imposée sur la face supérieure.

la face supérieure entraîne l'écoulement interne en rotation. L'écoulement est incompressible et laminaire avec $U_{inlet} = 1m/s$, une viscosité cinématique $\nu = 0.1m^2/s$ et $Re = aU_{inlet}/\nu = 10$. Le pas de temps est fixe avec $\Delta_t = 0.001s$ pour un temps de calcul total $t_f = 0.1s$. Son caractère laminaire et sa faible résolution spatiale font de cette simulation un candidat idéal pour mener à bien tous les tests de fonctionnement de CONES.

Les observations sont mesurées sur une simulation de référence du *lid-driven cavity flow* pour la même configuration. L'unique changement est la valeur U_{inlet} qui est multipliée par 5 pour cette simulation de référence $U_{inlet}^{obs} = 5$. Cette vitesse constitue la valeur cible vers laquelle l'optimisation menée à l'aide de CONES doit se rapprocher. Deux configurations de placement des capteurs d'observations sont utilisés pour l'exemple dans les sous-sections suivantes. Ces sections exposent plusieurs cas d'utilisation de l'outil selon différents hyper-paramètres.

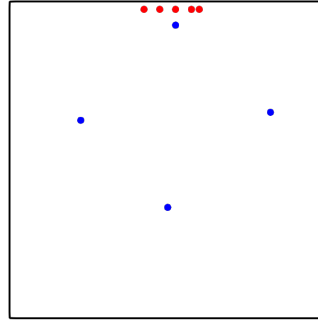


FIGURE 4.3 : Groupes d'observations utilisées. Groupe 1 en rouge, groupe 2 en bleu

4.4.1 Optimisation avec un EnKF classique

Deux groupes de capteurs différents ont été utilisés pour mesurer la vitesse instantanée dans le domaine, ils sont indiqués en rouge (groupe 1) et bleu (groupe 2) sur la Fig. 4.3.

Le groupe 1 contient des observations très proches de la condition limite imposée. Il procure donc une information très précise de la valeur de vitesse cible. Pour le groupe 2, l'information est moins qualitative au vu du positionnement plus éloigné de la face supérieure. Il est donc plus difficile d'atteindre la valeur cible de vitesse avec ce groupe d'observations. Pour utiliser le bon groupe d'observations lors du lancement du calcul, il suffit de renseigner les bonnes informations dans les fichiers *obs_coordinates.txt* et *obs_field.txt*. Une fois cette opération réalisée, le fichier *conesDict* doit être rempli avec la configuration correspondant au cas souhaité. 4 membres sont utilisés (`ensemble = 4`) pour les besoins de l'exemple. Bien que très limité, ce nombre est suffisant pour atteindre ici des résultats satisfaisant. Le calcul s'effectue en série sur 4 processeurs dédiés aux membres d'ensemble auxquels s'ajoute le processeur utilisé par l'algorithme du EnKF. Ce cas d'étude peut ainsi être testé sur une large gamme de machines, équipées d'au moins 8 processeurs. Le nombre de membres peut cependant être augmenté pour améliorer la précision et la robustesse du système dans le cas où davantage de processeurs sont disponibles. Chaque membre peut également être décomposé en sous-domaines grâce à la fonction d'OpenFOAM dédiée (*decomposePar*). Attention cependant, une décomposition coïncidente avec le positionnement des capteurs peut légèrement dégrader la solution dans le cas de la cavité. En effet, dans ce cas de figure, les interpolations calculées par OpenFOAM sont moins précises. Il en résulte une diminution de la fiabilité du système à cause de l'importance donnée à chaque observation et de la forte réponse de ce dernier aux changements paramétriques.

Les phases d'*analyses* de l'algorithme sont réalisées toutes les deux *prévisions* (`observationWindow = 2`). Le nombre d'observations dépend du groupe choisi (`numberObsProbesVelocity = 4` ou `5`). U_{inlet} est le seul paramètre optimisé (`numberParameters = 1`). L'observation est stationnaire (`obsTimeDependency = false`). Les observations sont uniquement constituées de vitesse (`obsType = U`) et seules les composantes U_x et U_y sont utilisées puisque le cas est 2D (`obsVelocityComponents = (1, 1, 0)`). Le bruit gaussien utilisé pour étendre les observations est un bruit blanc avec un écart type fixé à 20% des valeurs de chaque observation (`obsNoiseType = "relVal"`, `velocityObsNoise = 0.2`). Aucune localisation ainsi qu'aucune inflation n'est utilisée pour l'instant. Les champs de vitesses initiaux des membres d'ensemble sont identiques. En revanche, le paramètre U_{inlet} est initialement différent pour chaque membre et modifie donc le

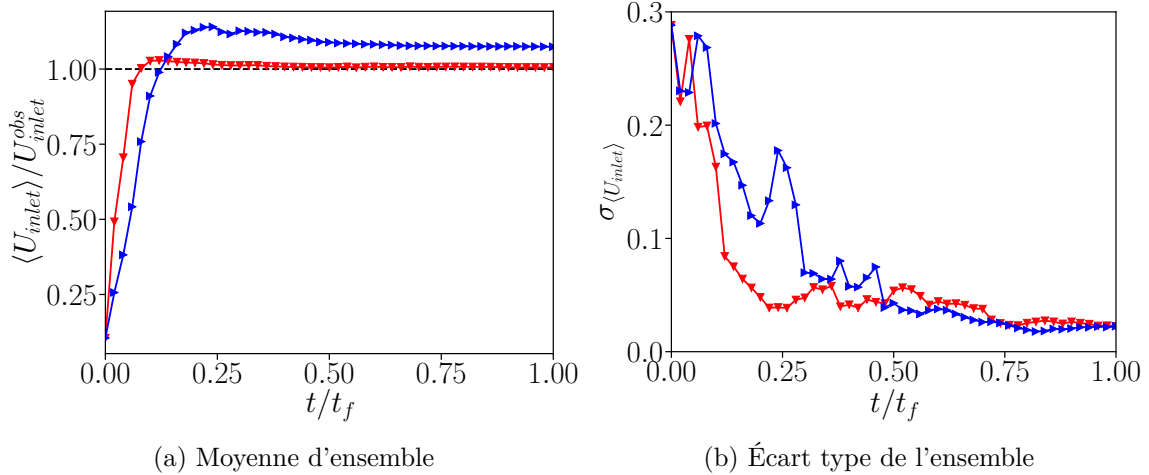


FIGURE 4.4 : Évolution de la vitesse imposée sur la face supérieure. En rouge évolutions pour le groupe 1 et en bleu évolutions pour le groupe 2.

champ de vitesse du domaine en conséquence. Les valeurs de paramètres sont uniformément distribuées entre 0 m.s^{-1} (exclu) et 1 m.s^{-1} pour ce cas simple. La figure 4.4 montre l'évolution de la moyenne des $N_e \times U_{inlet}$ ainsi que leur écart type. Comme attendu, la valeur atteinte pour le groupe 1 par tous les membres d'ensemble est identique à la valeur imposée pour l'observation. Cependant, le groupe 2 n'atteint pas la valeur cible. Les informations de vitesses instantanées de ce dernier ne sont pas assez représentatives pour atteindre cette valeur. Cela signifie qu'une gamme de valeurs de vitesse d'entraînement peuvent conduire à une même valeur de vitesse instantanée au niveau des capteurs d'observations du groupe 2 (en prenant en compte la confiance donnée aux observations).

4.4.2 Effets des hyper-paramètres

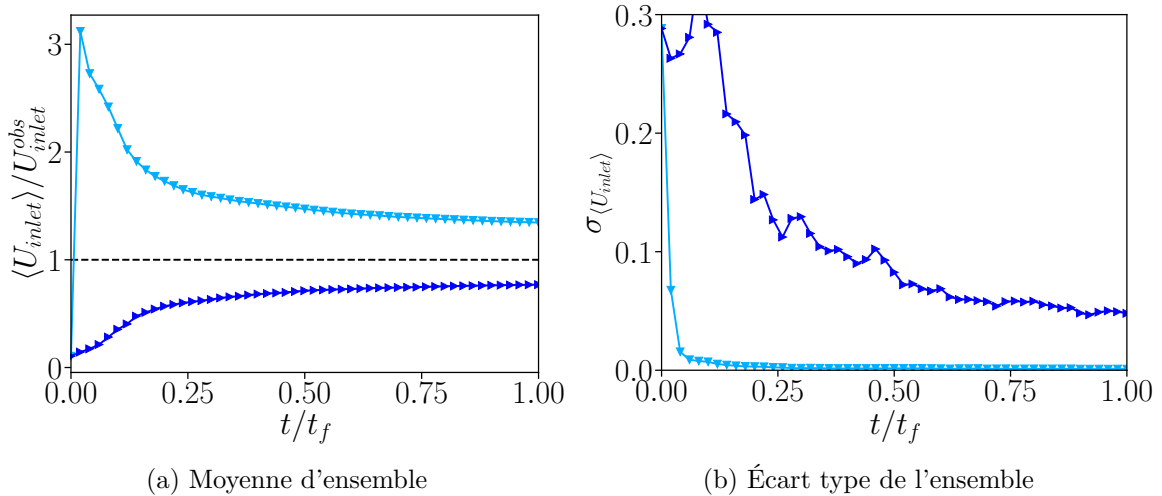
Le choix des hyper-paramètres dépend généralement du cas test étudié, certaines valeurs sont cependant récurrentes et utilisées dans un grand nombre d'études (e.g. le nombre de membres dans l'ensemble). La cavité nous permet de mettre en lumière l'influence de ces hyper-paramètres relativement simplement. Tous les tests effectués dans cette section et la suivante sont résumés dans le Tab. 4.3.

La Fig. 4.5 montre l'influence du bruit gaussien utilisé pour étendre les observations. Pour rappel, ce bruit, choisi de manière empirique, correspond aux valeurs de la diagonale de la matrice de covariance des erreurs sur les mesures $R = \sigma_\alpha I$. Il est utilisé dans l'équation 3.55 pour créer une matrice d'observations de taille cohérente. Par exemple, un pourcentage très faible traduit une grande confiance dans les observations. Ceci peut être responsable d'une trop forte réponse du système aux valeurs des observations, i.e. une prédiction des corrections par l'EnKF trop élevées.

La courbe en bleu clair correspond au profil paramétrique temporel pour $\sigma_\alpha = 1\%$, en bleu foncé pour $\sigma_\alpha = 50\%$. Une valeur trop faible entraîne une sur-prédiction du paramètre U_{inlet} , pouvant s'apparenter à de l'*overfitting* ou sur-apprentissage. La variance de l'ensemble est également très faible, à cause de cette grande confiance dans les observations. Cela ne

Groupe d'observations	σ_α	λ_p	type de localisation	Figures associées
2	1%	0%	aucune	4.5
2	50%	0%	aucune	4.5
2	20%	0%	aucune	4.6 et 4.7
2	1%	1%	aucune	4.6
2	20%	0%	uniquement de covariance	4.7
2	20%	0%	hyper-localisation	4.7 et 4.8
2	1%	0%	hyper-localisation	4.8
2	1%	1%	hyper-localisation	4.8

TABLE 4.3 : Configuration hyper-paramétrique des tests effectués sur la Cavité


 FIGURE 4.5 : Évolution temporelle du paramètre optimisé. En bleu clair $\sigma_\alpha = 1\%$ et en bleu foncé $\sigma_\alpha = 50\%$.

permet pas au système de rechercher un optimum global une fois qu'il a convergé vers une valeur. Au contraire, la courbe en bleu foncé montre le phénomène inverse dit "*underfitting*" ou sous-apprentissage. Bien que la variance de l'ensemble reste élevée, la précision de la prédiction est amoindrie. Autrement dit, en prenant en compte le bruit donné aux observations, un large intervalle de valeurs de vitesses des membres d'ensemble correspond aux valeurs des mesures d'observations, et donc aucun changement n'est nécessaire lors de la mise à jour de la matrice d'état étendue (de la description paramétrique). Le problème d'*underfitting* n'a pas de solutions simples, le choix de cet hyper-paramètre peut donc paraître primordial. Cependant, une large gamme de valeurs fonctionne dans les cas étudiés, un intervalle de 5 à 30% semble par exemple convenir pour la présente simulation.

En revanche, les résultats obtenus par *overfitting* peuvent être nettement améliorés. L'inflation est une méthode qui permet d'augmenter la variance du système et donc de lui permettre de cibler l'optimum global i.e. améliorer la précision de la prédiction. Un problème majeur dans le cas de l'*overfitting* est la très forte et rapide diminution de variance de la description paramétrique (voir la courbe bleu clair sur la Fig. 4.5b). La Fig. 4.6 montre en bleu clair l'optimisation

réalisée avec $\sigma_\alpha = 1\%$ ainsi qu'une inflation stochastique du paramètre $\lambda = 1\%$ (`inflationType = "stochastic"`, `parametersInflation = 0.01`). La deuxième courbe correspond au premier cas réalisé avec le groupe 2 sur la Fig. 4.4 pour lequel $\sigma_\alpha = 20\%$. L'influence de l'inflation est nettement visible au vu des valeurs d'écart type obtenues. L'augmentation de la variance permet donc au système de cibler l'optimum global dicté par l'observation et d'éviter le phénomène d'*overfitting*. Néanmoins, ces hyper-paramètres sont à magner avec précaution. Une trop forte réponse du système peut conduire à des valeurs temporairement aberrantes (par exemple lors de la première phase d'*analyse* sur la Fig. 4.6a) et imposer de fortes contraintes au solveur physique pouvant conduire à la divergence du calcul dans des cas plus exigeants (par exemple pour le chapitre 5, chapitre 6 et chapitre 7).

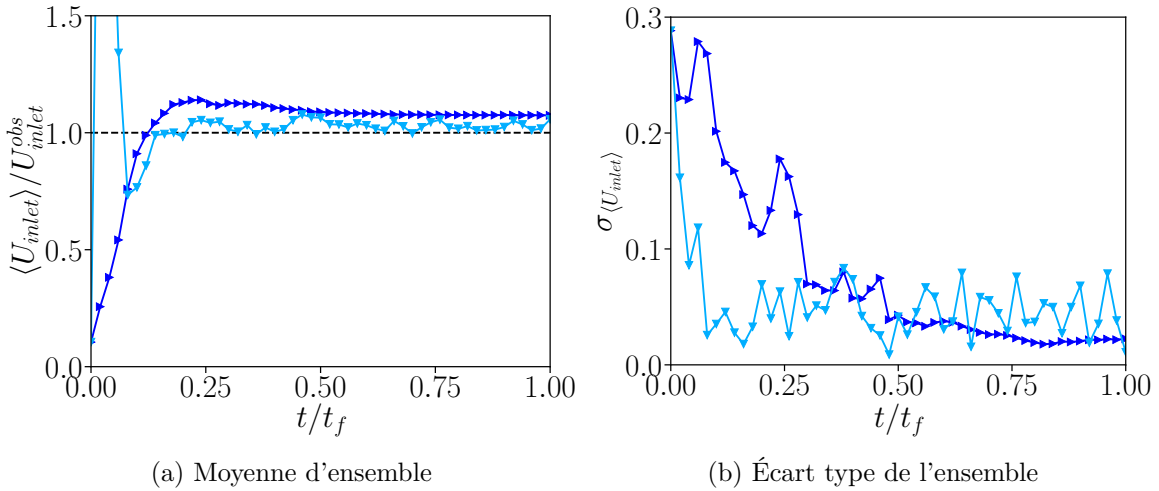


FIGURE 4.6 : Évolution temporelle du paramètre optimisé. En bleu clair $\sigma_\alpha = 1\%$ et $\lambda_p = 1\%$. En bleu foncé $\sigma_\alpha = 20\%$.

4.4.3 (Hyper)-localisation

Cette partie aborde les techniques de localisations du filtre de Kalman d'ensemble implémentées dans CONES. La localisation permet entre autres de réduire les corrélations de composantes du gain de Kalman pour des cellules dont les variables physiques sont normalement non corrélées (cf. chapitre 3). Plusieurs techniques peuvent être utilisées dans l'outil tel que la localisation de covariance, la localisation physique ou encore l'hyper-localisation. Le choix de ces méthodes de localisation se fait via les "switchs" correspondants dans le *conesDict* et les informations du fichier *clippingCells.txt* comme expliqué dans la Sec. 4.3.

Il n'est pas nécessaire d'utiliser une quelconque méthode de localisation pour la Cavité. En effet, la diffusion de la quantité de mouvement dans le domaine se fait très rapidement via la condition limite U_{inlet} imposée et mise à jour en tant que paramètre dans CONES. L'influence de ce genre de méthodes localisant les corrections effectuées dans le domaine n'a donc que peu d'impact sur la dynamique générale du système, bien que ces dernières puissent atténuer les corrections extrêmes conduisant à la divergence du calcul. En revanche, l'hyper-localisation a un fort impact sur l'inertie temporelle de la description paramétrique. En effet, une nouvelle

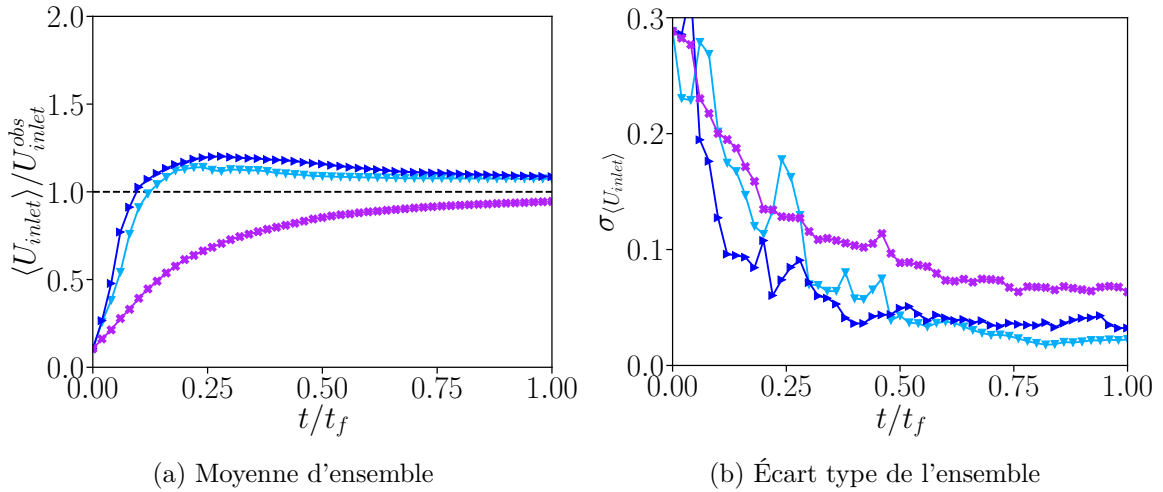


FIGURE 4.7 : Évolution temporelle du paramètre optimisé pour $\sigma_\alpha = 20\%$. En bleu clair aucune localisation, en bleu foncé localisation de la covariance, en violet hyper-localisation.

valeur du paramètre U_{inlet} est calculée pour chaque zone hyper-localisée (une valeur par capteur d'observations). La moyenne de ces valeurs est ensuite utilisée pour les nouvelles étapes de prévisions. Ceci a pour conséquence de conserver un fort niveau de variance sur le paramètre optimisé (comparé à un EnKF classique). La figure 4.7 montre l'évolution temporelle de U_{inlet} pour trois méthodes de localisations. Comme attendu, les résultats utilisant la localisation de covariance montrent un comportement très similaire au cas sans localisation. En revanche, l'hyper-localisation atténue la réponse du système, comme le montre la figure 4.7a. Ceci entraîne une meilleure stabilité ainsi qu'une meilleure précision finale pour un temps de réponse rallongé. La variance paramétrique est également plus élevée (cf. Fig. 4.7b). Ces caractéristiques sont bienvenues dans la plupart des cas, le EnKF ayant une forte proportion à corriger trop fortement les paramètres, entraînant un effondrement de la variance.

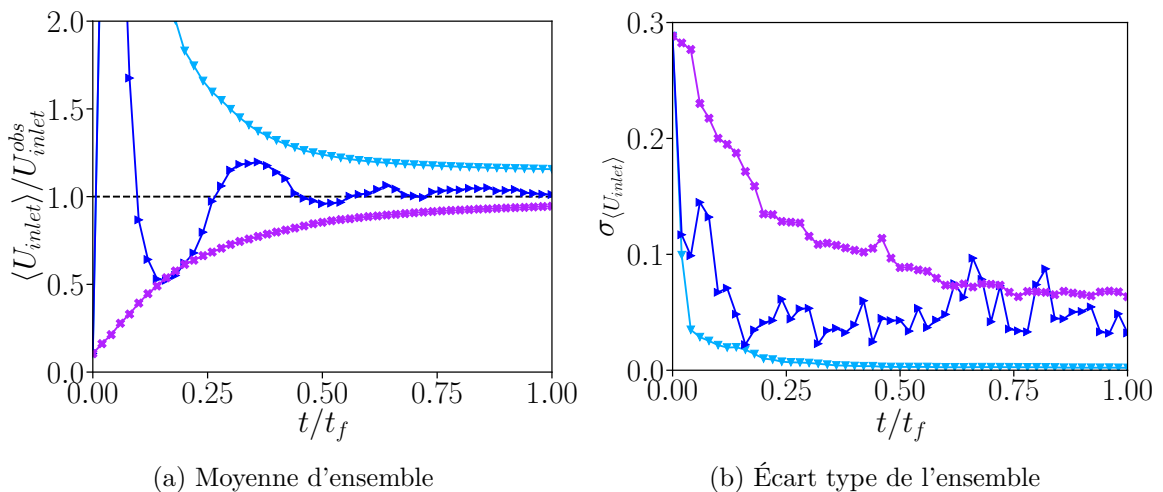


FIGURE 4.8 : Évolution temporelle du paramètre optimisé avec hyper-localisation. En bleu clair $\sigma_\alpha = 1\%$, en bleu foncé $\sigma_\alpha = 1\%$ et $\lambda_p = 1\%$, en violet $\sigma_\alpha = 20\%$.

La Fig. 4.8 montre le comportement du paramètre avec hyper-localisation dans le cas de l'*over-fitting* évoqué dans le Sec. 4.4.2. Le cas de l'hyper-localisation réagit de manière identique au cas classique. Une confiance trop forte dans les observations entraîne une chute de la variance du paramètre et donc une valeur finale imprécise (courbes bleu clair). Encore une fois, l'utilisation d'un faible niveau d'inflation sur le paramètre permet au système de cibler l'optimum global imposé par les observations. De plus, le niveau de variance reste en moyenne inférieur au cas $\sigma_\alpha = 20\%$ alors qu'une faible inflation est imposée sur la courbe bleu foncé.

L'EnKF hyper-localisé (HLEnKF) est donc un outil puissant qui permet une réduction significative du temps d'exécution (cf. Sec. 4.5) ainsi qu'un changement du comportement dynamique de l'optimisation paramétrique du EnKF. L'évolution temporelle de cette dernière est certes ralentie mais la précision est grandement améliorée et le système est stabilisé. La mise à jour du paramètre se fait pour chaque capteur, la valeur utilisée est la moyenne calculée sur l'ensemble des capteurs. Cela a également pour conséquence de conserver un niveau de variance plus élevé que dans un filtre de Kalman classique. Cette méthode est intégrée dans CONES via le switch correspondant. La création du fichier *clippingCells.txt* est également simplifiée via les scripts fournis avec CONES. L'utilisation du HLEnKF dans CONES est donc simple et apporte une amélioration des performances de l'optimisation dans certains cas.

4.5 Consommation et temps d'exécutions

Cette section est dédiée à l'évaluation du temps d'exécution de l'outil. L'étape de calcul du gain de Kalman au sein de l'algorithme EnKF est la plus coûteuse en termes de temps d'exécution des phases d'analyse. Les autres étapes de l'algorithme et les opérations de transfert de données au sein de CONES ont un impact plus faible sur la complexité temporelle. Rappelons d'abord la complexité algorithmique des deux opérations matricielles de base intervenant dans le calcul, en supposant que les opérations arithmétiques des éléments individuels des matrices aient une complexité unitaire $\mathcal{O}(1)$:

- Multiplication pour des matrices de taille (n,m) et (m,p) : $\mathcal{O}(nmp)$
- Inversion pour une matrice de taille (n,n) : $\mathcal{O}(n^3)$

Le gain de Kalman est donné par l'équation suivante (voir chapitre 3) :

$$\mathbf{K}_{k+1} = \mathbf{X}_{k+1}^f (\mathbf{S}_{k+1})^T \left[\mathbf{S}_{k+1} (\mathbf{S}_{k+1})^T + \mathbf{R}_{k+1} \right]^{-1} \quad (4.1)$$

Avec N le nombre de degrés de liberté du système, N_e le nombre de membres dans l'ensemble et N_o le nombre d'observations. La taille de chaque matrice impliquée dans le calcul est :

- \mathbf{X}_{k+1} : (N, N_e)
- \mathbf{S}_{k+1} : (N_o, N_e)
- \mathbf{R}_{k+1} : (N_o, N_o)

Ainsi, les opérations intermédiaires présentent les complexités suivantes :

- $\mathbf{X}_{k+1}^f(\mathbf{S}_{k+1})^T = \mathbf{A} : \mathcal{O}(NN_eN_o)$
- $\mathbf{S}_{k+1}(\mathbf{S}_{k+1})^T = \mathbf{B} : \mathcal{O}(N_o^2N_e)$
- $B + \mathbf{R}_{k+1} = \mathbf{C} : \mathcal{O}(N_o^2)$
- $[\mathbf{C}]^{-1} = \mathbf{D} : \mathcal{O}(N_o^3)$
- $AD : \mathcal{O}(NN_o^2)$

Par conséquent, la complexité algorithmique totale du gain de Kalman est la somme de la complexité de chaque opération :

$$\mathcal{O}(K) = \mathcal{O}(N_o^3 + [N + N_e + 1]N_o^2 + NN_eN_o) \quad (4.2)$$

Le résultat est un polynôme du troisième degré dépendant du nombre d'observations, soulignant l'importance de cette valeur. Réduire le nombre d'observations utilisées dans l'assimilation peut considérablement réduire le temps d'exécution. En outre, pour des calculs de mécanique des fluides numérique, le nombre de composantes de la matrice d'état peut être très élevé. Une façon de réduire ce nombre consiste à effectuer l'EnKF sur un domaine réduit, contenant toutes les observations, c'est le principe de la localisation physique. Une autre façon de réduire encore davantage le coût de calcul consiste à utiliser l'hyper-localisation. L'objectif est de réaliser un EnKF réduit pour chaque capteur utilisé dans l'observation. Une précision importante est que le nombre d'observations n'est pas équivalent au nombre de capteurs dans le cas où les observations sont vectorielles ou tensorielles. Notons c le nombre de capteurs, o_h le nombre d'observations pour chaque EnKF réduit, égal au nombre de composantes des scalaires, vecteurs ou tenseurs utilisées pour les observations (peut être égal à 1, 2, 3, 4, 5 ou 6) et N_i le nombre de cellules du domaine réduit correspondant à chaque capteur. La complexité algorithmique de chaque calcul de gain de Kalman réduit est donnée par :

$$\mathcal{O}(K) = \mathcal{O}(o_h^3 + [N_i + N_e + 1]o_h^2 + N_iN_eo_h) \quad (4.3)$$

Dans le cas où il existe une zone hyper-localisée par capteur (pas de chevauchement de zones), la complexité totale de l'algorithme devient :

$$\mathcal{O}(K) = \mathcal{O}(co_h^3 + [\sum_{i=1}^c N_i + cN_e + c]o_h^2 + (\sum_{i=1}^c N_i)N_eo_h) \quad (4.4)$$

Cette complexité souligne le gain potentiel en termes de temps d'exécution que peut entraîner le HLEnKF. Dans le cas d'une observation scalaire, ou encore en ne considérant qu'une seule composante d'un vecteur ou d'un tenseur pour l'observation, la complexité algorithmique du gain de Kalman devient linéaire. Elle dépend en grande partie du nombre de degrés de liberté de la matrice d'état, c'est-à-dire du nombre de mailles considérées pour l'inférence du domaine.

Le Tab. 4.4 résume les temps d'exécution d'une phase d'assimilation de données complète pour différentes configurations utilisant le cas du chapitre 6. Le temps mesuré ici correspond à une phase d'*analyse* complète et non au calcul du gain seulement. Par ailleurs, ces configurations entraînent une amélioration de la prédiction des paramètres inférés similaire dans tous les cas.

En outre, la validation de l'algorithme du HLEnKF a été menée sur le cas 7 et est disponible en annexe B. Pour la configuration de base cas 1, le domaine est constitué de $N = 350000 \times 3$ degrés de libertés, c'est-à-dire le nombre de valeurs dans le domaine physique pour 3 composantes de vitesses, $N_o = 408 \times 3$ observations et $N_e = 40$ membres dans l'ensemble. La complexité algorithmique associée est $\mathcal{O}(1, 59 \times 10^{12})$ pour un temps d'analyse de 405 s. Dans le cas où seules les composantes axiales de vitesses observées sont utilisées, cas 2, $N_o = 408$ observations, la nouvelle complexité algorithmique est $\mathcal{O}(1, 92 \times 10^{11})$ et le temps est divisé par 6. La réduction des degrés de liberté cas 3 entraîne une réduction quasi-linéaire du temps d'exécution pour atteindre 14.5 s avec une complexité algorithmique de $\mathcal{O}(1, 45 \times 10^{10})$. Le tableau met également en lumière le temps d'exécution pour différentes configurations utilisant le HLEnKF dans les cas 4 à 7. La simple division en deux zones de calculs grâce au HLEnKF dans le cas 4 permet de diviser le temps de calcul par presque 2.5. Le meilleur temps d'exécution est atteint pour le cas 6 pour lequel 19 zones de calculs sont utilisées pour un temps d'exécution 28 fois inférieur au cas 2 et 176 fois inférieur au cas 1. Il est cependant à noter que le cas 7, utilisant 176 zones de calcul, présente un temps d'exécution légèrement plus élevé de 8.8 s, soit seulement 7 fois inférieur au cas 2. Cette augmentation provient en grande majorité de l'optimisation algorithmique de CONES qui est sujette à de futures améliorations.

Cas	Description	N	N_o	N_e	Zones de calcul	Temps d'exécution(s)
1	Domaine complet	$350,000 \times 3$	1224	40	1	405
2	Complet + Observations axiales uniquement	$350,000 \times 3$	408	40	1	63.5
3	Domaine réduit 1	$78,700 \times 3$	408	40	1	14.5
4	HL - Domaine réduit 1 + division en 2 zones	$78,700 \times 3$	408	40	2	6
5	HL - Domaine réduit 2	$58,800 \times 3$	408	40	4	4.5
6	HL - Domaine réduit 3	$50,300 \times 3$	408	40	19	2.3
7	HL - Domaine réduit 4	$24,500 \times 3$	408	40	176	8.8

TABLE 4.4 : Complexité du gain et temps d'exécution d'une phase d'*analyse* complète pour le cas du canal turbulent [chapitre 6](#).

Écoulement incompressible stationnaire autour d'un corps non profilé

Sommaire

5.1	Introduction	87
5.2	Configuration de l'observation et de la simulation RANS	89
5.2.1	Modèle de bâtiment et essais en soufflerie	89
5.2.2	Observation : données expérimentales de la soufflerie	90
5.2.3	Simulation RANS	91
5.3	Procédure d'assimilation de données	92
5.3.1	Cas A : EnKF classique	96
5.3.2	Cas B : EnKF avec localisation de la covariance	97
5.3.3	Cas C : EnKF avec localisation et inflation	98
5.4	Résultats	100
5.4.1	Champ de vitesse	100
5.4.2	Champ de pression	103
5.5	Conclusions	106

5.1 Introduction

La première configuration étudiée durant ce travail de thèse est le cas bien connu du *high-rise building* [50]. C'est la première à avoir été inférée à l'aide de CONES. Le cas a en particulier été choisi à cause de la mise à disposition de simulations LES de référence et de données

expérimentales réalisées en soufflerie. La simulation RANS 3D stationnaire de l'écoulement incompressible autour du bâtiment est augmentée par l'intégration de données expérimentales à l'aide de CONES. Cette configuration constitue une étape intermédiaire avant l'étude d'écoulements plus complexes via l'utilisation de simulations LES dans le [chapitre 6](#). Ce travail a par ailleurs fait l'objet d'une publication [141].

Les configurations réalistes dans les environnements urbains montrent une interaction non linéaire de plusieurs phénomènes complexes qui ne peuvent pas être isolés [134]. Le cas du *high rise building* montre de telles caractéristiques malgré sa géométrie plutôt simple. La topologie d'écoulement complexe produite par les effets simultanés de la séparation des couches limites, des instabilités déclenchées par des couches de cisaillement ou encore par les sillages turbulents et autres régions de recirculation est difficile à capturer avec précision, même avec des stratégies avancées axées sur l'utilisation de données. Parmi les travaux rapportés dans la littérature, Ben-Ali et al. [8] et Zhao et al. [154] proposent une amélioration des modèles RANS classiques en utilisant des données expérimentales pour piloter leur procédure. Ce type d'application est encore rare dans la littérature pour la mécanique des fluides car la plupart des analyses reposent sur l'utilisation de modèles de résolutions différentes (voir [chapitre 3](#)). L'avantage principal de l'utilisation de données numériques issues d'une simulation haute-fidélité repose sur le contrôle complet des conditions aux limites et initiales, permettant d'exclure certains biais pouvant émerger de mesures expérimentales. Ben-Ali et al. [8] ont inféré le comportement de la modélisation de la turbulence dans OpenFOAM en utilisant des mesures expérimentales de pression sur la surface du bâtiment. Ils ont développé une méthode adjointe pour effectuer une telle optimisation. Leurs résultats ont montré que la précision de la modélisation RANS pouvait être améliorée. Cependant, cette amélioration est également le fruit du développement de techniques numériques avancées, y compris la modification des équations dynamiques pour les quantités physiques associées à la turbulence (énergie cinétique turbulente \mathcal{K} , taux de dissipation de l'énergie ε ou le taux spécifique de dissipation d'énergie ω). L'optimisation des constantes de modèle, qui sont accessibles à un utilisateur standard du code, ne produirait pas d'améliorations significatives dans ce cas. D'autre part, Zhao et al. [154] ont utilisé un code EnKF pour déduire le comportement de telles constantes de modèle, toujours en utilisant le logiciel OpenFOAM. Pour effectuer l'optimisation, ils ont utilisé des mesures de vitesse moyennées dans le temps disponibles sur 230 capteurs répartis sur la zone de passage de l'écoulement autour du bâtiment. Ils ont montré une calibration efficace des constantes du modèle. Toutefois, une telle répartition des données est difficile à obtenir dans des conditions réalistes, où les capteurs sont principalement regroupés dans des régions locales où les mesures peuvent être effectuées de manière efficace. Par exemple, Dans le cas de la mesure par fil chaud, en raison d'une interférence avec l'équipement, seul un nombre limité de fils peuvent être utilisés pour effectuer des mesures simultanément. Certains emplacements privilégiés de capteurs peuvent être exclus en raison de la limitation de l'équipement. Par exemple, les écoulements très turbulents entraînant de fortes recirculations dans le domaine peuvent entraver la précision des mesures [55]. De même, le placement de capteurs de pression près des bords d'un modèle peut également être restreint en raison de limitations pratiques. Ces considérations mettent en évidence les multiples problèmes de placement des capteurs, dans lequel les aspects algorithmiques et pratiques doivent être traités avec soin.

Le travail présenté dans ce chapitre utilise des observations hétérogènes. Elles sont fournies sous la forme de données moyennées dans le temps provenant de plusieurs prises de pression sur la surface du bâtiment ainsi que d'une quantité limitée de mesures de vitesse obtenues par fils chauds. L'hétérogénéité de l'observation présente une difficulté supplémentaire du fait des différences techniques des outils de mesure et d'acquisition. Cette observation hétérogène, qui partage des caractéristiques avec les données employées dans les deux études précédemment discutées, donne l'occasion d'effectuer une optimisation de modèles RANS en utilisant des données réalistes disponibles dans un environnement urbain. L'augmentation par assimilation de données est effectuée en optimisant plusieurs constantes libres globales qui déterminent le comportement de la fermeture des modèles RANS. Il sera démontré que les approches basées sur l'EnKF sont adaptées à l'intégration robuste des données expérimentales dans des modèles CFD d'ordre réduits. Le chapitre est organisé de la manière suivante. La section 5.2 expose la configuration des mesures expérimentales ainsi que les caractéristiques de la simulation RANS de base. La section 5.3 présente la configuration de l'assimilation de données. Les différentes techniques sont décrites et comparées, en sélectionnant la technique la plus performante. Les résultats obtenus sont comparés aux résultats d'une simulation LES ainsi qu'aux données expérimentales haute-fidélité dans la section 5.4. Enfin, les observations finales sont faites et les perspectives futures sont discutées dans la section 5.5.

5.2 Configuration de l'observation et de la simulation RANS

5.2.1 Modèle de bâtiment et essais en soufflerie

Les essais ont été menés dans la soufflerie à couche limite atmosphérique de l'Université Ruhr-Bochum, en Allemagne. La soufflerie a une section transversale de 1.6×1.8 m et la longueur de la section d'essai est de 9.4 m. Le modèle est de section carrée aux côtés $B = 0.133$ m, et de hauteur $H = 0.4$ m, représentant un bâtiment de 120 m de hauteur à échelle réelle. Le bâtiment a un toit plat, et la direction du vent 0° est étudiée de sorte que la vitesse asymptotique soit alignée sur la direction privilégiée de l'écoulement x . La Fig. 5.1 a) montre le modèle de bâtiment en bois monté sur une plate-forme rotative dans la soufflerie.

Le profil de vent incident moyen, mesuré à vide au centre de la plate-forme rotative dans la soufflerie, correspond à celui d'une loi en puissance de 0.2, comme indiqué sur la Fig. 5.1 b). La vitesse moyenne à la hauteur du modèle est $u_{ref} = 16$ m/s, tandis que l'intensité de la turbulence dans la direction de l'écoulement est $I_u = 13\%$. Ceci est représentatif d'un terrain de catégorie II [33] simulant des conditions réalistes de l'écoulement autour de grands bâtiments isolés. Cette configuration peut être utilisée pour approximer la topologie de l'écoulement dans des zones urbaines comportant un grand bâtiment dominant entouré par de petits édifices. Une telle configuration est typique des zones périurbaines en périphérie des grandes villes.

Les mesures en soufflerie comprennent des mesures de la pression à l'aide de 64 prises de pression sur le toit et 26 prises sur les façades, ainsi que 28 mesures de vitesse par fil chaud au-dessus du toit, comme le montre la figure. 5.2.

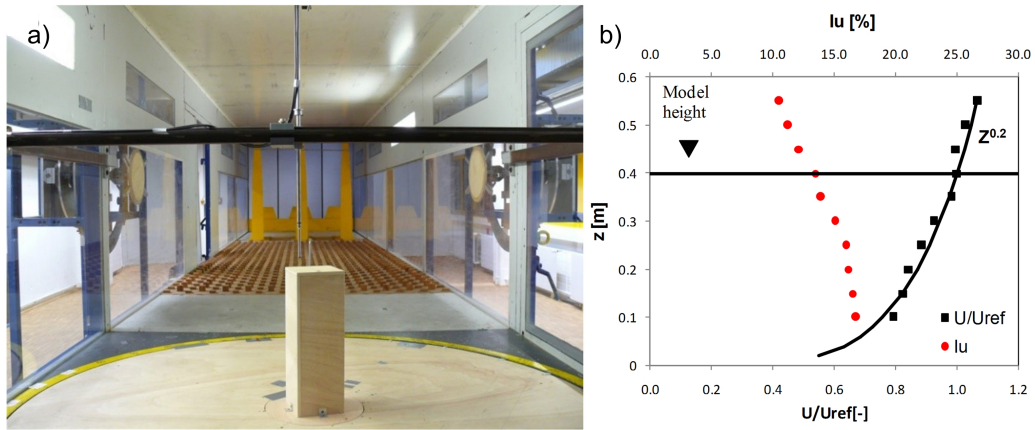


FIGURE 5.1 : Section d'essai de la soufflerie utilisée pour produire les données expérimentales. a) Visualisation du modèle de bâtiment et de son installation dans la soufflerie et (b) forme du profil de vitesse moyenne et de l'intensité de la turbulence dans la direction de l'écoulement x .

5.2.2 Observation : données expérimentales de la soufflerie

Les données échantillonnées sont hétérogènes car différents capteurs sont utilisés pour capturer les caractéristiques du champ de vitesse sur le toit et les mesures de pression sur la surface du bâtiment. Les vitesses au-dessus du toit sont principalement mesurées à deux hauteurs différentes ($z/H = 1.1, 1.15$) à des endroits marqués par des cercles noirs sur la Fig. 5.2 d). En outre, au-dessus de deux emplacements spécifiques, comme le centre du toit (fig. 5.2 d), neuf hauteurs sont considérées avec un écart de $(z - H)/H = 0.025$ entre chaque capteur. Les mesures sont effectuées à l'aide d'un anémomètre à fil chaud, qui se compose de deux fils croisés permettant de mesurer à la fois les composantes de vitesse en direction de l'écoulement ainsi que la vitesse verticale. Toutes les données de vitesse sont échantillonnées à la fréquence de 2000Hz.

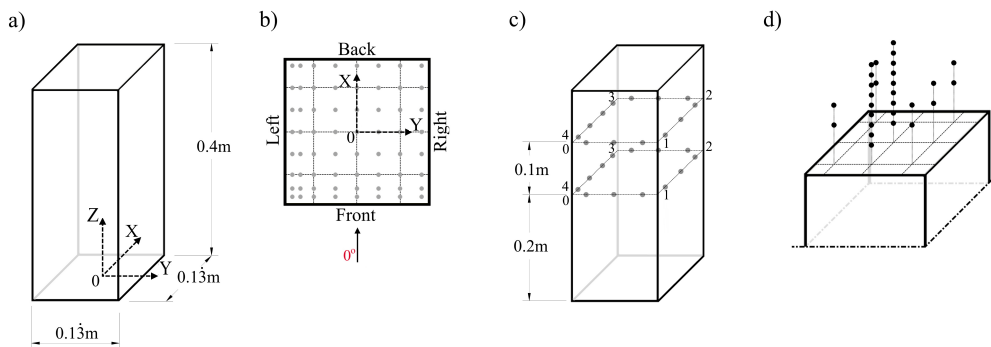


FIGURE 5.2 : Géométrie du bâtiment de grande hauteur avec a) les dimensions principales et le système de coordonnées ; b) la vue supérieure avec l'emplacement des prises de pression ; c) les façades avec l'emplacement des prises de pression ; d) les observations de vitesse mesurées au-dessus du toit.

Outre les mesures de vitesse, la pression en surface est également échantillonnée à différents

endroits, comme indiqué sur la Fig. 5.2 b) et c) marquée par des cercles gris clair. Les pressions locales à la surface sont obtenues à une fréquence d'échantillonnage de 1000Hz à l'aide d'un système d'acquisition multicanal. Les effets des tubes sont compensés numériquement [93]. Plus de détails sur les essais en soufflerie et l'analyse de l'écoulement autour du bâtiment, en particulier pour l'écoulement au-dessus du toit, sont présentés par Hemida et al. [50]. Les données expérimentales sont ici utilisées pour améliorer les capacités prédictives de deux modèles RANS stationnaires (standard $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ SST). Par conséquent, les séries de données temporelles disponibles pour les composantes de vitesse et pour la pression ont été moyennées dans le temps.

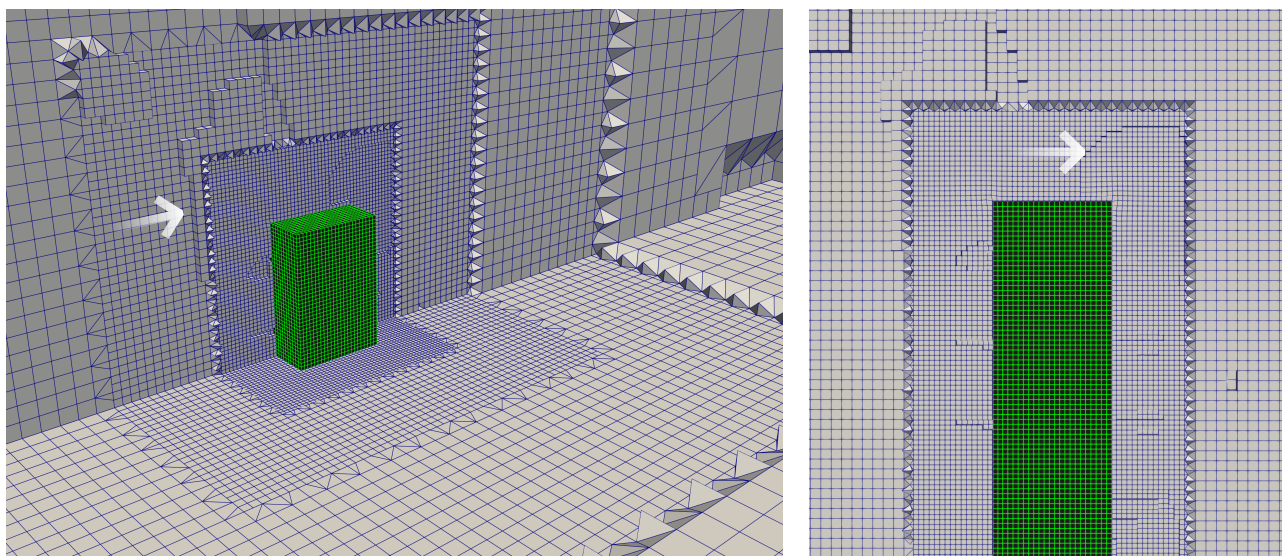


FIGURE 5.3 : Vue du maillage utilisé pour les calculs RANS. La gauche montre le plan vertical central et un plan horizontal à $H/2$. La droite montre le plan vertical central. La direction de l'écoulement asymptotique est indiquée par la flèche blanche.

5.2.3 Simulation RANS

Le cas considéré de bâtiment est une représentation numérique des essais en soufflerie. Les dimensions du domaine numérique sont choisies en respectant les recommandations données par Tominaga et al. [134]. La longueur du domaine en direction (x) en amont est de $5H$. Les dimensions totales du domaine sont longueur (x) \times largeur (y) \times hauteur (z) = $15.5 \times 4.5H \times 4H = 6.2\text{m} \times 1.8\text{m} \times 1.6\text{m}$. Pour la direction z , la hauteur a été choisie pour correspondre à la hauteur de la soufflerie.

Un maillage structuré est utilisé près de la surface du bâtiment, comme indiqué sur la Fig. 5.3. La distance du centre de la première maille adjacente par rapport à la surface du bâtiment conduit à une moyenne de $y^+ = 141$ et un minimum de $y^+ = 40$, ce qui assure que les centres des mailles à la paroi sont situés dans la couche logarithmique. Le nombre total de mailles utilisées pour discrétiser le domaine est de 513 266 mailles.

Une étude d'indépendance du maillage est réalisée en comparant les résultats avec ceux obtenus à l'aide d'un maillage plus raffiné. Ce dernier est composé de 4.3×10^6 mailles et est

caractérisé par une résolution spatiale deux fois plus élevée près du modèle de bâtiment que le maillage grossier. La pression moyenne prédite par les simulations grossière et raffinée est comparée aux emplacements des prises de pression de la Fig. 5.2. La comparaison a montré que 86% des points sur la surface du bâtiment ont une différence relative inférieure à 10%.

Deux modèles de turbulence sont étudiés : $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ SST. Les conditions aux limites d'entrée pour le champ de vitesse ainsi que pour les caractéristiques statistiques de la turbulence sont définies à l'aide des équations suivantes :

$$u_x(z) = u_{ref} \left(\frac{z}{z_{ref}} \right)^{0.2} \quad (5.1)$$

$$u_y(z) = u_z(z) = 0 \quad (5.2)$$

$$\mathcal{K}(z) = a(I_u(z)u(z))^2 \quad (5.3)$$

$$\varepsilon(z) = \frac{u_{ABL}^{*3}}{\kappa(z + z_0)} \quad (5.4)$$

$$\omega(z) = \frac{\varepsilon(z)}{\beta' \mathcal{K}(z)} \quad (5.5)$$

Pour le modèle $\mathcal{K} - \varepsilon$, la condition aux limites d'entrée pour la vitesse moyenne $\bar{\mathbf{u}}$, de l'énergie cinétique turbulente \mathcal{K} et du taux de dissipation de la turbulence ε sont déterminés à partir des profils de la composante verticale de vitesse incidente dans la soufflerie. L'énergie cinétique turbulente \mathcal{K} est calculée en utilisant l'Eq. 5.3 à partir de \mathbf{u} et de l'intensité de turbulence mesurée I_u ($I_u(z) = \frac{\sigma_u(z)}{\bar{\mathbf{u}}(z)}$) mesurés, où $\sigma_u(z)$ est l'écart type de la composante de vitesse dans la direction de l'écoulement. a est un paramètre pouvant varier entre 0.5 et 1.5 [98, 103, 134], et dans cette étude, $a = 1$ est choisi, comme recommandé par [134]. Les Eq. 5.4 et Eq. 5.5 utilisent la constante de Von Karman $\kappa = 0.42$ et $\beta' = 0.09$, respectivement. La longueur de rugosité aérodynamique est définie comme $z_0 = 0.002$ m, et la vitesse de frottement comme $u_{ABL}^* = \kappa \frac{u_{ref}}{\ln\left(\frac{H+z_0}{z_0}\right)}$. De même, dans le cas du modèle de turbulence $\mathcal{K} - \omega$ SST, le taux de dissipation turbulente spécifique ω est calculé sur la base du taux de dissipation de la turbulence Eq. 5.4 et de l'énergie cinétique de la turbulence Eq. 5.3. Les conditions aux limites d'entrée sont paramétrées à l'aide de la bibliothèque groovyBC pour OpenFOAM. La condition de sortie est à pression imposée avec une pression relative constante égale à zéro ainsi qu'un gradient de vitesse nul.

L'algorithme SIMPLE [39] a été utilisé pour le couplage pression-vitesse. Des choix classiques ont été effectués pour les schémas numériques. Des schémas de premier ordre ont été utilisés pour les termes de convection, tandis que des schémas centrés de second ordre sont utilisés les termes visqueux. L'interpolation de la pression du centre de la cellule vers le centre des faces a été obtenue grâce à des schémas linéaires de second ordre natifs à OpenFOAM.

5.3 Procédure d'assimilation de données

CONES est ici utilisé pour étudier la configuration de l'écoulement autour du bâtiment présenté dans le cas test Sec. 5.2. En particulier, les outils d'AD sont utilisés pour optimiser la valeur de

certaines constantes globales qui déterminent les performances du modèle de turbulence. Ces coefficients peuvent être définis manuellement par l'utilisateur sans aucune modification structurelle du solveur. Les résultats actuels visent à fournir de nouvelles recommandations générales pour l'utilisation de la modélisation de la turbulence, en particulier pour les applications industrielles, étant donné que les caractéristiques physiques du cas test étudié sont observées pour la plupart des écoulements autour de corps non-profilés tridimensionnels. Deux études principales sont menées. Dans la première, les cinq coefficients globaux du modèle $\mathcal{K} - \varepsilon$ [66] sont optimisés dans le but de minimiser la divergence entre les résultats RANS et l'observation expérimentale de haute-fidélité fournie. La deuxième analyse partage le même objectif, mais le modèle utilisé est le $\mathcal{K} - \omega$ SST [148]. Dans ce cas, l'optimisation par AD vise les neuf coefficients de modèle introduits au chapitre 2.

Le premier aspect clé à prendre en considération pour la représentation du modèle est de déterminer un état *a priori* approprié pour les champs de vitesse et de pression, ainsi que pour la description paramétrique. Pour cette dernière, les valeurs obtenues par Margheri et al. [69] utilisant la propagation d'incertitudes sont préférées aux valeurs classiques obtenues par Launder et Sharma [66]. Ces valeurs de base, qui sont affichées le Tab. 5.1, sont la moyenne initiale de l'ensemble N_e de simulations. Les valeurs des paramètres pour chaque simulation de cet ensemble sont initialement déterminées en utilisant une distribution gaussienne bornée telle que $\mathcal{N}(\mu_N, \sigma_N^2)$, où μ_N est la valeur moyenne du paramètre et σ_N est choisi pour fournir une variabilité initiale suffisamment grande de l'espace paramétrique basé sur le travail de Margheri et al. [69]. Ceci garantit une distribution initiale suffisamment large des paramètres afin que l'EnKF puisse cibler avec succès une configuration optimisée. L'ensemble est composé de $N_e = 40$ membres. Ce nombre est choisi en tenant compte de travaux similaires se basant sur un modèle de CFD pour la partie modèle du EnKF [59, 87, 84].

L'observation est obtenue à partir de données moyennées dans le temps mesurées sur les capteurs décrits dans la Sec. 5.2.1. Parmi ceux-ci, les données obtenues par fil chaud sont la mesure des composantes u_x et u_z de la vitesse. Le nombre total d'observations utilisées dans la procédure d'assimilation de données est donc de 146. Les données sont lues au début des phases d'analyses dans le format suivant : $\mathbf{y} = [u_{x1} \dots u_{x28} \ u_{z1} \dots u_{z28} \ p_{29} \dots p_{118}]^T$. Le vecteur d'observations ne change pas tout au long du calcul puisque l'écoulement est stationnaire. Il est stocké dans la mémoire RAM et directement utilisé à chaque phase d'analyse. L'une des principales caractéristiques des approches basées sur le filtre de Kalman est qu'un niveau de confiance approprié doit être choisi quant à la précision de l'observation. L'exactitude des données expérimentales actuelles est estimée à environ 5% pour les prises de pressions et les mesures par fil chaud. On suppose également que les incertitudes des mesures expérimentales ne sont pas corrélées. Par conséquent, la matrice de covariance de l'observation R_{k+1} est considérée comme constante et exprimée comme $R = \sigma_m I$, où σ_m est la variance décrivant l'incertitude sur les mesures.

Enfin, les calculs d'AD sont effectués à l'aide de solveurs CFD stationnaires. Il n'y a donc pas de véritable évolution temporelle pour le cas test étudié dans ce chapitre. Initialement, les simulations sont exécutées en utilisant les différentes descriptions produites par la distribution gaussienne *a priori*. Ensuite, la procédure d'assimilation optimise le champ de vitesse et les paramètres pendant la phase d'analyse de l'algorithme. L'augmentation de l'état obtenue

$\mathcal{K} - \varepsilon$ Parameters	default values	Prior of the EnKF		Optimized values
		μ_N	σ_N	
C_μ [-]	0.09	0.1	0.01	0.032
$C_{\varepsilon 1}$ [-]	1.44	1.575	0.1	0.165
$C_{\varepsilon 2}$ [-]	1.92	1.9	0.1	4.080
$\sigma_{\mathcal{K}}$ [-]	1.0	1.0	0.1	0.476
σ_ε [-]	1.3	1.6	0.1	0.1

$\mathcal{K} - \omega$ SST Parameters	default values	Prior of the EnKF		Optimized values
		μ_N	σ_N	
$\sigma_{\mathcal{K} 1}$ [-]	0.85	0.7	0.05	0.134
$\sigma_{\mathcal{K} 2}$ [-]	1.0	1.0	0.05	0.004
$\sigma_{\omega 1}$ [-]	0.5	0.625	0.05	0.812
$\sigma_{\omega 2}$ [-]	0.856	0.856	0.05	0.103
α_1 [-]	0.5556	0.575	0.05	0.013
α_2 [-]	0.44	0.44	0.05	0.069
β_1 [-]	0.075	0.09	0.005	0.45
β_2 [-]	0.0828	0.0828	0.005	0.008
β^* [-]	0.09	0.09	0.005	0.191

TABLE 5.1 : Coefficients globaux des modèles de turbulence RANS à optimiser via l'AD. Colonne de gauche à droite : valeurs classiques dans la littérature, caractéristiques de la distribution gaussienne bornée utilisée pour générer les *a priori*, valeurs optimisées.

par l'analyse, utilisée comme condition initiale pour l'étape suivante, est capable d'accélérer la convergence vers la prochaine phase d'analyse et de stabiliser le calcul avec une légère augmentation des ressources nécessaires. Les simulations sont ensuite ré-exécutées pour un nombre suffisamment important d'itérations afin que le solveur propage efficacement l'effet de la mise à jour des constantes de modèle de turbulence et obtienne une solution stationnaire convergée. Ensuite, une nouvelle phase d'analyse est effectuée et le cycle se poursuit jusqu'à ce que l'optimisation paramétrique ait atteint un niveau de convergence approprié.

Trois implémentations différentes de l'EnKF ont été testées. L'objectif est d'évaluer la robustesse du modèle d'AD, ainsi que de choisir la meilleure stratégie en termes de rapport entre précision et coût du calcul. Cette étude d'efficacité de la procédure a été effectuée uniquement pour le modèle $\mathcal{K} - \varepsilon$. Trois calculs d'AD indépendants sont effectués. Les variations ne traitent pas des détails du modèle ou de l'observation, mais des hyper-paramètres de la procédure d'AD. Plus précisément, les cas analysés sont les suivants :

- Cas A : EnKF classique.
- Cas B : EnKF avec localisation de la covariance.
- Cas C : EnKF avec localisation de la covariance et inflation.

Des analyses préliminaires ont été effectuées afin d'identifier une configuration initiale appropriée pour les hyper-paramètres qui favorise la performance de l'algorithme d'assimilation.

Parameter	$\mathcal{K} - \varepsilon$ model default values	Prior of the EnKF		
		μ_N	σ_N for cases A,B	σ_N for case C
C_μ [-]	0.09	0.1	0.01	0.005
$C_{\varepsilon 1}$ [-]	1.44	1.575	0.1	0.05
$C_{\varepsilon 2}$ [-]	1.92	1.9	0.1	0.05
$\sigma_{\mathcal{K}}$ [-]	1.0	1.0	0.1	0.05
σ_ε [-]	1.3	1.6	0.1	0.05

TABLE 5.2 : Comparaison entre les constantes conventionnelles du modèle RANS $\mathcal{K} - \varepsilon$ et les paramètres initiaux employés pour l'EnKF ($N_e = 40$).

Les distributions Gaussiennes initiales indiquées dans le Tab. 5.2 sont bornées entre σ_N et $7\sigma_N$ selon le paramètre considéré. Ces bornes ont été empiriquement définies en fonction de la sensibilité des coefficients aux variations. Par exemple, $C_{\varepsilon 1}$ est borné par $1.25\sigma_N$, mais σ_ε est borné par $7\sigma_N$. L'état physique initial (champ de vitesse) de chaque membre de l'ensemble est obtenu à partir d'une seule simulation en utilisant les valeurs des constantes de modèle données Ref. [69].

Pour les cas A et B, la fenêtre d'observation est de 150 itérations de solveur CFD entre chaque phase d'analyse. Pour le cas C, ce nombre est réduit à 100 pour tenir compte des phases d'analyse supplémentaires nécessaires en raison de l'inflation. Le nombre d'itérations a été choisi en observant les résultats des analyses préliminaires, qui ont montré qu'au moins 50 itérations étaient nécessaires pour obtenir une convergence satisfaisante de la nouvelle configuration paramétrique sur les quantités physiques.

Pour la localisation dans les cas B et C, le domaine est également coupé dans un volume suffisamment grand autour des observations, selon les propositions effectuées par Moldovan et al. [83] pour la géométrie BARC. Considérant que la taille de la matrice d'état est liée au nombre de degrés de liberté calculés par le modèle CFD, une localisation physique peut réduire les coûts globaux du calcul et assurer la stabilité de l'algorithme. L'estimation de l'état n'est plus appliquée qu'au volume de contrôle montré sur la Fig. 5.4, contenant 179 568 éléments, ce qui signifie 35% du nombre total de mailles. En outre, les coefficients l de la localisation de covariance sont spécifiquement sélectionnés pour chaque direction de l'espace de sorte que la discontinuité du champ de vitesse à la limite de la région de calcul soit égale ou inférieure à 0.3%, assurant la continuité des solutions physiques. En tenant compte de ce critère, $l_x = l_z = 0.0195$ et $l_y = 0.0438$ dans l'équation 3.65.

L'évolution temporelle de l'optimisation des coefficients de modèle pour les trois cas étudiés est maintenant commentée. Il est important de souligner que, malgré certaines différences importantes observées dans l'optimisation des constantes de modèle avec les trois stratégies, la prédiction de l'écoulement est très similaire. Cela implique que les différences observées sont dans une région où la sensibilité de la solution aux paramètres est faible, c'est-à-dire que de relativement grandes variations paramétriques correspondent à de petits changements physiques.

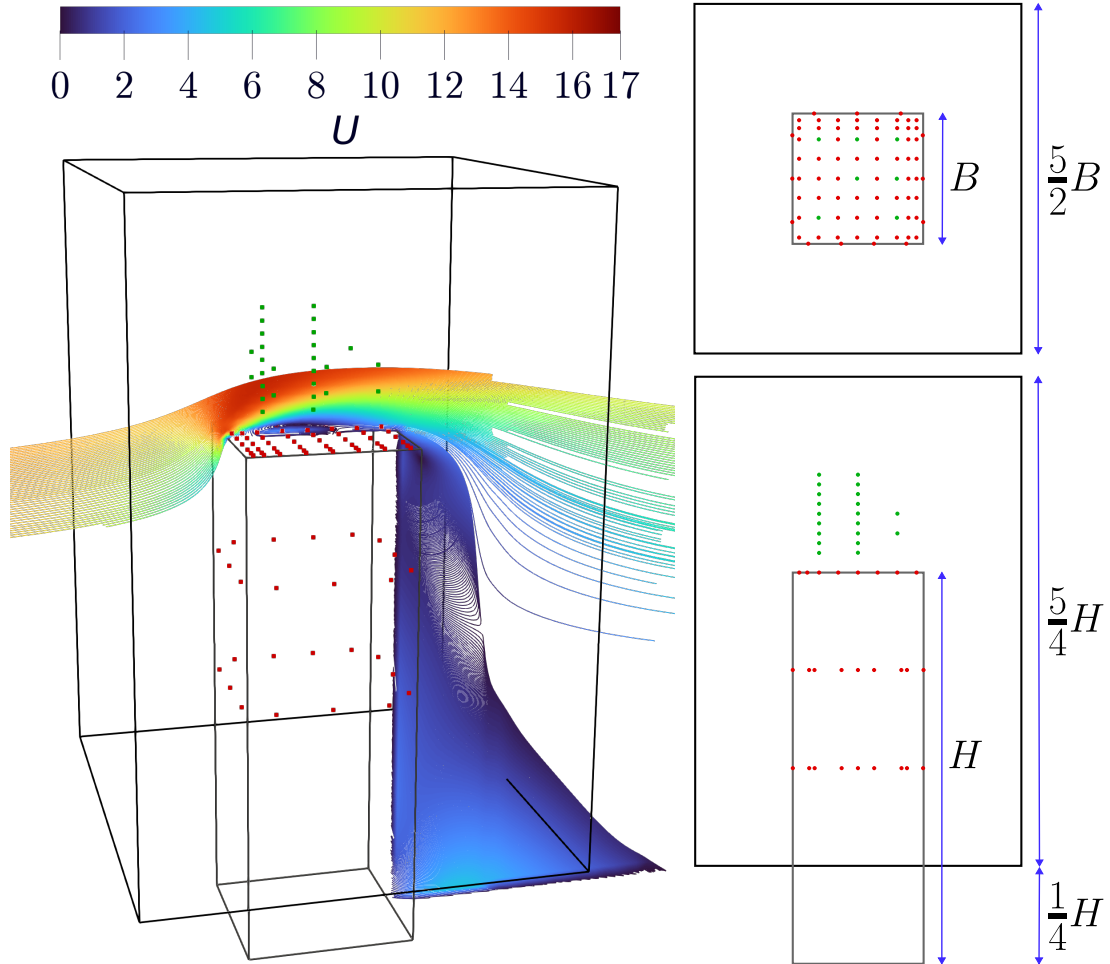


FIGURE 5.4 : Région rectangulaire utilisée pour la localisation : les capteurs de pression sont représentés en rouge et les captures de vitesse sont affichées en vert.

5.3.1 Cas A : EnKF classique

L'EnKF classique est utilisé pour ce premier cas. Le calcul se termine lorsqu'une convergence appropriée des paramètres est atteinte, ce qui dans ce cas est atteint après 100 phases d'analyse (c'est-à-dire un total de 15000 itérations CFD). L'évolution de la valeur moyenne des cinq paramètres du modèle $\mathcal{K} - \varepsilon$ est exposée sur la Fig. 5.5. Les résultats finaux obtenus par l'EnKF diffèrent sensiblement des valeurs de référence. La rapidité de convergence des paramètres entre eux est également significativement différente. En particulier, l'évolution de σ_ε mérite quelques commentaires. Ce coefficient contrôle l'amplitude du terme de diffusion turbulente dans l'équation de ε , $D_\varepsilon = \nu_t / \sigma_\varepsilon + \nu$, associé à la non-homogénéité de l'écoulement (voir chapitre 2). L'optimisation effectuée par l'EnKF vise des valeurs très faibles de σ_ε au cours du calcul, ce qui augmente le poids de D_ε dans l'équation. Cependant, il peut arriver que l'optimisation entraîne des valeurs négatives de σ_ε pour certains membres à cause du bruit attribué aux observations et propagé par le gain de Kalman, ce qui entraîne une divergence de la simulation CFD. Par conséquent, une contrainte a été imposée afin que les valeurs ne puissent pas être inférieures à

une très faible valeur positive. Les autres paramètres, tel que C_μ et $C_{\varepsilon 1}$, convergent à une valeur proche de 1/3 de l'estimation initiale. $\sigma_{\mathcal{K}}$ n'expose pas de grandes variations et $C_{\varepsilon 2}$ est significativement plus grand. De plus, ce dernier paramètre ne semble pas complètement converger. La comparaison des résultats obtenus avec les trois configurations, affichées Tab. 5.3, indiquent une grande variabilité pour ce coefficient. L'analyse des grandeurs physiques semble indiquer que, pour $C_{\varepsilon 2} > 4$, la solution présente une très faible sensibilité aux variations de ce paramètre.

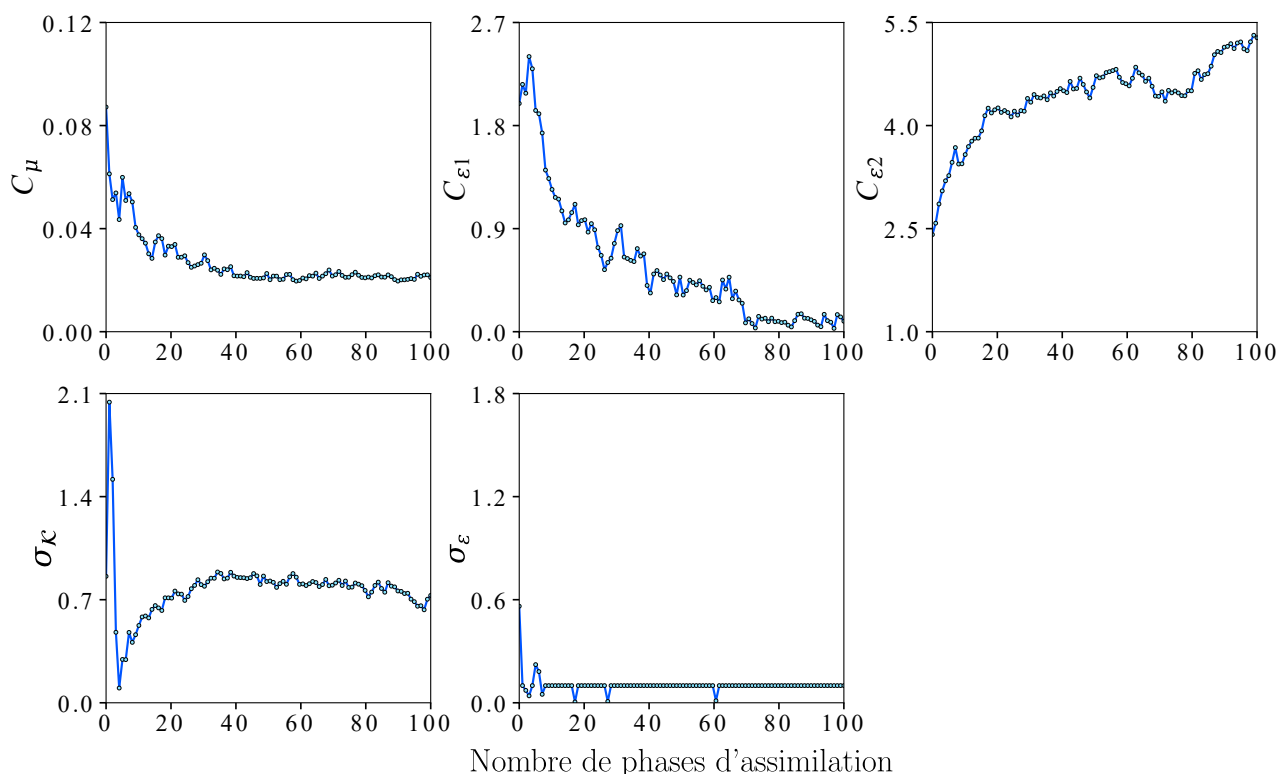


FIGURE 5.5 : Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF sans inflation et localisation.

5.3.2 Cas B : EnKF avec localisation de la covariance

Pour ce cas, le calcul est effectué en localisant le gain de Kalman comme indique chapitre 3. Une localisation physique est également effectuée comme précisé dans la section précédente. L'évolution des cinq coefficients est exposée sur la Fig. 5.6. Les tendances générales et en particulier l'évolution de σ_ε sont similaires à celles observées pour le cas A. Les remarques fournies pour le cas A, en particulier pour l'évolution de σ_ε , sont également valables pour cette configuration. Par ailleurs, un aspect supplémentaire à prendre en compte ici est la diminution substantielle du coût du calcul. La localisation, en particulier physique mais aussi de covariance, permet d'atteindre une précision au moins similaire pour un coût plus faible. Ceci suggère que ce genre de techniques peuvent être très efficaces pour des calculs impliquant un très grand nombre de degrés de liberté.

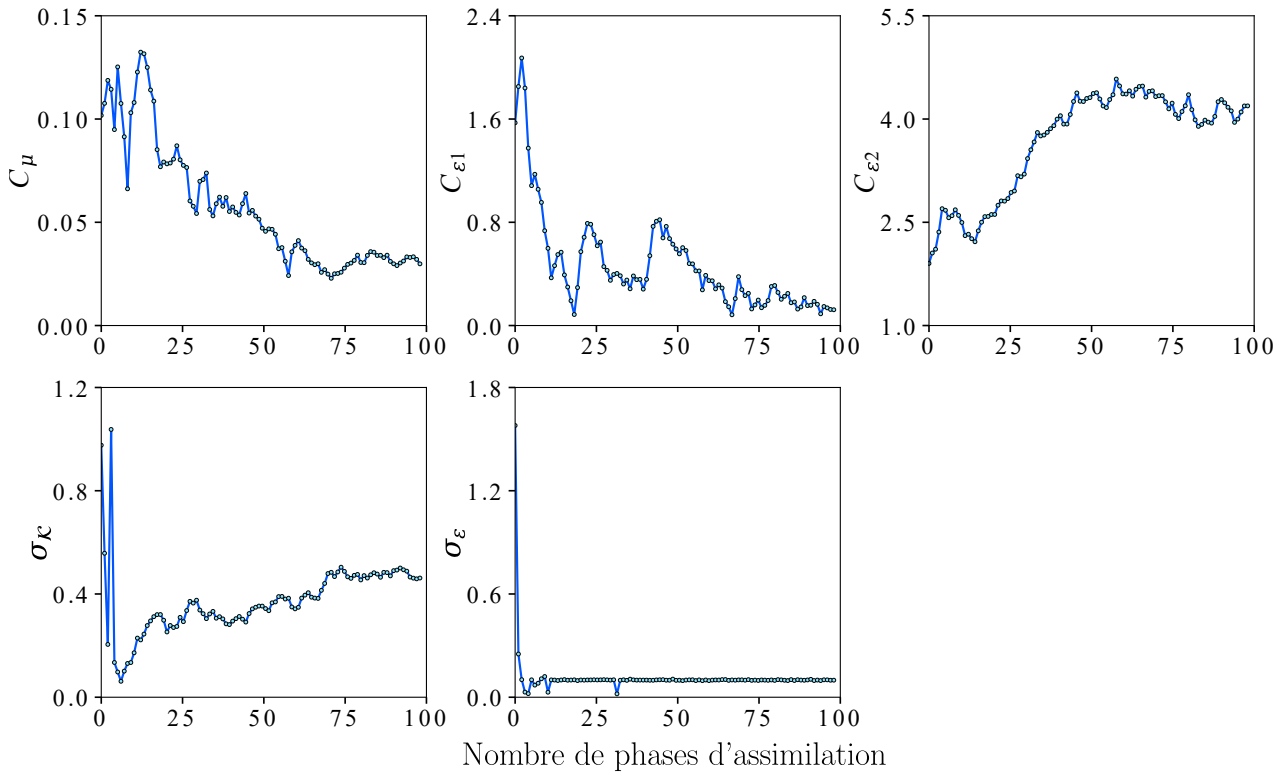


FIGURE 5.6 : Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF sans inflation mais avec localisation.

5.3.3 Cas C : EnKF avec localisation et inflation

L'AD est ici effectuée ici avec une inflation déterministe appliquée sur les paramètres de modèle et ainsi que les méthodes de localisation utilisées dans le cas B. C'est la configuration la plus avancée en termes de complexité de l'algorithme d'assimilation. L'évolution des cinq paramètres est exposée sur la Fig. 5.7. Pour assurer la robustesse de la simulation au cours des premières étapes de calcul, la valeur d'inflation λ est progressivement augmentée de 1.05 à 1.3 puis supprimée pour faire converger les paramètres ($\lambda = 1.05$ pour $k \in [1, 40]$, $\lambda = 1.1$ pour $k \in [41, 120]$; $\lambda = 1.2$ pour $k \in [121, 160]$). Certains coefficients tels que C_μ et σ_K montrent une plus grande sensibilité aux variations de λ , soulignant l'importance de l'inflation dans l'identification d'un espace paramétrique approprié pour l'optimisation. Pour cette raison, la convergence est atteinte beaucoup plus tard dans ce cas. En outre, la valeur seuil choisie pour σ_ε est légèrement augmentée ici, afin d'éviter certains problèmes de stabilité rencontrés du fait de la variabilité plus élevée associée à l'inflation paramétrique.

Il convient de faire quelques remarques sur les performances des trois configurations. Malgré des différences dans les techniques utilisées et les résultats à première vue différents obtenus pour l'optimisation des paramètres, la prédiction des grandeurs physiques est assez similaire. Les valeurs des paramètres du modèle ont probablement convergé vers un optimum robuste, où la sensibilité de la solution à d'autres variations paramétriques est très faible. Cet aspect, qui nécessite une enquête plus approfondie, peut indiquer qu'une optimisation robuste peut être

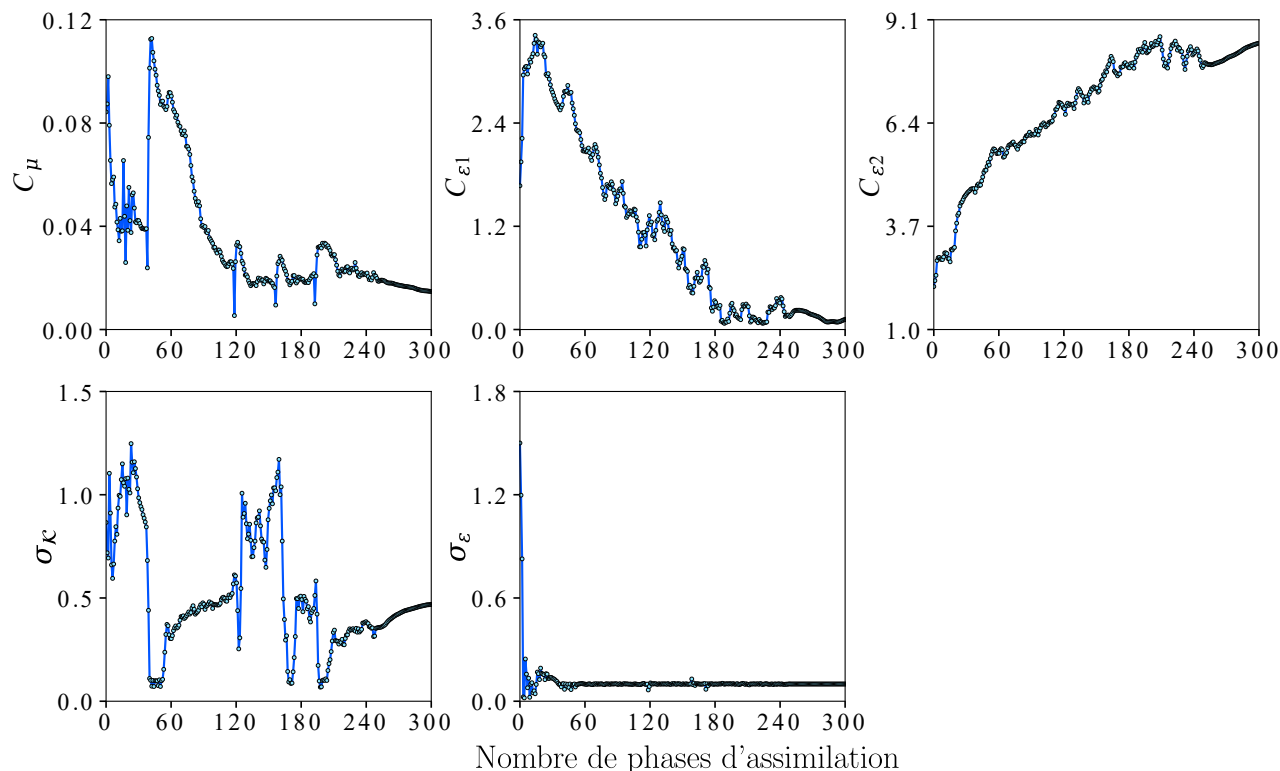


FIGURE 5.7 : Évolution des coefficients du modèle $\mathcal{K} - \varepsilon$. La stratégie d'AD utilisée est l'EnKF avec inflation et localisation.

Parameter	Optimized values		
	Case A	Case B	Case C
C_μ [-]	0.021	0.032	0.015
$C_{\varepsilon 1}$ [-]	0.091	0.165	0.152
$C_{\varepsilon 2}$ [-]	5.278	4.080	8.574
$\sigma_{\mathcal{K}}$ [-]	0.729	0.476	0.477
σ_ε [-]	0.1	0.1	0.1

TABLE 5.3 : Coefficients de modèle optimisés $\mathcal{K} - \varepsilon$ obtenus avec les différentes stratégies basées sur l'EnKF.

obtenue en établissant un intervalle de confiance approprié pour l'observation. L'application de la localisation s'est également avérée efficace ici. La réduction des degrés de liberté du processus d'AD a réduit les ressources nécessaires pour chaque phase d'analyse sans entraîner une dégradation des résultats. Par ailleurs, probablement en raison des caractéristiques de l'espace paramétrique optimal trouvé, les techniques d'inflation n'ont pas amélioré les résultats. L'EnKF avec localisation et sans inflation a donc été utilisé dans l'analyse physique des résultats présentée dans la section suivante. Les coefficients optimisés des modèles $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ SST à l'aide de cette procédure sont donnés dans le Tab. 5.1.

5.4 Résultats

L'analyse se concentre d'abord sur la précision de la prédiction du champ de vitesse. Cette prédiction est essentielle pour les applications environnementales. Elle est utile pour l'évaluation d'indicateurs urbains pertinents tels que la force du vent au niveau des piétons, la dispersion des polluants, la ventilation naturelle, la ventilation urbaine et l'énergie éolienne en milieu urbain. Malgré les lacunes décrites dans l'introduction, la modélisation RANS est considérée comme un outil d'analyse pertinent dans ce domaine [127]. La deuxième partie de la section est consacrée à l'analyse du champ de pression. Pour cette quantité physique, qui est pertinente pour les applications d'ingénierie structurelle axées sur les contraintes aérodynamiques des bâtiments, la simulation des grandes échelles haute fidélité est considérée comme plus appropriée. Néanmoins, il sera démontré que la modélisation RANS piloté par les données peut également atteindre un niveau de précision satisfaisant du champ de pression, en fonction du nombre et du positionnement des données haute-fidélité utilisées comme observation.

Les résultats des calculs d'AD sont comparés aux données disponibles. Les comparaisons incluront les résultats obtenus à partir des simulations antérieures (modèles RANS $\mathcal{K} - \varepsilon$ standard et $\mathcal{K} - \omega$ SST), les résultats expérimentaux moyennés dans le temps, une simulation LES de référence validée [61], les calculs RANS utilisant les modèles optimisés obtenus par Ben-Ali et al. [8] et Zhao et al. [154]. Ben-Ali et al. ont par ailleurs développé un certain nombre de stratégies d'AD plus complexes qui englobent l'optimisation des constantes globales mais aussi le développement de termes de correction pour les équations dynamiques de la turbulence. Pour la présente étude, il a été décidé d'utiliser le plus simple et moins précis de leurs modèles, à savoir celui ciblant uniquement l'optimisation des constantes de modèle. Cette décision a été prise pour fournir une base d'évaluation cohérente entre les résultats actuels et les conclusions fournies par Zhao et al. Les résultats obtenus en utilisant le modèle de Ben-Ali et al. étant presque identiques à la simulation $\mathcal{K} - \varepsilon$ *a priori*, elle ne sera pas montrée.

5.4.1 Champ de vitesse

La vitesse est une variable explicite dans les solveurs ségrégués incompressibles. Par conséquent, la performance des stratégies d'assimilation de données peut être évaluée par l'amélioration qualitative obtenue pour la prédiction de cette quantité.

La Fig. 5.8 montre la comparaison du profil de vitesse dans le sens de l'écoulement u_x et du profil de vitesse verticale u_z pour plusieurs emplacements correspondant aux positions du fil chaud, pour les analyses d'AD utilisant les modèles $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ SST. La comparaison comprend les versions *a priori* et optimisées des modèles RANS, ainsi que les résultats expérimentaux utilisés comme observation. Pour les deux modèles, la précision du champ u_x prédit via AD est sensiblement améliorée pour chaque emplacement. Des différences très mineures par rapport aux essais en soufflerie peuvent être observées, ce qui est conforme au niveau de confiance qui a été prescrit dans le bruitage de l'observation. Par ailleurs, la prédiction de la vitesse verticale u_z est très similaire à la valeur *a priori*. Cela n'est pas surprenant, étant donné que pour cette variable, les résultats *a priori* et les données expérimentales sont également similaires et du même ordre de grandeur que la confiance donnée à l'observation. Un résultat intéressant peut être observé pour le point 20, où la différence est maximale entre la

valeur *a priori* et expérimentale pour u_z . La prédiction se rapproche des valeurs expérimentales, confirmant que l'EnKF est capable de fournir une prédiction statistiquement plus précise de l'écoulement, dans la limite de confiance donnée aux différentes sources d'information. L'amélioration de la prédiction est encore plus nette dans le cas du modèle $\mathcal{K} - \omega$ SST optimisé. Cela est en particulier visible dans la zone proche du toit pour les points 20 et 36. La précision de la prédiction de la vitesse verticale u_z est également significativement améliorée. Par exemple, le profil du modèle $\mathcal{K} - \omega$ SST *a priori* au-dessus du point 36 montre une forte composante normale de vitesse, tandis que le modèle optimisé par AD et les données expérimentales montrent une prédiction de l'écoulement quasi-parallèle au toit.

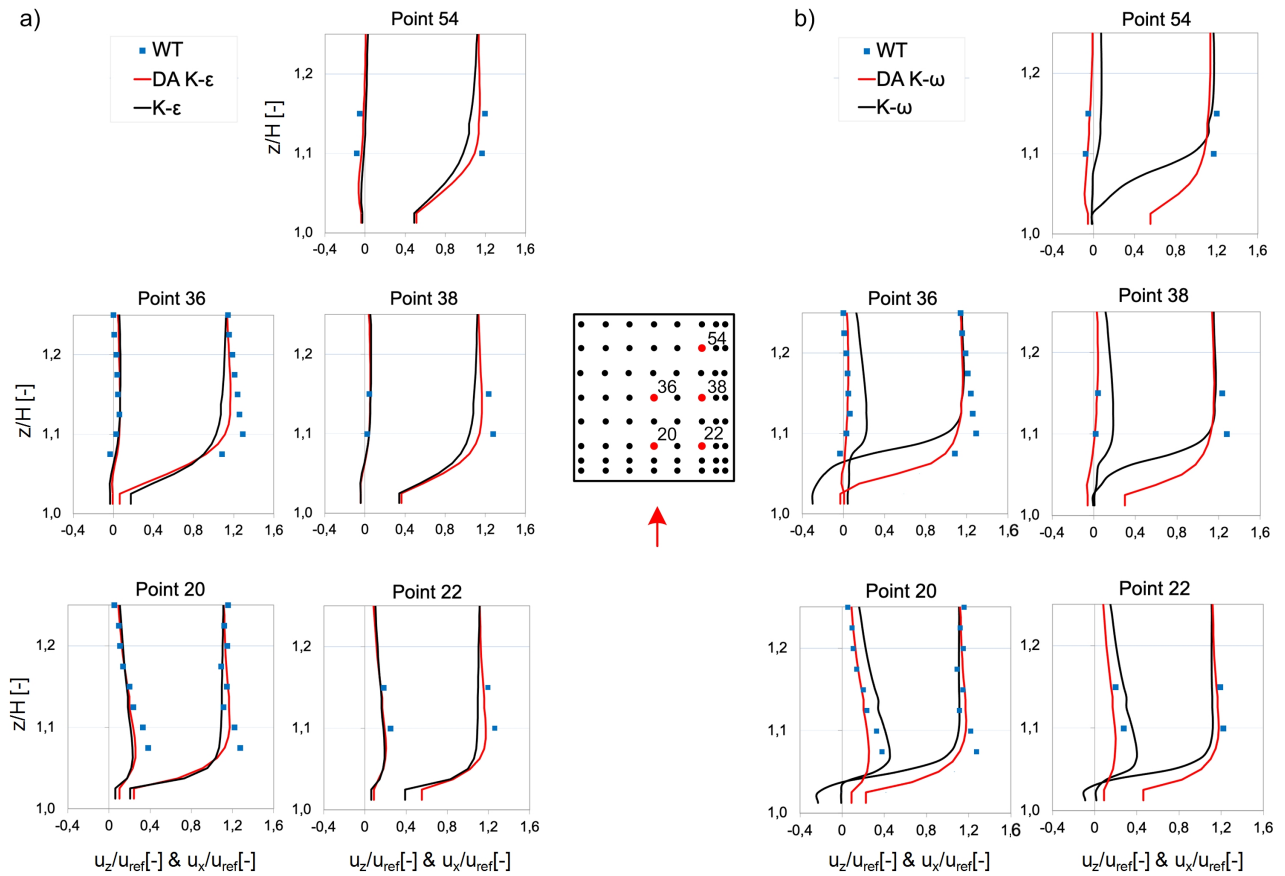


FIGURE 5.8 : Profils de vitesse verticale et dans le sens de l'écoulement au-dessus des emplacements marqués en rouge sur le toit (Points : 20, 22, 36, 38, 54) : a) comparaison entre les données de soufflerie (WT), le modèle $\mathcal{K} - \varepsilon$ et le modèle $\mathcal{K} - \varepsilon$ augmenté par AD, b) comparaison entre les données de soufflerie (WT), le modèle $\mathcal{K} - \omega$ et le modèle $\mathcal{K} - \omega$ augmenté par AD. Ici, u_{ref} représente la vitesse obtenue à 1 m en amont du bâtiment à la hauteur H dans chaque cas test ($u_{ref} \approx 13$ m/s dans les simulations et $u_{ref} \approx 16$ m/s dans la soufflerie).

Les caractéristiques du champ de vitesse sont évaluées plus en détail dans la Fig. 5.9, où les lignes de courant sur un plan vertical $x - z$ au centre du bâtiment sont représentées. Ici, les simulations *a priori* et augmentées sont comparées à une simulation LES de référence, ainsi qu'aux résultats obtenus à partir du modèle $\mathcal{K} - \varepsilon$ optimisé par Zhao et al. [154]. Une augmentation

qualitative de la vitesse est visible juste au-dessus de la séparation du toit pour les cas assimilés par rapport aux simulations RANS *a priori*. Ce résultat, plus proche de l'écoulement prédit par la LES de référence, est associé à la prédiction améliorée de l'écoulement observé sur la Fig. 5.8. Le comportement de la bulle de recirculation derrière le bâtiment est maintenant étudié. Dans cette région, les simulations par AD effectuées ne considèrent que des observations de pression réparties sur deux couronnes, positionnées dans la moitié supérieure du bâtiment, comme le montre la Fig. 5.2 c). Cependant, la taille de cette région est également fortement affectée par les caractéristiques de l'écoulement au sommet du bâtiment. Une réduction significative de la bulle de recirculation derrière le bâtiment peut être observée. Elle est le fruit de l'effet combiné des informations de vitesse et de pression disponibles au niveau des capteurs. Les simulations RANS *a priori* surestiment la taille de la bulle de recirculation, en particulier pour les résultats obtenus avec le modèle $\mathcal{K} - \omega$ SST, où aucun rattachement de l'écoulement sur le toit n'est observé. Les simulations pilotées par les données utilisant des observations hétérogènes actuelles et le modèle $\mathcal{K} - \varepsilon$ de Zhao et al. présentent dans ce cas une réduction significative, même en comparaison avec le LES. La meilleure représentation globale de la région de recirculation est fournie par le modèle SST $\mathcal{K} - \omega$ augmenté, qui présente une précision convaincante en particulier près du sommet du bâtiment. Cependant, son comportement vers la mi-hauteur du bâtiment n'est pas en accord avec le LES de référence. Afin d'améliorer encore la précision de la prédiction, il pourrait être envisagé d'utiliser conjointement différents modèles de turbulence, qui pourraient être ajustés spatialement de manière locale. Des travaux allant dans ce sens ont récemment été proposés par Cherroud et al. ([16]) dans le cadre de la reconstruction par apprentissage automatique de modèles de turbulence. Ce type de stratégie peut être intégré dans l'EnKF assez simplement, mais nécessiterait des ressources de calcul accrues.

Un zoom de l'écoulement, exposé sur la Fig. 5.10, fournit des informations supplémentaires sur la prédiction de la région de recirculation au-dessus du toit du bâtiment. En particulier, le point de rattachement du modèle RANS $\mathcal{K} - \varepsilon$ standard, représenté par une flèche rouge, est bien plus en amont que la valeur prédite par la LES de référence, représentée par la flèche verte. D'autre part, le modèle $\mathcal{K} - \omega$ SST ne présente aucun rattachement de la couche limite, ce qui est responsable de la grosse bulle de recirculation se propageant derrière le bâtiment. Les simulations par AD présentées améliorent significativement la prédiction de la longueur de la bulle de recirculation du toit. D'autre part, le modèle piloté par les données de Zhao et al. semble obtenir un rattachement presque instantané de la couche limite sur le toit.

Enfin, le champ d'énergie cinétique turbulente, normalisé par le carré de la fluctuation de vitesse caractéristique u'_{ref} , est représenté sur la Fig. 5.11. Les deux modèles RANS utilisés *a priori* ne parviennent pas à fournir la quantité adéquate de \mathcal{K} dans la grande région de recirculation derrière le bâtiment. Le modèle $\mathcal{K} - \varepsilon$ a également tendance à surestimer l'énergie cinétique turbulente devant le bâtiment. Les simulations augmentées par AD proposées fournissent une prédiction de \mathcal{K} significativement meilleure. Le modèle piloté par les données de Zhao et al. a également de manière générale amélioré la prédiction de cette quantité, mais des valeurs élevées pour \mathcal{K} sont observées devant le bâtiment. Cette surestimation est l'un des facteurs responsables du rattachement instantané de la bulle de recirculation observé sur le toit.

En résumé, la modélisation RANS augmentée, pilotée par les données et basée sur des observations hétérogènes, s'est avérée fournir une prédiction significativement améliorée. Bien que

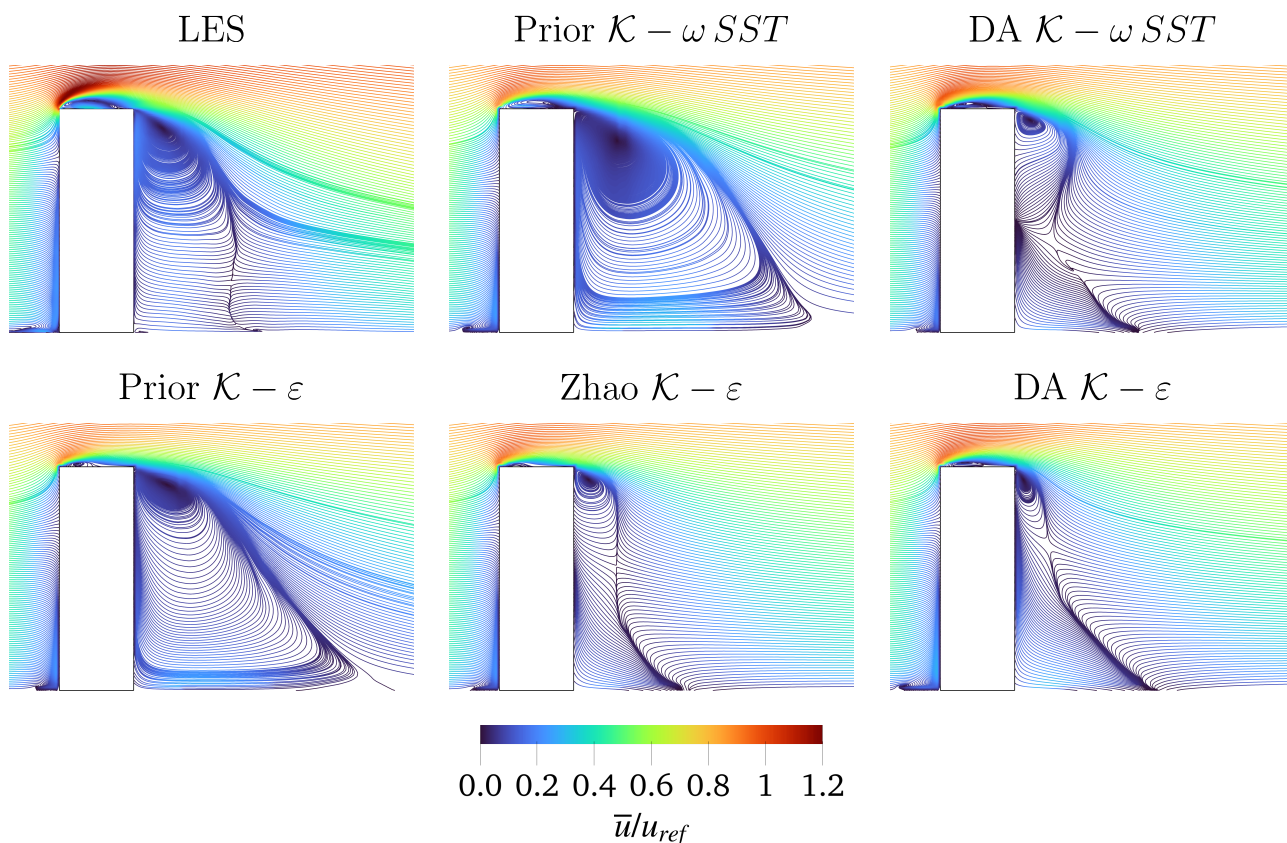


FIGURE 5.9 : Lignes de courant colorées par le champ de vitesse moyenné dans le temps obtenu sur un plan $x - z$ au centre du bâtiment. Des comparaisons sont effectuées entre une simulation des grandes échelles (LES validée) [142], un modèle $\mathcal{K} - \omega$ SST standard et augmenté par AD, $\mathcal{K} - \varepsilon$ optimisé par Zhao et al. [154], $\mathcal{K} - \varepsilon$ standard et augmenté par AD. Ici, u_{ref} représente la vitesse obtenue 1 m en amont du bâtiment à la hauteur H dans la soufflerie.

des différences mineures dans les résultats aient été observées entre la version par AD actuelle du modèle $\mathcal{K} - \varepsilon$ et le modèle de Zhao et al., ce dernier impliquait un nombre significativement plus important de capteurs, qui étaient répartis de manière plus homogène dans le domaine physique. Par conséquent, les présents résultats soulignent l'importance de la qualité de l'information en termes de localisation des capteurs, au contraire de la quantité d'observations s'avérant être un facteur secondaire si le placement des capteurs est efficace.

5.4.2 Champ de pression

Le comportement du champ de pression est maintenant étudié. Cette quantité physique est nettement plus difficile à prédire pour la simulation numérique incompressible car l'équation de Poisson utilisée par le solveur CFD utilise le champ de pression simultanément comme variable physique et comme contrainte pour garantir l'incompressibilité de l'écoulement. Par conséquent, l'analyse de cette quantité est cruciale pour évaluer la stabilité et la précision des algorithmes. Le coefficient de pression moyen est défini comme $C_p \left(C_p = \frac{\bar{p} - p_{ref}}{0.5 \rho u_{ref}^2} \right)$, où p_{ref} , ρ et

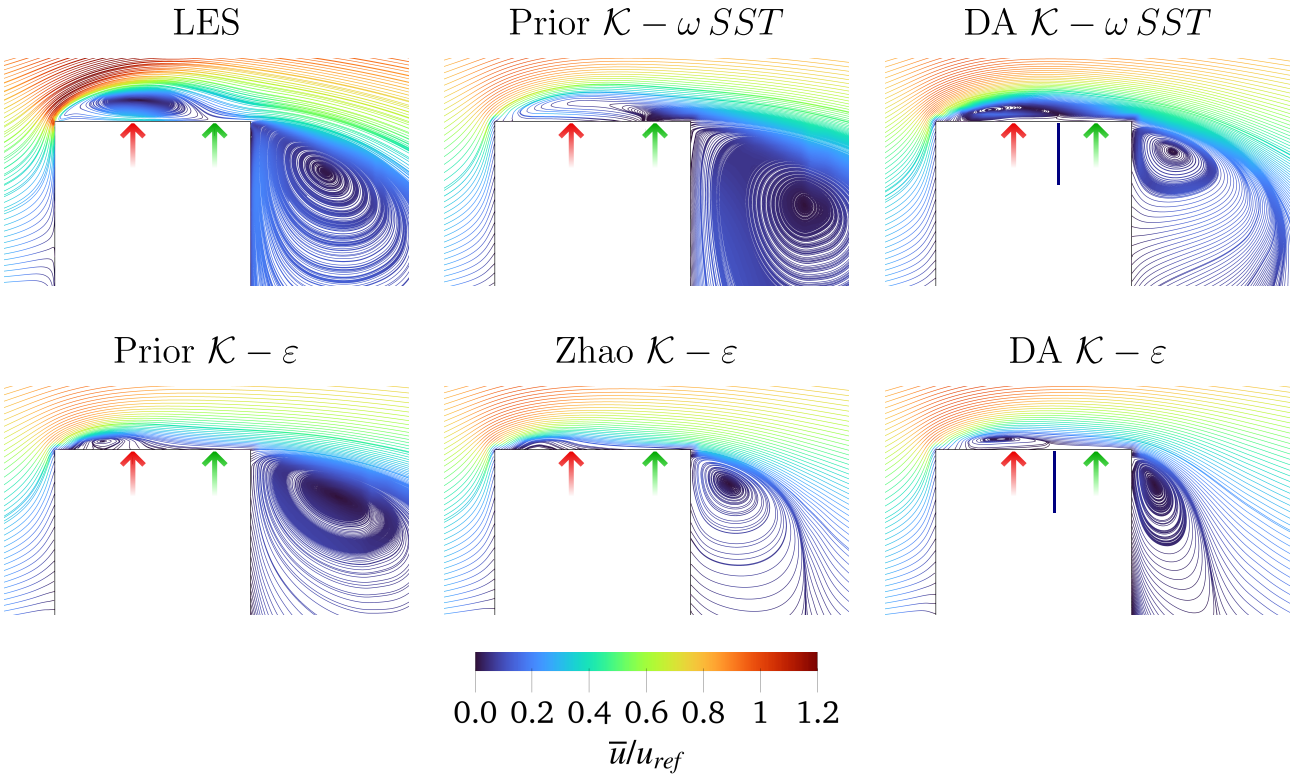


FIGURE 5.10 : Zoom sur les lignes de courant au-dessus du toit. La flèche rouge indique la position de rattachement de la bulle de recirculation pour le modèle $\mathcal{K} - \varepsilon$ standard et la flèche verte fournit la même information pour la simulation LES validée. La ligne bleue indique la position de rattachement pour la simulation présentée dans chaque image. Ici u_{ref} représente la vitesse obtenue 1 m en amont du bâtiment à la hauteur H dans la soufflerie.

u_{ref} sont respectivement la pression de l'écoulement à l'infini, la densité de l'air et la vitesse de référence (calculée à 1 m en amont du bâtiment à la hauteur H dans chaque simulation). Dans la Fig. 5.12, le coefficient de pression moyen est présenté en termes de comparaison des mesures de performance avec les données expérimentales. De cette façon, les différences locales entre chaque simulation numérique et les données expérimentales sont regroupées en groupes ne dépassant pas un certain seuil d'erreur (10%, 20% et 30% dans ce cas). L'analyse de ce critère pour le seuil d'erreur 10% peut conduire à tort à la conclusion que les simulations *a priori* (par exemple, 23% des occurrences pour le modèle $\mathcal{K} - \omega SST$) se comportent mieux que les simulations augmentées par AD (16% des occurrences). Cette information est cependant trompeuse, car un grand nombre d'occurrences pour les simulations avec AD se situent juste en dehors de cet intervalle. En fait, lorsque les seuils d'erreur plus élevés sont pris en compte, les simulations avec AD surpassent les RANS *a priori*. Pour un seuil d'erreur de 20%, une amélioration allant jusqu'à 17% des occurrences est observée avec l'utilisation de l'AD. L'amélioration se situe entre 10% et 30% lorsqu'un seuil d'erreur de 30% est considéré. Pour ce dernier seuil, le modèle $\mathcal{K} - \omega SST$ augmenté par AD atteint plus de 70% ce qui certifie la précision globale de la méthode. Les résultats sont également représentés sous forme d'histogramme dans la Fig. 5.13. Cette représentation confirme que les résultats obtenus par les simulations par AD ont

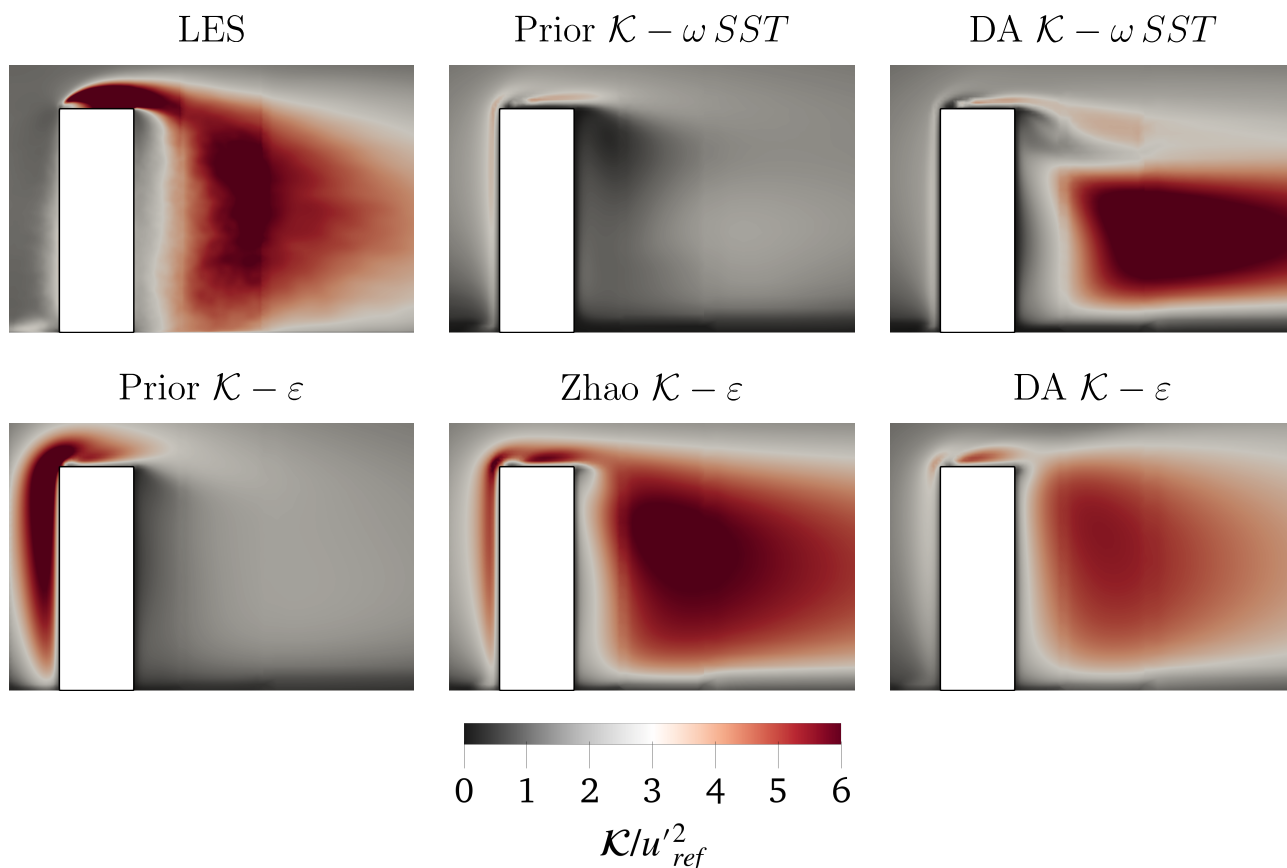


FIGURE 5.11 : Énergie cinétique turbulente \mathcal{K} obtenue sur un plan $x - z$ au centre du bâtiment. Des comparaisons sont effectuées entre une simulation des grandes échelles (LES validée) [142], un modèle $\mathcal{K} - \omega$ SST standard et augmenté par AD, $\mathcal{K} - \varepsilon$ optimisé par Zhao et al. [154], $\mathcal{K} - \varepsilon$ standard et augmenté par AD. Ici, $u'_{ref} = I_u u_{ref}$ est obtenu à partir de l'intensité turbulente dans le sens de l'écoulement $I_u = 13\%$ et u_{ref} mesurée à 1 m en amont du bâtiment à la hauteur H .

tendance à se regrouper vers des régions d'erreur plus faibles par rapport aux RANS classiques. Cependant, cet aspect mérite une étude beaucoup plus approfondie car les résultats obtenus avec les modèles pilotés par les données de Ben-Ali et al. et Zhao et al., qui ne sont pas présentés ici, montrent une précision dégradée pour le champ de pression prédit. Cela pourrait impliquer que l'observation hétérogène peut être utile pour améliorer la prédiction globale du solveur numérique, probablement en raison de la corrélation entre les variables physiques pilotées par une dynamique non linéaire.

Pour finir, la comparaison du coefficient de pression moyen C_p calculé au niveau des prises de pression sur le toit lors des essais en soufflerie et des calculs RANS est présentée dans la Fig. 5.14. Les résultats pour le modèle $\mathcal{K} - \omega$ SST indiquent à nouveau une amélioration significative de la prédiction. La version RANS utilisée comme *a priori* présente une valeur quasi-constante de C_p pour toute la longueur x/B , car la couche limite ne se rattache pas sur le toit. D'autre part, le modèle $\mathcal{K} - \omega$ SST augmenté par AD présente une prédiction significativement améliorée, plus proche des données expérimentales. En effet, ces deux distributions présentent une forme de « bosse », typique d'un écoulement avec une région séparée suivie d'un rattachement [49].

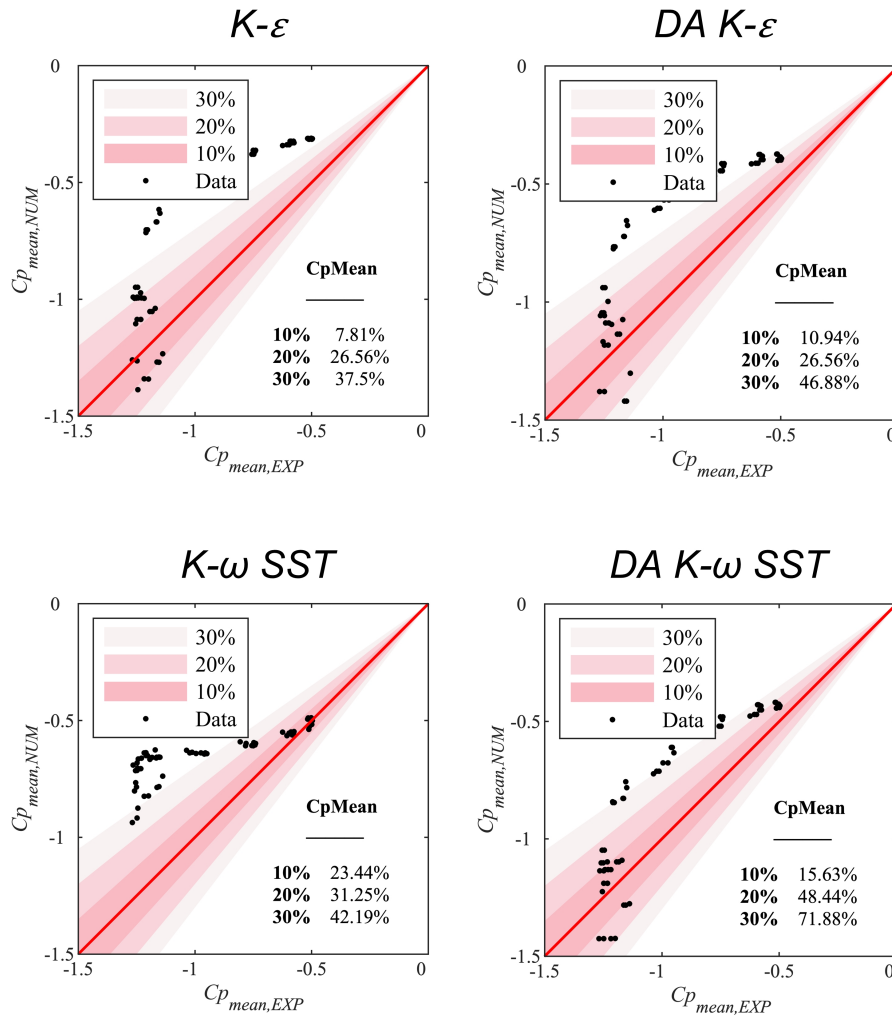


FIGURE 5.12 : Nuage de points du coefficient de pression moyen C_p pour les simulations numériques avec comparaison des mesures de performance avec les données expérimentales. Rangée supérieure : modèle $K - \varepsilon$, rangée inférieure : modèle $K - \omega$ SST. Colonne de gauche : modèle RANS *a priori*, colonne de droite : RANS augmenté par AD.

Cette observation est conforme à la topologie d'écoulement observé au-dessus du toit sur la Fig. 5.10. Pour la simulation avec AD utilisant le modèle $K - \varepsilon$, des améliorations sont également observées, mais celles-ci semblent être mineures par rapport au modèle RANS *a priori*. Par conséquent, l'ampleur de ces améliorations n'est pas aussi importante que pour le calcul basé sur le modèle $K - \omega$ SST.

5.5 Conclusions

CONES a pour la première fois été utilisée pour réaliser une étude basée sur les données de l'écoulement autour d'un bâtiment de grande hauteur. Plus précisément, des échantillons expérimentaux hétérogènes, sous forme de données provenant de prises de pression et de fils chauds,

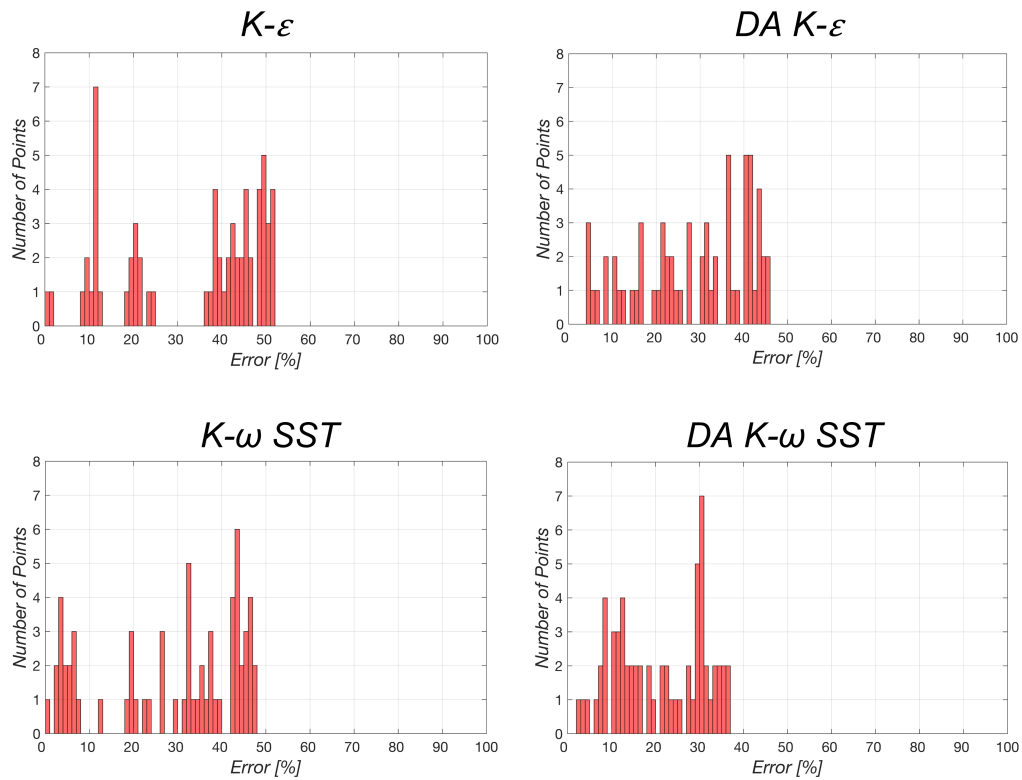


FIGURE 5.13 : Histogrammes représentant l'erreur dans la prédiction du coefficient de pression C_p par rapport aux essais. Les résultats numériques sont échantillonnés aux emplacements des prises de pression. Rangée supérieure : modèle $\mathcal{K} - \varepsilon$, rangée inférieure : modèle $\mathcal{K} - \omega$ SST. Colonne de gauche : modèle RANS *a priori*, colonne de droite : modèle RANS augmenté par AD.

ont été intégrés aux simulations CFD RANS, réalisées à l'aide d'OpenFOAM. Le couplage a été réalisé à l'aide du filtre de Kalman d'ensemble (EnKF), et de techniques avancées telles que la localisation et l'inflation. L'état étendu inféré via l'EnKF a également été utilisé pour améliorer les caractéristiques prédictives du modèle en optimisant les cinq/neuf constantes globales des modèles de turbulence $\mathcal{K} - \varepsilon/\mathcal{K} - \omega$ SST, respectivement, utilisés pour fermer les équations.

Les résultats ont montré une amélioration globale des quantités physiques étudiées. Ces derniers sont également assez équivalents selon les différentes stratégies d'AD. La localisation physique et la localisation de covariance, qui ont été évaluées pour le modèle $\mathcal{K} - \varepsilon$, semblent efficaces pour l'étude des écoulements complexes. La réduction des degrés de liberté du problème d'AD n'a pas affecté la qualité des résultats, tout en réduisant globalement le temps nécessaire aux procédures pilotées par les données. En revanche, l'utilisation de l'inflation n'a pas produit de meilleurs résultats, notamment en raison de l'augmentation des ressources de calcul requises.

L'analyse du champ de vitesse montre que l'EnKF permet de réduire considérablement l'erreur dans le sens de l'écoulement et dans le sens vertical, selon la confiance accordée à l'observation. Les effets de l'inférence paramétrique sont également observés dans la région de recirculation derrière le bâtiment. Dans ce cas, la précision des résultats est affectée par le modèle RANS choisi pour l'optimisation. La topologie physique de l'écoulement devient

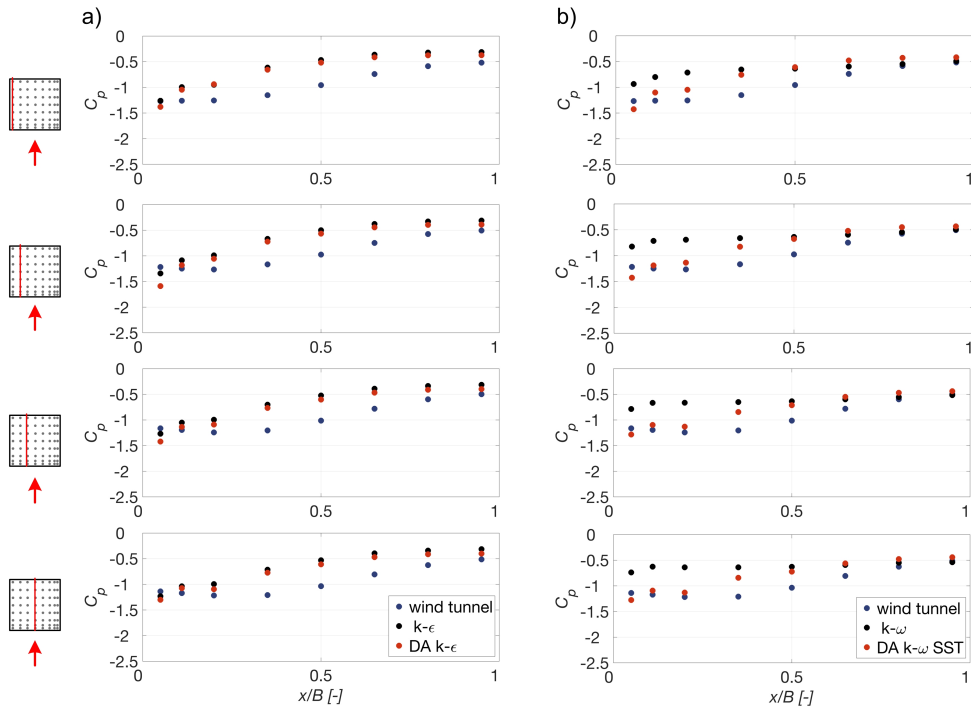


FIGURE 5.14 : Coefficient de pression moyen aux points de prise de pression le long des lignes rouges marquées sur le toit : a) comparaison entre les données de la soufflerie (WT), du modèle $\mathcal{K} - \epsilon$ et du modèle $\mathcal{K} - \epsilon$ augmenté par AD ; b) comparaison entre les données de la soufflerie (WT), du modèle $\mathcal{K} - \omega$ SST et du modèle $\mathcal{K} - \omega$ SST augmenté par AD.

similaire à la LES de référence validée avec les données expérimentales, même si la taille de la bulle de recirculation apparaît trop réduite. Pour le champ de pression, des améliorations sont observées au vu de la diminution globale de l'erreur dans la prédiction du coefficient de pression moyen. Dans la zone du toit, l'amélioration du comportement statistique de la pression est liée à la précision accrue dans l'estimation du rattachement de la bulle de recirculation de cette région. Les résultats présentés ont été obtenus à l'aide d'outils CFD open source accessible à tous les utilisateurs, utilisant une structure ségréguée et une description globale des coefficients contrôlant la modélisation de la turbulence. L'utilisation de solveurs dits couplés, plus sophistiqués, pourrait fournir de meilleurs résultats lorsque utilisés dans des outils d'AD faisant intervenir des données de pression comme observation.

Écoulement incompressible statistiquement stationnaire académique

Sommaire

6.1	Introduction	109
6.2	Cas test et configuration de la procédure d'AD	111
6.2.1	Écoulement turbulent en canal plan, $Re_\tau \approx 550$	111
6.2.2	Stratégie d'assimilation de données	115
6.3	Prédiction des caractéristiques statistiques avec l'AD séquentielle	118
6.4	Synchronisation des champs de l'écoulement	127
6.5	Conclusion	131

6.1 Introduction

Le [chapitre 5](#) a permis de mettre à l'épreuve CONES sur l'inférence d'applications RANS 3D stationnaires. L'outil a montré son potentiel pour la calibration de constantes liées à la fermeture de modèles de turbulence via l'utilisation de l'état étendu du filtre de Kalman d'ensemble. Ce nouveau chapitre se consacre à l'étude de l'inférence d'une simulation des grandes échelles (LES) [\[101, 114\]](#) d'un écoulement interne incompressible. Le régime de l'écoulement est permanent, ce dernier est donc statistiquement stationnaire. Cette analyse représente la deuxième étape du travail mené dans ce manuscrit. L'étape finale sera atteinte dans le [chapitre 7](#) avec l'inférence d'une simulation LES instationnaire compressible.

La LES est sans doute la stratégie la plus étudiée des dernières décennies. L'une de ses caractéristiques les plus intéressantes est sa capacité à représenter naturellement les caracté-

ristiques tridimensionnelles instationnaires de l'écoulement. Cette propriété clé, qui n'est pas obtenue par la plupart des fermetures utilisées pour simuler la turbulence, est essentielle pour la prédiction des écoulements statistiquement stationnaires [64], des écoulements périodiques [31, 136] et des écoulements fortement accélérés, qui sont statistiquement instationnaires [71]. Pour toutes ces classes d'écoulements, la LES a le potentiel de prédire avec précision des interactions écoulement/paroi complexes qui peuvent évoluer dans l'espace et le temps comme celles trouvées dans les moteurs à combustion interne [29, 30]. De plus, la LES peut capturer des événements rares localisés dans le temps qui affectent l'organisation structurale de l'écoulement et ses propriétés. Un exemple de tels événements rares qui ont été largement étudiés ces dernières années est la prédiction des orages dans les écoulements externes pour l'ingénierie environnementale / urbaine [120, 121].

La représentation des caractéristiques instantanées de l'écoulement présente également un grand potentiel pour les applications LES dans le cadre du paradigme du jumeau numérique. La plupart des études dans la littérature sur la mécanique des fluides couplent le système physique à des modèles d'ordre réduit [107, 131, 17, 85, 3, 68]. Ainsi la communication et le contrôle sont limités aux macro-caractéristiques statistiques de l'écoulement. Les applications LES dans ce contexte sont potentiellement révolutionnaires car le couplage en temps réel d'un système réel avec la LES est cohérent en termes de représentation physique. La mise en œuvre réussie d'un jumeau numérique basé sur la LES pourrait anticiper les événements extrêmes via la simulation numérique et empêcher les événements catastrophiques pour le jumeau physique. Mons et al. [87] ont réalisé une optimisation avancée du modèle de Smagorinsky [119], l'une des fermetures de sous-maillages les plus utilisées dans la littérature, pour le cas test de l'écoulement en canal plan. Dans leur travail, la procédure d'AD s'appuie sur les caractéristiques statistiques de l'écoulement pour l'optimisation. Bien que les résultats obtenus augmentent considérablement la précision globale du solveur LES, cette procédure n'est pas adaptée à l'optimisation à la volée dans le cadre d'un jumeau numérique. Un certain nombre de travaux d'AD ont également ciblé la synchronisation numérique et la reconstruction d'écoulements instantanés turbulents à partir de données limitées (voir chapitre [chapitre 3](#)).

Ce chapitre propose une analyse approfondie de l'application de l'EnKF à la LES en termes de (1) optimisation du modèle de sous-maillages et (2) augmentation de l'état pour la synchronisation de l'écoulement. Le cas test de l'étude est l'écoulement turbulent en canal, qui a déjà été analysé à l'aide de techniques d'AD [87, 145]. Le point de nouveauté ici est que l'optimisation et l'augmentation d'état sont effectuées à la volée, informant progressivement les membres de l'ensemble (simulations LES) avec des données DNS résolues dans le temps qui sont échantillonnées sur un nombre limité de capteurs près de la paroi. L'objectif est d'évaluer la robustesse de la procédure, à la fois en termes d'optimisation et de reconstruction de l'écoulement, lorsque des données éparses en temps et en espace sont utilisées. Le couplage à la volée des simulations LES et des données DNS est effectué via CONES. Le chapitre est structuré de la manière suivante. Dans la section 6.2, le cas d'étude et la configuration de l'assimilation de données seront introduits. Dans la section 6.3, les résultats de l'optimisation du modèle de sous-maillages sont discutés. Dans la section 6.4, l'impact global de la méthodologie d'AD sur l'écoulement instantané prédit et la corrélation avec les données DNS disponibles sont étudiés. Enfin, la section 6.5 conclut ce chapitre.

6.2 Cas test et configuration de la procédure d'AD

6.2.1 Écoulement turbulent en canal plan, $Re_\tau \approx 550$

Le cas test choisi pour réaliser la procédure d'AD est l'écoulement turbulent en canal plan pour $Re_\tau = u_\tau h / \nu = 536$. Ici $u_\tau = \sqrt{\tau_w / \rho}$ est la vitesse de frottement et τ_w est la contrainte de frottement à la paroi. h est la demi-hauteur du canal et ν est la viscosité cinématique. Ce cas test académique, qui est piloté par des mécanismes de cisaillement à la paroi et exclut naturellement les aspects complexes associés aux gradients de pression moyens favorables/défavorables [101], est néanmoins un défi pour la LES [79]. Une interaction non linéaire complexe se produit entre deux sources d'erreur principales, à savoir celles associées à la discrétisation numérique et à la fermeture de sous-maillages. Ces mécanismes sont responsables d'une très grande sensibilité de l'écoulement à des variations relativement faibles de la discrétisation du maillage et du modèle de sous-maillages sélectionné. Par conséquent, ce cas test est un excellent candidat pour étudier les objectifs présentés dans l'introduction. Pour rappel, le but de ce cas d'étude n'est pas de débattre du potentiel de la LES à prédire avec précision cette configuration d'écoulement, car il a été montré dans la littérature qu'une très grande sensibilité des résultats à la configuration numérique est observée [79]. Les choix de résolution de maillage effectués ici fournissent une précision relativement faible de la simulation *a priori*, comme cela sera montré par la suite. Le but de la présente étude est de vérifier si l'AD séquentielle utilisant des mesures instantanées de vitesses peut optimiser les modèles conventionnels de la littérature et si la reconstruction de l'état peut se synchroniser avec les données obtenues aux capteurs. Ce dernier point représente un élément essentiel dans la communication efficace entre un jumeau physique et son homologue numérique et sera abordé dans la Sec. 6.4. Les résultats obtenus à partir des simulations des grandes échelles réalisées dans ce travail seront comparés aux données DNS sur le même cas test précédemment réalisé par l'équipe de recherche [72].

Les caractéristiques géométriques sont présentées sur la Fig. 6.1. La taille du domaine étudié est $3\pi h \times 2h \times \pi h$. x est la direction de l'écoulement, y la direction normale et z la direction de l'envergure. Une condition de non-glissement est imposée sur les parois supérieure et inférieure. Une condition limite périodique est appliquée sur les quatre côtés latéraux. Un terme source, déjà intégré dans le solveur d'OpenFOAM, est inclus dans les équations dynamiques pour préserver le débit massique global dans le temps. Plus précisément, le terme source vise la conservation de la vitesse moyenne $u_b = \iiint_{V_D} u_x dV' / V_D$, où V_D est le volume du domaine physique étudié. La valeur cible utilisée pour toutes les simulations est $u_b = 0.899u_c$, où u_c est la vitesse moyenne au centre du canal obtenue par la DNS. La viscosité cinématique ν est la même pour les calculs DNS et LES. Le nombre de Reynolds volumique obtenu par la DNS est égal à $Re = 2hu_b / \nu = 20124$.

Deux LES standards sont réalisées en utilisant le modèle sous-maillages bien connu Smagorinsky [119] (voir chapitre 2), avec et sans l'inclusion de la correction de Van Driest à la paroi. Elles sont respectivement appelées VD-LES (LES Smagorinsky avec Van Driest) et BL-LES (LES Smagorinsky dite *baseline* sans Van Driest). Cette fermeture de sous-maillages a été choisie pour effectuer une comparaison cohérente avec le travail de Mons et al. [87], qui ont optimisé ce modèle en utilisant une procédure d'AD non séquentielle s'appuyant sur l'observation de quantités statistiques. Les simulations sont exécutées par le solveur pimpleFoam d'OpenFOAM

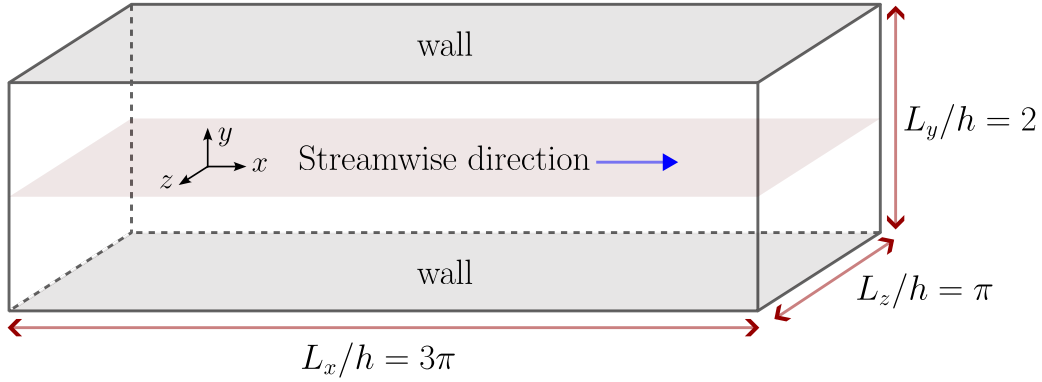


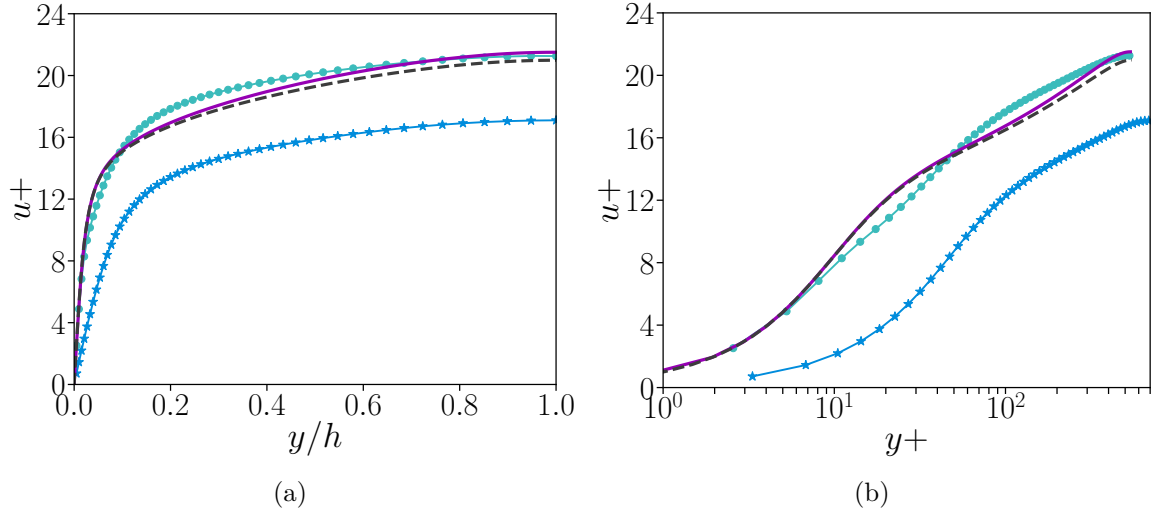
FIGURE 6.1 : Taille du domaine d'étude.

utilisant l'algorithme PIMPLE (voir Sec. 2.2). Un schéma implicite du second ordre a été utilisé pour l'avancement temporel de la solution et des schémas implicites centrés du second ordre ont été utilisés pour la discrétisation des dérivées spatiales. Le maillage est composé de 350 000 cellules, dont les détails sont rapportés dans le Tab. 6.1 aux côtés de ceux de la DNS de référence. La taille des éléments de maillage est adimensionnée par rapport à l'échelle visqueuse à la paroi $\delta_\nu = \nu/u_\tau$. L'exposant \star est utilisé lorsque les normalisations sont effectuées en utilisant le u_τ calculé par la DNS. L'exposant $+$ est lui utilisé lorsque u_τ est obtenu à partir de chaque simulation LES. Δx^\star et Δz^\star sont obtenus en utilisant une distribution uniforme. Une expansion géométrique est utilisée pour contrôler la taille des éléments Δy dans la direction normale pour garantir une résolution plus élevée proche de la paroi. La taille du plus petit élément Δy_1^\star au niveau de la paroi et du plus grand élément Δy_c^\star au niveau de la ligne centrale du canal sont reportées. La taille des éléments de maillage utilisés pour le calcul des simulations LES de base est légèrement supérieure aux valeurs typiques observées pour des simulations LES dans ce cas, qui sont estimées à $\Delta x^\star \approx 50$, $\Delta y_1^\star \approx 1$ et $\Delta z^\star \approx 20$ [87]. Ce choix a été fait afin de (1) évaluer les capacités de la méthode d'AD à fournir une estimation de l'état précise et une inférence paramétrique même dans des conditions sous-résolues et (2) obtenir des exécutions plus rapides de l'algorithme d'AD en utilisant un ensemble de simulations suffisamment grand. Les conditions initiales pour les simulations BL-LES et VD-LES ont été définies à l'aide d'un champ interpolé à partir d'une solution DNS. Les simulations ont tout d'abord calculé pour une durée de 50 temps d'advectifs t_A , calculés tel que $t_A = h/u_c$, afin de dissiper le champ initial. Ensuite, les statistiques moyennes de premier et second ordres ont été calculées sur une fenêtre temporelle de $900t_A$. Le pas de temps pour l'avancement temporel de la solution est constant et égal à $\Delta t = 0.02t_A$. Le temps associé à la diffusion de quantité de mouvement via la viscosité sur une cellule est $t_d = \Delta_x^2/\nu \approx 200$. Le nombre de Fourier associé est égal à $FO = \Delta_t/t_d \approx 10^{-4} \ll 1$. Ainsi, le pas de temps choisi est suffisamment petit pour capturer la diffusion de quantité de mouvement sur une cellule.

Les résultats des simulations BL-LES et VD-LES sont maintenant comparés aux résultats DNS et aux résultats d'une DNS de référence supplémentaire disponible dans la littérature pour un nombre de Reynolds $Re_\tau = 550$ [25]. La Fig. 6.2 montre le profil de vitesse moyen normalisé dans le sens de l'écoulement $u^+ = \overline{u_x}/u_\tau$. Les moyennes (indiquées par le symbole $\bar{\cdot}$) sont effectuées dans le temps ainsi que dans le sens de l'écoulement et dans le sens de l'envergure,

Type	L_x	L_y	L_z	N_x	N_y	N_z	Δx^*	Δy_1^*	Δy_c^*	Δz^*	Cells
LES	3π	h	π	70	100	50	73	2.6	27.8	34	3.5×10^5
DNS	6π	h	2π	1024	256	512	9.9	0.96	11.2	6.6	13.4×10^7

TABLE 6.1 : Détail des maillages utilisés pour les simulations LES ainsi que pour la DNS de référence.


 FIGURE 6.2 : Profil de vitesse moyenné dans le sens de l'écoulement u^+ pour les simulations de base Smagorinsky LES (★), avec van Driest (●) et DNS (—) comparé aux données de la littérature (—).

afin d'obtenir une meilleure convergence statistique. L'écart entre la prédiction BL-LES et les résultats DNS est significatif. L'un des éléments clés affectant ce manque de précision est la prédiction erronée de la contrainte de cisaillement à la paroi τ_w et donc de la vitesse de frottement u_τ résultant en $Re_\tau = 686$ pour la simulation BL-LES. Pour ce paramètre, un écart de 28% avec les résultats DNS est observé. La principale source d'erreur pour la prédiction de cette quantité est liée à la fermeture de sous-maillages utilisée, pour laquelle ν_{SGS} ne tend pas vers zéro à l'approche de la paroi, comme le montre la Fig. 6.3. La simulation VD-LES, pour laquelle une correction de paroi est obtenue via la fonction de van Driest, donne $Re_\tau = 530$. Pour cette simulation, ν_{SGS} tend vers zéro à la paroi comme prévu. Un écart important avec les données DNS est également observé pour la prédiction des composantes du tenseur des contraintes de Reynolds résolu $\overline{u'_i u'_j}$. Les quantités $u'_x u'_x{}^+ = \overline{u'_x u'_x} / u_\tau^2$, $u'_y u'_y{}^+ = \overline{u'_y u'_y} / u_\tau^2$, $u'_z u'_z{}^+ = \overline{u'_z u'_z} / u_\tau^2$ et $u'_x u'_y{}^+ = \overline{u'_x u'_y} / u_\tau^2$ sont présentées sur la Fig. 6.4. Ni l'amplitude ni la position du pic ne sont prédites avec précision par les simulations BL-LES et VD-LES, cette dernière montrant néanmoins de meilleurs résultats.

Les résultats obtenus via les simulations LES indiquent donc qu'en utilisant la configuration numérique décrite avec le modèle de sous-maillages Smagorinsky, une prédiction précise des moments statistiques du champ n'est pas obtenue. Dans la sous-section 6.2.2, la procédure d'AD utilisée pour améliorer la prédiction de l'écoulement en utilisant cette configuration LES est détaillée.

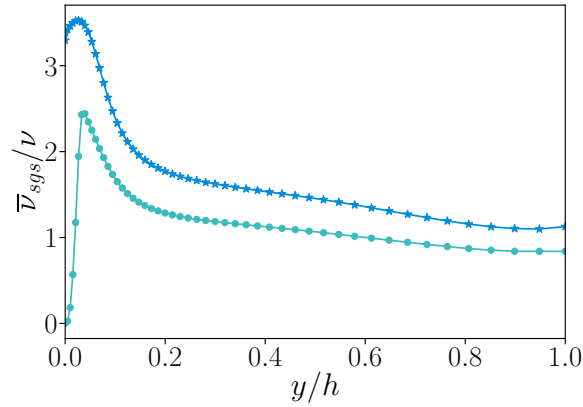


FIGURE 6.3 : Distribution du rapport $\bar{\nu}_{SGS}/\nu$ dans la direction verticale pour les LES avec (●) et sans van Driest (*).

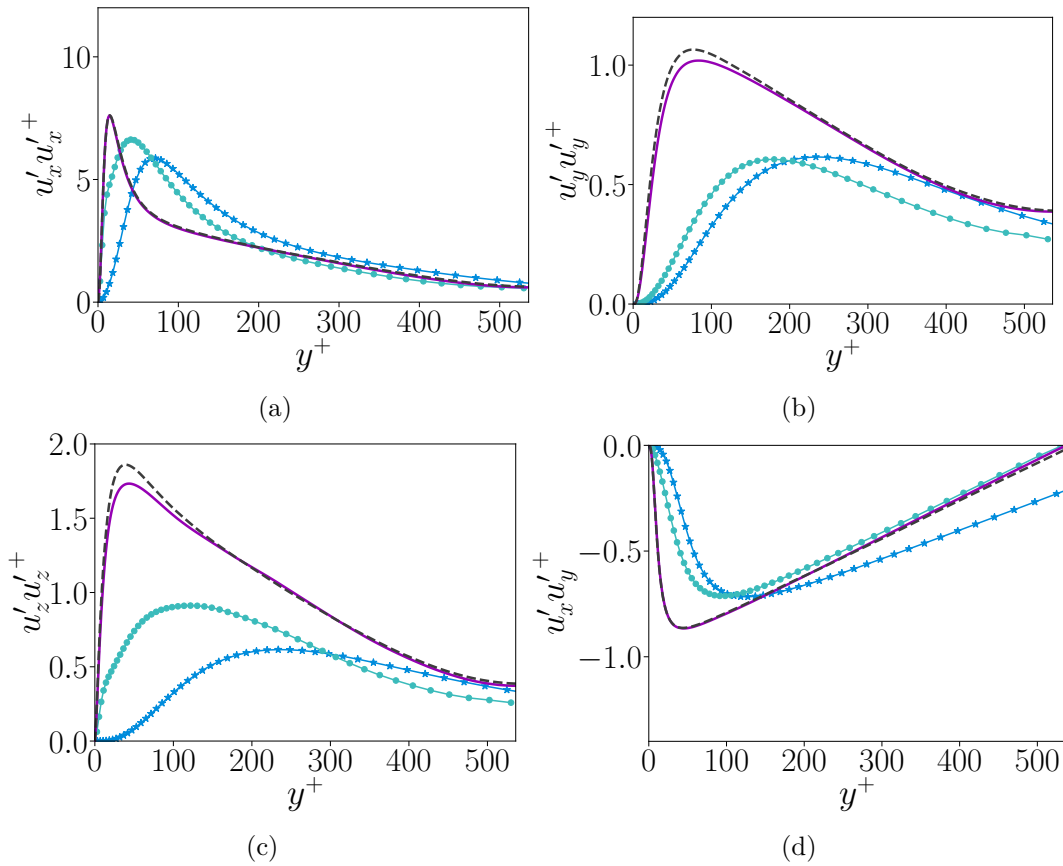


FIGURE 6.4 : Composantes du tenseur de Reynolds pour les simulations de base Smagorinsky LES (★), avec van Driest (●) et DNS (-) comparé aux données de la littérature (---).

6.2.2 Stratégie d'assimilation de données

Les simulations d'AD réalisées dans ce chapitre visent à fournir des états augmentés instantanés du cas test étudié. Cet objectif sera atteint en couplant à la volée la prédiction numérique du solveur LES avec des informations localisées échantillonnées à partir de la DNS de référence. Comme discuté dans l'introduction, l'étude vise à évaluer si (1) la procédure d'AD est capable d'obtenir une optimisation satisfaisante du modèle LES utilisé et (2) si la reconstruction de l'état est efficace pour synchroniser le modèle et l'observation, à proximité des capteurs et loin de ceux-ci. Pour le premier point, la configuration des simulations numériques a été choisie pour effectuer une comparaison étroite avec le travail de Mons et al. [87], qui utilisaient une stratégie d'AD non séquentielle basée sur des ensembles avec observation des moments statistiques du champ de vitesse. Dans le présent travail, les capteurs capturent le champ de vitesse instantané, ce qui conduit à un problème d'optimisation plus complexe en raison des caractéristiques turbulentes de l'écoulement. Il est donc intéressant de vérifier si les deux stratégies conduisent à des résultats similaires. La configuration LES présentée dans la Sec. 6.2.1 est inférée via des observations haute-fidélité synthétiques. Des échantillons résolus dans le temps du champ de vitesse instantané de la DNS de référence sont utilisés à cette fin. Les échantillons sont collectés sur 10 800 capteurs dans le domaine physique situés sur l'intervalle $0.48 \leq y^+ \leq 56.4$ c'est-à-dire dans la sous-couche visqueuse, dans la région tampon et dans la plage inertielle. L'échantillonnage dans le temps est effectué à un intervalle constant de $\Delta t_{DA} = 0.04t_A$.

La mise en place de la procédure de l'EnKF est maintenant détaillée. La taille $[N_{ext}, N_e]$ de la matrice d'état \mathbf{U} est donnée par $N_e = 40$ (nombre de membres de l'ensemble) et $N_{ext} = N + N_\theta$. Ici N_θ est le nombre de paramètres optimisés par la procédure de l'EnKF et il est différent pour les trois procédures pilotées par les données qui seront présentées par la suite. $N = 3n_{cells}$ est égal à trois fois le nombre de mailles utilisés dans la procédure d'AD. En effet, le nombre de degrés de liberté considérés dans la procédure d'AD correspond aux trois composantes du champ de vitesse pour chacun des éléments de maillage n_{cells} . La valeur de n_{cells} est strictement liée à la localisation physique réalisée en découpant le domaine numérique analysé. Cette procédure, illustrée sur la Fig. 6.5, consiste à exclure du calcul de l'AD les éléments de maillage compris dans l'intervalle $0.18 < y/h < 1.82$. Ces éléments sont relativement éloignés des capteurs et donc le risque de corrélation parasite affectant la stabilité de l'algorithme d'AD est élevé. Par ailleurs, le domaine exclu représente environ 56% du nombre total de cellules utilisées par le modèle LES. Le gain en termes de coût du calcul pour le gain de Kalman est également réduit d'environ 56% comme indiqué dans le Tab. 4.4. La localisation de covariance est également appliquée au calcul afin que les discontinuités associées à la mise à jour des éléments contenus dans le découpage physique soient lissées. La structure de la matrice L utilisée pour la localisation de la covariance est celle présentée dans l'équation 3.65, où le paramètre $l = 0.175$ dans les directions de l'écoulement et de l'envergure et $l = 0.000985$ dans la direction normale à la paroi.

L'observation est obtenue à partir de 408 capteurs sélectionnés parmi les 10 800 disponibles. La contrainte $x \in [0.6\pi, 2.4\pi]$, $z \in [0.25\pi, 0.75\pi]$ a été appliquée dans la sélection pour prendre en compte les différentes tailles de domaine pour les LES et la DNS et pour exclure les problèmes potentiels émergeant des conditions aux limites périodiques. L'emplacement des capteurs est indiqué par les points rouges sur la Fig. 6.5. Comme discuté précédemment, les trois compo-

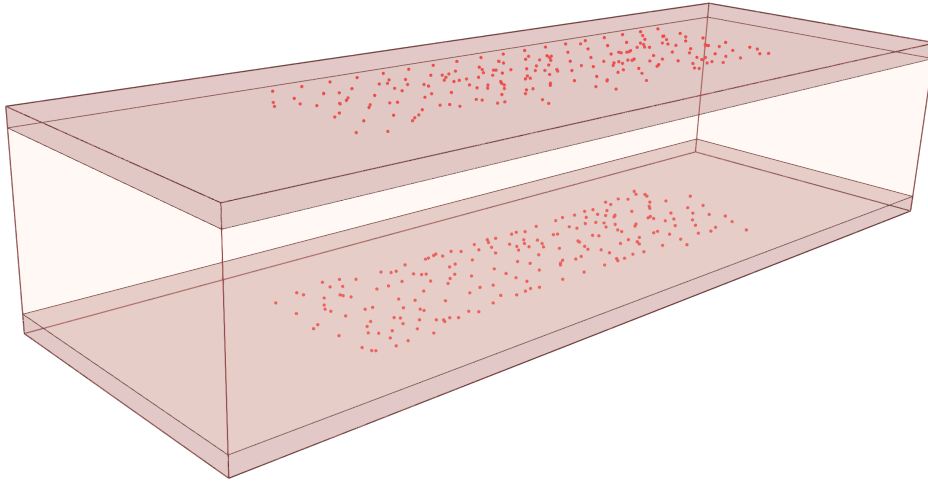


FIGURE 6.5 : Position des capteurs utilisés pour obtenir l’observation utilisée dans l’assimilation de données. La région en rouge correspond au découpage physique pour l’EnKF, c’est-à-dire la région dans l’espace où l’augmentation d’état est effectuée.

santes du champ de vitesse instantané sont échantillonnées. Cependant, dans les configurations suivantes, le vecteur d’observations est composé de 408 échantillons de vitesse dans le sens de l’écoulement uniquement. Cette décision exploite les caractéristiques du solveur numérique, qui tend localement à préserver la condition solénoïdale du champ de vitesse [75]. Ainsi, les composantes de vitesse normale et dans le sens de l’envergure sont automatiquement déterminées lorsque la vitesse dans le sens de l’écoulement est fournie comme observation. Ce point est important car il implique que la taille du vecteur d’observation est trois fois plus petite, ce qui permet une réduction significative des ressources de calcul nécessaires pour effectuer l’EnKF. Ce point est analysé plus en détail dans l’annexe A.1, où les résultats d’une procédure d’AD supplémentaire (DA-LESA), utilisant les trois composantes du champ de vitesse pour chaque capteur, sont comparés aux réalisations d’AD discutées dans le cœur du chapitre. La confiance dans les données DNS est déterminée par la matrice \mathbf{R} présentée dans le chapitre 3. La matrice est diagonale et exprimée comme $\mathbf{R} = \sigma_m^2 \mathbf{I}$, où σ_m quantifie l’incertitude des mesures. Une précision de 20% est appliquée en pourcentage aux valeurs de chaque observation. Cela implique que la variance du champ de vitesse oscille entre 0.003 et 0.188 selon la distance par rapport à la paroi du capteur considéré. Ces dernières remarques impliquent également que le poids donné à chaque observation est le même et ne dépend pas d’une seule valeur absolue, comme utilisé dans l’annexe A.1. La matrice d’observations est donc construite telle que :

$$\alpha_{i,k+1} = \alpha_{k+1}(1 + \epsilon_{i,k+1}), \text{ with } \epsilon_{i,k+1} \sim \mathcal{N}(0, \mathbf{R}_{k+1}) \quad (6.1)$$

La procédure globale d’assimilation de données est maintenant présentée. Les membres d’ensemble sont initialisés avec un état *a priori* en termes de champ physique initial et de description paramétrique du modèle de sous-maillages. Le premier est un instantané des champs de vitesse des simulations BL-LES ou VD-LES présentées dans la Sec. 6.2.1 en fonction du cas test correspondant. Il est le même pour chaque membre de l’ensemble. Ce choix a été effectué

pour accélérer l'étalonnage du modèle d'AD, les différences dans les solutions de chaque membre émergent alors à cause de la variabilité des caractéristiques du modèle de sous-maillages. Les conditions initiales pour la description paramétrique du modèle de sous-maillages sont différentes pour les trois simulations d'AD effectuées et seront décrites dans les sections 6.2.2.1, 6.2.2.2 et 6.2.2.3. Une fois l'état initial fourni, la procédure fait avancer dans le temps les LES membres de l'ensemble pour un total de $300t_A$, en effectuant une phase d'analyse tous les $0.12t_A$. Ce choix, qui implique qu'un seul échantillon d'observation sur trois est utilisé dans la procédure d'AD, donne un total de 2500 phases d'analyse. Si l'on considère que le pas de temps pour les simulations LES est $\Delta t = 0.02t_A$, cela indique qu'une phase d'assimilation de données est effectuée toutes les six étapes de prévision. Aucune inflation d'état n'est utilisée dans les procédures d'AD. Cependant, une inflation stochastique *paramétrique* variable dans le temps est incluse pour améliorer l'efficacité de l'optimisation. Aucune inflation n'a été utilisée de $t_A = 0$ à $t_A = 12$ afin de converger vers un optimum local pour l'optimisation des paramètres. Ensuite, une inflation relativement forte a été incluse pour $t_A = [12; 24]$ avec $\lambda = 10\%$, suivi de $\lambda = 5\%$ pour $t_A = [24; 36]$. Ce choix a permis d'éviter un effondrement rapide de la variance du système et en même temps d'améliorer la précision de l'optimisation. Finalement, $\lambda = 1\%$ a été utilisé pour calculer les moyennes des moments statistiques. Les moyennes statistiques sont calculées dans l'intervalle $t \in [50, 300]$, afin de dissiper en toute sécurité les niveaux élevés de variance précédemment utilisés pour la convergence de la description paramétrique.

Les trois principales simulations d'AD sont maintenant présentées en détail, mettant en évidence les différences entre les procédures.

6.2.2.1 Procédure d'AD 1 (DA-LES1) : optimisation du coefficient C_k basé sur la simulation BL-LES

Dans cette première configuration (appelée DA-LES1), le vecteur des paramètres à optimiser est constitué d'un élément, qui est la constante de modèle C_k de la fermeture de sous-maillages du modèle Smagorinsky. Cela équivaut à optimiser le coefficient bien connu C_S , qui a été étudié dans la littérature, notamment dans le cadre de la quantification d'incertitudes [74, 76]. Comme indiqué précédemment, la valeur de cette constante globale est mise à jour à chaque phase d'analyse. Les valeurs initiales de C_k pour les simulations de l'ensemble $N_e = 40$ sont déterminées à l'aide d'une distribution gaussienne bornée $\mathcal{N}(\mu_u, \sigma_u^2)$. En considérant les données de la littérature [74], $\mu_u = 0.094$ et $\sigma_u = 0.03$ ont été choisis pour obtenir un espace paramétrique suffisamment large. La distribution gaussienne est limitée à des valeurs dans la plage $\mu_u \pm 2\sigma_u$, afin d'éviter une paramétrisation initiale non physique qui pourrait conduire à la divergence de l'algorithme.

6.2.2.2 Procédure d'AD 2 (DA-LES2) : décomposition spatiale du coefficient C_k

Suite aux résultats de DA-LES1, une optimisation plus complexe est ciblée pour améliorer les capacités prédictives du solveur LES. En exploitant les caractéristiques d'homogénéité du cas test dans la direction de l'écoulement x et dans la direction de l'envergure z , l'optimisation de cette deuxième configuration (appelée DA-LES2) cible le comportement d'une expression fonctionnelle pour $C_k(y)$. Plus précisément, les coefficients libres d'une décomposition gaussienne

de C_k sont considérés comme des variables à optimiser :

$$C_k = \sum_{i=1}^n \exp \left(a_i - \frac{(y - y_i)^2}{\sigma_i^2} \right) \quad (6.2)$$

Pour chacune des n fonctions gaussiennes utilisées dans la décomposition, les paramètres libres à déterminer sont a_i (intensité du pic), σ_i (largeur de la fonction) et y_i (position du pic). Les fonctions sont considérées comme symétriques par rapport à la demi-hauteur du canal, en raison de la symétrie statistique dans la direction normale à la paroi y . La décomposition est effectuée en utilisant $n = 5$ fonctions gaussiennes. Cela conduit à un total de 15 paramètres dans le vecteur de contrôle θ à optimiser via l'EnKF. La distribution a priori moyenne pour ces fonctions est présentée sur la Fig. 6.6a. Cette distribution initiale est choisie de manière que le pic de trois de ces fonctions soit plus proche de la paroi, afin de fournir une représentation appropriée de ν_{SGS} dans cette région. Pour chaque membre de l'ensemble, la valeur des 15 coefficients libres est déterminée en utilisant une distribution gaussienne bornée ($\pm 2\sigma$) tel que $a_i = \mathcal{N}(-4.5, 0.3^2)$, $\sigma_i = \mathcal{N}(0.18, 0.04^2)$ et $y_i = \mathcal{N}(0.15, 0.05^2)$. Il convient de souligner que cette stratégie peut en principe fonctionner comme une correction de ν_{SGS} en fonction de la distance à la paroi, comme la fonction de van Driest. En effet, les fonctions exponentielles peuvent conduire à $C_k \rightarrow 0$ pour $y \rightarrow 0$, ce qui imiterait le comportement de la fonction de van Driest qui opère sur la valeur de Δ à proximité de la paroi.

6.2.2.3 Procédure d'AD 3 (DA-LES3) : optimisation du coefficient C_k basé sur la simulation VD-LES

Pour la troisième procédure (appelée DA-LES3), le vecteur des paramètres à optimiser est constitué d'un élément, qui est la constante de modèle C_k de la fermeture de sous-maillages du modèle Smagorinsky, de manière similaire à la procédure DA-LES1. Ainsi, les valeurs initiales de C_k pour les simulations de l'ensemble $N_e = 40$ sont obtenues en utilisant la même distribution gaussienne bornée $\mathcal{N}(\mu_u, \sigma_u^2)$. La fonction de paroi de van Driest est dans ce cas incluse dans le modèle de sous-maillages optimisé. Cette analyse vise à vérifier si l'inclusion d'une fonction de paroi affectera le processus d'optimisation d'AD et la précision du modèle résultant.

6.3 Prédiction des caractéristiques statistiques avec l'AD séquentielle

Dans cette section, une attention particulière est portée sur l'optimisation de la fermeture du modèle de sous-maillages. Les résultats de DA-LES1 et DA-LES2 sont étudiés pour observer comment la procédure d'AD affecte dynamiquement la valeur du paramètre C_k ainsi que pour évaluer les effets de l'optimisation paramétrique sur le comportement statistique de l'écoulement. Un point important qui doit être souligné est que de tels moments statistiques ne sont pas directement observés par les algorithmes d'AD. En fait, contrairement aux études récentes dans la littérature [87], la procédure d'AD ne s'appuie que sur des champs instantanés de l'écoulement obtenus à partir du modèle et échantillonnés comme observation.

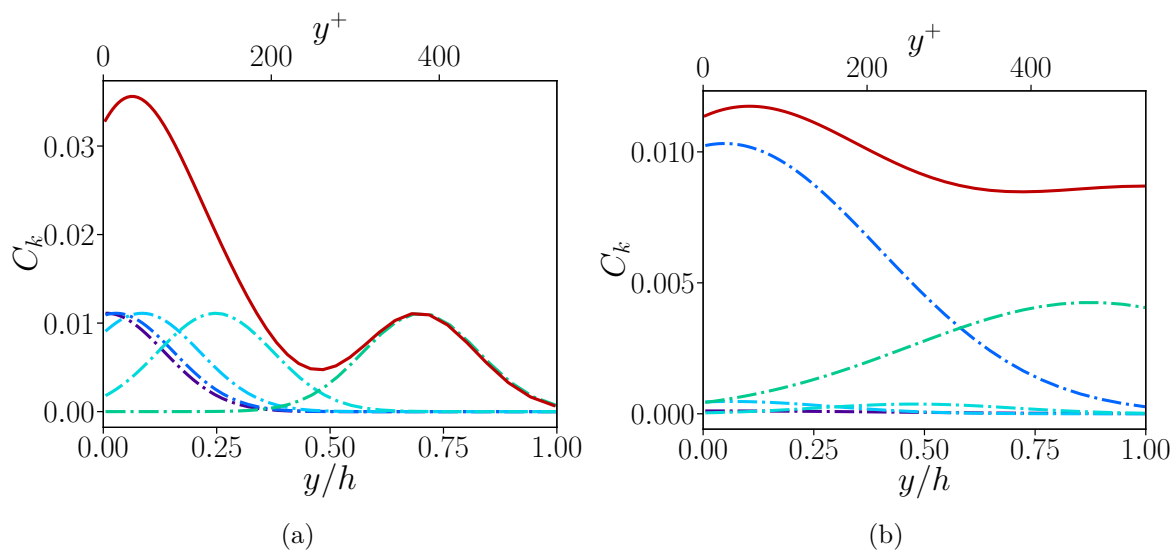


FIGURE 6.6 : En bleu : profils des cinq fonctions gaussiennes utilisées pour représenter C_k . En rouge : distribution de C_k . Gauche : distribution initiale *a priori*. Droite : distribution optimisée par AD

Premièrement, le comportement optimisé du paramètre C_k est étudié. Pour des simulations classiques utilisant les valeurs prescrites par le code CFD, $C_k = 0.094$, $C_\varepsilon = 1.048$ correspond à $C_S \approx 0.17$. Une fois la convergence paramétrique de l'optimisation d'AD obtenue, les coefficients présentent une évolution temporelle très faible. Les résultats obtenus à partir de la configuration DA-LES1, qui vise une optimisation globale de C_k , montrent que la moyenne temporelle de la valeur optimisée pour cette procédure est $C_k \approx 0.014$. Ce résultat, qui correspond à $C_S \approx 0.04$, est 7 fois plus petit que la valeur par défaut de C_k fournie par le code. L'incertitude associée au nombre limité de membres d'ensemble a été évaluée en répétant les phases initiales d'AD en utilisant différentes distributions aléatoires pour C_k . Ces résultats indiquent que les valeurs optimisées se situent dans la plage $C_k \in [0.012, 0.017]$. Dans ces plages, les variations des quantités physiques prédites sont très faibles et se situent dans le seuil de confiance des observations (c'est-à-dire les valeurs de la matrice \mathbf{R}).

Les résultats de la procédure DA-LES2 sont présentés sur la Fig. 6.6, où le profil de C_k et chaque fonction de la décomposition gaussienne spatiale sont représentés. Le profil final de C_k en rouge est à nouveau significativement inférieur à la distribution utilisée *a priori*. Les valeurs vont de 0.008 dans la région centrale de l'écoulement à une valeur maximale d'environ 0.012 atteinte près de la paroi autour de $y^+ = 60$. De plus, les contributions des modes de la décomposition gaussienne pour décrire le profil augmenté de C_k semblent être très différentes. Deux modes principaux gouvernent la forme de C_k . Le premier mode présente une décroissance lente et quasi-linéaire se déplaçant de la ligne centrale vers la paroi. Le second présente un maximum dans la région proche de la paroi ($y^+ \approx 30$). L'amplitude des trois autres modes est significativement plus faible et leur rôle est principalement de lisser le profil de C_k . Malgré la complexité plus élevée de cette stratégie, la distribution dans la direction y de C_k est quasi-constante et $C_k \approx 0.01$ c'est-à-dire très similaire à la valeur globale obtenue dans la procédure DA-LES1.

L'amélioration des capacités prédictives via l'optimisation d'AD est maintenant évaluée en comparant les moments statistiques du champ de vitesse avec les résultats DNS disponibles ainsi qu'avec les LES de base. Tout d'abord, la prédiction de la vitesse de frottement u_τ , qui est l'une des caractéristiques clés de ce cas test, est considérablement améliorée pour toutes les procédures d'AD. En effet, la vitesse de frottement pour la DNS ciblée est $u_\tau = 0.048$ et la simulation LES de base BL-LES prédit un $u_\tau = 0.0614$, ce qui représente une sur-prédiction de 28%. DA-LES1 et DA-LES2 prédisent presque la même vitesse de frottement, avec $u_\tau^{DL1} = 0.04617$ et $u_\tau^{DL2} = 0.04619$. Dans ces cas, la vitesse de frottement est légèrement sous-estimée par rapport à la DNS, mais l'écart n'est que de 4%. Cette augmentation de la précision s'accompagne d'une réduction significative de la viscosité de sous-maillages dans la région proche de la paroi, qui ne s'adapte pas correctement pour la LES Smagorinsky. Étant donné que les valeurs obtenues pour C_k dans la région proche de la paroi avec les deux procédures d'AD sont presque identiques, il n'est pas surprenant d'observer des variations minimales dans la prédiction de u_τ . Des conclusions similaires peuvent être tirées de l'analyse de la Fig. 6.7, où le profil de vitesse moyenne normalisé u^* par rapport à y^* est présenté. Les moyennes pour les procédures d'AD sont effectuées de telle sorte que $u^* = \langle \overline{u_x} \rangle / u_\tau$ où $\langle \cdot \rangle$ est l'opérateur de moyenne d'ensemble et $\overline{\cdot}$ est l'opérateur de moyenne temporelle. La zone grise représente la région où la localisation physique est appliquée pour l'assimilation des données. Les résultats obtenus via les deux procédures d'AD montrent une amélioration globale de la prédiction du champ de vitesse, réduisant en moyenne l'écart avec les données DNS. Cette observation est une conséquence directe de la prédiction améliorée pour u_τ . Le comportement apparemment plus précis de la BL-LES près du centre du canal est en fait une compensation des erreurs entre l'erreur numérique locale et la prédiction erronée de u_τ , qui peut être observée sur la Fig. 6.7a. En effet, avec une prédiction plus précise de u_τ , les résultats de la BL-LES se rapprocheraient fortement des résultats obtenus par AD. Des écarts mineurs peuvent être observés entre les procédures DA-LES1 et DA-LES2, qui sont sans doute associés au taux de convergence de l'EnKF utilisant 40 membres de l'ensemble.

Les composantes normalisées du tenseur des contraintes de Reynolds résolu $u'_i u'_j{}^* = \langle \overline{u'_i u'_j} \rangle / u_\tau^2$ sont présentées sur la Fig. 6.8. Là encore, pour les configurations par AD, une amélioration globale de la précision de la prédiction de ces quantités est observée. Pour toutes les composantes, l'emplacement du pic est prédit avec précision. L'amplitude des composantes présente également une amélioration générale, qui dépend cependant de la composante considérée. Alors qu'un très bon accord avec les données DNS est observé pour $u'_z u'_z{}^*$, une légère diminution de la précision est en revanche obtenue pour $u'_y u'_y{}^*$. Les résultats presque identiques obtenus avec les deux procédures DA-LES1 et DA-LES2 suggèrent que les variations de C_k dans la direction y pour cette dernière n'affectent pas la prédiction de l'écoulement. Ces procédures ont sans doute atteint le potentiel d'amélioration des performances du Smagorinsky, dont la représentation des échelles de sous-maillages est affectée par de fortes limitations intrinsèques [87].

L'analyse des quantités physiques normalisées par le u_τ calculé pour chaque simulation (suffixe +) conduit à des conclusions similaires. Les profils de vitesse moyenne dans le sens de l'écoulement, qui sont présentés sur la Fig. 6.9, confirment le manque global de précision de la simulation de base, qui est maintenant encore plus amplifié par l'erreur significative dans la prédiction de u_τ . Les composantes du tenseur de contrainte de Reynolds résolu, qui sont reportées sur la Fig. 6.10, fournissent également des indications très similaires.

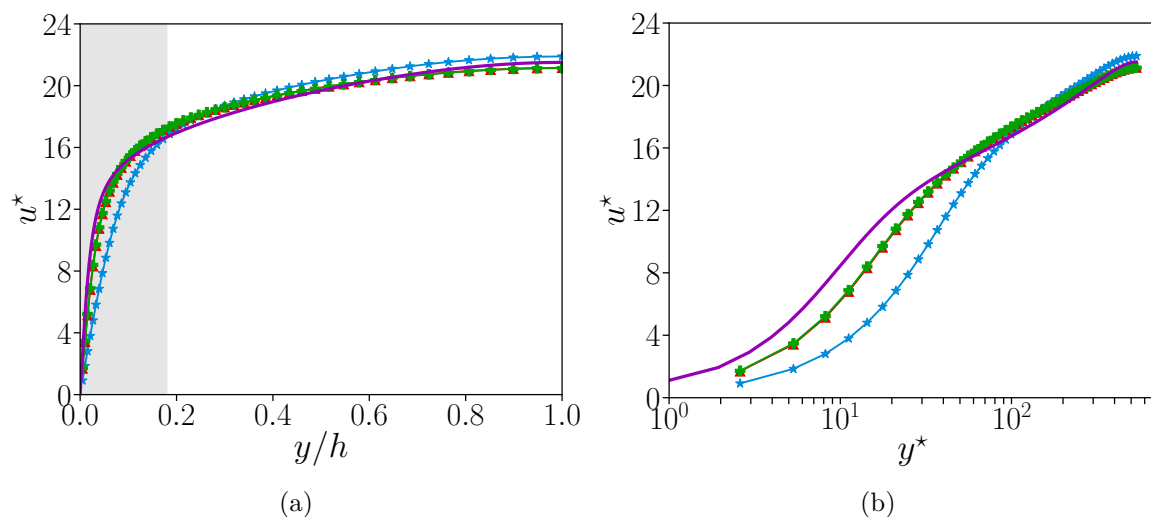


FIGURE 6.7 : Profils de vitesse sur le demi-canal pour DA-LES1 (\blacktriangleright), DA-LES2 ($+$), BL-LES (\star) et DNS ($-$). L'adimensionnement est réalisée en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l'AD.

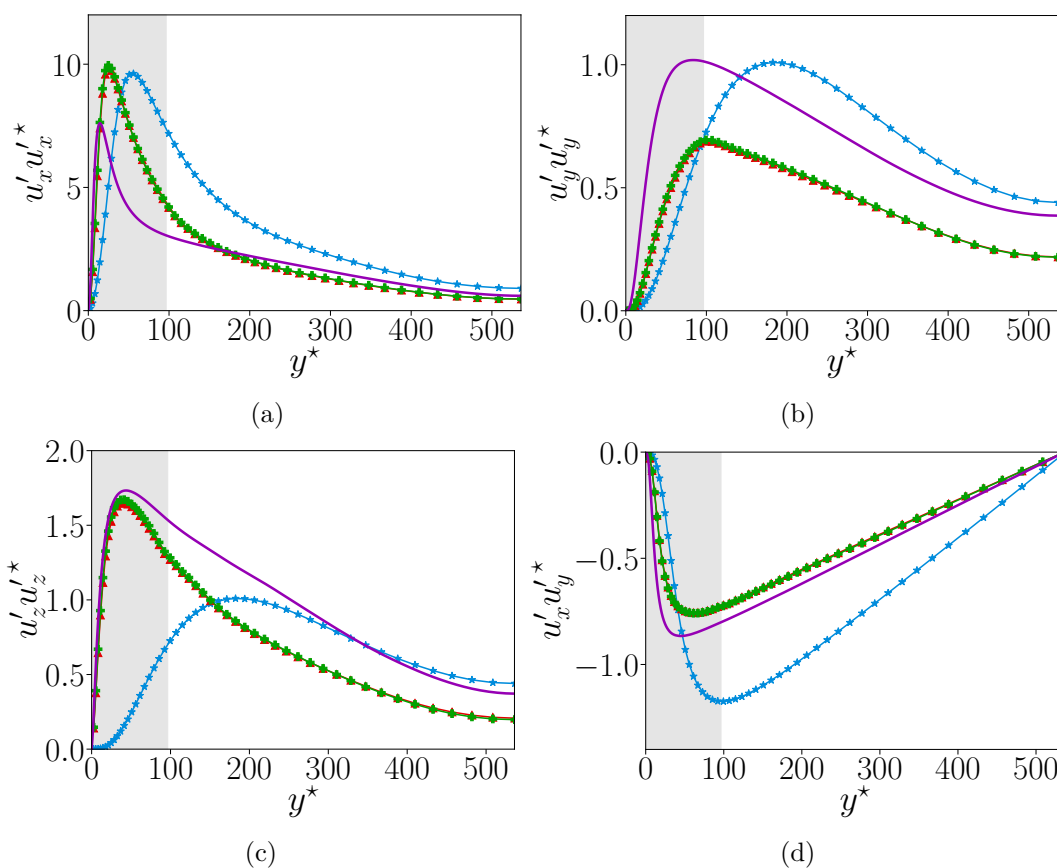


FIGURE 6.8 : Composantes du tenseur de contrainte de Reynolds pour DA-LES1 (\blacktriangleright), DA-LES2 ($+$), BL-LES (\star) et DNS ($-$). L'adimensionnement est réalisée en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l'AD.

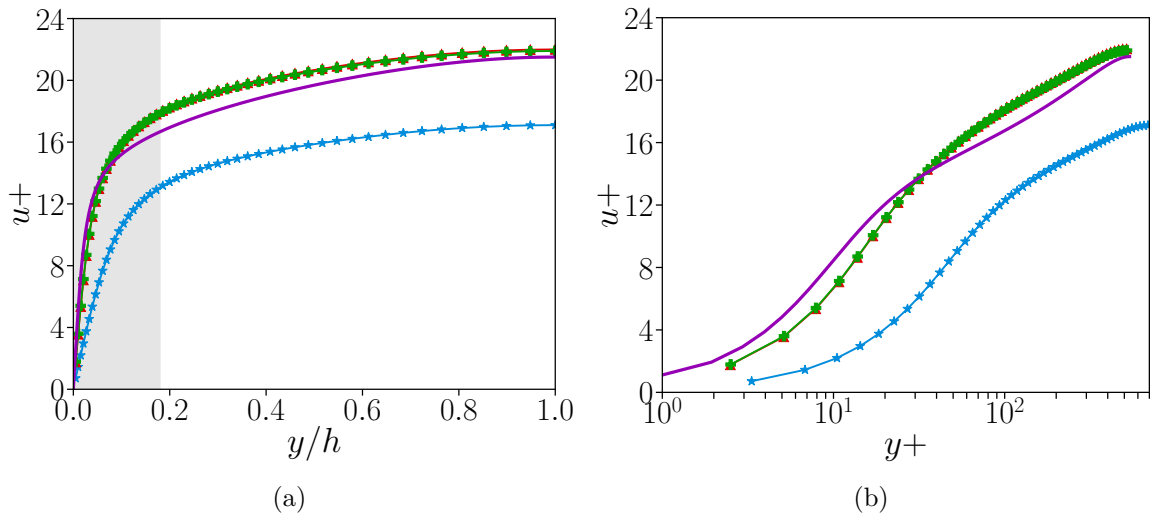


FIGURE 6.9 : Profils de vitesse sur le demi-canal pour DA-LES1 (\blacktriangleright), DA-LES2 ($+$), BL-LES (\star) et DNS ($-$). La zone grise représente la région où la localisation physique est appliquée pour l'AD.

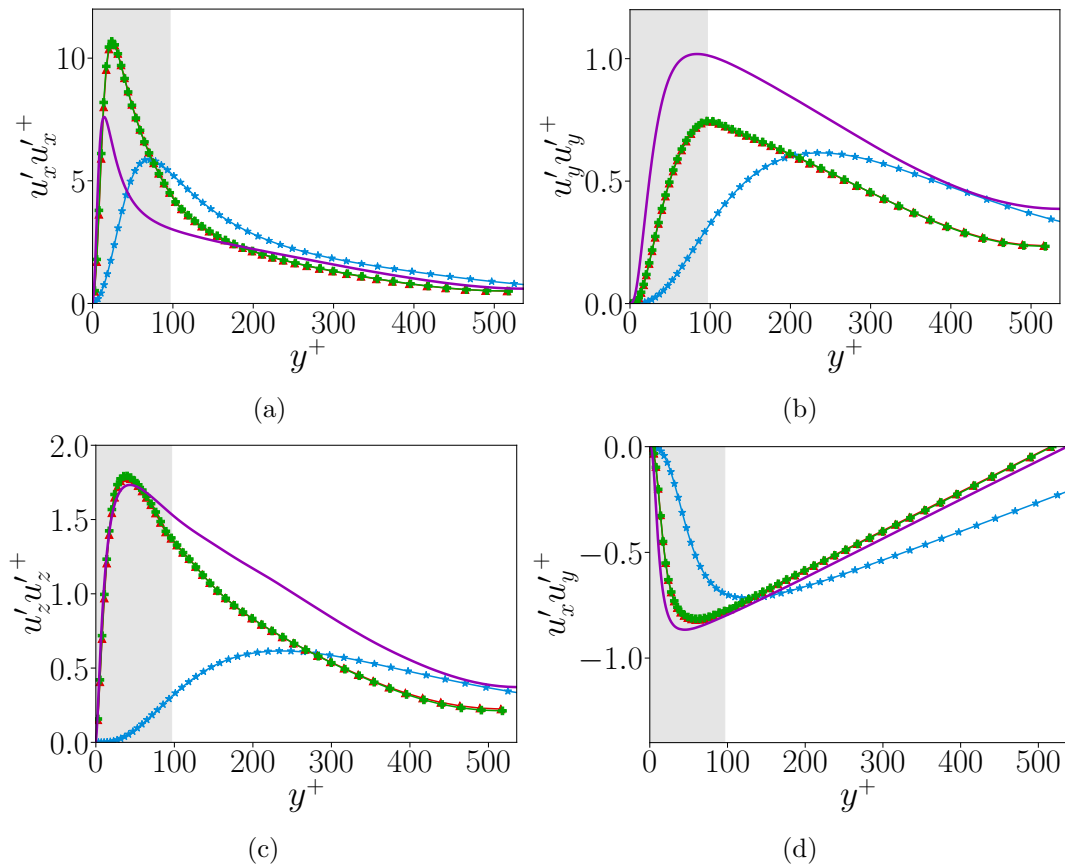


FIGURE 6.10 : Composantes du tenseur de contrainte de Reynolds pour DA-LES1 (\blacktriangleright), DA-LES2 ($+$), BL-LES (\star) et DNS ($-$). La zone grise représente la région où la localisation physique est appliquée pour l'AD.

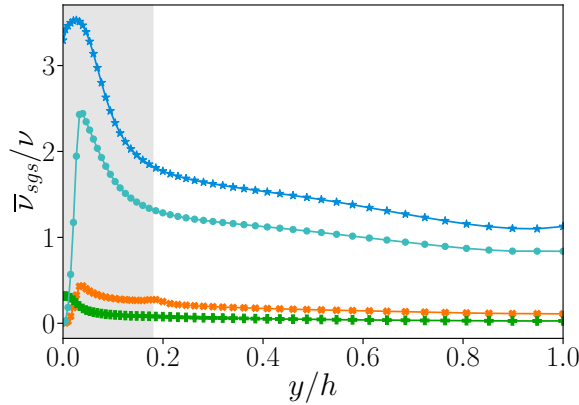


FIGURE 6.11 : Distribution du rapport $\bar{\nu}_{SGS}/\nu$ dans la direction verticale pour BL-LES (*), VD-LES (●), DA-LES2 (+) and DA-LES3 (×)

Les résultats de la procédure DA-LES3 sont maintenant discutés. Ici, l’algorithme d’AD est utilisé pour optimiser un modèle plus complexe, à savoir le modèle de Smagorinsky couplé à la fonction de paroi de van Driest. Tout d’abord, les profils moyens dans le temps de ν_{SGS} pour les trois procédures d’AD sont présentés sur la Fig. 6.11. Tous les modèles augmentés par AD tendent à réduire de manière significative la viscosité artificielle introduite dans les équations discrétisées par rapport au modèle classique. De plus, la distribution globale pour DA-LES2 et DA-LES3 est très similaire, ce qui indique que l’algorithme d’optimisation est robuste. La principale différence est que l’inclusion de la fonction de paroi dans DA-LES3 est responsable de $\nu_{SGS} \rightarrow 0$ à proximité de la paroi, tandis qu’une valeur faible mais non nulle est obtenue pour DA-LES2. L’optimisation pour la procédure DA-LES3 conduit à une optimisation paramétrique moyenne temporelle $C_k \approx 0.021$, qui est 4.5 plus petite que la valeur par défaut de C_k mais 50% plus élevée que la valeur optimisée par DA-LES1. Cependant, cette différence semble avoir un impact mineur sur la prédiction des moments statistiques du champ de vitesse, qui sont présentés sur les Fig. 6.12 et 6.13. Les résultats indiquent qu’une précision très similaire est obtenue par les procédures DA-LES3 et DA-LES2, même si le modèle sous-jacent est différent. De plus, il semble que les modèles d’AD soient globalement plus performants que la simulation VD-LES, dont les résultats sont cependant plus précis que ceux obtenus par la simulation BL-LES. La principale différence pratique observée entre les configurations DA-LES2 et DA-LES3 est la rigidité de l’algorithme lors de l’optimisation. En raison de l’inclusion de la fonction de paroi, certains hyper-paramètres de l’AD dans la configuration DA-LES3 ont dû être relaxés pour assurer la convergence. Cette observation souligne les difficultés à réaliser une optimisation efficace pour des simulations complexes et indique en outre que des descriptions de fermeture plus simples devraient être utilisées lorsque cela est possible, étant donné que la précision obtenue pour les procédures DA-LES2 et DA-LES3 est équivalente mais que la première utilise un modèle plus simple.

Enfin, une analyse spectrale du champ de vitesse \mathbf{u} est réalisée dans la Fig. 6.14. Cette variable a été échantillonnée dans le temps grâce à quatre sondes situées dans l’intervalle $y^* \in [1.45, 46.74]$. Les spectres de puissance ont été obtenus en utilisant une transformée de Morlet [135] pour la simulation BL-LES, la configuration DA-LES2 et la DNS. Les spectres sont tracés

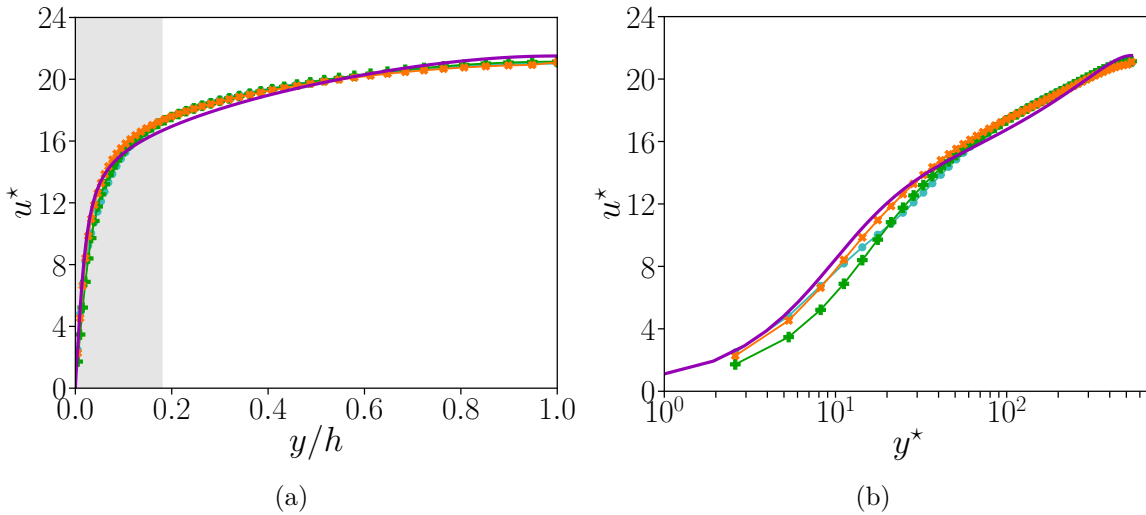


FIGURE 6.12 : Profils de vitesse de demi-canal pour DA-LES2 (+), DA-LES3 (\times), VD-LES (\bullet) et DNS ($-$). L'adimensionnement est réalisé en utilisant la vitesse de frottement u_τ obtenue dans la simulation DNS. La zone grise représente la région où la localisation physique (clipping) est appliquée à des fins de DA.

par rapport au nombre d'onde sans dimension, $\kappa^+ = \kappa\nu/u_\tau$ avec $\kappa = 2\pi f/u_c$. f est ici l'ensemble des fréquences utilisées pour la transformée de Morlet. Sur la première ligne, les données pour la composante u_x sont présentées aux endroits où l'observation est disponible et les données de ce capteur sont utilisées dans la phase d'analyse d'AD (indiquée par U-DA dans la légende). La comparaison des simulations BL-LES et par AD montre que la précision des spectres a été améliorée pour chaque y^* étudiée. Le meilleur résultat est observé à proximité de la paroi comme le montre la Fig. 6.14a. Pour cet emplacement, l'amplitude du signal est améliorée d'environ un ordre de grandeur. La comparaison des spectres DNS et DA-LES2 indique également un décalage du nombre d'onde pour lequel la puissance spectrale commence à diminuer rapidement. Ce décalage est très proche du rapport des résolutions de mailles dans le sens de l'écoulement (voir le Tab.6.1). Pour la simulation BL-LES, cette chute d'énergie commence à des nombres d'onde plus faibles que pour DA-LES2. Cette observation peut être justifiée par l'écart dans la prédiction de u_τ (qui est utilisé pour obtenir κ^+) ainsi que par la fermeture du modèle Smagorinsky, qui fournit un effet dissipatif indésirable aux grandes échelles. Les résultats sur la deuxième ligne des Fig. 6.14d et 6.14e montrent les spectres de puissance pour les composantes transversale et verticale. Ils sont obtenus à un endroit où un capteur DNS est disponible et utilisé pour la phase d'analyse d'AD, mais l'information assimilée (vitesse dans le sens de l'écoulement) n'est pas celle étudiée ici. De manière similaire à ce qui a été observé pour les spectres de vitesse dans le sens de l'écoulement, une amélioration globale est obtenue pour la procédure DA-LES2. Ce résultat confirme l'effet bénéfique global de l'AD sur l'ensemble du champ, et pas seulement pour les variables pour lesquelles l'observation est disponible. L'analyse est complétée par les résultats de la Fig. 6.14f montrant le spectre de la vitesse dans le sens de l'écoulement échantillonné à un endroit non utilisé dans la phase d'analyse d'AD (N-DA). Là encore, le spectre montre une amélioration similaire à ce qui a été observé sur

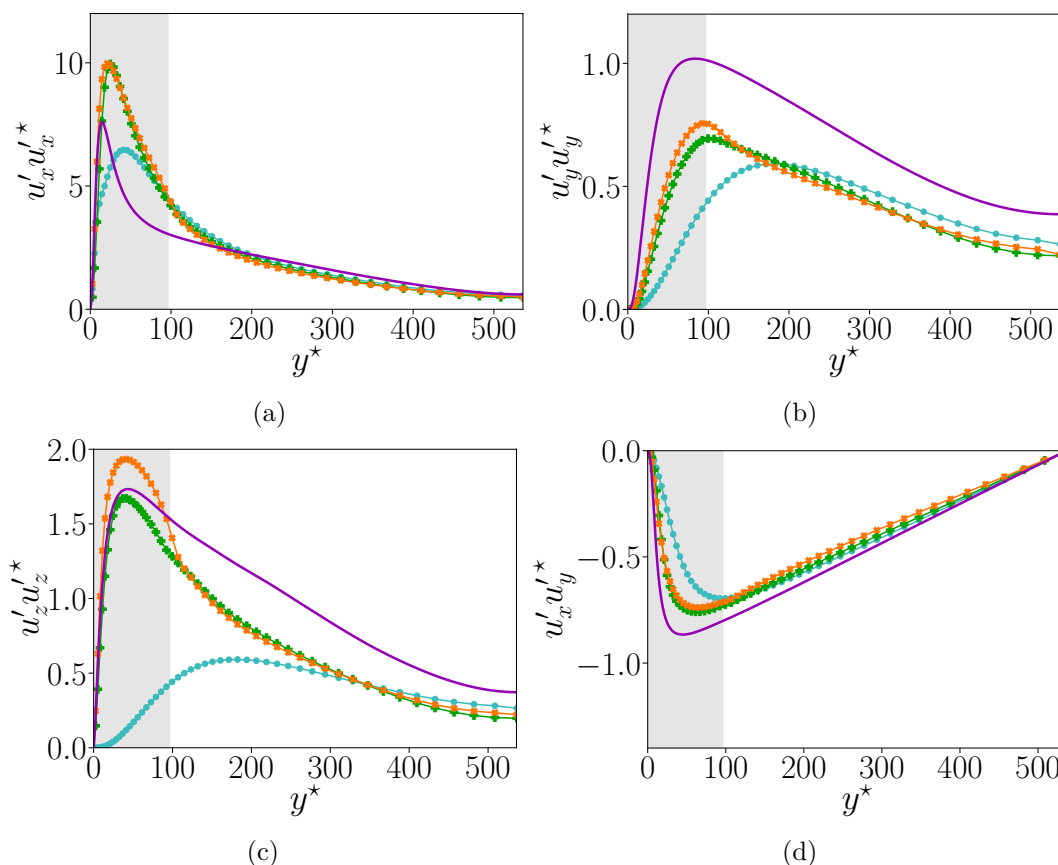


FIGURE 6.13 : Composantes du tenseur de contrainte de Reynolds pour DA-LES2 (+), DA-LES3 (×), VD-LES (•) et DNS (–). L’adimensionnement est réalisé en utilisant la vitesse de frottement u_τ obtenue dans la DNS. La zone grise représente la région où la localisation physique est appliquée pour l’AD.

les capteurs activement utilisés dans la procédure d’AD, indiquant que l’optimisation de la fermeture de sous-maillages est globalement bénéfique notamment pour réduire la dissipation de l’énergie résolue aux grandes échelles.

En résumé, l’assimilation de données séquentielle utilisant des mesures instantanées est capable d’améliorer la précision de la LES via l’étalonnage de la fermeture de sous-maillages. Un point intéressant est que les résultats actuels sont similaires aux conclusions de Mons et al. [87]. Cependant, l’optimisation effectuée par Mons et al. a été obtenue via l’observation de quantités physiques utilisées pour évaluer les performances des simulations LES. Dans le cas présent, le processus d’optimisation est plus complexe, en raison de la nature instantanée de l’observation ainsi que de sa rareté dans l’espace et le temps. Ainsi, les présents résultats ouvrent des perspectives d’optimisation en temps réel de la CFD des grandes échelles à l’aide d’outils basés sur l’EnKF, une fois que les architectures de calcul seront suffisamment robustes pour de telles applications. De la même manière que ce qui a été observé par Mons et al. [87], l’optimisation paramétrique peut atténuer mais pas éliminer l’écart entre la LES utilisant le modèle de Smagorinsky et la DNS, en raison des limitations intrinsèques du modèle de sous-maillages. Bien que ce problème soit difficile à résoudre, il semble que l’AD séquentielle ait un

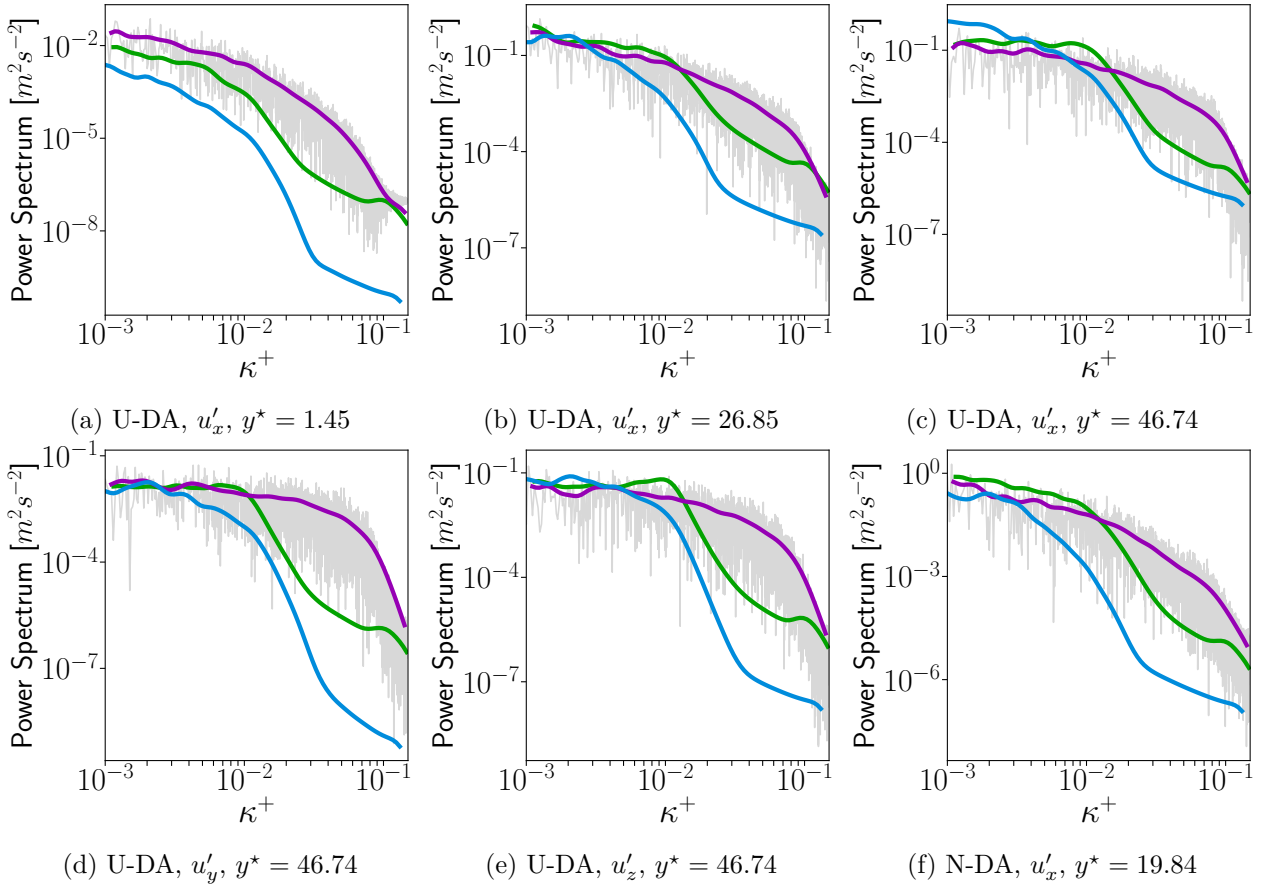


FIGURE 6.14 : Spectres de puissance du champ de vitesse obtenus par transformée en ondelettes de Morlet. Les résultats sont présentés pour la DA-LES2 (—), BL-LES (—) et la DNS (—). Pour cette dernière, des résultats supplémentaires obtenus par FFT sont présentés en gris.

potentiel plus élevé pour déterminer en temps réel les formes structurales du modèle de sous-mailles que les approches non séquentielles. Enfin, les stratégies utilisées dans cette analyse indiquent que la meilleure précision est obtenue pour des valeurs très faibles de la constante de modèle C_k . Bien que la configuration DA-LES2 fournisse une distribution spatiale plus sophistiquée de ce paramètre, les valeurs sont suffisamment faibles pour considérer que l'effet dynamique de la fermeture de sous-mailles devient globalement et localement mineur, comme le montrent les profils obtenus pour les procédures DA-LES1 et DA-LES3. Ces résultats sont cohérents avec des travaux récents dans la littérature présentant des comparaisons approfondies entre les fermetures de sous-mailles explicites et implicites [22].

Pour conclure, le comportement optimisé de la constante de modèle C_k est obtenu via un calibrage qui utilise des informations de haute-fidélité provenant d'un écoulement en canal plan turbulent. Son comportement est donc probablement sous-optimal s'il est appliqué à différentes classes d'écoulement. Cependant, l'application de l'AD dans ce cas est envisagée dans le cadre d'une future application aux jumeaux numériques. Pour ce type d'application, il n'est pas nécessaire d'avoir une méthode universelle précise, mais il serait raisonnable de rechercher une fermeture flexible capable de modifier ses performances (locales et globales) tout en obtenant

des données en temps réel. Ces résultats mettent en évidence le potentiel de l'AD pour cette classe d'applications.

6.4 Synchronisation des champs de l'écoulement

Les capacités de synchronisation de l'algorithme d'AD sont maintenant étudiées. La synchronisation correspond à la capacité de l'algorithme à réduire progressivement l'écart entre la solution instantanée du modèle et l'observation, à la fois à proximité et loin des capteurs. En cas de succès, les seules corrections d'état appliquées par la phase d'analyse d'AD sont dues à l'accumulation d'erreurs dans les étapes de prévision, en raison du manque de précision du modèle. Bien que la synchronisation ne soit pas nécessaire pour obtenir une prédiction précise des moments statistiques, tels que ceux étudiés dans la Sec. 6.3, elle a une importance cruciale pour l'analyse des caractéristiques instantanées des écoulements non stationnaires. En effet, dans le paradigme du jumeau numérique, une synchronisation efficace permet au modèle d'identifier les événements extrêmes et ainsi d'éviter les occurrences critiques pour l'équivalent physique. Les outils basés sur l'EnKF peuvent naturellement agir sur la synchronisation de l'écoulement instantané. Grâce à la flexibilité de la quantité observée et à la corrélation locale entre les variables physiques, leur efficacité dans cette tâche est supposée supérieure aux algorithmes de type *Nudging*, récemment utilisé pour analyser la synchronisation dans le canal plan turbulent [145]. Cependant, lors des calculs d'AD, la variabilité de l'ensemble a tendance à diminuer relativement rapidement, empêchant potentiellement une synchronisation efficace. Pour éviter ce problème, l'inflation doit être correctement optimisée.

Dans cette section, un certain nombre de calculs par AD sont effectués pour étudier les effets de l'inflation sur le taux de synchronisation. Pour cette nouvelle procédure, l'attention est focalisée sur les toutes premières phases d'analyse de l'AD, et les résultats sont étudiés sur deux temps advectifs t_A . Les analyses d'AD sont maintenant effectuées tous les deux pas de temps, soit $0.04t_A$, ce qui correspond à un total de 50 phases de mise à jour du champ de vitesse sur la fenêtre temporelle étudiée. Une fréquence de mise à jour aussi élevée a été imposée pour garantir que les erreurs dues à la rareté temporelle des données soient négligées [73]. L'estimation de l'état est obtenue via la prédiction de l'écoulement de 40 membres, qui sont initialisés à l'aide d'un champ de vitesse différent mais partagent la même valeur de C_k obtenue par le modèle de sous-maillages optimisé via la procédure DA-LES2. Les 40 champs de vitesse utilisés comme états *a priori* ont été générés en exécutant une seule simulation avec le même modèle de sous-maillages optimisé obtenu dans la Sec. 6.3 et en échantillonnant un champ complet tous les $10t_A$.

L'inflation est ici appliquée uniquement sur l'estimation de l'état via l'approche stochastique décrite chapitre 3. L'inflation appliquée à la description paramétrique du modèle de sous-maillages est ici fixée à zéro pour exclure les effets dus aux différents comportements de la fermeture LES. De plus, la matrice de covariance \mathcal{R} est également la même pour chaque procédure d'AD et elle est fixée à $\mathbf{R} = \sigma_m^2 \mathbf{I}$ avec $\sigma_m = 5\%$ appliqué en pourcentage aux valeurs de chaque observation, comme expliqué dans la Sec. 6.2.2. Plus de détails sur ces deux derniers hyper-paramètres sont fournis dans l'annexe A.2, où l'inflation paramétrique est montrée comme ayant un effet négligeable pour les besoins de cette étude.

ID du capteur	x/h	y/h	z/h	y^*	Utilisé pour l'AD
1	7.372	0.0027	2.349	1.45	oui
2	3.967	1.9631	1.233	19.84	oui
3	7.464	1.9501	1.748	26.85	non

TABLE 6.2 : Détails des capteurs utilisés pour étudier la synchronisation.

Premièrement, l'efficacité de la synchronisation est évaluée en analysant les informations instantanées. Le champ de vitesse obtenu par les membres de l'ensemble est évalué à trois endroits, qui sont sélectionnés parmi les 10 800 capteurs précédemment utilisés dans la DNS de référence. Les détails sur les capteurs sont donnés dans le Tab. 6.2. Deux des capteurs sont utilisées dans l'algorithme d'AD, tandis que le dernier n'est pas directement utilisé. Néanmoins, les données obtenues pour ce dernier peuvent être utilisées à des fins de comparaison.

L'évolution de la vitesse instantanée dans le sens de l'écoulement u_x normalisé par la vitesse moyenne de la ligne médiane u_c est illustrée sur la Fig. 6.15 pour les trois capteurs. La vitesse échantillonnée à partir de la DNS, qui est utilisée comme observation, est indiquée en bleu. Les données échantillonnées au même endroit à partir des membres de l'ensemble de la procédure d'AD sont indiquées en noir. Dans ce cas, la ligne noire correspond à une moyenne d'ensemble. Les zones d'ombre représentent visuellement le niveau de confiance/la variabilité des données. Plus précisément, la zone bleue est connectée aux valeurs incluses dans la matrice de covariance \mathbf{R} avec une taille de $2\sigma_m$. De même, la zone grise représente l'intervalle de confiance à 95% de l'ensemble, soit deux fois son écart type. Cette quantité est initialement déterminée par la distribution des états *a priori* pour $t_A = 0$, et elle est progressivement affectée par l'inflation appliquée à l'état physique au fur et à mesure que les phases d'analyse sont effectuées. Les trois capteurs ont été sélectionnés pour mettre en évidence différentes caractéristiques de l'écoulement. Les capteurs 1 et 2 correspondent aux capteurs utilisés dans les procédures d'AD, mais elles sont situées à des distances différentes de la paroi ($y^* = 1.45$ et $y^* = 19.84$, respectivement). En revanche, le capteur 3 est situé à $y^* = 26.85$ et n'est pas utilisé dans l'AD. La première ligne sur la Fig. 6.15 montre que si aucune inflation d'état n'est utilisée, la variabilité initiale du modèle due aux états *a priori* s'effondre très rapidement avec un rétrécissement drastique de la zone grise. Les zones grises et bleues présentent une superposition très limitée, ce qui empêche les réalisations du modèle de se synchroniser avec l'observation. Sur la deuxième, troisième et quatrième lignes de la Fig. 6.15, une inflation d'état de plus en plus importante est utilisée pendant les phases d'analyse. Une augmentation de la zone grise associée à la variabilité du modèle est visible, qui ne diminue pas pour des temps de simulation plus longs. L'analyse des résultats pour les capteurs 1 et 2 indique clairement qu'un seuil de 15% d'inflation gaussienne de l'état semble suffisant pour obtenir une synchronisation convaincante du champ de vitesse en correspondance avec les capteurs. Ce seuil pourrait être encore plus bas si des algorithmes plus sophistiqués pour l'inflation d'état sont utilisés. Des améliorations significatives avec l'augmentation de l'inflation sont également observées pour le capteur 3, même si la synchronisation n'est pas complètement obtenue. Pour partiellement conclure, une fois qu'une superposition

significativement importante des zones de confiance est obtenue pendant une durée suffisamment longue, une bonne synchronisation est obtenue. Un comportement similaire a également été observé par Tandeo et al. [128] mais pour un modèle unidimensionnel.

Dans un deuxième temps, une estimation de la racine de l'erreur quadratique moyenne normalisée (indiqué par Φ) pour le champ de vitesse est utilisée pour fournir une évaluation globale des capacités de l'algorithme d'AD à synchroniser le modèle LES avec les données disponibles de la DNS. Ce calcul est effectué en considérant les données des 408 capteurs utilisés dans l'algorithme d'AD et pour 408 capteurs qui n'ont pas été utilisés pour l'optimisation. La définition de Φ pour le pas de temps k est :

$$\Phi_k = \sqrt{\frac{\sum_{j=1}^{N_o} (\langle s_{j,k} \rangle - \alpha_{j,k})^2}{N_o}} / \alpha_k^{moy} \quad (6.3)$$

avec $\langle s_k \rangle$ la moyenne de l'ensemble et $\alpha_k^{moy} = \sum_{j=1}^{N_o} \alpha_{j,k} / N_o$ et N_o le nombre d'observations (capteurs)

Les résultats sont présentés sur la Fig. 6.16 pour les quatre cas précédemment analysés, soit 0%, 5%, 15% et 25% d'inflation d'état. La courbe rouge correspond à une valeur limite Φ_{lim} calculée en utilisant les résultats de la procédure sans inflation contre 1000 jeux de valeurs d'observations (pour différents temps). Par conséquent, pour un nombre infini d'observations et de capteurs, Φ_{lim} sur les Fig. 6.16 (a) et (b) devrait être identiques. Φ_{lim} est ici différent en raison du nombre limité d'observations et des coordonnées hétérogènes des jeux de capteurs. Les résultats de la Fig. 6.16 (a) correspondent à l'écart moyen observé sur les 408 emplacements où les capteurs sont utilisés pour l'AD. Les procédures d'AD sont significativement meilleures dans les premières étapes, grâce à la variabilité initialement fournie par le choix des états *a priori*. Cependant, les résultats ont tendance à se dégrader assez rapidement pour la procédure d'AD sans inflation d'état et devraient présenter des erreurs très similaires à Φ_{lim} après un temps suffisamment long. D'autre part, les trois expériences d'AD avec inflation d'état non nulle se comportent de manière très similaire. Leur amplitude est significativement plus petite que Φ_{lim} et elle ne semble pas se détériorer dans la fenêtre temporelle analysée. La Fig. 6.16 (b) montre les valeurs d'erreur en correspondance de capteurs où les données DNS sont disponibles, mais elles ne sont pas utilisées dans la procédure d'AD. La diminution de l'erreur lors de l'utilisation de différents niveaux d'inflation est ici qualitativement similaire à ce qui a été discuté précédemment pour la Fig. 6.16 (a) même si, dans ce cas, les résultats pour les procédures d'AD sont plus proches de Φ_{lim} . De plus, les résultats semblent être plus sensibles à la valeur de l'inflation d'état, car une inflation très forte semble avoir de moins bons résultats qu'une inflation d'état modérée. Les perturbations semblent suffisamment fortes pour introduire un effet de bruit indésirable sur la prédiction de l'écoulement, dégradant la précision globale. Un autre problème potentiel avec les hyper-paramètres, qui n'est pas étudié dans le présent travail, est associé à la longueur caractéristique utilisée pour la localisation de la covariance. Si la longueur sélectionnée est grande, des corrélations parasites peuvent apparaître en raison d'erreurs d'échantillonnage. D'autre part, une courte longueur peut empêcher une représentation précise de la corrélation entre les variables, fonctionnant comme un filtre sur les interactions non locales multi-échelles observées dans les écoulements turbulents.

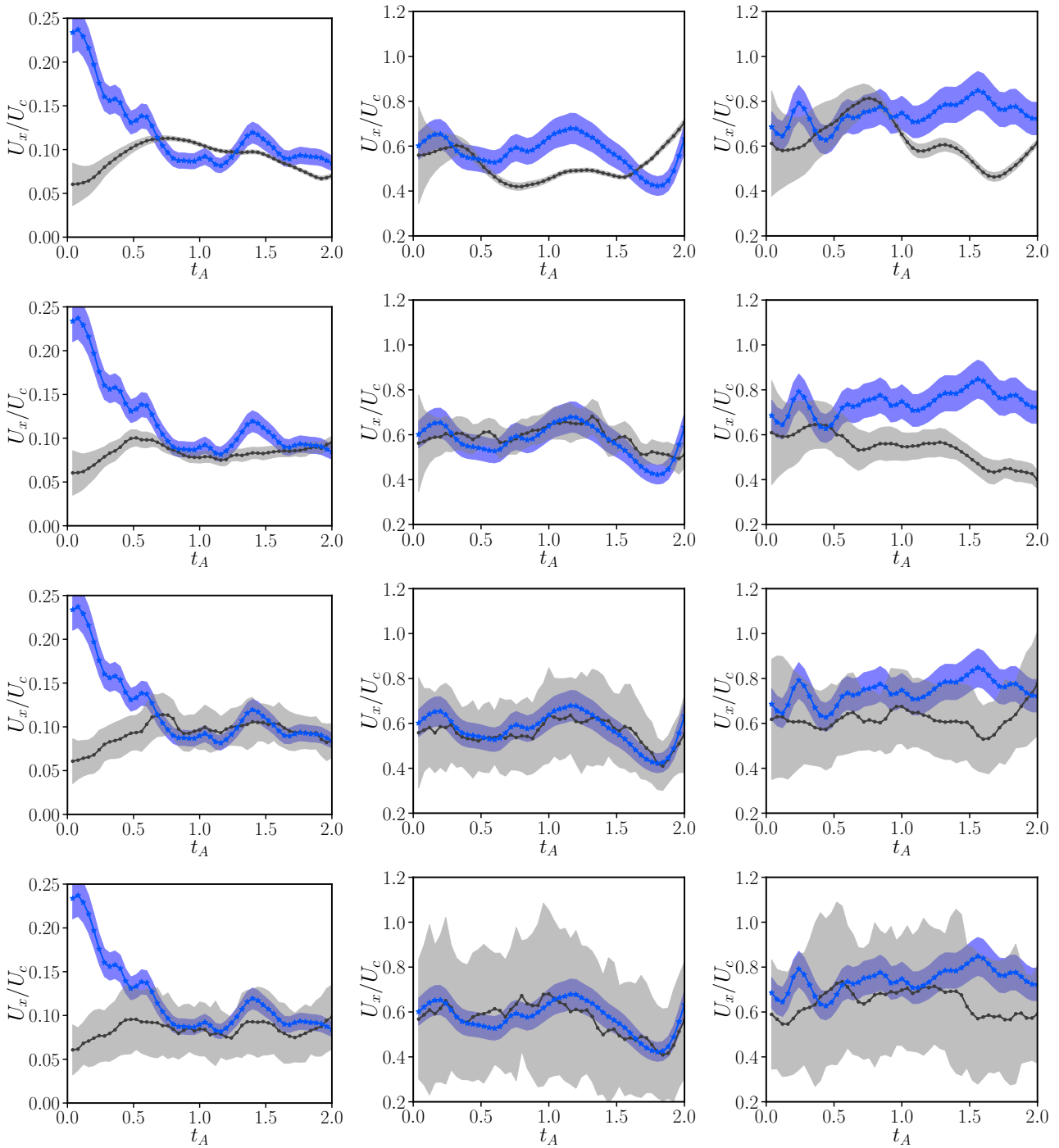


FIGURE 6.15 : Synchronisation du champ de vitesse. Les données en bleu représentent l'observation, tandis que les lignes noires indiquent la prédiction de l'ensemble. Les résultats sont présentés pour (colonne de gauche) le capteur 1, (colonne centrale) le capteur 2 et (colonne de droite) le capteur 3. La valeur de l'inflation d'état est égale à (première ligne) 0%, (deuxième ligne) 5%, (troisième ligne) 15% et (quatrième ligne) 25%.

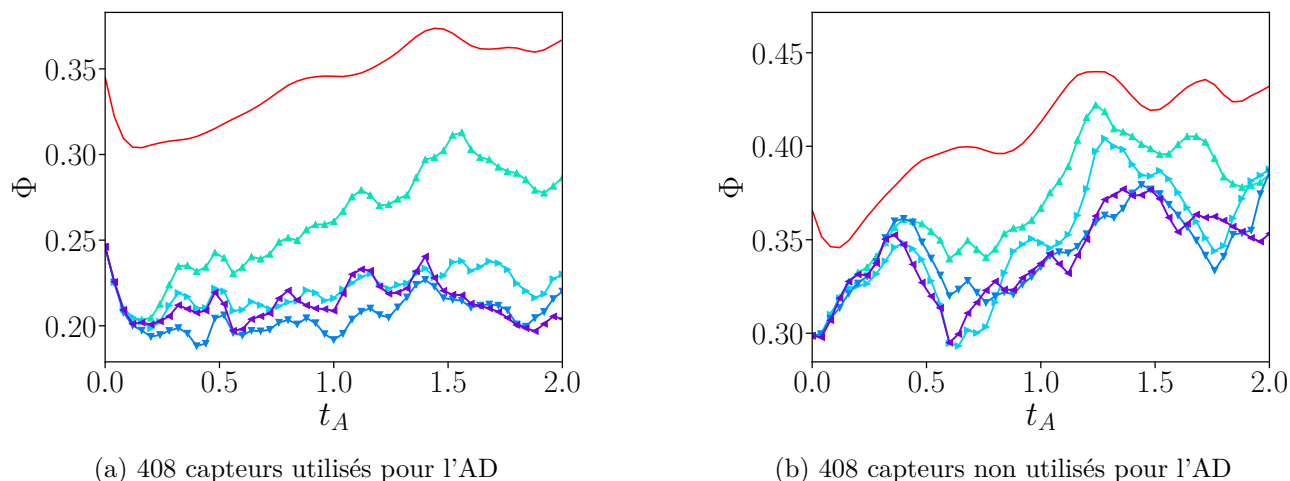


FIGURE 6.16 : Courbes de Φ : Du vert au violet - 0%, 5%, 15% et 25% d'inflation. La courbe rouge montre la limite moyenne obtenue en comparant les valeurs des simulations DA-LES à 1000 temps d'observations.

En résumé, ces résultats confirment que l'effet de l'EnKF n'est pas seulement local, mais grâce aux interactions d'échelle capturées par le modèle LES sous-jacent, une amélioration globale de la prédiction de l'écoulement instantané est obtenue. L'analyse de la quantité globale Φ souligne à quel point l'assimilation de données peut être importante pour fournir une estimation réussie de l'état instantané, ce qui pourrait être encore plus important qu'une optimisation paramétrique précise pour la prédiction d'événements extrêmes rares et pour l'optimisation d'écoulements non stationnaires présentant une forte évolution temporelle de ses caractéristiques.

Il convient de souligner que les propriétés de synchronisation de l'AD semblent universelles et qu'elles ne devraient pas être liées aux caractéristiques du cas test étudié, contrairement à l'optimisation des coefficients du modèle qui a été présentée dans la section précédente (Sec. 6.3).

6.5 Conclusion

Le potentiel de la stratégie d'assimilation de données séquentielle basée sur l'EnKF pour de futures applications de type jumeaux numériques a été étudié pour la première fois dans ce travail de thèse. L'algorithme piloté par les données est couplé à un ensemble de réalisations utilisant un modèle de simulation des grandes échelles. L'attention de ce chapitre est principalement consacrée à la représentation correcte des caractéristiques instantanées, qui peuvent être essentielles pour prédire et anticiper les événements extrêmes affectant les applications industrielles. Pour effectuer l'analyse, les simulations LES ont été combinées aux informations instantanées localisées de haute-fidélité obtenues à partir d'une DNS. Tout d'abord, les procédures d'AD ont utilisé des valeurs instantanées du champ de vitesse pour optimiser le comportement paramétrique du modèle de Smagorinsky utilisé pour la fermeture des équations des échelles de sous-maillages. Trois stratégies ont été proposées pour obtenir une valeur optimisée de la constante

de modèle C_k . Malgré la différence de complexité, les stratégies fournissent un résultat similaire, i.e. une réduction significative de l'intensité de C_k et du modèle de sous-maillages. L'optimisation paramétrique réduit l'écart entre les moments statistiques des champs de vitesses et des données DNS, mais ne l'élimine pas. La raison derrière cette observation est associée aux limitations structurelles du modèle de Smagorinsky, dont la nature intrinsèquement dissipative n'est pas en mesure de prendre pleinement en compte les effets des échelles filtrées et leurs interactions avec l'écoulement résolu, comme observé par Mons et al. [87].

Le modèle d'AD a ensuite été utilisé pour analyser l'efficacité de la reconstruction et de la synchronisation des écoulements avec les données échantillonnées haute-fidélité disponibles. Il a été démontré que l'AD est capable d'améliorer significativement la corrélation entre les résultats du modèle et l'observation, mais l'efficacité d'une telle synchronisation est régie par la variance de l'ensemble pouvant être piloté par l'inflation d'état. Cet hyper-paramètre est une caractéristique clé essentielle de l'algorithme d'AD qui mérite des études plus spécifiques à l'avenir. De même, les effets de la localisation physique et de covariance, qui ont été exclus dans la présente analyse, sont étudiés plus profondément dans l'Annexe B traitant de la validation de l'algorithme hyper-localisé du EnKF à l'aide du canal plan.

Écoulement interne compressible et instationnaire d'une géométrie de moteur

Sommaire

7.1	Introduction	133
7.2	Configurations numériques	135
7.2.1	Présentation de la géométrie et validation dans le cas stationnaire	135
7.2.2	Simulation de référence haute fidélité (LES-HF)	137
7.2.3	Simulation basse fidélité (LES-LF)	141
7.3	Calibration de la condition d'entrée	142
7.3.1	Mise en place de l'assimilation de données	142
7.3.2	Structure hyper-paramétrique et conditions initiales	146
7.3.3	Résultats	146
7.4	Synchronisation et reconstruction de structures locales	152
7.4.1	Détails de configuration	152
7.4.2	Capacités de synchronisation	157
7.5	Conclusion	163

7.1 Introduction

Le [chapitre 6](#) a démontré l'importance du niveau de variance de l'ensemble dans la précision de la correction apporté par le filtre de Kalman d'ensemble. Cette étude a de plus exposé les capacités de synchronisation du EnKF pour un écoulement statistiquement stationnaire

incompressible académique modélisé à l'aide de la LES. La simulation des grandes échelles est également capable de prédire les phénomènes instationnaires de configurations transitoires, par exemple pour la représentation de phénomènes non cycliques [109, 108] et d'écoulements transitoires [102] dans les moteurs à combustion interne. L'assimilation de données séquentielle est maintenant utilisée dans le cas d'une simulation LES compressible d'écoulement complexe et instationnaire de moteur.

La simulation d'un moteur à quatre cylindres typique complet requiert une quantité de ressources peu envisageable dans le cadre de cette étude. Le banc à volute stationnaire présente un grand intérêt dans la modélisation de l'aérodynamique d'un moteur. Cette géométrie est en particulier utilisée par les motoristes dans les premières phases de développement pour optimiser les conduites d'admission ou encore la géométrie de soupape [129]. Elle est caractérisée par une seule soupape insérée dans une conduite d'admission éjectant le fluide dans une chambre cylindrique. La distance de levée de soupape est fixe, cette dernière est donc immobile. L'écoulement du banc à volute stationnaire se caractérise par une turbulence élevée anisotrope et inhomogène, des effets de compressibilité pouvant modifier l'écoulement, des phénomènes de parois ou encore le vrillage de l'écoulement (qui n'est pas étudié ici) [1]. L'écoulement expulsé par la conduite d'admission autour de la soupape forme un jet soumis à de forts cisaillements. L'écoulement vient impacter la paroi du cylindre de chaque côté de la soupape, ce qui génère des structures rotationnelles. Un sujet d'actualité concerne l'étude des variations cycle à cycle dans le moteur à combustion interne [29, 30]. Les variabilités cycliques de combustion sont dues à plusieurs facteurs, parmi lesquels l'un des plus complexes à identifier provient de la variabilité du jet d'admission, du fait de la nature intrinsèquement non linéaire et complexe de la turbulence instationnaire combiné à leurs interactions avec les parois et de perturbations extérieures. L'étude précise de ces événements et de leur origine reste difficile. La prédiction exacte des conditions aux limites, quand elle est à l'origine de ces aléas, pourrait permettre leur étude via le prisme numérique.

Dans le domaine du pilotage par les données, relativement peu d'études se consacrent à la calibration de conditions aux limites via l'assimilation d'informations locales. Différentes méthodes sont utilisées pour la prédiction de conditions d'entrées telles que : la minimisation d'une fonction de coût par rapport à des données expérimentales pour la simulation d'écoulements à surface libre avec une méthode LBM [97], une optimisation basée sur les gradients pour une condition d'entrée de type turbulence synthétique d'une simulation LES de configuration urbaine [65] ou encore l'utilisation du *machine learning* pour reconstruire une condition d'entrée de tube d'aspiration de turbine hydraulique à partir d'informations aval de l'écoulement pour des simulations RANS et LES [139]. Dans le domaine de l'assimilation de données, seules deux études semblent se consacrer à l'inférence d'une condition d'entrée. Sousa et al. [122] utilisent un filtre de Kalman d'ensemble pour inférer la condition d'entrée d'une configuration urbaine à partir de mesures réelles de données anémométriques pour des simulations RANS. Moussie et al. [92] ont pour objectif de réduire les écarts entre la simulation numérique et les mesures expérimentales grâce à l'EnKF en modifiant une condition d'entrée de turbulence synthétique pour un corps non profilé. Aucune de ces études ne se consacre à l'inférence de conditions d'entrée instationnaires et à la reconstruction de structures locales représentatives d'événements marginaux. Ces derniers points sont cruciaux pour le développement d'applications respectant

les principes du paradigme du jumeau numérique. Ils sont essentiels pour améliorer l'efficacité et la robustesse du jumeau physique grâce à l'étude de son équivalent numérique.

Dans une première section 7.2, la géométrie du banc à volute stationnaire est validée dans un cas stationnaire à l'aide de données expérimentales disponibles [130]. Le banc à volute stationnaire est par la suite utilisé pour étudier le calibrage d'une condition d'entrée instationnaire sinusoïdale à partir d'informations locales de l'écoulement grâce à l'assimilation de données séquentielle dans la Sec. 7.3. L'annexe B a permis de valider l'utilisation de l'algorithme séquentiel fortement localisé du filtre de Kalman d'ensemble hyper-localisé (HLEnKF) pour ce cas complexe et exigeant. Ce dernier est mis à contribution pour fournir une synchronisation locale des champs de vitesse à partir des données observées dans la Sec. 7.4. Ce chapitre est clos par une conclusion de ces derniers résultats ainsi que l'énoncé de nouvelles perspectives pour l'assimilation de données séquentielle dans le paradigme du jumeau numérique dans la Sec. 7.5.

7.2 Configurations numériques

La géométrie du banc volute stationnaire est dans un premier temps explicitée. Le cas de référence haute-fidélité est ensuite validé par comparaison d'un écoulement stationnaire à des données expérimentales disponibles. L'équivalent basse-fidélité est enfin abordé.

7.2.1 Présentation de la géométrie et validation dans le cas stationnaire

Le domaine d'étude est une géométrie simplifiée de moteur appelé le banc volute stationnaire. Il a notamment été utilisé pour diverses études faisant intervenir la simulation des grandes échelles [2, 94, 130, 46]. Ces travaux s'inspirent de l'expansion asymétrique de Dellenback et al. [26] utilisée en 1988 pour la prise de mesures d'un écoulement turbulent vrillé. Le domaine d'étude correspondant à la géométrie étudiée est présenté schématiquement sur la Fig. 7.1. Il se compose d'une conduite d'entrée dans laquelle sont insérés un guide et une soupape fixe ainsi que d'une chambre cylindrique recevant le jet de soupape. Cet écoulement est représentatif de l'admission dans un moteur à allumage commandé. Ce type de géométrie est largement utilisé dans les phases d'études préliminaires des motoristes pour l'optimisation des dimensions des conduites, des formes de soupapes ou encore des niveaux de mise en rotation de l'écoulement [129]. Toutes les spécifications géométriques sont indiquées sur le schéma. Les valeurs principales sont le diamètre de la soupape $D_v = 40$ mm, le rayon de la chambre de sortie $R = D_p/2 = 60$ mm, sa longueur $L_p = 500$ mm ainsi que la longueur de la conduite d'entrée $L_e = 104$ mm. La tête de la soupape est à une distance $L_{vh} = 10$ mm de la fin de la conduite d'entrée.

Pour mener à bien l'étude menée dans ce chapitre, une base de données de référence est dans un premier temps créée à partir d'une simulation haute-fidélité. Cette base de données sera par la suite utilisée pour l'assimilation de données sur des simulations basse-fidélité avec un modèle de turbulence identique. Le solveur utilisé est *rhoPimpleFoam* d'OpenFOAM pour la simulation d'écoulements compressibles avec l'algorithme PIMPLE (voir section 2.2). Un schéma implicite du second ordre a été utilisé pour l'avancement temporel de la solution. Des schémas implicites centrés du second ordre ont été utilisés pour la discrétisation des dérivées spatiales. Un schéma du second ordre 75% centré et 25% avant est utilisé pour la divergence de la vitesse et de

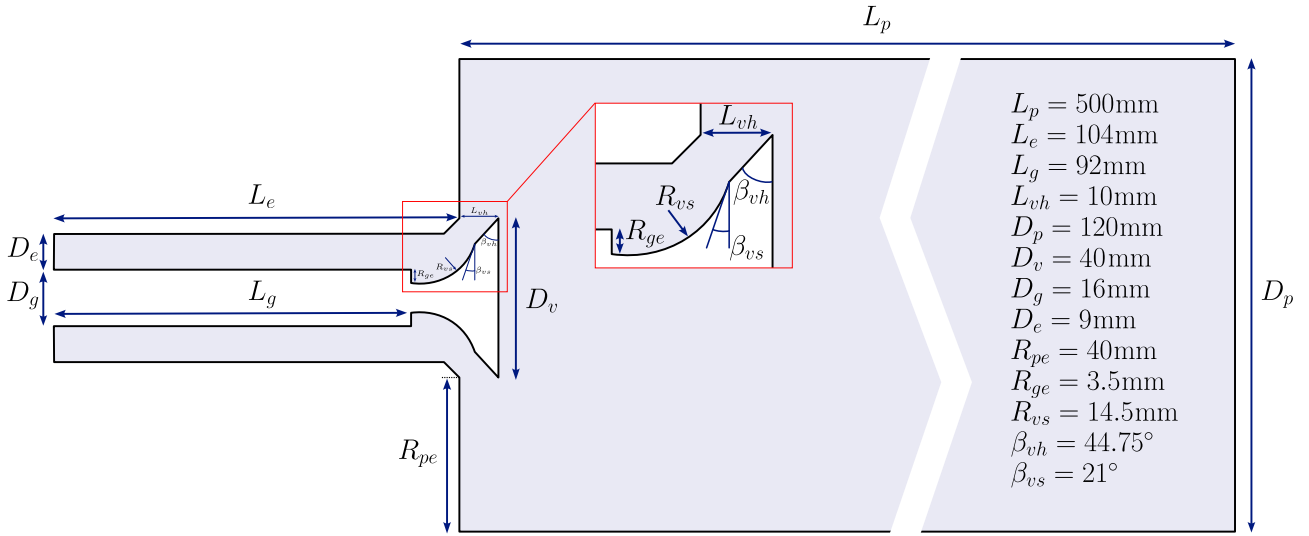


FIGURE 7.1 : Schéma de la géométrie simplifiée de moteur étudiée comportant une seule soupape.

l'énergie afin d'améliorer la stabilité du système. Aucune combustion n'est simulée dans le domaine. Seul l'écoulement de l'air est étudié. Le fluide est considéré comme un gaz parfait. Ses propriétés initiales, prises entre 20 °C et 25 °C, sont imposées dans les fichiers de configuration d'OpenFOAM. Elles sont les suivantes : masse molaire $M^{mol} = 28.9 \text{ g mol}^{-1}$, chaleur spécifique à volume constant $C_v = 718 \text{ kJ K}^{-1}$, viscosité dynamique $\mu = 1.82 \times 10^{-5} \text{ kg m}^{-1} \text{ s}^{-1}$ ainsi qu'un nombre de Prandtl $Pr = 0.708$. L'écoulement est compressible, la masse volumique est déduite à chaque pas de temps grâce à l'équation des gaz parfaits. La condition aux parois est une condition de non-glissement. Pour la vitesse, la condition de sortie *pressureInletOutletVelocity* spécifie un gradient nul basé sur la pression dynamique sans recirculation de l'écoulement. La condition *waveTransmissive* est imposée en sortie pour le champ de pression avec une pression de référence de $p_{atm} = 101\,325 \text{ Pa}$. Elle permet d'éviter la réflexion des ondes sur le patch de sortie du domaine. Le gradient de pression est nul en entrée. Une température fixe de 295 K est imposée en entrée, et le gradient de température est nul en sortie.

Thobois et al. [129] ont réalisé des mesures de vitesses par *Laser Doppler Anemometry* (LDA) sur un banc volute stationnaire identique. Ces données sont utilisées pour la validation de la configuration de référence dans le cas d'un écoulement stationnaire. La vitesse débitante axiale en entrée du domaine est donc la même, soit $u_b = \iiint_{V_D} u_x dV' / V_D = 65 \text{ m s}^{-1}$. Ceci entraîne un débit massique équivalent $Q_m = 0.055 \text{ kg s}^{-1}$ ainsi qu'un nombre de Reynolds calculé à partir de l'épaisseur de la conduite annulaire d'admission de $Re = 37500$. Le maillage de cette simulation haute-fidélité, montré sur la Fig. 7.2, est constitué de 131 764 400 éléments. Il a été généré grâce à l'outil *blockMesh* d'OpenFOAM. Il est donc structuré et composé uniquement d'hexaèdres. La taille des éléments de maille est calibrée de sorte que la distance à la paroi r adimensionnée par l'échelle visqueuse soit unitaire dans la conduite d'admission $\Delta r_1^+ = r / \delta_\nu = r u_\tau / \nu \approx 1$, avec $u_\tau = \sqrt{\tau_w / \rho}$ la vitesse de frottement à la paroi, τ_w la contrainte de frottement et ρ la masse volumique. La taille des mailles augmente ensuite progressivement dans le cylindre. Le calcul est réalisé grâce à une simulation LES avec le modèle de turbulence WALE explicité dans le chapitre 2. Pour la validation dans le cas stationnaire, le calcul

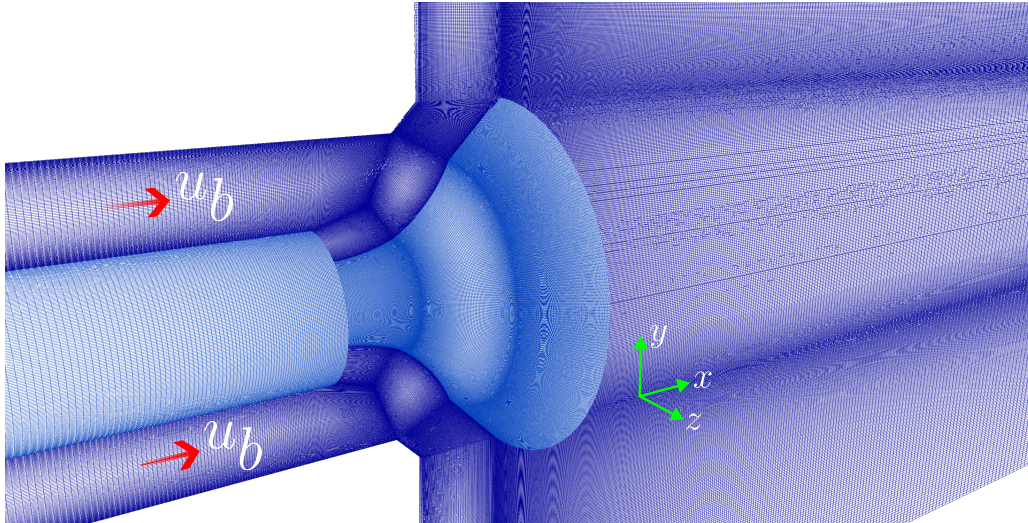


FIGURE 7.2 : Zoom du maillage structuré utilisé pour la simulation haute-fidélité de référence. Le maillage est affiché en bleu clair sur la soupape et bleu foncé sur un plan en $z = 0$ pour le domaine d'étude.

des moyennes de vitesse est exécuté pour une durée de 8 temps advectifs avec $t_A = D_e/u_b$. Bien qu'étant assez faible, ce temps de calcul de moyenne est suffisant pour la validation de la configuration. Par ailleurs, le cas stationnaire ne sera pas utilisé pour la production de données pour l'assimilation de données. Il est donc préférable d'utiliser le moins de ressources possible pour l'étape de validation de la configuration. La figure 7.3 montre le profil radial de vitesse axiale moyenne normalisé par la vitesse débitante \bar{u}_x/u_b . Les moyennes $\bar{\cdot}$ sont temporelles et azimutales. Les données expérimentales sont données pour une section $x = 0.124$ m en orange mais soumises à de fortes incertitudes de positionnement. Les données numériques sont données en bleu à gauche et à droite sur la Fig. 7.3 pour une section $x = 0.124$ m et $x = 0.126$ m. La très grande correspondance des profils pour cette dernière montre que la configuration haute-fidélité est cohérente du point de vue topologique de l'écoulement, en particulier dans la taille et la position des structures de recirculations.

7.2.2 Simulation de référence haute fidélité (LES-HF)

La configuration de simulation a été validée dans un cas stationnaire par comparaison avec les données expérimentales disponibles. Pour rappel, l'objectif de l'étude est la calibration de conditions d'entrées et l'inférence de phénomènes instationnaires. Une condition instationnaire est donc imposée dans la simulation de référence.

La condition d'entrée instationnaire est une onde sinusoïdale de vitesse telle que :

$$U_{inlet} = a_{ref} \sin(2\pi f_{ref} t + \phi_{ref}) + U_o \quad (7.1)$$

U_o (pour "offset") correspond à la composante moyenne de l'onde. Celle-ci est égale à la vitesse débitante établie lors du cas stationnaire de sorte que l'onde sinusoïdale oscille autour de la condition stationnaire validée, $U_o = u_b$. La phase ϕ_{ref} est nulle. L'amplitude représente une

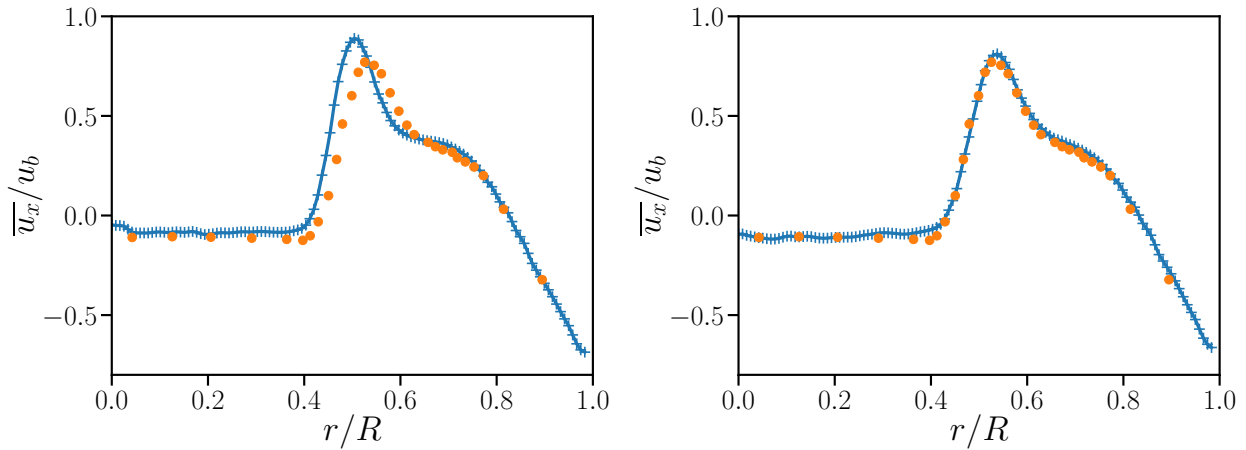


FIGURE 7.3 : Profils radiaux de vitesses axiales pour deux section $x = 0.124$ mm à gauche et $x = 0.126$ mm à droite. En orange : données expérimentales de Thobois et al. [130]. En bleu : données issues de la simulation LES haute-fidélité

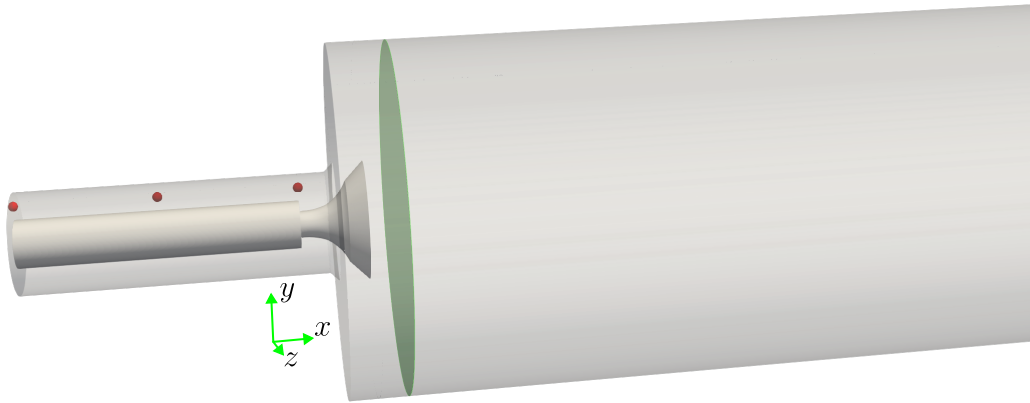


FIGURE 7.4 : Placement des capteurs utilisés pour mesurer l'évolution de la condition d'entrée dans la conduite d'admission. La section 2D $x = 0.124$ m régulièrement utilisée dans ce travail est indiquée en vert.

oscillation de 15% autour de la vitesse débitante, $a_{ref} = 9.75 \text{ m s}^{-1}$. La fréquence de référence $f_{ref} = 800 \text{ Hz}$ est environ 6 à 10 fois plus élevée que les fréquences caractéristiques des signaux de pression et de vitesse dans les tubulures d'entrée liées à la propagation d'ondes acoustiques [100]. Ce choix permet de simuler numériquement un nombre plus élevé de périodes comparé aux valeurs de fréquences réalistes pour un coût équivalent. L'ordre de grandeur de ces fréquences est cependant similaire, les dynamiques attendues dans l'écoulement sont donc similaires. La période de l'onde sinusoïdale est $t_p = 1.25 \times 10^{-3} \text{ s}$. Le pas de temps de la simulation est $\Delta_t = 2 \times 10^{-8} \text{ s}$.

Les figures 7.5 montrent l'évolution de vitesse et de la pression sur les 3 capteurs positionnés en début, milieu et fin de la conduite d'admission indiqués sur la Fig. 7.4. Les évolutions de vitesses et de pressions montrent des dynamiques très différentes selon leur position dans la conduite. En particulier, les dynamiques de vitesses sont bien plus élevées à la fin de la conduite

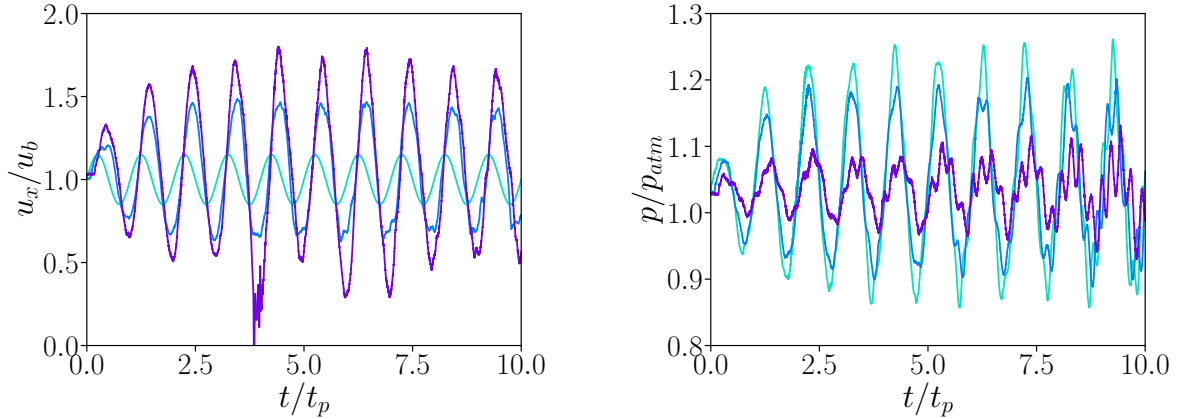


FIGURE 7.5 : Évolution de la vitesse axiale u_x adimensionnée par la vitesse débitante u_b et de la pression p adimensionnée par la pression atmosphérique p_{atm} en fonction du temps. Les valeurs sont données pour 3 capteurs disposés le long de la conduite d'admission de la LES-HF en $x = 0$ bleu clair, 0.046 bleu foncé, 0.092 violet, au centre de la conduite d'entrée en $z = 0$ et $y = 0.0125$.

ou l'amplitude des variations atteint 60 à 70% de la vitesse débitante. De même, la pression semble exposer des variations similaires. Ceci implique une très grande variation du champ de vitesse dans la conduite et dans le cylindre pour des variations faibles de la condition d'entrée. De plus, l'évolution de l'amplitude de l'onde sinusoïdale traverse un régime transitoire jusqu'à la période 6 avant de se stabiliser. Seules les périodes 7 à 10 seront donc utilisées pour l'assimilation de données Sec. 7.3 et 7.4.

La figure 7.6 présente le champ de vitesse dans le domaine pour une section en $z = 0$ ainsi que la pression indiquée sur la surface du guide et de la soupape pour la 9ème période. Des zones de recirculations sont visibles de part et d'autre du jet. La première est facilement identifiable et se situe dans les coins de la chambre cylindrique. La deuxième, plus large, correspond à la recirculation de l'écoulement vers l'intérieur du cylindre, i.e. dans la zone centrale en aval de la soupape. L'onde sinusoïdale imposée en entrée est également visible sur la figure au vu de la diminution et augmentation de vitesse présente dans la conduite d'admission.

La LES-HF est utilisée comme simulation de référence pour l'assimilation de données. Des mesures sont effectuées en aval de la soupape pour créer une base de données cohérente et exploitable comme observation pour l'assimilation de données. Cet échantillonnage est donc réalisé à partir de la 6ème période à partir de laquelle l'oscillation de l'onde d'entrée semble avoir atteint un régime permanent. Pour se rapprocher du cadre de travail expérimental, c'est-à-dire de mesures PIV d'un système réel, les mesures sont réalisées sur des plans 2D en $x = 0.124$ m, $x = 0.144$ m, $x = 0.164$ m, $y = 0$ et $z = 0$. Pour couvrir une large gamme d'observations potentielles, un total de 56 226 capteurs est utilisé pour la prise de mesures. La Fig. 7.7 montre leur répartition. En outre, la densité de capteurs est plus élevée dans la zone de passage du jet sortant de la conduite d'admission. La vitesse est mesurée aux positions des capteurs tous les $\Delta_{tm} = 50\Delta_t = 1 \times 10^{-6}$ s. Ceci correspond à 1 250 mesures par période de l'onde sinusoïdale $t_p = 1\,250\Delta_{tm} = 62\,500\Delta_t$. D'après une étude visuelle des champs de la simulation haute-fidélité, les plus petites instabilités de Kelvin Helmholtz sont de l'ordre de 0.0005 m et 0.005 m pour les

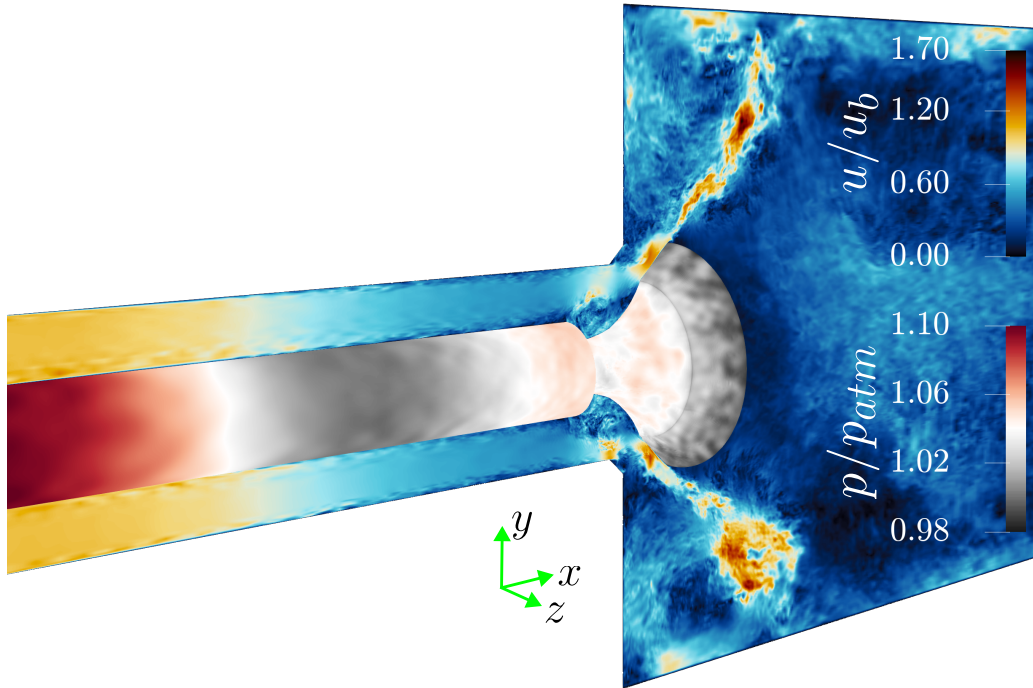


FIGURE 7.6 : Champ de vitesse de la simulation haute fidélité de référence en $t = 9t_p$ sur une section telle que $z = 0$. La pression est représentée sur la surface de la soupape.

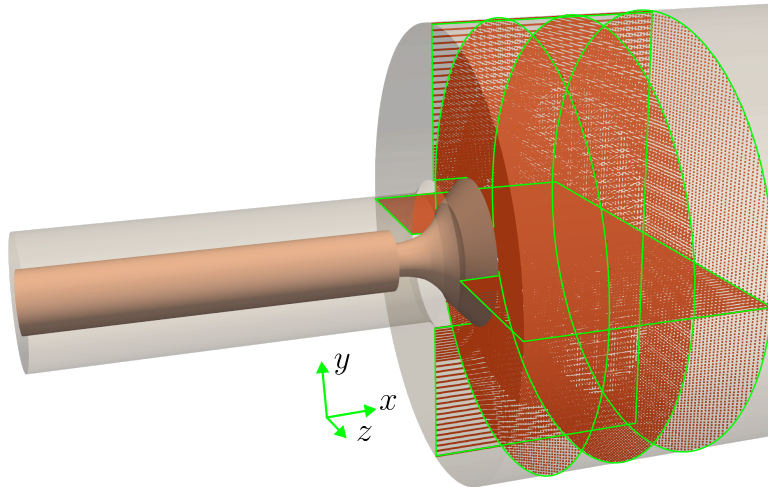


FIGURE 7.7 : Les points oranges représentent la position des capteurs pour la mesure des données. Ils sont répartis sur 5 plans : $x = 0.124$, $x = 0.144$, $x = 0.164$, $y = 0$ et $z = 0$

plus grandes structures cohérentes. Le temps advectif en fonction de la vitesse débitante pour ces instabilités est d'environ 8×10^{-6} s pour les plus petites et 8×10^{-5} s pour les plus grandes. Suivant ce constat, au minimum 8 mesures sont réalisées dans le temps advectif des plus petites instabilités et jusqu'à 80 pour les plus grandes. La résolution temporelle de la prise de mesure est donc suffisante pour capter ces plus petites structures intervenant dans la dynamique de l'écoulement.

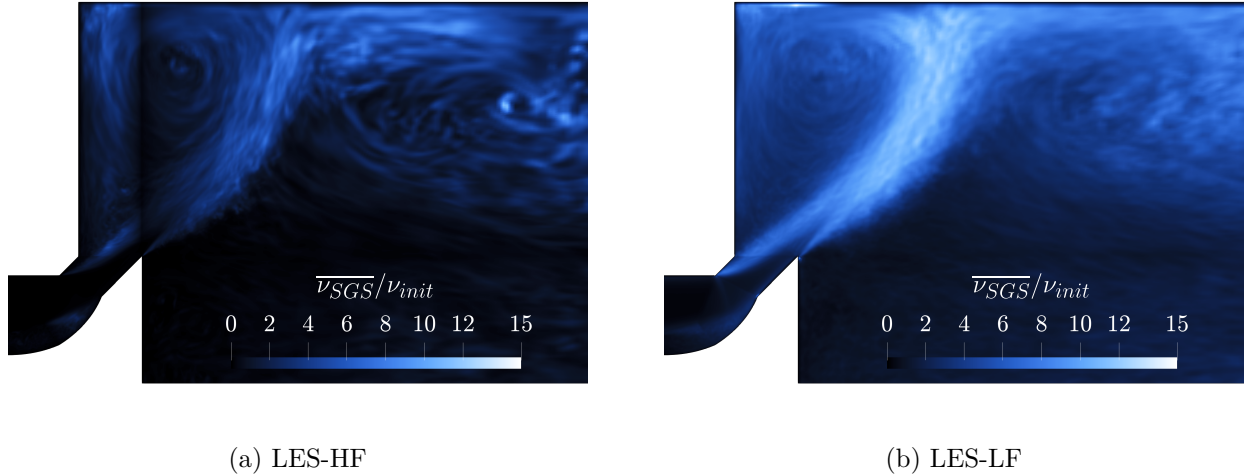


FIGURE 7.8 : Champs de viscosités de sous-maille moyens adimensionnés par la viscosité cinématique initiale ν_{SGS}/ν_{init} sur un plan 2D en $z = 0$ pour les configurations haute et basse fidélité.

7.2.3 Simulation basse fidélité (LES-LF)

Une simulation moins résolue est également menée. Cette simulation sera la base de l'assimilation de données qui utilisera les mesures issues de la simulation haute-fidélité LES-HF. La LES-LF (*low fidelity*) est donc également une simulation LES utilisant le modèle WALE. Son maillage est constitué de 8 872 800 éléments. La même structure de maillage que pour la LES-HF est utilisée. Cette fois, la taille des éléments de maille est calibrée de sorte que la distance à la paroi r adimensionnée par l'échelle visqueuse soit 4 à 5 fois supérieur dans la conduite d'admission $\Delta r_1^+ = r/\delta_\nu = ru_\tau/\nu \approx 4 - 5$ par rapport à la LES-HF. Cette différence de résolution entraîne une estimation de la viscosité de sous-maille par le modèle plus élevée, comme le montre la Fig. 7.8. Le champ moyen de viscosité de sous-maille $\overline{\nu_{SGS}}$ adimensionné par la viscosité cinématique initiale ν_{init} est environ 2 fois plus élevé en aval de la soupape pour la LES-LF par rapport à la LES-HF. La dynamique de l'écoulement est donc légèrement différente et quelque peu dégradée.

En effet, une fois un régime d'oscillations permanent atteint, les oscillations de vitesses dans la conduite d'admission présentent des variations plus élevées que pour l'équivalent haute-fidélité, en témoigne la Fig. 7.9. Cette fois, les maximums de vitesses atteints en fin de conduite d'admission représentent une variation de +100% par rapport à la vitesse débitante. Le minimum atteint par l'onde sinusoïdale représente une amplitude de 50 à 60%. Cette variation d'amplitude du signal sinusoïdal est particulièrement visible à la fin de la conduite d'admission sur le capteur précédent le jet de soupape. Cette différence essentielle pour le développement du jet dans le cylindre met en évidence l'erreur attribuable à la différence de comportement du modèle de turbulence du fait de la différence dans la résolution du maillage, i.e. dans les grandeurs filtrées par la LES et dans l'estimation de la viscosité turbulente.

La LES-LF est utilisée comme simulation de base pour l'assimilation de données dans les Sec. 7.3 et 7.4. En dehors des objectifs énoncés en introduction pour cette étude, il sera intéressant de voir si l'assimilation de données est capable de compenser la prédiction détériorée de la LES-LF dans la conduite d'admission en appliquant un écart de calibrage de la condition d'entrée

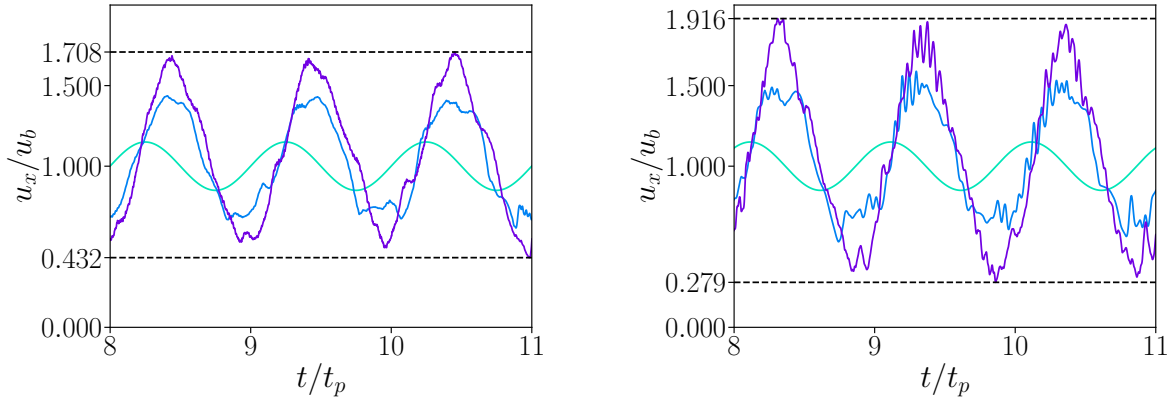


FIGURE 7.9 : Évolution de la vitesse axiale u_x adimensionnée par la vitesse débitante u_b en fonction du temps pour les configurations HF à gauche et LF à droite. Les valeurs sont données pour 3 capteurs disposés le long de la conduite d'admission de la LES-LF en $x = 0$ bleu clair, 0.046 bleu foncé, 0.092 violet, au centre de la conduite d'entrée en $z = 0$ et $y = 0.0125$.

par rapport aux valeurs de référence de la simulation LES-HF.

7.3 Calibration de la condition d'entrée

L'assimilation de données est effectuée sur la simulation basse fidélité (comparée à la référence) LES-LF qui sera la base de l'étude. Les données sont mesurées sur la simulation haute-fidélité LES-HF utilisant la même géométrie ainsi que le même modèle.

7.3.1 Mise en place de l'assimilation de données

Comme énoncé dans l'introduction de ce chapitre, la stratégie d'assimilation de données vise dans un premier temps à calibrer la condition d'entrée d'une simulation utilisant la configuration de la LES-LF. Cette calibration est réalisée en faisant l'hypothèse que la condition de référence est inconnue. C'est généralement le cas pour des simulations de parties réduites de systèmes réels pour lesquels les conditions locales d'entrée d'évènements extrêmes sont difficiles à mesurer. Une deuxième hypothèse implique l'existence des données d'observations, ce qui est rarement le cas pour les systèmes abordés dans ce chapitre. En revanche, cette hypothèse est valide en considérant l'étude d'un système expérimental et s'intègre parfaitement dans le paradigme du jumeau numérique qui se base sur une production de données de plus en plus grande pour les systèmes réels. La calibration d'entrée est réalisée grâce à l'algorithme d'assimilation de données du filtre de Kalman d'ensemble hyper-localisé explicité au [chapitre 3](#). Dans le cadre de cette calibration, seule l'estimation des paramètres est activée. L'estimation de l'état, c'est-à-dire l'inférence du champ de vitesse, n'est pour l'instant pas considérée. En effet, la modification des champs de vitesse pour des champs initialement très différents et éloignés des données de référence entraînerait des gradients élevés de vitesses pour les membres de l'ensemble pouvant conduire à la divergence du système. L'estimation de l'état sera étudiée dans un second temps

dans la Sec. 7.4 après la calibration effective des conditions d'entrée des membres d'ensemble. Plusieurs ingrédients sont nécessaires pour la procédure d'AD :

- *Un modèle* : ici le modèle numérique utilisé pour l'avancée temporelle de la solution est le LES-WALE utilisé pour la simulation LES-LF. C'est également le même que pour la LES-HF mais l'estimation de la viscosité de sous-maille est différente étant donné la différence de résolution de maillages.
- *Des observations* : les observations sont une sélection de la base de données créée à partir des 56 226 capteurs mesurant la vitesse sur la simulation de référence. Les capteurs choisis seront explicités pour chaque configuration d'AD. Les capteurs utilisés diffèrent notamment dans la Sec. 7.4 où la résolution et le placement des observations peuvent jouer un rôle.
- *C.O.N.E.S.* : l'outil développé est ici utilisé avec le solveur rhoPimpleFoam d'OpenFOAM pour la simulation d'écoulements compressibles avec l'algorithme Pimple (voir [algorithme 1](#)).

35 membres sont utilisés pour l'inférence des paramètres d'entrée. La taille $[N_{ext}, N_e]$ de la matrice d'état \mathbf{U} est donnée par $N_e = 35$ et $N_{ext} = N + N_\theta$. $N = 3n_{cells}$ est égal à trois fois le nombre de mailles utilisés dans la procédure d'AD. En effet, le nombre de degrés de liberté considérés dans la procédure correspond généralement aux trois composantes du champ de vitesse pour chacun des éléments de maillage n_{cells} . Il est cependant à noter que l'estimation paramétrique seule ne nécessite que la prise en compte des paramètres à optimiser dans la procédure d'assimilation de données, i.e. $N_{ext} = N_\theta$. Concernant N_θ , le nombre de paramètres optimisés par la procédure du HLEnKF ne correspond pas aux paramètres de l'onde sinusoïdale de référence LES-HF. En effet, la forme réelle d'une condition d'entrée pour ce type de configuration dans des conditions particulières est rarement connue. L'ambition donnée à la calibration de la condition d'entrée est donc de pouvoir reconstituer n'importe quelle condition de référence qui serait inconnue. Dans ce contexte, la description paramétrique attribuée aux 35 membres de l'ensemble doit permettre de reconstituer une large gamme de dynamiques. Pour cela, une somme d'ondes sinusoïdales telle que dictée par les séries de Fourier semble idéale. Pour une configuration de ce calibre, le coût du calcul est très dépendant du nombre de membres dans l'ensemble. Ce nombre de membres doit cependant être suffisamment élevé pour prédire avec précision la description paramétrique donnée par les données observées. Une somme de quatre ondes sinusoïdales est donc utilisée pour la calibration de la condition d'entrée telle que :

$$U_{inlet} = \sum_{i=1}^4 a_i \sin(2\pi f_i t + \phi_i) + U_o \quad (7.2)$$

Ceci conduit à un total de 13 paramètres à optimiser via la procédure d'AD, i.e. quatre paramètres pour les quatre amplitudes des ondes sinusoïdales, quatre fréquences, quatre phases ainsi qu'une valeur d'*offset* pour calibrer la moyenne de l'onde résultante. 400 capteurs d'observations de la LES-HF sont choisis pour la calibration de la condition d'entrée (d'autres capteurs sont choisis dans les sections suivantes). Les trois composantes de la vitesse sont observées, ce

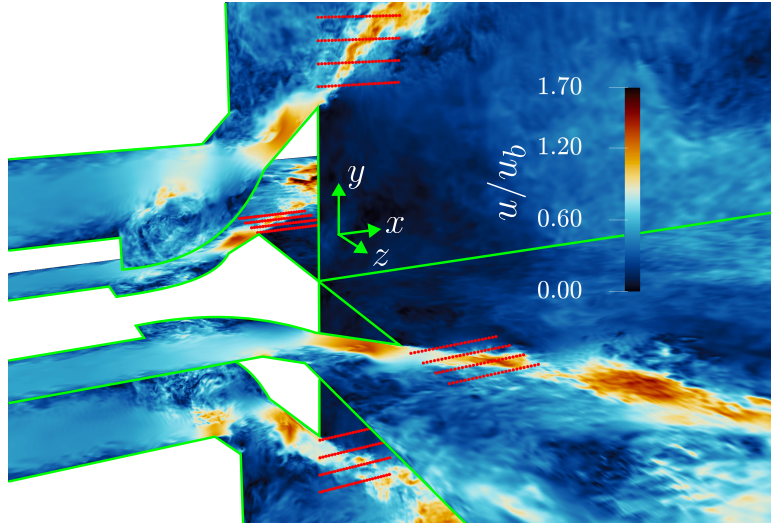


FIGURE 7.10 : Champ de vitesse de la simulation haute fidélité de référence en $t = 9t_p$ pour deux sections tel que $y = 0$ et $z = 0$ délimitées en vert. Les points rouges correspondent au capteurs d'observations sélectionnées pour la calibration de la condition d'entrée. Ces capteurs se situent sur des plans 2D $y = 0$ et $z = 0$ sur le chemin emprunté par le jet de soupape.

qui conduit à un total de 1200 observations pour la procédure de calibration. Les observations choisies se situent directement dans le jet en sortie de conduite d'admission, sur les plans $y = 0$ et $z = 0$ explicités dans la Sec. 7.2.2 comme le montre la Fig. 7.10. Seules les périodes 7 à 10 de la simulation de référence LES-HF sont utilisées pour l'assimilation de données. Elles sont donc utilisées en boucle au-delà de ces 4 périodes disponibles. L'utilisation du HLEnKF implique l'inférence d'une zone restreinte autour des capteurs d'observations. Pour des capteurs si rapprochés, les zones d'hyper-localisation sont coïncidentes, ce qui diminue l'intérêt d'une telle procédure. Néanmoins, seule l'estimation paramétrique est exécutée pour la procédure de calibration de la condition d'entrée. Le chevauchement des zones n'a donc que peu d'importance étant donné que l'état physique n'est pas corrigé.

Un élément à prendre en considération pour la procédure d'optimisation est le temps advectionnel depuis l'entrée du domaine jusqu'aux zones observées. Ce temps, calculé avec la vitesse débitante de référence, est supérieur à une période de référence $t_A^{inletDA} \approx 0.00175 \text{ s} > t_p$. Cette situation problématique est illustrée sur la Fig. 7.11. À chaque mise à jour de la description paramétrique par l'algorithme d'AD, l'information nouvelle transportée par le champ de vitesse vers les zones d'observations parcourt toute la conduite d'admission en un temps supérieur à la période de référence. Ceci limite le nombre de phases d'assimilation par périodes. Un nombre trop élevé de phases d'assimilation par période conduit à la divergence du système du fait de la comparaison successive d'informations statistiquement similaires. Dans ce cas, l'algorithme du *ensemble Kalman smoother* [36], qui permet la prise en compte d'informations passées et futures pour l'inférence du système, fournirait probablement de meilleurs résultats, mais pour un coût de calcul plus élevé. Idéalement, l'intervalle de temps de la fenêtre d'assimilation devrait donc être supérieur à $t_A^{inletDA}$ avec le HLEnKF. Toutefois, il reste possible d'utiliser quelques phases d'assimilations par période. Seules les premières phases d'assimilation sont véritable-

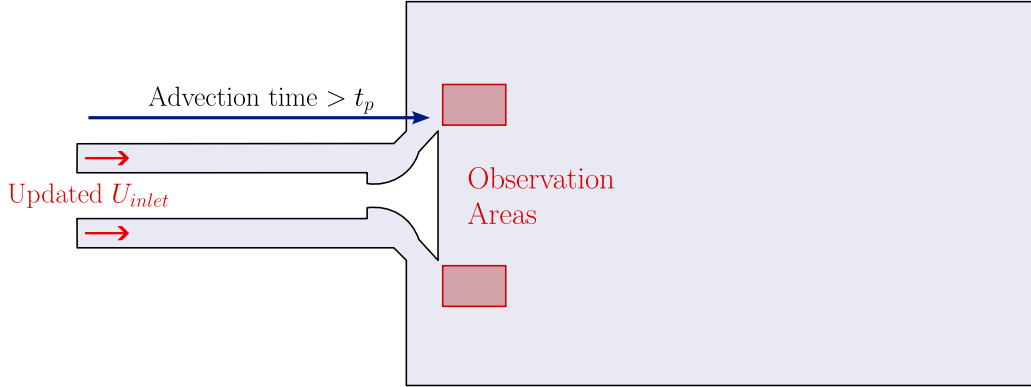


FIGURE 7.11 : Schéma illustrant le temps advectif de la transmission de l'information d'entrée mise à jour aux zones d'observations.

Distributions	<i>a priori</i> 1	<i>a priori</i> 2
Amplitudes	$a_1 \sim \mathcal{N}(8, 1)$, $a_2 \sim \mathcal{N}(5, 2)$ $a_2 \sim \mathcal{N}(2, 4)$, $a_4 \sim \mathcal{N}(0, 5)$	$a_1 \sim \mathcal{N}(9.75, 1)$, $a_2 \sim \mathcal{N}(0, 2)$ $a_2 \sim \mathcal{N}(0, 3)$, $a_4 \sim \mathcal{N}(0, 4)$
Fréquences	$f_i \sim \mathcal{N}(800, 100)$	$f_1 \sim \mathcal{N}(700, 100)$, $f_2 \sim \mathcal{N}(900, 100)$ $f_3 \sim \mathcal{N}(1100, 100)$, $f_4 \sim \mathcal{N}(1300, 100)$
Phases	$\phi_i \sim \mathcal{N}(0, 0.5)$	$\phi_i \sim \mathcal{N}(0, 0.5)$
Offset	$U_o \sim \mathcal{N}(65, 10)$	$U_o \sim \mathcal{N}(65, 10)$

TABLE 7.1 : Distributions *a priori* utilisées pour l'estimation paramétrique de la procédure d'assimilation de données.

ment critiques, puisqu'elles utilisent une information non inférée. Au-delà du temps advectif considéré, l'information comparée aux zones d'observations est une information glissante, correspondant à l'information inférée par la phase d'assimilation qui précède le temps $t - t_A^{inletDA}$. Pour cette étape de calibration de condition d'entrée, 5 phases d'assimilations sont effectuées par période. Ceci correspond à l'utilisation de 1 observation sur 250. Pour la même raison, la confiance donnée aux observations via la matrice de covariance d'erreur de mesure \mathbf{R} est relaxée durant les premières phases de calcul. La matrice est diagonale et exprimée comme $\mathbf{R} = \sigma_m^2 \mathbf{I}$, où σ_m quantifie l'incertitude des mesures. Une précision de 17.5% est appliquée en pourcentage aux valeurs de chaque observation durant les trois premières périodes inférées. Par la suite, $\sigma_m = 5\%$. L'inflation stochastique (voir la Sec. 3.3.3.1) est utilisée pour augmenter la variance de la description paramétrique au cours du calcul. L'onde résultante de la condition d'entrée des membres d'ensemble dépend des 4 fonctions sinusoidales utilisées pour la former, qui dépendent elles-mêmes des paramètres les décrivant. L'inflation de ces derniers a donc une très grande influence sur la variance de l'onde résultante. Une valeur faible d'inflation est donc utilisée après la première période inférée, soit $\lambda = 0.5\%$. À partir de la 9ème période inférée, l'inflation est abaissée à $\lambda = 0.375\%$.

7.3.2 Structure hyper-paramétrique et conditions initiales

L'état initial des 35 membres d'ensemble est uniquement défini par la variation de la description paramétrique. La distribution initiale des paramètres est décrite par des distributions gaussiennes aléatoires. Deux essais principaux sont réalisés. Ils sont résumés par le Tab. 7.1. La sélection de paramètres est réalisée aléatoirement via les distributions gaussiennes bornées par plus ou moins deux écarts types. Ces distributions sont à choisir avec précaution dans le paradigme de l'EnKF. La précision de la prédiction est soumise au choix des distributions *a priori* de l'assimilation, en particulier celles-ci doivent idéalement permettre de cibler une gamme de résultats dans laquelle se trouve la condition de référence [4]. La distribution nommée *a priori* 1 se concentre sur la variabilité des amplitudes des ondes. Ce paramètre est le plus délicat à configurer. L'addition d'ondes d'amplitudes équivalentes peut entraîner un décalage conséquent de l'onde résultante avec la référence. Les amplitudes de la distribution nommée *a priori* 2 sont au contraire calibrées pour se recentrer sur la condition de référence. L'espace des fréquences est lui élargi pour permettre de cibler une large gamme de fréquences. Dans les deux cas, les paramètres de phases et d'offset ne sont pas le sujet d'étude puisque (1) une distribution centrée sur 0 suffit à corriger le faible déphasage pouvant exister dans la condition de référence (2) la calibration de l'offset n'est pas soumise à l'interaction des ondes, sa calibration est donc relativement simple comparée aux autres paramètres. Par ailleurs, la calibration d'une condition d'entrée moyenne telle que l'offset présenté ici est similaire à la calibration de la condition de l'exemple présenté dans le chapitre 4 pour lequel le EnKF et le HLEnKF montrent d'excellents résultats malgré une configuration fortement réduite.

Les deux distributions sont exposées sur les Fig. 7.12. L'onde résultante de la description paramétrique est tracée en gris pour chaque membre de l'ensemble. Leur moyenne est indiquée en bleu. Le tracé des courbes est réalisé grâce à un script python utilisant les paramètres imposés sur les simulations CFD. Elles sont un indicateur de l'évolution des conditions d'entrée sans aucune modification de la part de l'algorithme d'assimilation de données. Comme prévu, l'*a priori* 1 présente des variations d'amplitudes moyennes plus élevées que pour l'*a priori* 2. Ce dernier montre en revanche des dynamiques de plus hautes fréquences que l'*a priori* 1. L'*a priori* 2 est également utilisé dans la Sec. Annexe C.1.2 où l'estimation paramétrique est réalisée à partir d'un jeu d'observations différent. Pour toutes les procédures, les champs initiaux correspondent aux champs de la LES-LF à $t = 10t_p$. Les calculs sont exécutés durant $2t_p$ pour propager les conditions d'entrées initiales des 35 membres de l'ensemble. Par la suite, 5 phases d'assimilation de données sont exécutées par période, la première phase se produisant à $t = 2.2t_p$.

7.3.3 Résultats

Les résultats de la calibration des conditions d'entrée pour l'*a priori* 1 et l'*a priori* 2 sont discutés dans cette section. Les figures 7.13 et 7.14 présentent les résultats de ces optimisations mesurées par les capteurs utilisées dans la conduite d'admission pour la configuration de la LES-HF 7.4. En particulier, la colonne de gauche correspond au capteur placé à l'entrée de la conduite, et donc à la condition directement imposée dans le fichier de configuration d'OpenFOAM. La colonne du milieu et celle de droite correspondent aux capteurs localisés au milieu

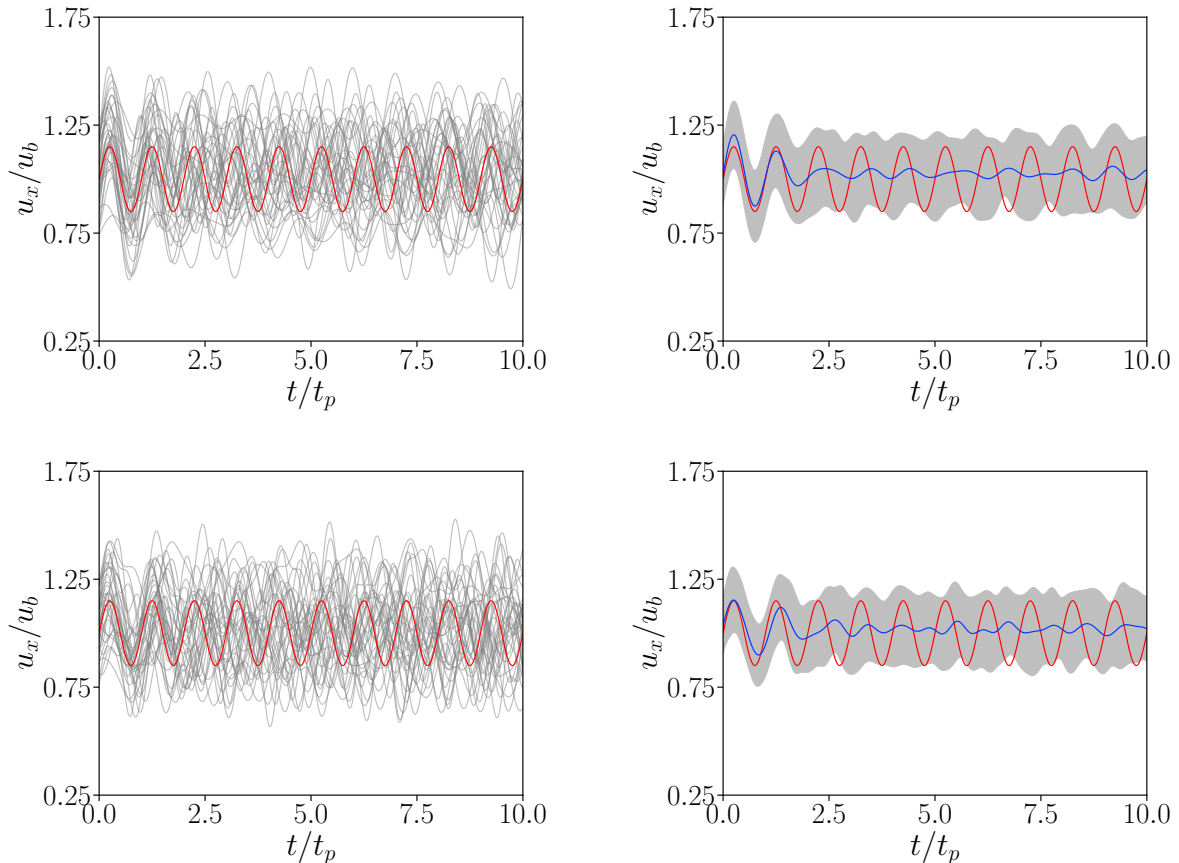


FIGURE 7.12 : Évolution de la condition d'entrée de vitesse axiale pour la distribution *a priori* 1 ligne du haut et *a priori* 2 ligne du bas. La courbe rouge correspond à la condition d'entrée de référence utilisée pour la LES-HF. La vitesse est adimensionnée par la vitesse débitante de référence u_b , le temps par la période de l'onde sinusoïdale de référence. Les courbes grises (gauche) correspondent aux 35 membres d'ensemble. La courbe bleue (droite) correspond à la moyenne des conditions des membres d'ensemble, la zone grisée indique l'écart type par rapport à la moyenne.

de la conduite et à la fin. Pour les deux procédures, l'action du HLEnKF est comme prévu visible à partir de la deuxième période à l'entrée de la conduite. Du fait du temps d'advection de l'information corrigée, ses effets sont décalés au milieu et à la fin de la conduite d'admission. Dans les deux procédures étudiées, l'optimisation réussie à converger vers la condition de référence. Un régime pseudo-stationnaire semble être atteint à partir de la période 5 ou 6. Ces résultats montrent que le filtre de Kalman d'ensemble hyper-localisé est capable de prédire une condition d'entrée sinusoïdale à partir de capteurs situés en aval de la soupape, malgré le temps d'advection élevé nécessaire à la propagation de l'information corrigée. En dépit de conditions initiales *a priori* relativement différentes, l'algorithme est capable de prédire l'onde sinusoïdale de la LES-HF dans les deux configurations étudiées résumées dans le Tab. 7.1. Une large gamme de valeurs *a priori* pour les amplitudes, les fréquences, les phases et l'offset permettent une calibration satisfaisante de l'onde sinusoïdale d'entrée. Ceci démontre la robustesse

de l'algorithme aux variations initiales des paramètres choisis pour les membres de l'ensemble.

En outre, les deux procédures semblent présenter le même phénomène de variation d'amplitude observé sur la configuration de base LES-LF. Ce phénomène est particulièrement visible pour l'optimisation de l'*a priori* 2 au capteur situé en entrée de la conduite d'admission sur la Fig. 7.14d. En effet, une erreur de 4 à 6% de l'amplitude maximale est visible entre la moyenne d'ensemble de l'onde sinusoïdale en bleue et la référence en rouge. Cet écart se réduit fortement au milieu de la conduite sur la Fig. 7.14e et est presque nul en fin de conduite d'admission sur la Fig. 7.14f. Pour rappel, la configuration basse fidélité LES-LF exposée dans la Sec. 7.2.3 montrait que l'amplitude de l'onde sinusoïdale augmentait davantage lors de sa propagation dans la conduite d'admission pour cette configuration basse-fidélité que pour la haute-fidélité. Il semble que la procédure d'assimilation de données compense en partie ce phénomène. L'amplitude est plus faible en entrée sur la Fig. 7.14d. Cette diminution se résorbe progressivement jusqu'à la zone d'expulsion du jet sur la Fig. 7.14f, i.e. aux abords des zones d'observations du HLEnKF. Il semble donc que la procédure d'assimilation soit non seulement capable d'identifier la condition d'entrée sinusoïdale de référence à partir d'informations locales mais aussi de compenser la prédiction erronée du modèle LES basse fidélité par l'altération de cette même condition de référence.

L'évolution temporelle de la moyenne d'ensemble des paramètres des quatre ondes sinusoïdales ainsi que du paramètre d'offset est donnée sur les Fig. 7.15 et 7.16 pour les deux configurations. Le paramètre d'offset (a), les amplitudes (b) ainsi que les fréquences (c) sont adimensionnés par les valeurs de référence utilisées dans la LES-HF. Tout d'abord, le paramètre d'offset (a) semble converger vers une valeur 2 à 4% inférieure à la référence dans les deux configurations. Ceci vient appuyer le constat dressé sur le décalage de la vitesse moyenne dans la conduite d'admission. Une calibration parfaite des ondes entraînerait la convergence de 3 amplitudes vers 0 pour ne conserver qu'une seule onde sinusoïdale identique à l'onde de référence. Ce n'est pas le cas ici. Dans les deux cas, les paramètres d'amplitudes convergent assez rapidement (6 à 8 périodes) vers des valeurs proches des moyennes des distributions gaussiennes utilisées pour créer les conditions *a priori*. La plus grosse variation est observée pour l'amplitude 3 de l'*a priori* 1 en vert sur la Fig. 7.15b. Pour ce paramètre, une variation d'environ 15% est observée par rapport à la valeur initiale. Les amplitudes de l'*a priori* 2 semblent converger vers des valeurs proches de 0 mais négatives pour certaines, entraînant une compensation des ondes. Cette compensation s'effectue de manière différente selon la configuration. Pour l'*a priori* 1, elle passe par une calibration dynamique des fréquences. Pour l'*a priori* 2 la compensation s'effectue de manière assez évidente par le déphasage des ondes, alors que les paramètres de fréquences convergent en 8 périodes environ.

La description paramétrique de chaque configuration peut être isolée pour certains pas de temps. Pour cela, les paramètres aux temps $t = 6t_p$, $t = 9t_p$ et $t = 12t_p$ sont séparément utilisés dans un script python pour tracer et visualiser l'onde résultante sur un temps plus long. Le tracé de ces ondes est donné sur la Fig. 7.17. La première ligne montre l'évolution de l'onde résultante par rapport à l'onde de référence pour l'*a priori* 1 et la deuxième pour l'*a priori* 2. Dans les deux cas, le tracé est effectué jusqu'à 100 périodes, ce qui permet d'apprécier la modulation de l'onde résultante dans le cas de la première configuration. En effet, ce phénomène de battement résulte de la superposition des 4 ondes aux fréquences proches mais amplitudes

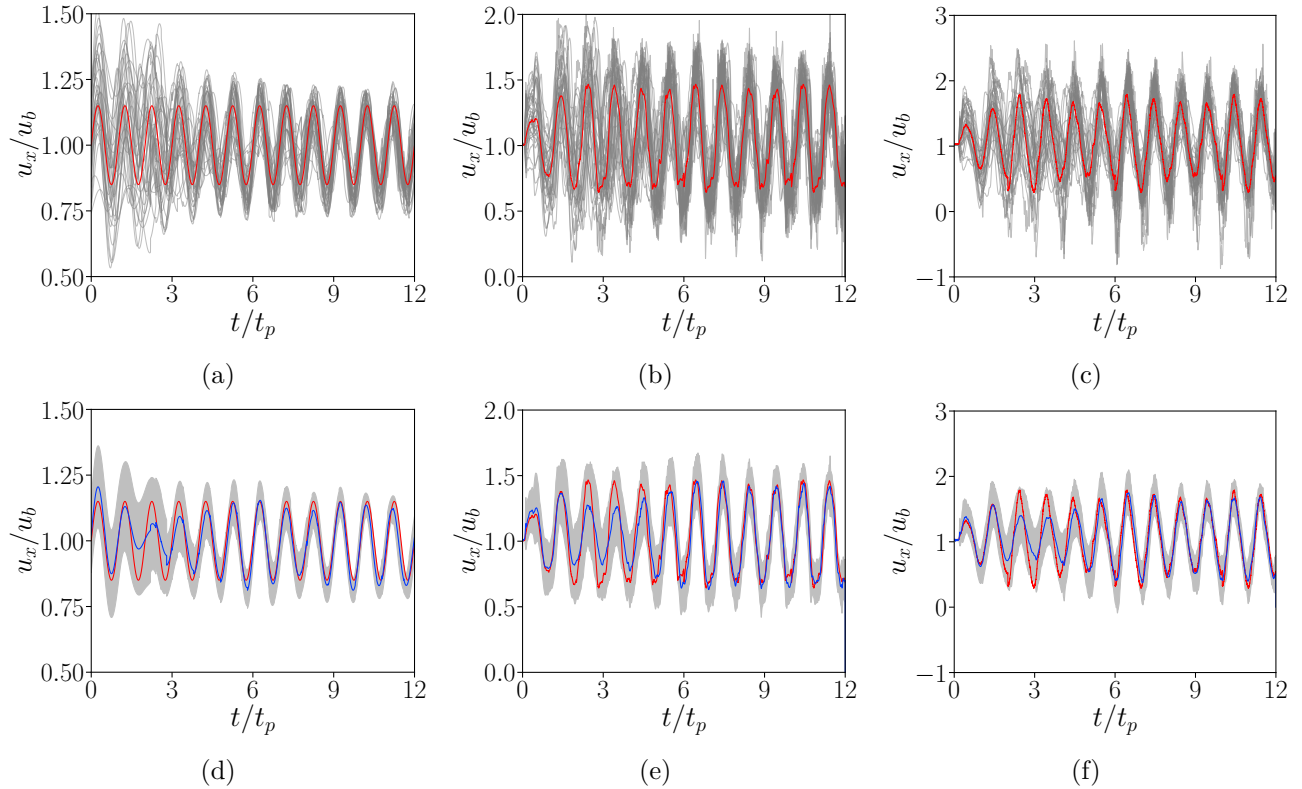


FIGURE 7.13 : Évolution de la condition d'entrée de vitesse axiale pour la distribution *a priori* 1. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleu : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.

différentes. L'amplitude de l'onde résultante est donc elle-même une onde constituée d'une somme de plusieurs fonctions trigonométriques. Par ailleurs, la topologie du battement n'est pas constante et évolue avec le temps, parfois vers des valeurs très éloignées de la référence, en témoigne les Fig. 7.17a, 7.17b et 7.17c. L'optimisation semble donc dynamique puisque les paramètres sont mis à jour à chaque phase d'assimilation vers des valeurs qui ne conviennent qu'à cet instant précis. Ce phénomène est très marginal pour l'*a priori* 2 ou l'onde résultante est constituée de fréquences très différentes et d'amplitudes proches. Dans ce cas, l'onde résultante évolue faiblement, l'optimisation semble statique. Puisque atteint par l'*a priori* 2, l'évolution de la description paramétrique de l'*a priori* 1 pourrait cependant converger vers une évolution statique après un certain nombre de périodes assimilées et avec une inflation paramétrique suffisante.

Pour résumer, le filtre de Kalman d'ensemble hyper-localisé est capable de calibrer la condition d'entrée de simulations LES-LF à partir d'observations de vitesses instantanées mesurées sur une simulation haute-fidélité LES-HF. Malgré l'advection lente de la condition d'entrée mise à jour vers les zones d'observations locales situées en aval de la soupape, le HLEnKF fournit une reconstruction convaincante de l'onde sinusoïdale de référence utilisée sur la LES-HF. La

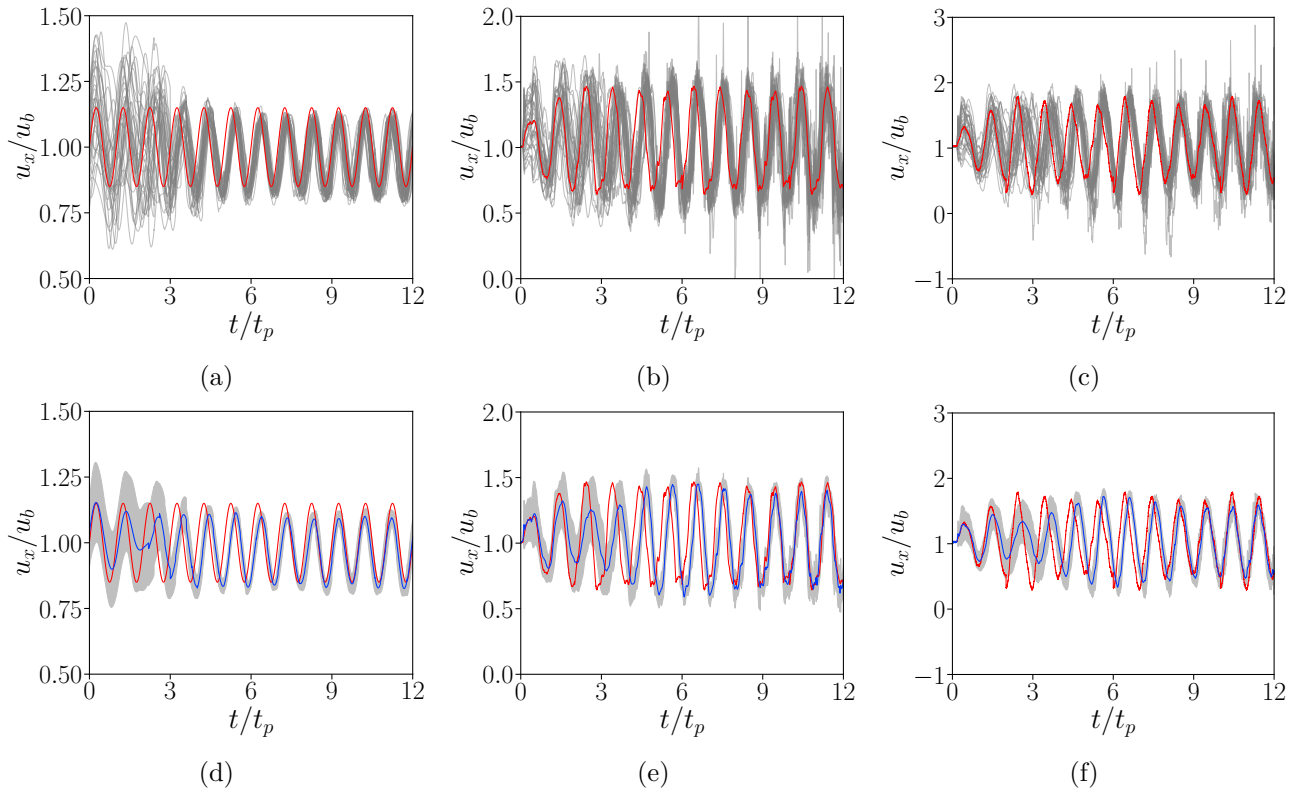


FIGURE 7.14 : Évolution de la condition d'entrée de vitesse axiale pour la distribution **a priori 2**. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleu : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.

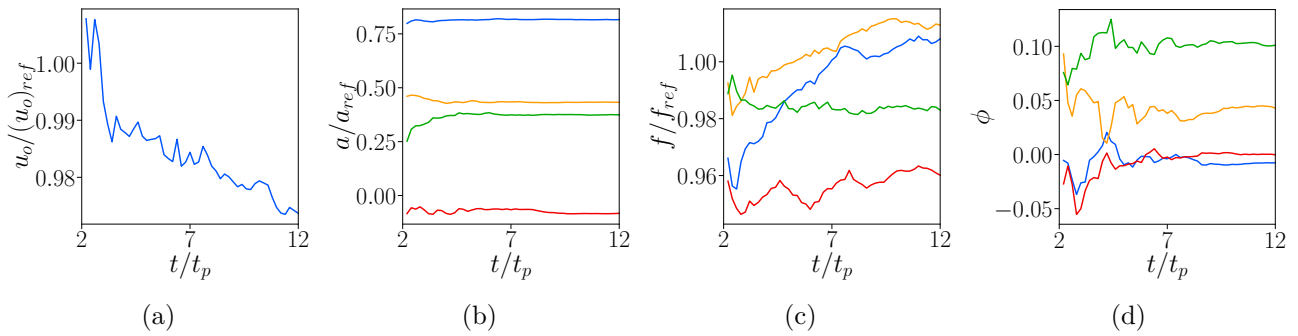


FIGURE 7.15 : Évolution des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution **a priori 1**. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4.

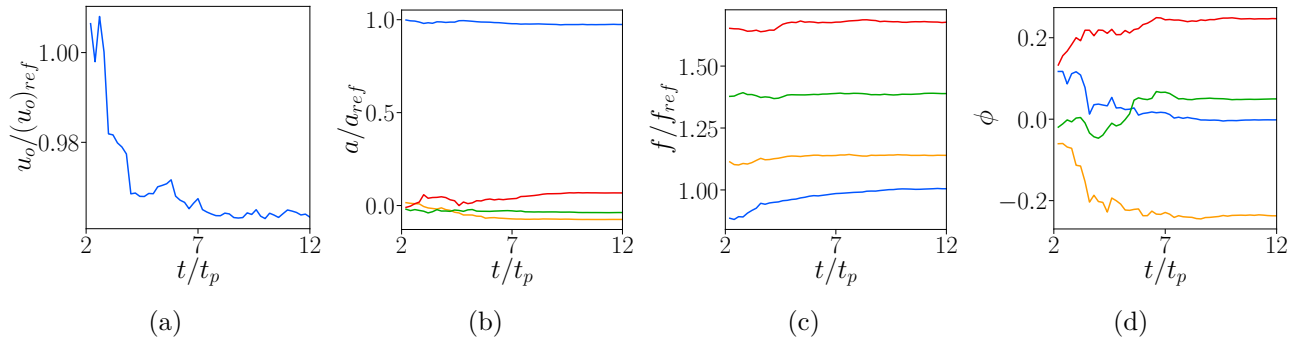


FIGURE 7.16 : Évolution des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution *a priori* 2. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4.

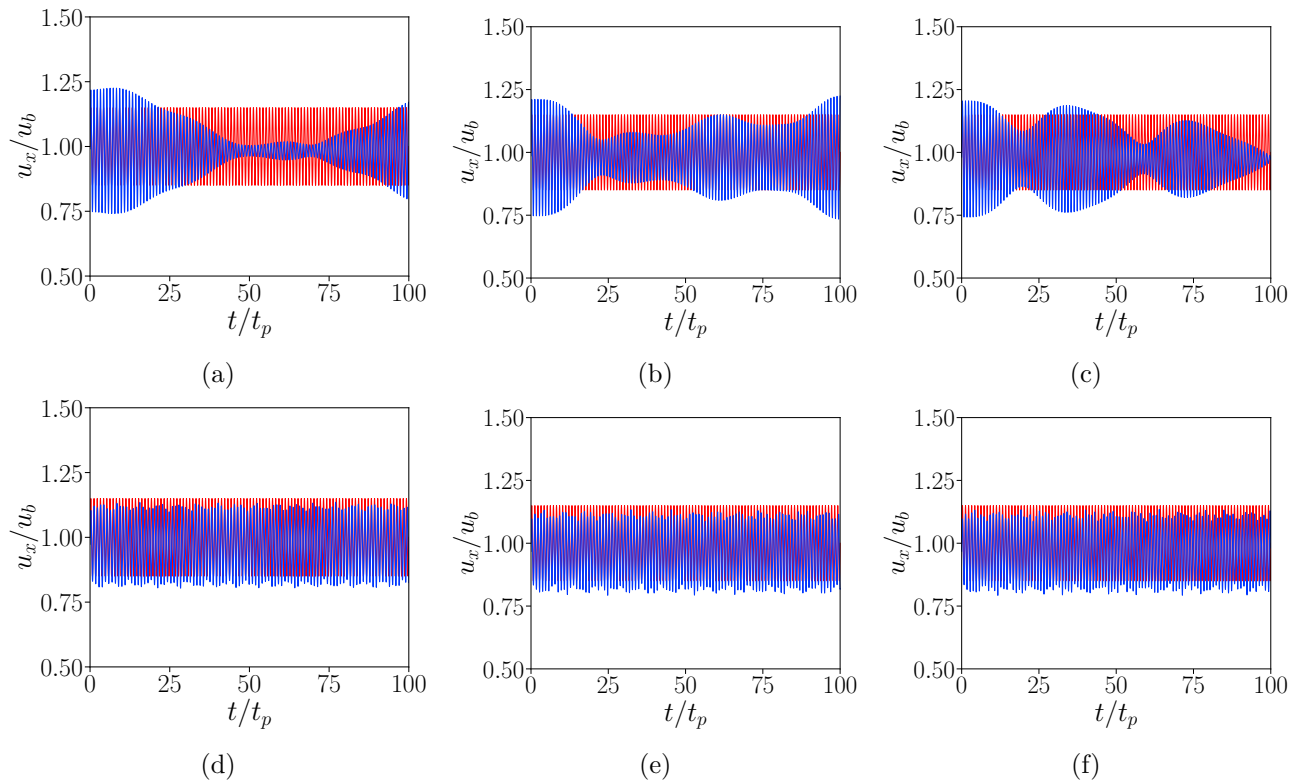


FIGURE 7.17 : Courbe bleu : onde sinusoïdale résultante de la condition d'entrée pour des jeux de paramètres isolés. Courbe rouge : onde de référence. Ligne du haut : *a priori* 1. Ligne du bas : *a priori* 2. De gauche à droite : jeux de paramètres aux temps $t = 6t_p/9t_p/12t_p$.

reconstruction de cette onde est différente pour les deux configurations. L'étude de l'évolution des paramètres permet de dresser un constat clair : le comportement de l'optimisation paramétrique est *dynamique* dans le cas de l'*a priori* 1 alors qu'il est *statique* dans le cas de l'*a priori* 2. Ces conclusions démontrent la grande robustesse de l'algorithme dans l'optimisation de la description paramétrique. Sa caractéristique séquentielle se montre particulièrement pertinente pour l'optimisation instationnaire de paramètres tout en démontrant sa capacité à converger vers une description optimale et stationnaire. L'EnKF et plus particulièrement le HLEnKF montrent un grand potentiel pour la calibration de conditions transitoires et locales en temps. D'autres variantes de l'algorithme de Kalman tel que le *ensemble Kalman smoother* (EnKS), pourrait cependant se montrer plus performant dans ce cas puisqu'il permet la prise en compte de temps passés et futurs lors de l'inférence de la simulation [36, 91]. En outre, cet algorithme est également susceptible d'être hyper-localisé. La caractéristique séquentielle du HLEnKF est à présent étudiée dans le cadre de la synchronisation locale du champ de vitesse à partir de l'estimation de l'état.

7.4 Synchronisation et reconstruction de structures locales

7.4.1 Détails de configuration

La Sec. 7.3 a démontré que le HLEnKF est capable de calibrer une condition d'entrée instationnaire à partir d'informations locales de vitesse. Le champ de vitesse des simulations calibrées aux coordonnées des capteurs utilisées pour les mesures d'observations est statistiquement similaire à la LES-HF grâce à la procédure d'assimilation. L'estimation de l'état peut à présent être envisagée sans risquer la divergence du système à cause de trop forts écarts dans le champ. En effet, l'estimation d'état du filtre de Kalman d'ensemble hyper-localisé permet de modifier le champ de vitesse dans la région des zones d'hyper-localisation du filtre. Cette caractéristique, si tant est que la précision de la prédiction de vitesse soit suffisante pour synchroniser les champs des membres d'ensemble avec l'observation, pourrait permettre de reconstruire les structures locales et particulières de cette même observation du système haute fidélité. Dans des cas réels, la reconstruction de structures du jumeau physique observé sur le jumeau numérique pourrait permettre la simulation, l'étude et la prévention de certaines conditions et événements extrêmes pouvant conduire à une diminution de l'efficacité du système physique ou nuire à son intégrité. Les objectifs de cette section sont donc multiples : (1) Synchroniser le champ inféré avec l'observation haute-fidélité pour diminuer l'écart entre ces deux champs (2) Évaluer l'effet de cette synchronisation sur la répartition énergétique du champ inféré par rapport à la référence. Ce deuxième point est essentiel pour prévenir et identifier l'évolution de structures énergétiques potentiellement néfastes pour le système réel.

La calibration effectuée à partir de l'*a priori* 2 est appelée LES-PE2 (pour "*parameter estimation prior 2*"). La LES-PE2 a été exécutée jusqu'à la période $t = 12t_p$. Bien qu'étant calibrées, les 35 simulations présentent une distribution naturelle de leurs champs de vitesse en aval de la soupape à cause des légères différences dans la description paramétrique de la condition d'entrée. Cette distribution naturelle est utilisée comme nouvelle distribution *a priori* pour la synchronisation des champs à partir de l'estimation d'état. Le calcul avec estimation d'état

est exécuté sur une période à partir du nouvel *a priori*, i.e. de $t = 12t_p$ à $t = 13t_p$. La correction du champ de vitesse ne dépend pas du temps advectif entre l'entrée de la géométrie et les zones d'observations. Toutes les observations disponibles peuvent donc être utilisées pour cette configuration, c'est-à-dire un maximum de 1250 phases d'assimilation de données sur la période inférée. L'estimation paramétrique est également utilisée pour mettre à jour la description paramétrique durant ces phases d'AD. Cependant, l'inflation stochastique appliquée aux paramètres est abaissée à $\lambda = 0.1\%$ à cause de l'augmentation de la fréquence d'assimilation pour conserver un niveau de variance de la description paramétrique équivalent à la procédure de calibration de la Sec. 7.3. Pour rappel, les données des périodes 7 à 10 de la LES-HF étaient bouclées et utilisées comme observations pour l'inférence de la LES-PE2. Les mesures utilisées pour cette nouvelle configuration sont donc celles correspondant à la neuvième période ($t = 8t_p$ à $t = 9t_p$) de la LES-HF. En outre, le champ de vitesse de la LES-HF était donné pour le temps $t = 9t_p$ sur la Fig. 7.6. Les capteurs utilisés dans cette procédure sont une nouvelle sélection de tous les capteurs utilisés pour les mesures de la LES-HF (Fig. 7.7). Le placement de ce nouveau jeu de capteurs diffère de la procédure de calibration de la Sec. 7.3. 582 capteurs sont sélectionnés et répartis principalement sur le plan $x = 0.124$ dans la région du jet de soupape. Quelques capteurs sur les plans $y = 0$ et $z = 0$ sont également sélectionnés, en particulier pour la comparaison qualitative des champs de vitesses avec la référence. Ce nouveau jeu de capteurs, appelé SE-set1 (*state estimation set 1*) est illustré sur la Fig. 7.18. Un jeu de capteurs plus dense, SE-set2, est également utilisé dans la Sec. Annexe C.2.2.

Concernant la taille de la matrice d'état \mathbf{U} , le nombre de degrés de liberté $N = 3n_{cells}$ est strictement liée à la procédure d'hyper-localisation utilisée pour l'inférence du système. La taille des zones d'hyper-localisation est calibrée par rapport à la longueur de localisation de covariance utilisée pour l'AD. Pour le problème étudié, cette longueur de localisation est liée à l'échelle intégrale, qui caractérise la distance à partir de laquelle la fluctuation de la vitesse en deux points n'est plus corrélée. La corrélation de vitesse point à point est calculée pour chaque capteur du plan de mesures $z = 0$ de la base de données observée, pour chaque temps disponible de la 9ème période. L'intégrale de la courbe de corrélation d'un point avec les autres correspond à l'échelle intégrale pour ce point. La moyenne des échelles intégrales de chaque point, pour chaque temps disponible, correspond donc à une échelle de corrélation de vitesse moyenne. Cette valeur est donc adaptée à la procédure de localisation du filtre de Kalman puisqu'il n'est plus utile de corriger les valeurs au-delà de cette distance de corrélation. L'échelle intégrale moyenne calculée sur une période à partir des mesures du plan $z = 0$ est $\mathcal{L} \approx 0.0017$ m. La taille des zones d'hyper-localisation est donc approximativement 0.0034 m. Pour une correction d'environ 1% à la frontière des zones hyper-localisées (et donc 100% au centre des zones) la valeur de l'échelle de corrélation serait $l = 0.00055$. Cependant, une densité plus élevée de capteurs pour une même échelle de localisation pourrait entraîner le chevauchement des zones avec le HLEnKF. CONES est programmé pour fusionner les zones qui s'intersectent. Dans le cas où les zones d'hyper-localisation forment un réseau continu, la fusion des zones résulterait en une seule zone principale. Pour éviter ce comportement, il est préférable de réduire légèrement la taille des zones d'hyper-localisation (et donc de l'échelle de localisation de covariance) pour éviter le chevauchement. Le choix relatif à l'échelle intégrale calculée est donc légèrement abaissé. La taille de zone est de 0.0025 m pour le jeu SE-set1. L'échelle de corrélation associée est égale à

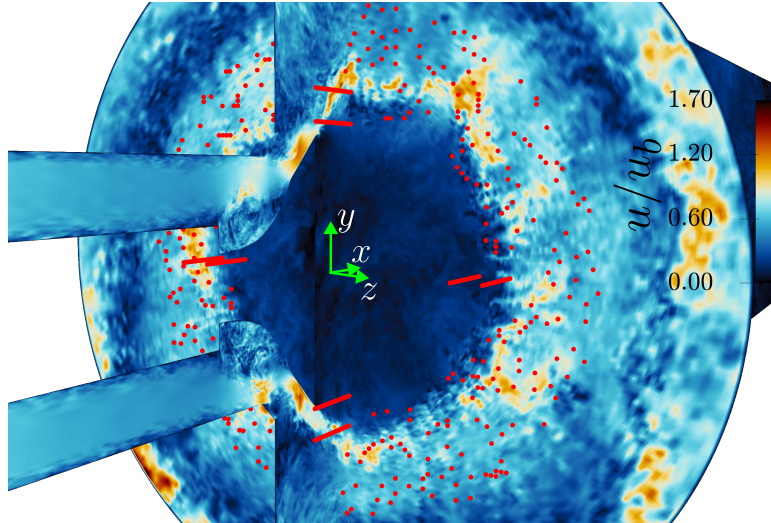


FIGURE 7.18 : Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les positions des capteurs du jeu SE-set1 d’observations sont indiquées par les points rouges.

Nom de la configuration	Estimation	Jeu d’observations	Fenêtre d’observation
LES-PE2	Paramètres	calibration	2500
LES-OW10	Paramètres & état	SE-set1	10
LES-OW50	Paramètres & état	SE-set1	50
LES-OW250	Paramètres & état	SE-set1	250

TABLE 7.2 : Résumé des configurations étudiées avec l’estimation d’état pour la synchronisation des champs comparés à la distribution *a priori* LES-PE2 utilisée pour ces configurations.

$l = 0.00041$. Ceci conduit à 63 274 mailles soit $N = 182\,822$ degrés de liberté dans la procédure répartis sur 134 zones hyper-localisées.

Les différentes configurations étudiées sont résumées sur le Tab. 7.2. En particulier, trois fenêtres d’observations différentes sont utilisées pour la configuration du jeu d’observation SE-set1. La première correspond à l’utilisation de toutes les observations disponibles en temps, soit une tous les $10\Delta_t$. Pour rappel, $\Delta_t = 1 \times 10^{-7}$ s pour la LES-LF. La fréquence d’assimilation est donc égale à 1000 kHz dans ce cas. La deuxième configuration correspond à l’utilisation d’une observation tous les $50\Delta_t$ et la troisième tous les $250\Delta_t$. Les fréquences d’assimilations associées sont alors de 200 kHz et 40 kHz. Or, une fréquence réaliste de mesures de vitesses par PIV se situe entre 1 kHz et 10 kHz [43]. Par ailleurs, la fréquence de l’onde sinusoïdale de référence choisie est minimum 6 fois supérieure aux fréquences caractéristiques de moteurs à combustion interne. La fréquence des grandes structures turbulentes caractéristiques de l’écoulement peuvent dépendre de la fréquence de l’onde. Elles seraient dans ce cas également 6 fois supérieures aux fréquences réalistes. Les fréquences de la deuxième et troisième configuration correspondraient donc à des valeurs d’environ 33.4 kHz et 6.7 kHz dans une configuration réaliste. La deuxième configuration est donc dans ce cas cadencée à une fréquence proche de l’ordre de grandeur des mesures PIV hautes fréquences. La troisième configuration entre dans l’intervalle de ces fréquences.

Un des outils disponibles pour évaluer la capacité de synchronisation des différentes confi-

gurations d'AD est l'estimation de la racine de l'erreur quadratique moyenne Φ :

$$\Phi_k = \sqrt{\sum_{j=1}^{N_o} (\langle \mathbf{s}_{j,k} \rangle - \boldsymbol{\alpha}_{j,k})^2} / N_o \quad (7.3)$$

avec $\langle \mathbf{s}_{j,k} \rangle$ la moyenne de l'ensemble des valeurs de vitesses aux coordonnées des observations, $\boldsymbol{\alpha}_{j,k}$ les observations et N_o leur nombre.

Celle-ci permet d'évaluer l'écart entre le champ de vitesse assimilé et la référence haute-fidélité aux coordonnées des observations. Le calcul est réalisé en considérant les 582 capteurs d'observations des configurations LES-OW250, OW50 et OW10. Pour cette dernière, deux niveaux de confiance dans les observations (5% et 20%) ainsi que deux niveaux d'inflations sont testés (0% et 5%). L'évolution de Φ pour ces essais est illustrée sur la Fig. 7.19. Celle-ci varie en fonction de l'évolution de l'onde sinusoïdale dans le domaine et est plus élevée pour les hautes vitesses entre $t/t_p = 12.5$ et $t/t_p = 13$. La courbe noire représente l'évolution de l'erreur pour la configuration sans aucune estimation d'état LES-PE2. C'est un marqueur de la variabilité naturelle des 35 simulations de l'ensemble. Les courbes rouge et orange correspondent aux configurations LES-OW250 et OW50 pour un niveau de confiance dans les observations de $\sigma_m = 5\%$. Pour OW50, l'erreur est globalement plus faible que pour le cas sans estimation d'état mais cette diminution varie au cours de la période. En revanche, la configuration OW250 présente une erreur Φ plus élevée que le cas sans estimation d'état LES-PE2 dans les zones à fort gradient de vitesse. Cela démontre que la cadence d'assimilation de la configuration LES-OW250 est trop faible pour synchroniser les champs observés. L'erreur diminue grandement pour les 4 variantes des configurations OW10. Les courbes vertes représentent un niveau de confiance de 20% dans les observations, les bleues 5%. La différence entre ces deux groupes de courbes est assez élevée entre $t/t_p = 12.6$ et $t/t_p = 12.8$ avec une différence d'environ 30% d'erreur au maximum. En outre, pour ces temps, l'erreur est réduite de 60% dans le meilleur des cas par rapport à la configuration LES-PE2. Ce résultat souligne l'importance du bruit donné aux observations par les différents niveaux de confiance. A l'inverse, l'inflation ne semble avoir aucun effet sur la capacité de synchronisation de l'algorithme. Ceci s'explique par deux facteurs : (1) une variabilité suffisante grâce à la distribution naturelle *a priori*. Cette dernière reste de plus constante en amont des zones d'observations. La variabilité de l'ensemble est donc naturellement maintenue à un niveau suffisant par l'advection naturelle des vitesses inférées. (2) L'effondrement de variance problématique des algorithmes du type EnKF est moins élevée avec le HLEnKF du fait de la forte localisation des corrections (estimation sur un état restreint et prises en compte d'observations uniquement locales).

Les gros écarts constatés entre les cas LES-OW10, LES-OW50 et LES-OW250 pour le même niveau de confiance dans les observations (courbes rouge, orange et bleue) sont dus au nombre de phases d'assimilations subies par chaque particule fluide. En effet, pour une large gamme de vitesse allant de $0.5u_b$ à $2u_b$ dans le jet en sortie de soupape, le temps advectif associé à la taille de la zone d'observation varie entre $\approx 77t_m$ et $\approx 19t_m$. Dans le cas LES-OW10, le nombre de phases d'assimilations subies par une particule fluide résidant dans la zone d'observation est donc entre 77 et 19 phases d'assimilations. Pour le cas LES-OW50, une particule fluide subit

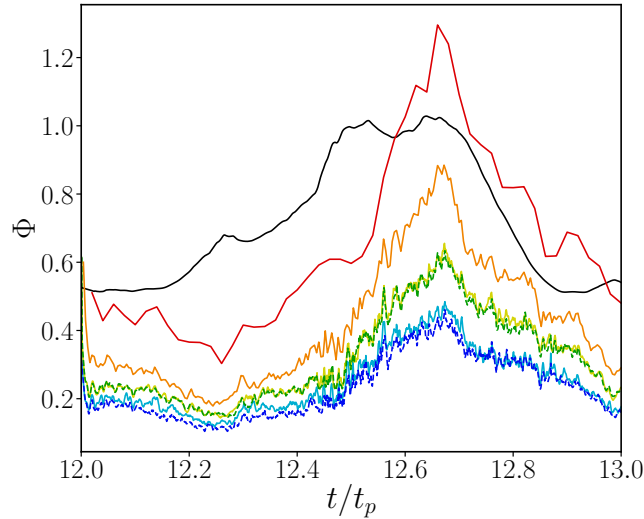


FIGURE 7.19 : Erreur quadratique moyenne Φ en fonction de la période inférée. Courbe noire : LES-PE2. Courbe rouge : LES-OW250 avec 5% de confiance sur les observations sans inflation. Courbe Orange : LES-OW50 avec 5% de confiance sur les observations sans inflation. Courbes vertes : LES-OW10 avec un niveau de confiance sur les observations de 20% pour deux niveaux d'inflations de 0% pour la couleur claire et 5% pour la couleur foncée. Courbes bleues : LES-OW10 avec un niveau de confiance sur les observations de 5% pour deux niveaux d'inflations de 0% pour la couleur claire et 5% pour la couleur foncée.

Vitesse du jet	LES-OW10	LES-OW50	LES-OW250
$0.5u_b$	76.9	15.4	3.1
$1u_b$	38.5	7.7	1.5
$1.5u_b$	25.7	5.1	1
$2u_b$	19.2	3.8	0.8

TABLE 7.3 : Nombre de phases d'assimilation subies par une particule fluide dans les zones d'observations en fonction de la vitesse du jet et des configurations étudiées.

logiquement 5 fois moins de phases d'AD, entre ≈ 15 et ≈ 4 . Enfin, pour le cas LES-OW250, le nombre de phases est encore divisé par 5 pour se situer entre ≈ 3 et 0.8. Pour ce dernier cas, certaines particules fluides traversant les zones d'observations ne subissent aucune correction pour les plus hautes vitesses. En considérant la faible correction apportée en bordure de zone à cause de la localisation de covariance, dans certaines conditions le cas LES-OW50 présente juste assez d'assimilations pour couvrir la correction des toutes les particules fluides transitant dans la zone, d'où la différence d'erreur Φ constatée entre les trois cas. Les temps advectifs des zones en fonction de la vitesse du jet et le nombre de phases d'assimilations associé sont résumés sur le Tab. 7.3.

Capteur	x (m)	y (m)	z (m)
1	0.124	-0.0267	0.0197
2	0.124	0.0205	-0.0254
3	0.124	0.0342	0.0205

TABLE 7.4 : Capteurs utilisés pour le contrôle de la synchronisation des vitesses inférées avec l'observation

7.4.2 Capacités de synchronisation

L'étude de la racine de l'erreur quadratique moyenne selon différents hyper-paramètres a montré qu'un niveau de confiance dans les observations de 5% et qu'une inflation nulle étaient adéquats. Les capacités de synchronisations des configurations pour ces hyper-paramètres sont maintenant abordées. Pour cela, différents capteurs utilisés dans les configurations d'AD comme observation sont isolés pour contrôler l'évolution des vitesses instantanées par rapport aux données haute-fidélité. Ces capteurs, résumés sur le Tab. 7.4, se situent sur le plan $x = 0.124$ mm. Les capteurs 1 et 2 sont situés dans le jet de sortie de soupape sur des zones opposées sur le plan alors que le capteur 3 est légèrement excentré dans une zone à plus fort cisaillement.

La Fig. 7.20 illustre les capacités de synchronisation de l'algorithme pour les composantes axiales de ces trois capteurs. Les figures de synchronisation des autres composantes ainsi que de la pression sont données en Sec. Annexe C.2.1. Elles présentent des résultats similaires aux composantes axiales. L'évolution de la composante u_x de vitesse instantanée observée normalisée par la vitesse débitante u_b de référence est représentée en bleu. Les données échantillonnées à partir des membres de l'ensemble de la procédure d'AD sont indiquées en noir. Dans ce cas, la ligne noire correspond à une moyenne d'ensemble. Les zones autour de ces courbes représentent visuellement le niveau de confiance/la variabilité des données. Plus précisément, la zone bleue correspond aux valeurs de confiance incluses dans la matrice de covariance d'erreur de mesures σ_m . De même, la zone grise représente l'intervalle de confiance à 66% de l'ensemble, c'est-à-dire son écart type. Les courbes rouge et verte illustrent l'évolution de la vitesse pour le membre 27 avant toute procédure d'AD et après l'application de la procédure correspondante à chaque ligne de figures. La première ligne correspond au cas de la configuration LES-PE2 sans estimation d'état. Comme remarqué précédemment, le niveau de variance de l'ensemble indiqué par les larges zones grises est relativement élevé. Ceci est dû à la variabilité naturelle des champs de vitesse de l'ensemble à cause des légères variations dans l'optimisation des conditions d'entrée de la configuration LES-PE2. La tendance d'évolution de vitesse est similaire à l'observation sans présenter de synchronisation. Par ailleurs, cette évolution prise de manière isolée présente de fortes variations similaires au cas sans assimilation, en témoigne les courbes rouges et vertes du membre 27. Les lignes 2 à 4 de figures présentent successivement des niveaux de synchronisation plus élevés. Logiquement, la synchronisation dépend fortement du nombre de phases d'assimilation de données subie par chaque particule fluide (indiquées dans le Tab. 7.3). Un nombre de phases d'AD plus faible est responsable d'une plus faible correction et donc d'une plus faible diminution de l'écart type de l'ensemble, en témoigne les configurations LES-OW250 et LES-OW50. Par ailleurs, le niveau de synchronisation de la LES-OW250 est peu représentatif des données observées. Les fortes corrections apportées à des intervalles de

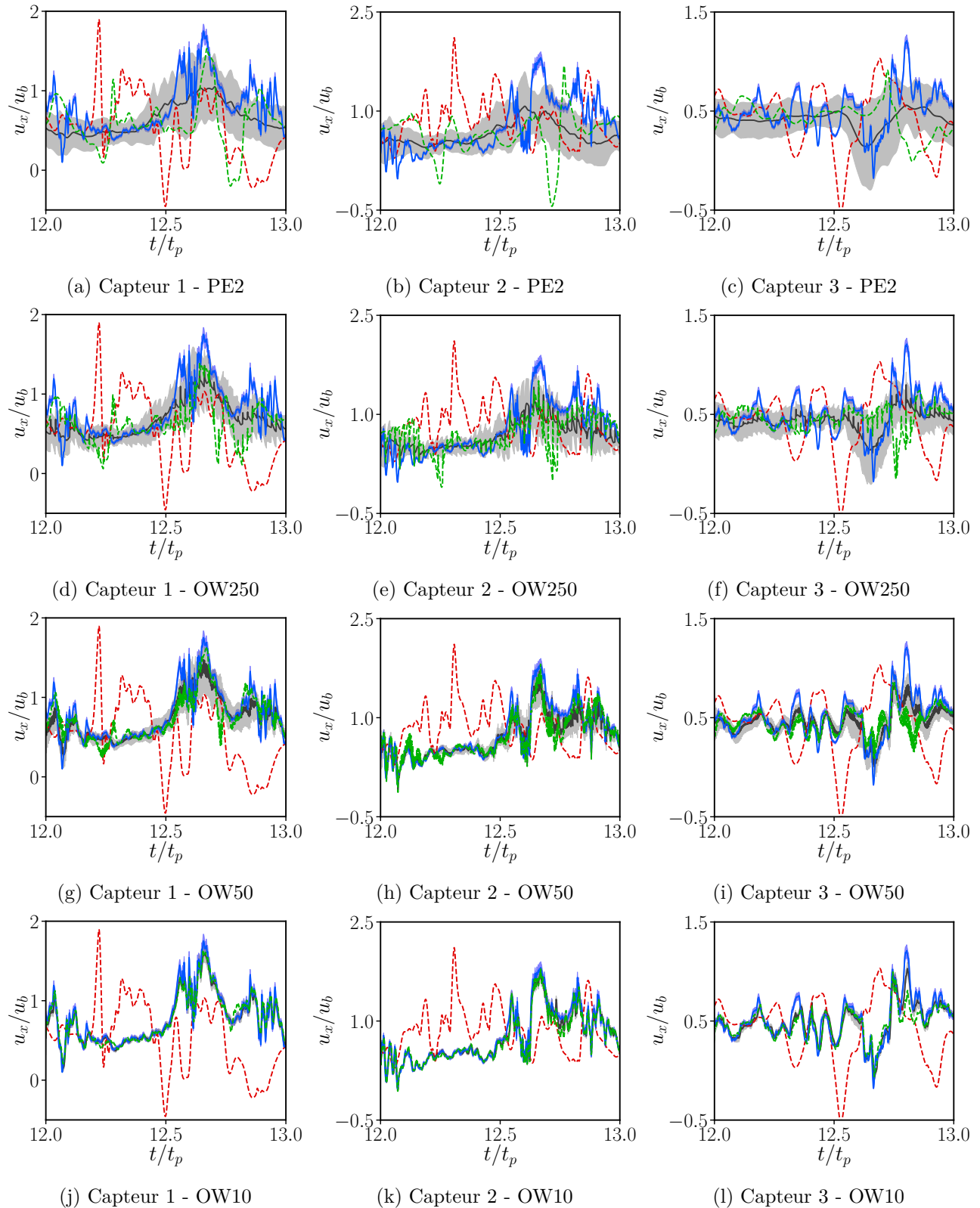


FIGURE 7.20 : Synchronisation de la composante axiale de la vitesse u_x adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (*a priori*) et les vertes dans le cas associé à chaque configuration (*a posteriori*).

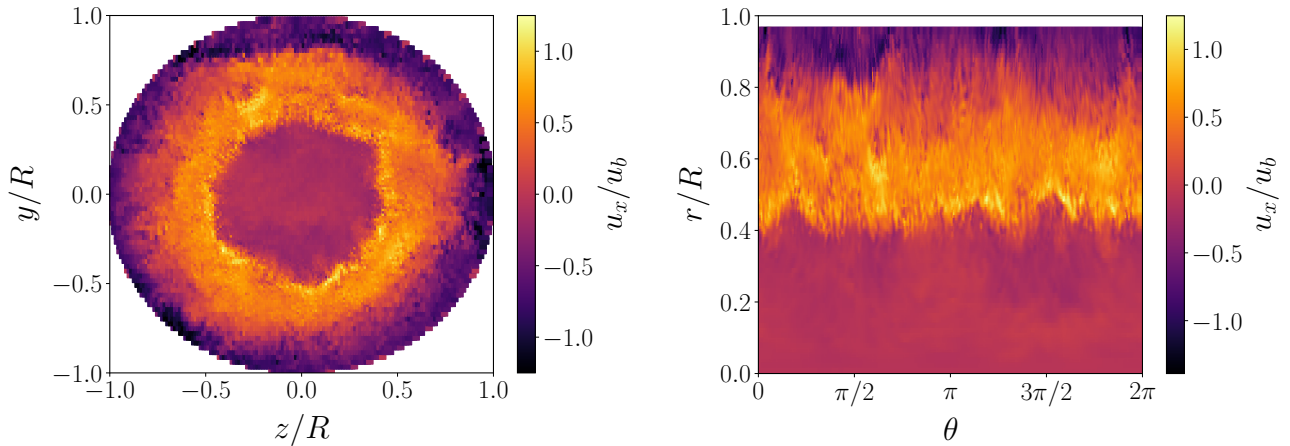


FIGURE 7.21 : Vitesses axiales de la simulation de référence LES-HF mesurées par les capteurs du plan $x = 0.124$. Gauche : mesures des capteurs. Droite : interpolation des valeurs sur une grille structurée en fonction de r/R et θ .

temps trop espacés entraînent de fortes oscillations non représentatives de l'écoulement de référence et peu réalistes. La LES-OW50 montre des résultats satisfaisants avec une capacité de synchronisation représentative des observations et un écart type resserré autour de ces données. Les meilleures capacités de synchronisation sont atteintes par la configuration LES-OW10. Pour les trois capteurs, la moyenne d'ensemble est presque parfaitement superposée aux valeurs observées. La précision du suivi diminue seulement localement dans les zones à fort gradient de vitesse entre $t = 12.5t_p$ et $t = 13t_p$. L'écart type de l'ensemble étant très faible, la précision de la synchronisation n'est pas seulement moyenne mais établie pour tous les membres d'ensemble. Le membre 27 montre par exemple un niveau de synchronisation équivalent à la moyenne avec une grande superposition de la courbe verte avec l'observation en bleu. Le HLEnKF se montre donc particulièrement efficace pour synchroniser le champ de vitesse avec les données haute-fidélité dans les zones d'observations. Cette synchronisation locale pourrait permettre la reconstruction d'évènements locaux en temps et en espace observés sur la configuration de référence haute-fidélité.

L'étude de l'évolution des vitesses instantanées dans les zones d'observations a montré une synchronisation locale des points observés. Le paradigme du jumeau numérique requiert un suivi global des topologies de structures turbulentes et des répartitions énergétiques dans le domaine afin de prévenir les évènements potentiellement préjudiciables. En particulier, le plan $x = 0.124$ m utilisé pour mesurer les données d'observations présente des caractéristiques uniques dans le domaine d'étude. Le jet de soupape traverse directement ce plan, puis l'écoulement est redirigé par la paroi du cylindre pour re-circuler en amont et passer à nouveau par le plan $x = 0.124$ m. La répartition énergétique de l'écoulement sur ce plan peut donc fournir une information pertinente de la précision de la prédiction fournie par le HLEnKF non seulement dans les zones corrigées mais sur la totalité de la zone d'influence du jet.

Le champ de vitesse axiale instantanée u_x de la simulation de référence LES-HF est illustré sur la Fig. 7.21. Les données de vitesses correspondent aux mesures effectuées par les 11 916

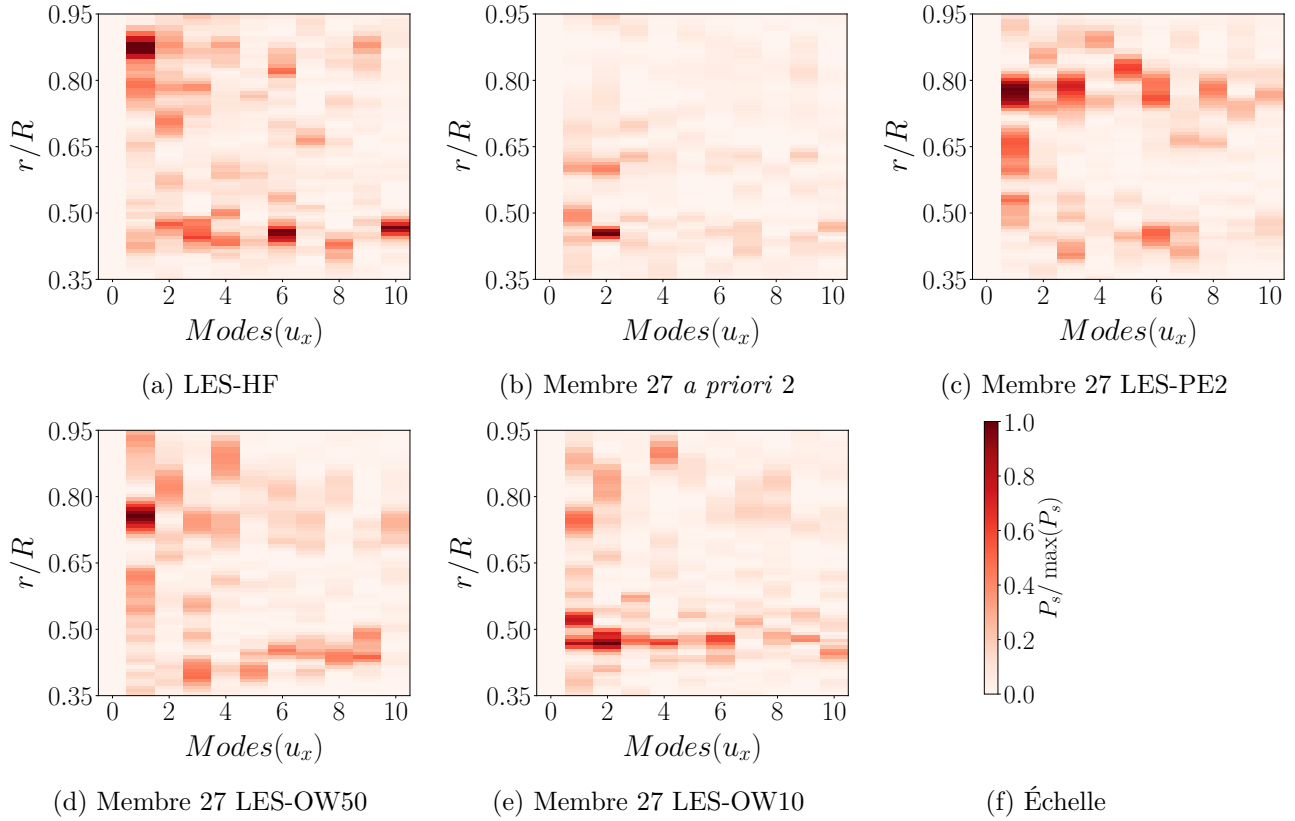


FIGURE 7.22 : Cartographie des modes principaux de la transformée de Fourier azimutale de la vitesse axiale $u_x - \overline{u_x}$ au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seul la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.

capteurs répartis sur le plan $x = 0.124$ m pour le temps $t = 9t_p$. Une transformée de Fourier azimutale de ce champ pourrait permettre d'isoler les modes caractéristiques de l'écoulement pour cette section et en particulier la répartition d'énergie. Pour cela, les 11 916 données de vitesses sont interpolées sur une grille cartésienne dépendant de r/R et θ . Ceci permet d'effectuer une transformée de Fourier azimutale à l'instant $t = 9t_p$ et pour chaque rayon disponible sur cette grille cartésienne. La comparaison de la répartition énergétique des modes de Fourier pour les prédictions des différentes approches permet de fournir une estimation de la qualité de la synchronisation du champ de vitesse et de la reconstruction des interactions multi-échelles de la turbulence. Pour cela, l'énergie des modes de Fourier des configurations analysées est comparée à la répartition énergétique des modes de la référence LES-HF. Cette comparaison est pertinente puisque le rôle de l'estimation d'état du HLEnKF est de reconstruire une structure de données se rapprochant le plus possible de l'observation. La prédiction vise la synchronisation d'une réalisation unique de l'écoulement de référence et pas seulement d'une réalisation similaire en moyenne mais dont l'état instantané diffère de cette dernière.

La Fig. 7.22 illustre la cartographie des 10 modes principaux de la transformée de Fourier de

la vitesse axiale sans sa moyenne azimuthale $u_x - \overline{u_x}$ pour la simulation de référence LES-HF et pour les simulations du membre 27 correspondant aux différentes configurations. En particulier, les résultats pour l'*a priori* 2, pour les configurations LES-PE2, LES-OW50 et LES-OW10 sont reportés. Les figures sont colorées en fonction de la puissance spectrale P_s (le carré du spectre d'amplitude) normalisée par la valeur maximale de puissance dans chaque cas. Ces spectres de puissances font état de l'énergie cinétique (par radians) transportée par chaque mode de Fourier. La cartographie de la simulation de référence présentent trois modes principaux : le mode 1 contient une grande quantité d'énergie dans la région de recirculation de l'écoulement, le mode 6 ainsi que le mode 10 présentent de grandes quantités d'énergies dans la région de cisaillement du jet. Les modes 2 et 3 sont également importants dans la représentation de ces deux zones. La simulation *a priori* 2 expose un niveau d'énergie bien plus important dans la zone de cisaillement ($r/R \approx 0.45$) pour le mode 2 contrairement aux modes 6 et 10. La prédiction d'énergie pour la zone de recirculation ($r/R \approx 0.9$) semble inexistante comparée à la LES-HF. La calibration de la condition d'entrée dans la configuration LES-PE2 vient corriger la distribution d'énergie dans cette cartographie. Elle semble cependant privilégier la zone de recirculation avec des niveaux d'énergies prépondérants pour les modes 1, 3, 5, 6, 8 dans cette région. Les simulations LES-OW50 et OW10 offrent la meilleure prédiction. Les distributions d'énergie semblent en adéquation avec la référence pour les régions de recirculation et cisaillement. En particulier, la configuration LES-OW10 reproduit correctement la présence énergétique de tous les modes pour la région de cisaillement malgré quelques décalage dans les niveaux obtenus. La prédiction de la distribution énergétique dans la région de recirculation semble moins précise malgré l'apparente similarité des modes 1, 2 et 4. Cette région n'étant pas directement affectée par une quelconque correction du HLEnKF, il est logique d'identifier une précision accrue de la prédiction pour la région de cisaillement du jet plutôt que pour la région de recirculation.

La transformée inverse pour un mode donné permet d'apprécier sa contribution dans le champ de vitesse. La Fig. 7.23 expose la transformée inverse du mode 10 des cartographies présentées. Logiquement, son influence est prépondérante dans la zone de cisaillement du jet pour $r/R \approx 0.45$ où les 10 oscillations de vitesses sont visibles. Celles-ci sont diffusées dans le cas de l'*a priori* 2 présentant des oscillations miroirs entre $r/R = 0.4$ et $r/R = 0.5$. La réalisation calibrée illustre la diffusion de l'énergie déjà présente sur la cartographie associée sur la Fig. 7.22c. La LES-OW10 réduit cet effet diffusif pour retrouver les 10 oscillations caractéristiques dans la zone de cisaillement à $r/R = 0.45$, bien que la représentation de celles-ci soit moins précise.

La Fig. 7.24 présente les cartographies des 10 premiers modes pour la transformée de Fourier de la composante u_y de vitesse. Seule cette composante est présentée, u_z ayant le même comportement décalé azimuthalement de $\pi/2$. Pour ces composantes, le mode 1 est prépondérant sur tous les autres. L'énergie est concentrée dans la région de cisaillement pour laquelle le jet n'a pas encore impacté la paroi du cylindre. Cette énergie s'étale de $r/R = 0.45$ à $r/R = 0.65$ pour la simulation de référence. Dans le cas de l'*a priori* 2 elle est recentrée pour des valeurs comprises entre $r/R = 0.45$ et $r/R = 0.55$. La calibration de la condition d'entrée entraîne une réalisation pour laquelle une sur-prédiction de l'énergie dans la région de recirculation et celle du jet est à noter pour les modes 1 et 3. Logiquement, cette sur-prédiction persiste légèrement

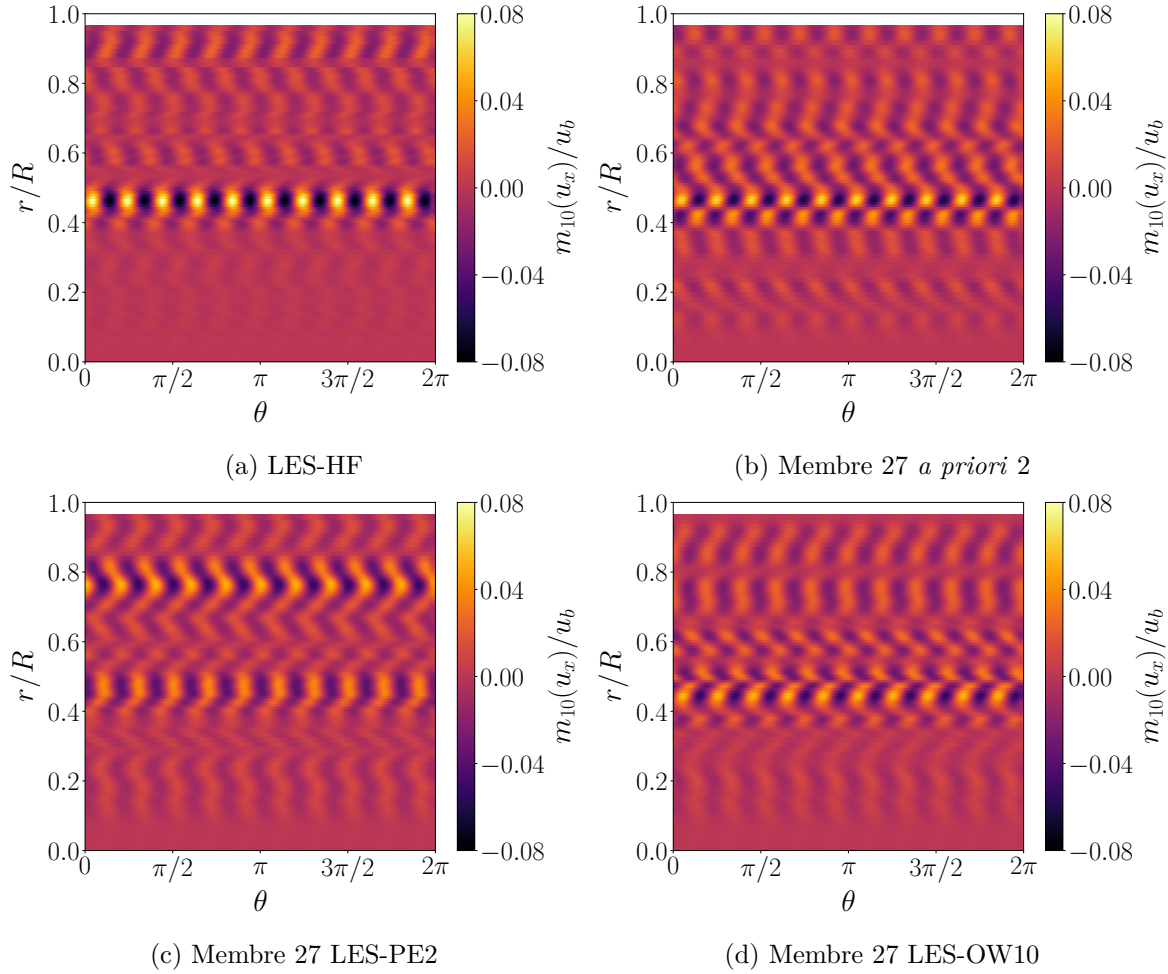


FIGURE 7.23 : Transformées de Fourier inverses du mode 10 de la composante $u_x - \overline{u_x}$ au temps $t = 13t_p$. Les vitesses sont normalisées par la vitesse débitante de référence u_b .

pour les configurations LES-OW50 et LES-OW10. L'étalement d'énergie dans la région de cisaillement pour le mode 1 est également légèrement plus grande que pour la référence dans le cas de la LES-OW10, avec des valeurs allant de $r/R = 0.35$ à $r/R = 0.65$. La configuration LES-OW50 présente en revanche un étalement similaire.

Pour résumer, l'étude de la répartition énergétique des modes de Fourier montre que le HLEnKF est capable de fournir une prédiction accrue des distributions énergétiques dans la zone inférée grâce aux mesures d'observations. Malgré l'aspect local des corrections, des améliorations sont également observées pour des rayons éloignés des zones d'inférence mais affectées par les corrections effectuées aux temps précédents et transportées par la recirculation de l'écoulement. La répartition d'énergie des composantes non axiales présentent cependant peu de différences entre toutes les configurations étudiées malgré l'excellente synchronisation de toutes les composantes de vitesse. Une comparaison qualitative des champs de vitesse pour la section $z = 0$ de la géométrie est fournie sur la Fig. 7.25. Celle-ci illustre en particulier le déphasage de l'*a priori* 2 par rapport à la configuration calibrée LES-PE2. En outre, la LES-OW10 présente des améliorations topologiques des structures de vitesses sans toutefois

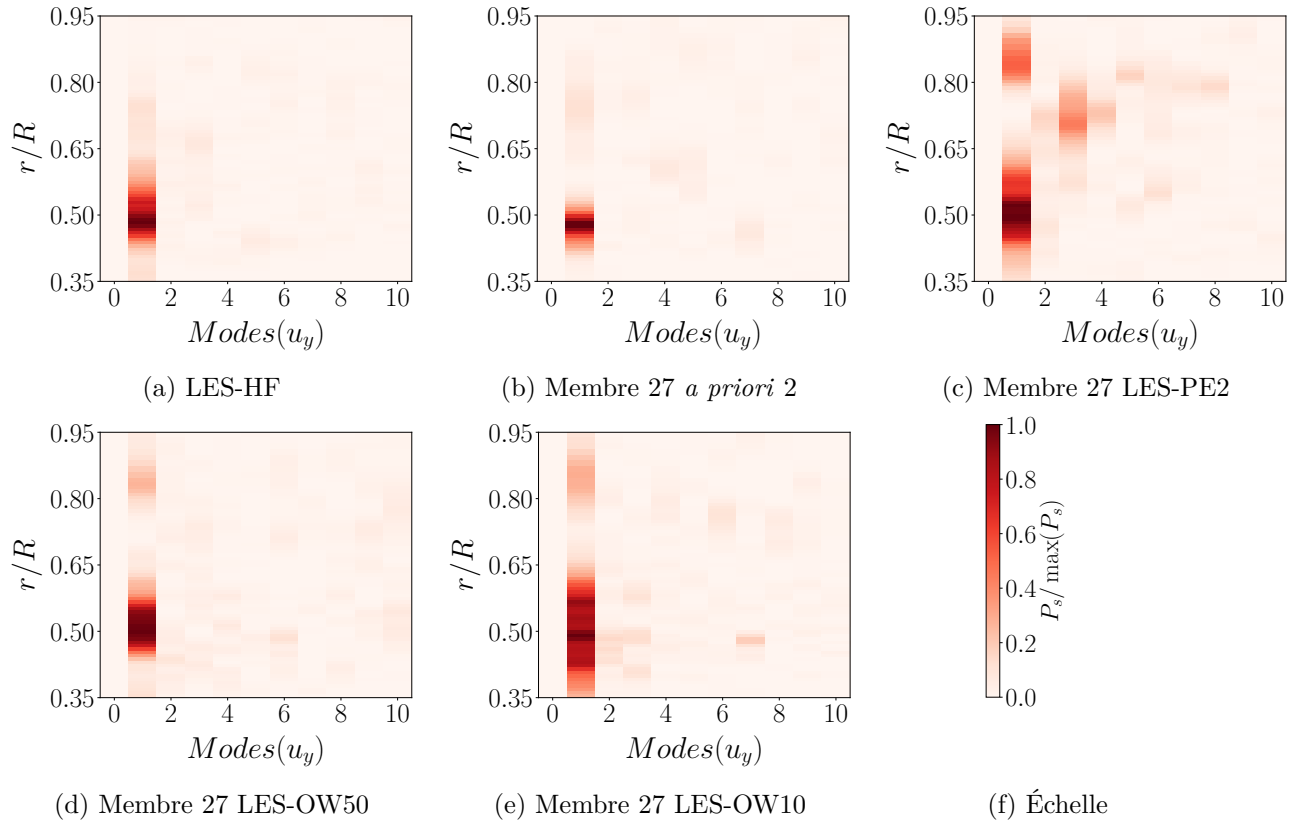


FIGURE 7.24 : Cartographie des modes principaux de la transformée de Fourier azimuthale de la vitesse u_y au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seul la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.

reconstituer la forme caractéristique de "pince" présente sur la simulation de référence LES-HF. Ces améliorations topologiques, la synchronisation excellente des vitesses dans les zones inférées ainsi que l'amélioration de la prédiction de la répartition énergétique de l'écoulement mettent en évidence le grand potentiel de l'assimilation de données séquentielle fortement localisée dans le paradigme du jumeau numérique.

7.5 Conclusion

Ce dernier cas d'étude a permis l'exploration de l'assimilation de données séquentielle pour une simulation 3D LES compressible et instationnaire. L'approche utilisée ainsi que le niveau de complexité en termes de degrés de liberté et de modélisation adoptée font du travail présenté dans ce chapitre un cas unique dans la littérature. L'outil CONES a permis de réaliser toutes les simulations nécessaires avec une grande facilité grâce à sa versatilité et son ergonomie améliorée. Le banc à volute stationnaire s'est montré être un cas d'étude particulièrement pertinent pour répondre aux deux objectifs fixés : (1) la calibration d'une condition d'entrée à partir d'informations aval et locales, (2) la synchronisation des vitesses avec ces informations.

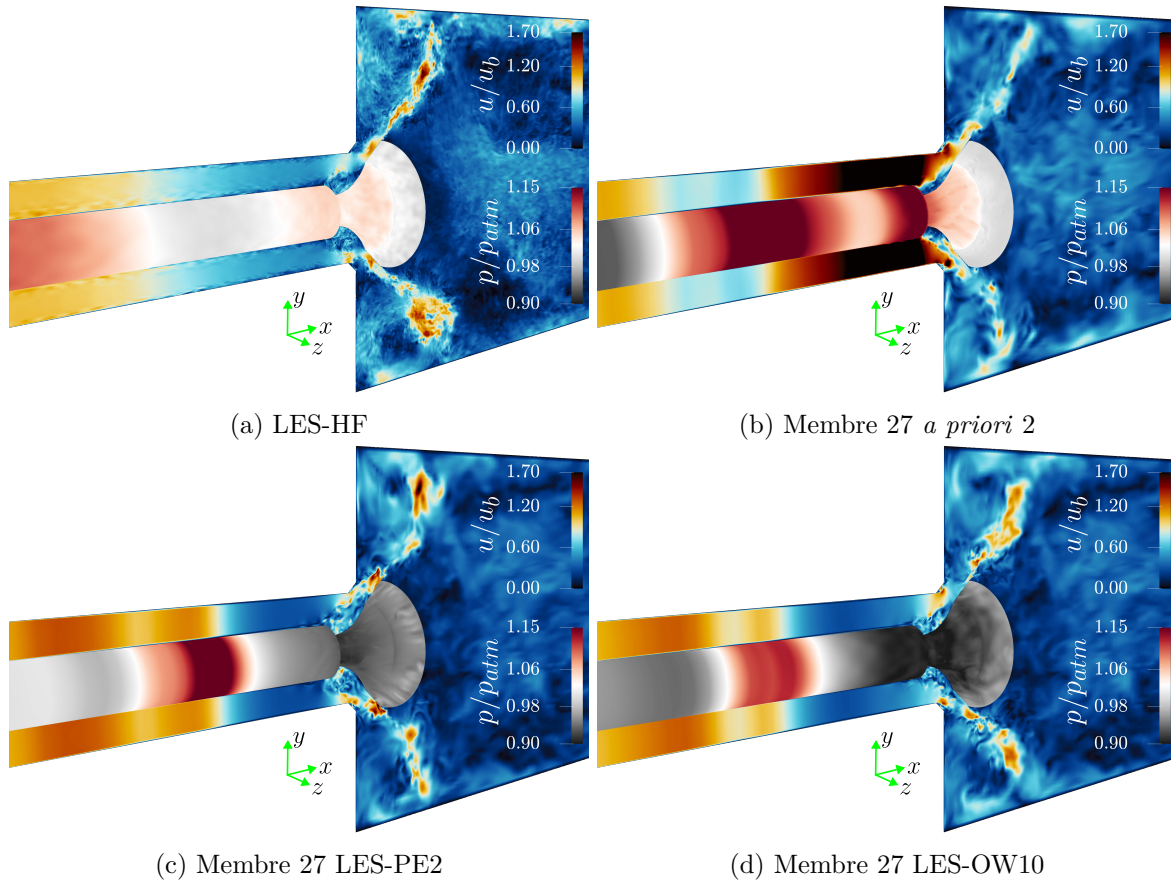


FIGURE 7.25 : Champs de vitesses des différentes configurations pour une section $z = 0$ de la géométrie. Les vitesses sont normalisées par la vitesse débitante de référence u_b .

Les données observées dans les simulations de plus basse fidélité ont été créées grâce à une simulation de référence haute-fidélité pour laquelle une condition d'entrée sinusoïdale dépendante du temps a été imposée sur la vitesse axiale. Le HLEnKF montre une calibration satisfaisante de cette condition d'entrée sur un ensemble de simulations basse-fidélité à partir de deux distributions paramétriques *a priori* différentes. Ces deux optimisations montrent des comportements différents. La première arbore une évolution dynamique de la description paramétrique faisant apparaître une onde résultante battante. L'amplitude est elle-même une onde complexe. La seconde optimisation est en revanche statique et présente une variation d'amplitude en entrée à l'origine d'une compensation de la prédiction erronée du modèle de turbulence. L'estimation paramétrique fournie par le HLEnKF montre donc un grand potentiel dans la calibration de conditions entièrement transitoires pour lesquelles la prédiction des modèles de turbulences conventionnels peuvent manifester certaines limitations dans la représentation de l'écoulement.

L'estimation d'état du HLEnKF fournit une synchronisation locale des champs de vitesse inférés avec les données observées. En particulier, l'étude des figures de synchronisations met en évidence la cadence d'assimilation nécessaire pour fournir une synchronisation satisfaisante, à partir de 200 000 Hz ici. La forte localisation ainsi que l'advection naturelle de l'écoulement préviennent la nécessité d'utiliser l'inflation stochastique de l'état. Dans ce cas, le HLEnKF

permet un excellent niveau de synchronisation des champs. L'étude de la distribution d'énergie cinétique des modes de Fourier fourni une information supplémentaire sur la qualité de la reconstruction du champ de vitesse. Bien qu'inchangée pour les composantes u_y et u_z , la répartition d'énergie pour la composante axiale se montre nettement améliorée par l'algorithme d'assimilation de données. Une amélioration est également visible dans la zone de recirculation qui n'a subi aucune correction directe mais contient les informations corrigées transportées par l'advection de l'écoulement.

Ces résultats montrent un grand potentiel pour l'intégration d'un algorithme d'assimilation de données séquentiel dans le paradigme du jumeau numérique. Pour aller plus loin, il est nécessaire de se rapprocher de conditions et géométries réalistes. En particulier, l'étude de conditions fortement transitoires et perturbées ainsi que l'utilisation de données réelles d'expérimentations semblent constituer les prochaines étapes de développement de ces applications futuristes.

Conclusion et perspectives

Retour sur le contexte

La baisse des émissions impose le développement de technologies durables de rupture pour le secteur des transports. La rapide évolution des technologies terrestres, maritimes et aériennes entraîne une complexification des systèmes physiques responsables d'une utilisation sous-optimale de ces systèmes. Le moteur à combustion interne est par exemple l'objet d'améliorations techniques tel que l'hybridation ou l'utilisation d'hydrogène. Ces nouveaux développements sont à l'origine de phénomènes physiques rendus plus difficiles à caractériser et à modéliser. De ce point de vue, la mécanique des fluides présente encore beaucoup de défis à explorer. La prédiction incorrecte du comportement de l'écoulement et de la combustion dans le moteur à combustion interne est à l'origine d'une mauvaise calibration, réduisant *in fine* l'efficacité énergétique d'une telle installation. Certains événements imprédictibles sont à l'origine de variations du fonctionnement du moteur résultant en une baisse d'efficacité ou en un risque pour l'intégrité du système lui-même. Les outils numériques disponibles pour la modélisation de tels phénomènes sont encore limités par les hypothèses simplificatrices nécessaires à la modélisation de systèmes trop coûteux sans cela.

La simulation des grandes échelles présente le meilleur compromis de modélisation d'écoulements turbulents. Elle permet la représentation partielle de phénomènes instationnaires sans atteindre ni le niveau de précision ni le coût de simulations numériques directes. Cependant, elle est incapable de représenter les événements extrêmes et conditions particulières à l'origine de variations pour le moteur sans pilotage particulier. L'assimilation de données incarne un candidat idéal pour le pilotage par les données de simulations des grandes échelles. La représentation et l'étude de variations et d'événements extrêmes via l'outil numérique, appelé jumeau numérique, pourraient permettre de prévenir les risques et baisses d'efficacité de l'équivalent réel du système, le jumeau physique. L'utilisation de l'AD pour la mécanique des fluides numérique comporte plusieurs objectifs : la calibration de fermetures de modèles de turbulences qui sont

souvent calibrés empiriquement et peu généralisables, la calibration de conditions aux limites pour à la fois compenser la prédiction erronée de modèles et atteindre une réalisation particulière d'écoulements impossible à obtenir sans influence extérieure et enfin la synchronisation de champs inférés à l'aide de données haute-fidélité du jumeau physique pour la reconstruction de structures particulières dans l'écoulement.

Quelle assimilation de données pour la mécanique des fluides ?

L'assimilation de données est une famille d'outils permettant l'inférence de modèles à partir d'échantillons haute-fidélité du système physique équivalent. Si elle a été largement utilisée dans le domaine des sciences météorologique, son application est encore relativement récente dans le domaine de la mécanique des fluides numérique. Deux grandes familles d'algorithmes se distinguent : l'assimilation de données variationnelle et séquentielle. Les revues de littérature réalisées pour ces deux familles d'algorithmes ont permis de dresser plusieurs conclusions. Les méthodes variationnelles semblent à l'heure actuelle plus populaires que leur équivalent séquentiel. Elles sont utilisées pour la reconstruction de champs et pour la calibration de fermetures de modèles de mécanique des fluides numérique, souvent via l'ajout de termes sources. Cependant, leurs applications privilégient l'inférence de régimes stationnaires ou de champs uniques de conditions initiales. L'assimilation de données séquentielle est une famille qui repose sur l'utilisation du filtre de Kalman et d'une quantité de variantes associées. En plus des utilisations citées pour l'équivalent variationnel, elle est utilisée pour les problèmes de contrôles de systèmes dynamiques et présente donc un avantage certain pour la reconstruction d'évènements marginaux tel que dicté par le paradigme du jumeau numérique.

C.O.N.E.S : Coupling OpenFOAM with Numerical Environments

Un des aspects importants de ce travail de recherche a été le développement d'un outil capable d'appliquer le filtre de Kalman d'ensemble aux problèmes de mécanique des fluides numérique. Pour cela, l'outil CONES a été développé afin de coupler la bibliothèque C++ OpenFOAM pour la résolution de problèmes de mécanique des fluides avec un code maison exécutant le filtre de Kalman d'ensemble. OpenFOAM présente une multitude de solveurs pour la résolution de problèmes variés. Sa composante open-source en faisait un choix idéal pour l'application de ce travail de thèse. Le premier cas étudié avec l'outil CONES est l'écoulement laminaire dans une cavité entraînant 2D. Ce problème constituait un candidat idéal pour le développement continu de CONES et constitue encore l'application par défaut pour le débogage et l'apprentissage de l'outil. Le comportement de la mise à jour paramétrique des différents types d'algorithmes mis en place dans CONES peut facilement être testé grâce à la cavité entraînant. En outre, son nombre de degrés de liberté est extrêmement faible facilitant sa mise en place sur des machines conventionnelles dédiées à la bureautique. CONES a également été mis en place sur plusieurs serveurs de calculs tels que les clusters nationaux Jean-Zay de l'IDRIS ainsi que sur la partition IRENE de Joliot-Curie sur le TGCC. La pérennité de l'outil ainsi que son développement continu sont assurés par de nouveaux travaux de recherche dédiés au développement et à la mise en place de nouveaux outils d'assimilation de données.

Augmentation d'une simulation RANS incompressible

Le niveau de complexité des trois applications principales de ce travail de thèse a été graduellement augmenté. Pour le premier cas d'étude, le pilotage par les données s'est concentré sur l'inférence d'une simulation RANS stationnaire incompressible à $\approx 513\,000$ mailles d'une configuration urbaine. Grâce aux partages de mesures expérimentales suffisantes pour l'inférence de la simulation par les données, l'outil a pour la première fois pu être mis à l'épreuve sur un cas concret. La description paramétrique des fermetures de modèles RANS $\mathcal{K} - \varepsilon$ et $\mathcal{K} - \omega$ SST a été inférée via l'utilisation de mesures expérimentales moyennées réalisées en soufflerie. Les mesures de vitesses par fil chaud au-dessus du corps non profilé de bâtiment ainsi que des prises de pressions sur la surface du corps ont permis d'optimiser les coefficients de fermeture des modèles de turbulence. La calibration de ces paramètres a entraîné une amélioration topologique des structures de recirculations en aval du bâtiment et sur son toit ainsi qu'une amélioration globale des quantités physiques étudiées. Le cas d'étude a en particulier mis en évidence l'importance de la qualité de l'information contenue dans les observations de sources hétérogènes plutôt qu'une grande quantité de données placées sur une grille structurée.

Augmentation d'une simulation LES incompressible

Ce deuxième cas d'étude concernait l'inférence d'un écoulement LES incompressible statistiquement stationnaire dans un canal turbulent à 350 000 mailles. Cette fois, la calibration du coefficient C_k du modèle de sous-maille Smagorinsky est effectuée à partir de mesures de vitesses instantanées d'une DNS haute-fidélité disponible. L'optimisation réalisée par le EnKF a convergé vers une valeur très faible mais non nulle du coefficient ce qui a réduit l'écart entre les moments statistiques des champs de vitesse inférés et des données DNS. La réduction de cet écart était cependant contrainte par les limites du modèle Smagorinsky. Dans un deuxième temps, le canal turbulent a permis l'étude de la synchronisation des champs de vitesse avec l'observation grâce à la correction apportée par l'estimation d'état du EnKF. Ceci a montré qu'une synchronisation satisfaisante des valeurs est possible avec un niveau de variance suffisant dans l'ensemble. La synchronisation locale des champs a également occasionné une réduction globale de l'écart entre les membres d'ensemble et la référence. Ce cas d'étude a notamment mis en évidence un des plus gros défauts du EnKF : l'effondrement de sa variance, entravant les capacités correctives de l'algorithme. La calibration fastidieuse de l'hyper-paramètre d'inflation a été nécessaire tout au long de la calibration de C_k et est au centre de l'étude de synchronisation des champs pour éviter la forte réduction de variance.

L'effondrement de variance du EnKF peut être mitigé grâce à la localisation qui vient atténuer les corrélations non physiques. Une stratégie d'hyper-localisation a donc été développée pour à la fois diminuer le coût de calcul, modérer le niveau de variance, et dans certains cas augmenter la précision de la prédiction. Ceci a été l'occasion de mettre à jour l'outil CONES pour le rendre plus simple d'utilisation, améliorer sa maintenabilité et son adaptabilité aux changements futurs. La variante développée appelée HLEnKF pour *hyper-localized ensemble Kalman filter* a pu être testé sur le canal turbulent où il a montré une optimisation du C_k optimale ainsi qu'une excellente synchronisation locale des champs sans aucune utilisation d'inflation et pour un coût de calcul plus faible.

Augmentation d'une simulation LES compressible

Le troisième cas étudié était de loin le plus difficile à mettre en place en raison de sa complexité. Le HLEnKF permettait de rendre le coût du calcul abordable et donc le cas d'étude envisageable. La géométrie du banc volute stationnaire a été adoptée pour mettre à l'épreuve l'inférence d'écoulements complexes de moteurs à combustion internes. Une simulation haute-fidélité de 131 millions de mailles a dans un premier temps permis de créer un jeu de mesures synthétiques par plans 2D. La simulation compressible et instationnaire inférée, plus basse fidélité, est constituée d'environ 8 700 000 mailles pour 35 membres dans l'ensemble. L'inférence de conditions aux limites a rarement été étudiée dans la littérature. La calibration de la description paramétrique de la condition d'entrée des simulations de l'ensemble a permis de reconstituer la condition imposée sur la simulation de référence. L'estimation paramétrique du HLEnKF a permis de reconstituer la condition de vitesse sinusoïdale pulsée sur les 35 membres d'ensemble malgré l'advection lente de l'information corrigée vers les zones d'observations. L'optimisation de la description paramétrique de l'entrée montre un comportement qui peut être statique ou dynamique. Ceci démontre le grand potentiel du HLEnKF pour la calibration de conditions d'entrées transitoires induisant l'apparition d'évènements marginaux dans l'écoulement. Une fois cette étape effectuée, il était possible de modifier le champ de vitesse sans risquer la divergence des calculs CFD. Le HLEnKF a montré une excellente synchronisation des champs de vitesses inférés dans les zones d'observation. L'inférence locale des champs s'est également montrée suffisante pour calibrer la distribution énergétique du champ de vitesse sur le plan considéré. L'influence de la correction est par ailleurs visible dans les zones de recirculations pour lesquelles aucune correction n'a été effectuée grâce à l'advection de l'information inférée dans l'écoulement. Ceci démontre que l'assimilation de données peut être utilisée comme filtre physique à travers duquel l'écoulement est inféré afin de reconstituer des évènements marginaux avants du fait de l'advection naturelle des corrections. Ce cas d'utilisation apparaît particulièrement pertinent pour des observations organisées sur des plans 2D à l'image de mesures PIV. Couplée à une calibration adéquate de conditions menant à l'émergence d'évènements extrêmes, la reconstruction fournie par le HLEnKF et plus largement par l'assimilation de données séquentielle semble adéquate pour l'application au paradigme du jumeau numérique.

Perspectives d'évolution

Une certaine quantité de développements est encore nécessaire pour atteindre le stade d'évolution de l'algorithme permettant une application concrète au paradigme du jumeau numérique. Ceux-ci sont listés ci-après, par ordre de difficulté :

- Concernant l'application de l'assimilation de données séquentielle aux systèmes industriels et aux régimes transitoires tels qu'étudié avec le banc à volute stationnaire, plusieurs évolutions sont nécessaires :
 - L'étude de conditions d'entrée perturbées et transitoires induisant directement l'apparition d'évènements extrêmes.
 - L'utilisation de données expérimentales sans connaissance de conditions aux limites préalable pour la validation de l'algorithme dans un cas concret.

- De manière générale, l'utilisation d'algorithmes tels que le *Ensemble Kalman Smoother* préférentiellement hyper-localisé pourrait permettre la prise en compte de temps successifs pour l'assimilation de données avec une faible résolution temporelle et améliorer les résultats.
- L'amélioration de l'hyper-localisation passe par plusieurs développements :
 - L'optimisation des hyper-paramètres de localisation permettrait la densification du réseau d'observations avec l'hyper-localisation.
 - La modification de la fusion des zones hyper-localisées lorsqu'elles s'intersectent pour la création de nouvelles zones en fonction du nombre d'observations disponibles. Ceci implique un lissage de la localisation de covariance et de l'inflation aux frontières des zones.
 - La localisation de l'inflation dans le cas de régimes transitoires sans périodicité.
- Ces derniers points suggèrent une complexification de la configuration hyper-paramétrique de l'assimilation de données séquentielle. L'objectif principal à long terme serait alors de rendre le paramétrage de l'algorithme quasi automatique tout au long de la procédure d'optimisation. Une sélection automatique des observations, de la localisation et de l'inflation permettrait ainsi de favoriser la réduction des interventions humaines nécessaire au choix des hyper-paramètres. Le domaine de l'intelligence artificielle pourrait permettre cette étape cruciale. L'utilisation de l'IA pour l'inférence d'écoulements semble limitée par l'absence de réalité physique dans l'optimisation. Elle pourrait en revanche trouver tout son potentiel d'utilisation dans le couplage à l'assimilation de données séquentielle en participant à son automatisation. Ceci permettrait de rendre l'assimilation de données, en particulier séquentielle, utilisable par tous et applicable dans des cas concrets industriels pour la progression de l'efficacité des systèmes réels.



Annexe : Chapitre 6

A.1 Utilisations d'informations physiques multiples dans la procédure d'AD

Afin de tester la sensibilité de l'algorithme d'AD à de multiples informations physiques disponibles sur un capteur, un test supplémentaire a été réalisé. Ce test, appelé DA-LESA, est presque identique à DA-LES1. La seule différence est que, pour chaque capteur, les trois composantes du champ de vitesse sont utilisées. Pour rappel, pour les tests DA-LES1, DA-LES2 et DA-LES3, seule la composante u_x du champ de vitesse a été utilisée comme observation dans la phase d'analyse d'AD. Par conséquent, pour DA-LESA, la matrice d'observation est composée de $408 \times 3 = 1224$ valeurs à chaque phase d'analyse d'AD. La matrice de covariance de l'erreur de mesure est exprimée par $\mathbf{R} = \sigma_m^2 \mathbf{I}$, où σ_m quantifie l'incertitude des mesures. Dans ce cas, σ_m est calculé en tenant compte d'une incertitude de 5% sur la vitesse maximale observée sur les données d'observations de la DNS pour imiter la précision de mesures expérimentales. Par conséquent, comme la vitesse maximale observée dans ces données est de $\approx 0.9 \text{ m.s}^{-1}$, $\sigma_m \approx 0.045$ et est dans ce cas la même pour tous les capteurs. Ce choix implique que la confiance dans les observations est plus faible à l'approche de la paroi, puisque les valeurs de vitesses diminuent. Cette décision est bénéfique pour obtenir un comportement robuste de l'EnKF, car de grandes divergences entre la DNS et la LES peuvent être observées proche paroi.

Les résultats de l'optimisation sont similaires à ceux des autres procédures d'AD, ce qui indique que la procédure de l'EnKF est robuste. La valeur optimisée de la constante du modèle est $C_k \approx 0.025$ qui est 3.7 plus petite que la LES de base et correspond à $C_S \approx 0.06$. Pendant la procédure d'AD, les valeurs présentent des oscillations dans l'intervalle $C_k \in [0.020, 0.030]$. La DA-LESA montre également une bonne amélioration dans la prédiction de la vitesse de frottement $u_\tau = 0.052$ avec une sur-prédiction de celle-ci de 8.3%, par rapport aux 28% de

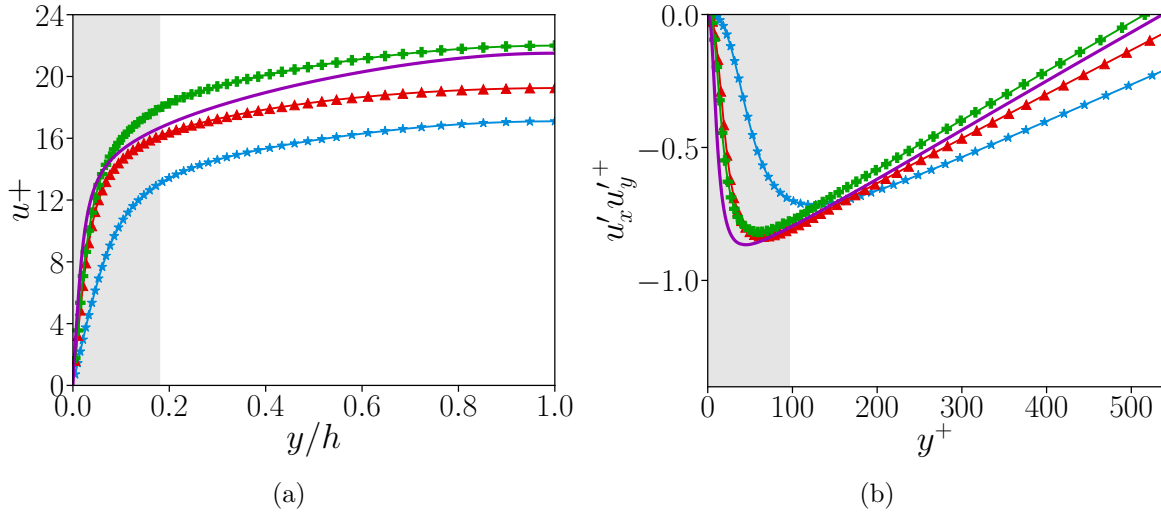


FIGURE A.1 : Profils de vitesse sur le demi-canal pour DA-LESA (\blacktriangleright), DA-LES1 ($+$), BL-LES ($*$) et DNS (\times)

BL-LES de base. La vitesse moyenne normalisée u^+ par rapport à y^+ et la contrainte de cisaillement résolue normalisée des procédures DA-LES1 et DA-LESA sont présentées sur la Fig. A.1. Les autres moments statistiques du champ de vitesse ne sont pas présentés ici par souci de concision, car ils fournissent des informations similaires. Les différences entre les procédures pour la prédiction des moments statistiques sont notables et principalement associées à la différence dans la prédiction de la vitesse de frottement, qui est moins précise pour la DA-LESA. Une raison possible est associée au niveau de confiance dans l'observation, qui a été fixé au même niveau pour les trois composantes du champ de vitesse. Dans la région proche de la paroi, la composante u_x est environ un ordre de grandeur plus grande que les deux autres composantes, et les incertitudes propagées dans le vecteur d'observation agissent comme un bruit aléatoire pour u_y et u_z . Le problème de la détermination d'une description hyper-paramétrique optimisée du niveau de confiance de l'observation, qui a dégradé la précision globale de l'assimilation de données dans ce cas, mérite de plus amples approfondissements.

A.2 Détails supplémentaires sur la synchronisation

La sensibilité de la synchronisation à l'inflation dans la description paramétrique du modèle et dans la variance de l'observation est ici discutée. La Fig. A.2a montre la racine de l'erreur quadratique moyenne normalisée Φ avec différents niveaux d'inflation sur les paramètres et sur l'état. Trois niveaux d'inflation sur les paramètres sont utilisés, des nuances claires aux nuances foncées : 0%, 2% et 5%. L'inflation d'état est fixée à 0% pour les couleurs bleues, 5% pour les couleurs vertes et 15% pour les couleurs orange-rouge. L'inflation sur les paramètres de modèle semble avoir un effet négligeable sur la synchronisation obtenue via l'AD par rapport à l'inflation d'état étant donné qu'il n'y a pas de différences apparentes dans les niveaux d'erreur pour un même groupe de couleurs. La Fig. A.2b considère différents niveaux de confiance dans l'observation (c'est-à-dire les valeurs dans la matrice \mathbf{R}) pour un niveau constant d'inflation

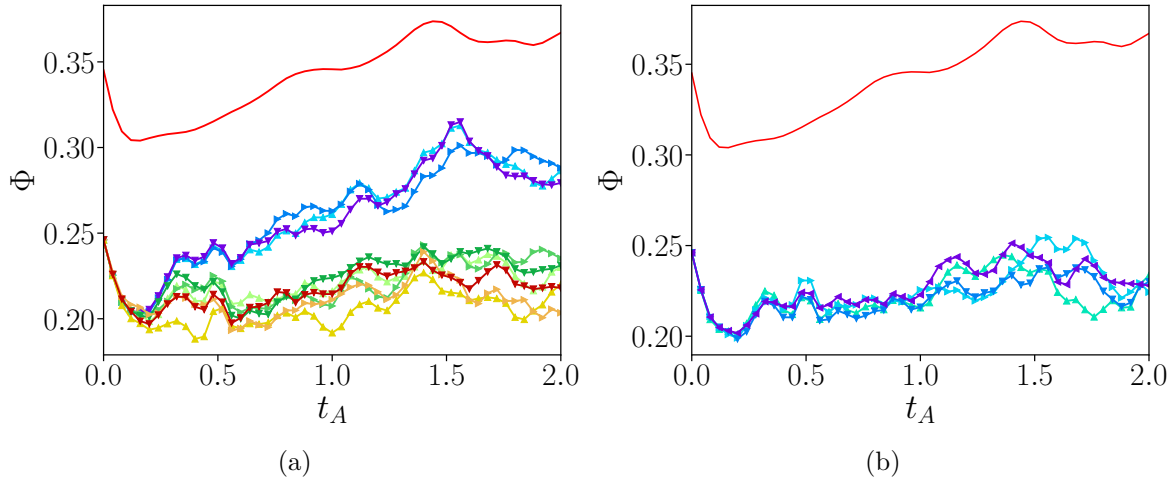


FIGURE A.2 : Test de sensibilité pour la fonction de coût Φ pour les 408 capteurs utilisés dans l'AD. (a) Différents niveaux d'inflation des paramètres sont utilisés, de la couleur claire à la couleur foncée : 0%, 2% et 5% d'inflation. (b) Différentes valeurs pour la confiance imposée dans l'observation sont présentées, du vert au violet : 0.5%, 1%, 5% et 10% de confiance

d'état. La fonction de coût Φ pour quatre niveaux de variance prescrite pour l'observation, 0.5%, 1%, 5% et 10%, est présentée pour la même inflation d'état (5%). Là encore, la précision de la synchronisation ne semble pas être affectée par le niveau de confiance dans les observations testées ici, qui se situe dans la gamme des recommandations pour une application robuste de l'EnKF. Une confiance très faible ou très élevée dans l'observation peut conduire à une mauvaise synchronisation ainsi qu'à une optimisation paramétrique inexacte, comme le montrent Tandeo et al. [128].

A.3 Consommation des calculs

L'assimilation de données réalisée pour le cas d'étude chapitre 6 a demandé 40 membres d'ensemble calculant chacun sur un seul cœur pour une consommation totale de 35 Go de RAM. La convergence des simulations ainsi que le calcul des moyennes statistiques a duré environ 3 jours. Le temps scalaire associé est donc d'environ 3 000 heures. Ces chiffres sont plus élevés que pour les simulations initiales qui ont duré 4 à 5 heures sur 20 cœurs. Le temps scalaire associé est d'environ 100 heures. Cependant, ces temps sont nettement inférieurs au calcul de référence DNS, pour lequel environ un million d'heures scalaires étaient utilisées pour dissiper les conditions initiales imposées et calculer les moyennes statistiques.

Annexe : Validation du filtre de Kalman d'ensemble hyper-localisé

Le travail présenté dans le [chapitre 6](#) a déjà fait l'objet d'une publication [140]. Ces résultats sont à présent utilisés pour la validation de la variante du filtre de Kalman développée dans le [chapitre 3](#), le filtre de Kalman d'ensemble hyper-localisé.

Cette validation est effectuée en deux étapes : (1) La comparaison de l'estimation paramétrique fournie par le HLEnKF par rapport au EnKF classique. (2) La comparaison des capacités de synchronisation des deux algorithmes. Ce dernier point est crucial pour l'étude d'évènements extrêmes dans le paradigme du jumeau numérique, comme cela était précisé dans la Sec. 6.4. À noter que le gain en termes de coût de calcul ne sera pas abordé ici puisqu'il est détaillé dans la Sec. 4.5 du [chapitre 4](#).

Dans un premier temps, l'optimisation de la constante du modèle Smagorinsky est comparée pour les deux algorithmes. La configuration utilisée est la même que pour la procédure DA-LES1 de la Sec. 6.2.2.1. Comme pour cette procédure, aucune inflation n'est utilisée de $t_A = 0$ à $t_A = 12$ soit la totalité du calcul effectué pour cette comparaison. Ce temps de calcul correspond à 100 phases d'assimilations de données, soit une toutes les $0.12t_A$ pour rappel. Dans la procédure DA-LES1, des niveaux d'inflations élevés sont utilisés dans la suite du calcul pour obtenir l'optimum global du coefficient C_k . Cette procédure, ainsi que les procédures DA-LES2 et DA-LES3, ont permis d'établir un intervalle pour lequel C_k est complètement convergé, soit entre $C_k = 0.01$ et $C_k = 0.025$. Cet intervalle est indiqué par la zone grise sur la Fig. B.1a. Il représente l'intervalle cible de l'optimisation. La courbe rouge représente l'évolution de C_k pour l'EnKF standard. Ce coefficient converge vers une valeur locale après $8t_A$. La Fig. B.1b montre les écarts types associés aux courbes commentées. L'effondrement de variance de l'ensemble pour la procédure du EnKF est responsable du manque de précision de l'optimisation. C'est pour cette raison qu'il est nécessaire d'augmenter virtuellement cette variance grâce à l'inflation dans la suite du calcul

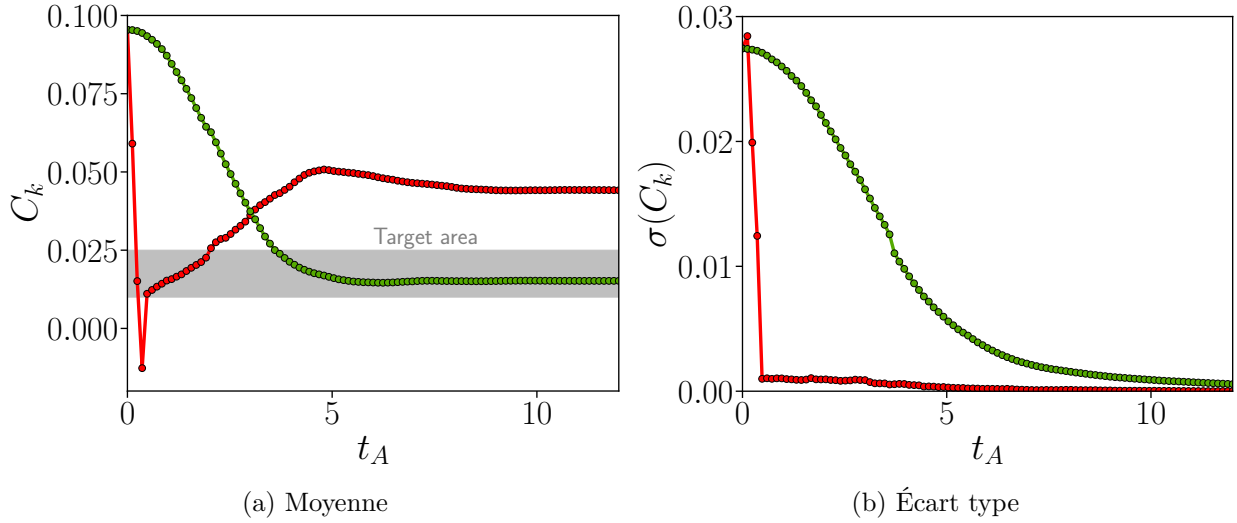


FIGURE B.1 : Évolution du coefficient C_k du modèle Smagorinsky au cours des 100 premières phases d’assimilations ($12t_A$) sans aucune inflation. La courbe rouge correspond à l’algorithme du EnKF et la courbe verte au HLEnKF. La zone cible grisée correspond aux valeurs convergées établies avec l’EnKF et de l’inflation dans la Sec. 6.3.

pour la procédure DA-LES1 dans le chapitre 6. La courbe verte représente l’évolution de C_k dans le cas du HLEnKF. Pour rappel, une valeur de C_k est calculé par chaque zone hyper-localisée. La valeur finale renvoyée vers le modèle de mécanique des fluides correspond à la moyenne de ces valeurs. Ceci a pour conséquence de ralentir la dynamique du système d’optimisation mais de grandement améliorer la précision de la prédiction de la valeur convergée, comme en témoigne la valeur finale atteinte pour $C_k \approx 0.015$ grâce au HLEnKF. Cette précision accrue est le fruit de la diminution lente de la variance, contrairement au EnKF classique. Dans ce cas précis, le HLEnKF prévient l’utilisation de l’inflation. Cette caractéristique est importante pour des procédures pour lesquelles il est souvent délicat d’évaluer les différents hyper-paramètres intervenant dans la configuration.

Dans un deuxième temps, les capacités de synchronisation du HLEnKF sont comparées à l’EnKF standard. Pour cela, la procédure utilisée dans la Sec. 6.4 est mise à contribution. La configuration est similaire : les champs de vitesses des 40 membres d’ensemble sont différents et issus de la configuration optimisée DA-LES2 et aucune estimation paramétrique n’est réalisée. Le jeu d’observations utilisé est identique. Le calcul est exécuté sur $2t_A$ avec une phase d’assimilation toutes les $0.04t_A$ soit 50 phases d’assimilations. La taille des régions hyper-localisées est calibrée pour éviter le chevauchement des zones. Seule l’échelle de localisation de covariance est donc modifiée. La structure de la matrice L est celle présentée dans l’équation 3.66, où le paramètre $l = 0.0494$ dans les directions de l’écoulement et de l’envergure et $l = 0.0247$ dans la direction normale à la paroi. Ces échelles de localisations sont utilisées à la fois pour le HLEnKF et l’EnKF standard dans les comparaisons des Fig. B.2 et B.3. La Fig. B.2 montre les valeurs de Φ pour quatre cas différents : l’EnKF standard sans localisation physique mais avec localisation de covariance ainsi que le HLEnKF pour 0% et 5% d’inflation. L’algorithme hyper-localisé montre une réponse presque immédiate dans les premières phases du calcul avec

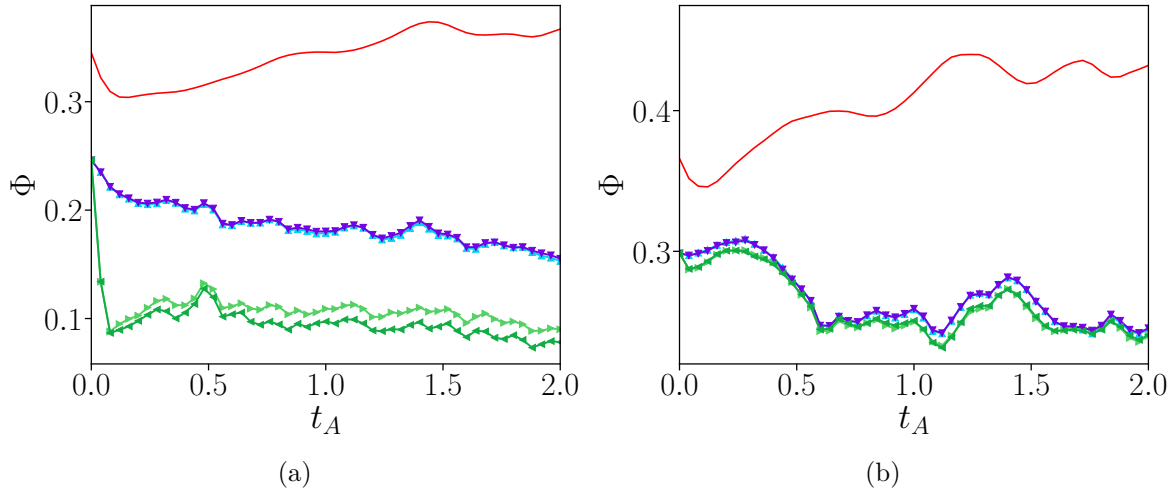


FIGURE B.2 : Test de sensibilité pour la fonction de coût Φ pour les 408 capteurs utilisés dans l'AD à gauche (a) et non utilisés à droite (b). Deux niveaux d'inflation des paramètres sont utilisés, de la couleur claire à la couleur foncée : 0% et 5% d'inflation. Les couleurs bleues correspondent aux résultats de l'EnKF et les couleurs vertes à ceux du HLEnKF.

un niveau d'erreur extrêmement faible comparé aux valeurs de l'EnKF pour les 408 capteurs utilisées dans l'AD. En revanche, les niveaux d'erreur atteints pour les capteurs non utilisés sont similaires, bien qu'inférieurs aux niveaux de la Fig. 6.16 Sec. 6.4. Cela s'explique par le fait que l'échelle de localisation est bien plus contraignante dans le cas présent que dans le cas original de la Sec. 6.4. Les corrections apportées par les procédures d'AD sont donc aussi faibles pour l'EnKF standard que pour le HLEnKF sur les capteurs non utilisés dans les AD. Les résultats semblent cependant légèrement meilleurs pour le HLEnKF.

Enfin, les Fig. de synchronisation B.3 peuvent permettre d'interpréter plus en détail ces résultats. Comme pour la Sec. 6.4, elles illustrent l'évolution de la vitesse instantanée dans le sens de l'écoulement u_x normalisée par la vitesse moyenne de la ligne médiane u_c pour les trois capteurs mentionnés. La vitesse échantillonnée à partir de la DNS, qui est utilisée comme observation, est indiquée en bleu. Les données échantillonnées au même endroit à partir des membres de l'ensemble de la procédure d'AD sont indiquées en noir. Dans ce cas, la ligne noire correspond à une moyenne d'ensemble. Les zones d'ombre représentent visuellement le niveau de confiance/la variabilité des données. Plus précisément, la zone bleue est connectée aux valeurs incluses dans la matrice de covariance \mathbf{R} avec une taille de $2\sigma_m$. La première ligne de cette série correspond au cas sans aucune assimilation de données. La deuxième ligne correspond aux valeurs corrigées par l'EnKF standard et la dernière par le HLEnKF. Les courbes non corrigées sont un indicateur du niveau de variance naturelle de l'ensemble dû à la distribution *a priori* des champs de vitesse choisis pour cette étude. Les deux procédures d'AD montrent des corrections très similaires sur le capteur non utilisé dans les AD (colonne de droite) avec des niveaux de variances diminuant peu par rapport au cas sans AD. Ces dernières remarques sont similaires aux conclusions des comparaisons de l'erreur Φ . En outre, les résultats semblent meilleurs sur ce capteur que dans le cas de la Sec. 6.4. Cela suggère qu'une correction très localisée et sans inflation est meilleure pour les capteurs non utilisés dans

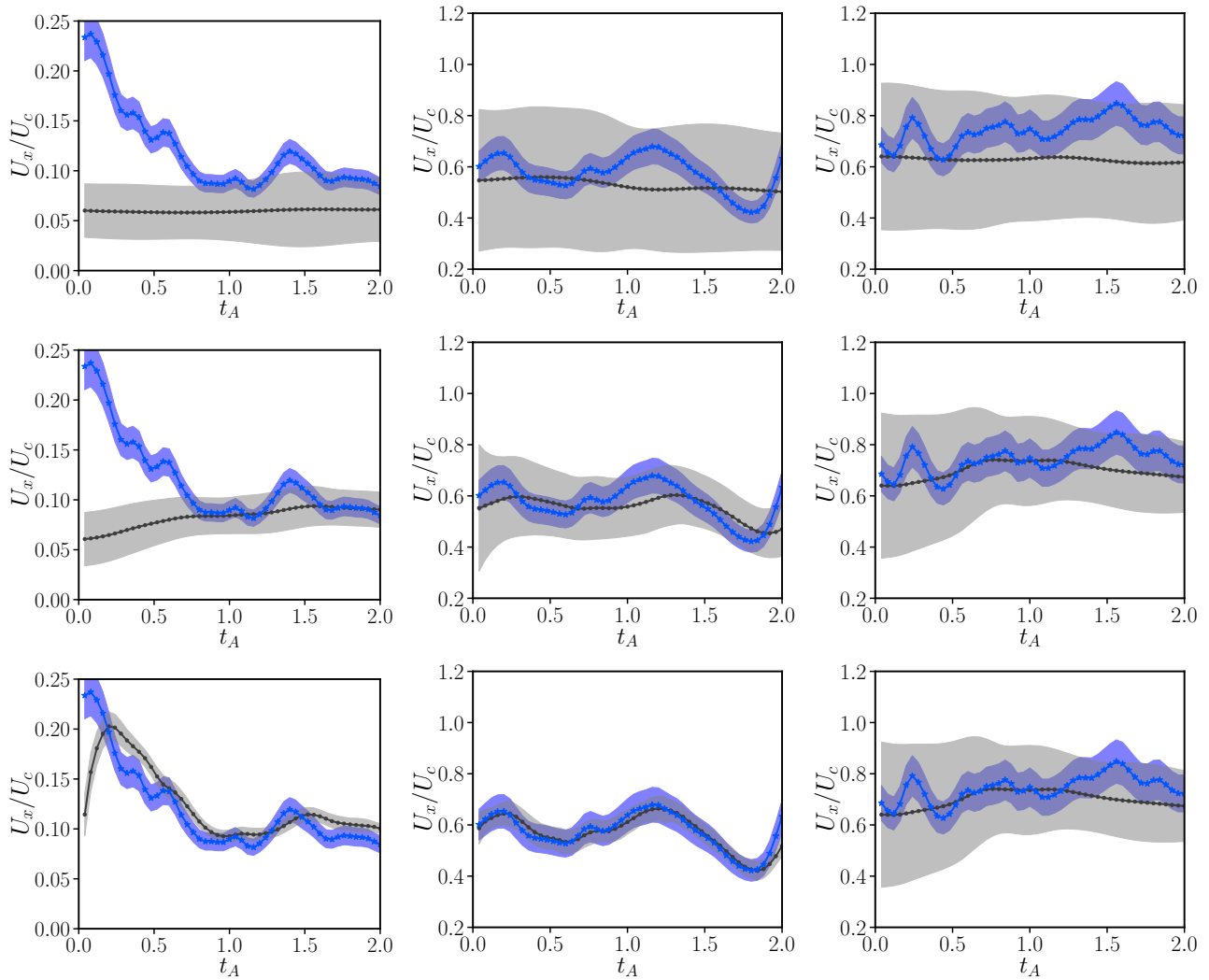


FIGURE B.3 : Synchronisation du champ de vitesse pour le HLEnKF sans inflation avec un niveau de localisation de covariance égal entre les procédures. Les données en bleu représentent l'observation, tandis que les lignes noires indiquent la prédiction de l'ensemble. Les résultats sont présentés pour (colonne de gauche) le capteur 1, (colonne centrale) le capteur 2 et (colonne de droite) le capteur 3. Aucune correction n'est effectuée sur la première ligne. La deuxième ligne correspond au EnKF sans localisation physique. La dernière ligne correspond au HLEnKF.

l'AD. Concernant les deux capteurs utilisés dans l'AD, les résultats du HLEnKF montrent une excellente faculté de synchronisation (colonne de gauche et du milieu). Les courbes de vitesses se superposent presque parfaitement, en particulier pour le capteur 2. Par ailleurs, le niveau de variance dans la procédure du HLEnKF diminue fortement et immédiatement dès la première phase d'assimilation mais conserve un niveau de variance équivalent au niveau de confiance donné aux observations. Au contraire, la procédure standard expose une diminution de variance classique dans le cas d'une procédure fortement localisée, avec une capacité de synchronisation peu satisfaisante. Ces derniers résultats soulèvent deux conclusions importantes : (1) le HLEnKF est capable d'une réponse presque instantanée dans le suivi de l'information contenue dans l'observation. (2) Le niveau de variance nécessaire est relativement faible et diminue peu, ce qui suggère une utilisation d'inflation limitée sur le long terme.

Pour résumer, le filtre de Kalman hyper-localisé présente un très grand potentiel pour l'estimation de paramètres et pour la synchronisation de champs de vitesse. Malgré la localisation fortement contrainte induite par l'algorithme, une amélioration globale de la prédiction de l'écoulement instantané est de nouveau obtenue. L'optimisation de la description paramétrique ainsi que l'estimation d'état sont moins dépendantes des niveaux d'inflation utilisés dans la procédure. Ce point pourrait permettre d'éviter l'utilisation et la calibration de cet hyper-paramètre, souvent fastidieuses pour des applications complexes. L'aptitude du HLEnKF à suivre les champs instantanés d'une simulation LES dans un canal plan turbulent permet d'envisager son utilisation dans une application aux régimes fortement transitoires, par exemple pour la reconstruction d'évènements extrêmes, locaux en temps et en espace.

Annexe : Chapitre 7

C.1 Calibration de la condition d'entrée

C.1.1 Évolution de l'écart type des paramètres

Les Fig. C.1 et C.2 montrent l'évolution de l'écart type des paramètres pour les optimisations de la description paramétrique de la condition d'entrée effectuée dans la Sec. 7.3 à partir de distributions *a priori* 1 et 2. Ces figures montrent l'effondrement de variance classique du filtre de Kalman [4]. En particulier, la valeur atteinte par tous les paramètres étudiés correspond au taux de variance imposé via l'inflation paramétrique stochastique de 0.5% puis 0.0375% utilisée pendant la procédure d'assimilation de données. Ceci démontre l'importance de l'inflation pour l'optimisation de paramètres grâce au filtre de Kalman.

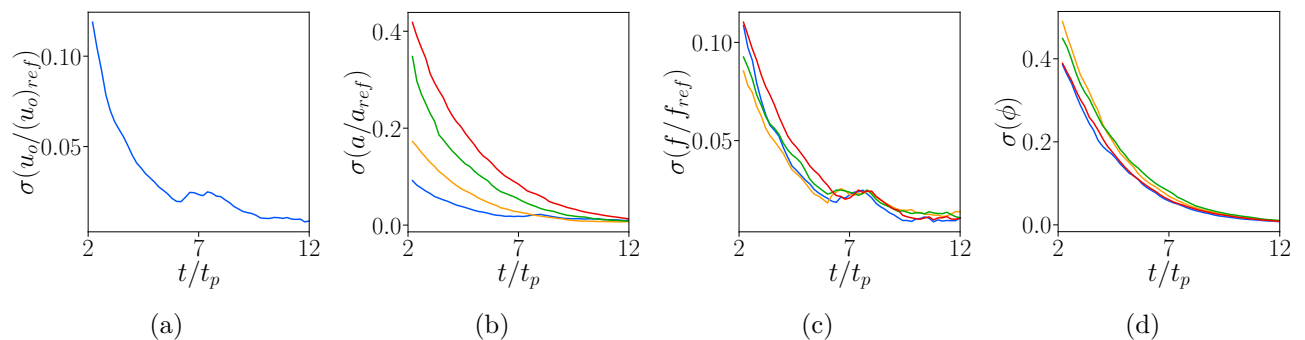


FIGURE C.1 : Évolution de l'écart type des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution *a priori* 1. Bleu : fonction sinusoïdale 1. Jaune : fonction sinusoïdale 2. Vert : fonction sinusoïdale 3. Rouge : fonction sinusoïdale 4.

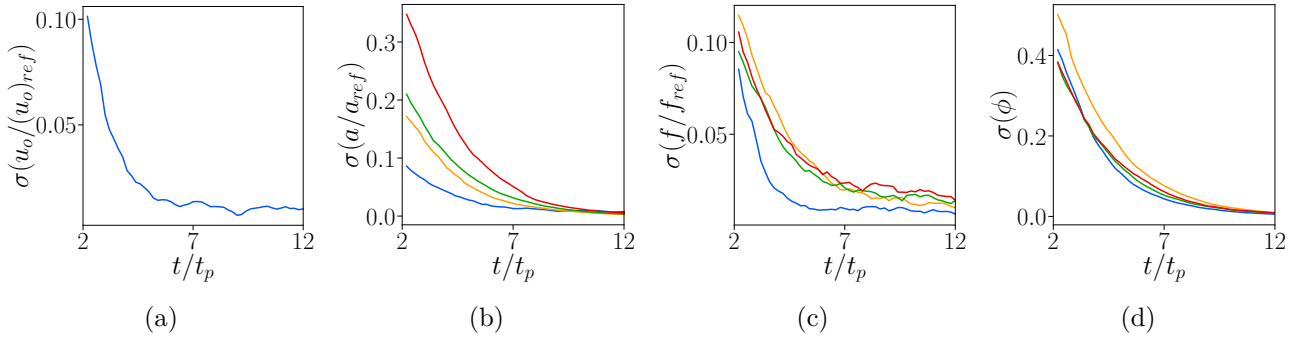


FIGURE C.2 : Évolution de l'écart type des paramètres décrivant la condition d'entrée au cours de la procédure d'assimilation pour la distribution *a priori* 2. Bleu : fonction sinusoidale 1. Jaune : fonction sinusoidale 2. Vert : fonction sinusoidale 3. Rouge : fonction sinusoidale 4.

C.1.2 Estimation paramétrique à partir d'un jeu d'observations différent

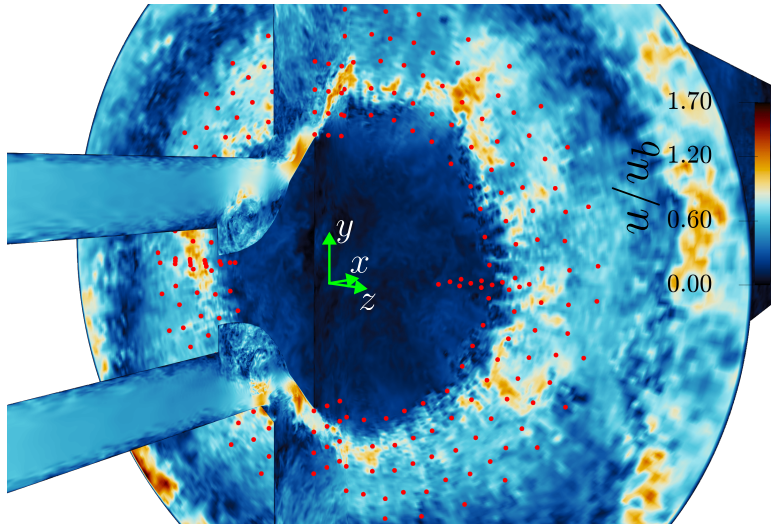


FIGURE C.3 : Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les points rouges correspondent au second jeu de capteurs d'observations sélectionnés pour la calibration de la condition d'entrée.

La calibration de la description paramétrique réalisée dans la Sec. 7.3 est ici exécutée à partir de l'assimilation d'un jeu d'observations différent. Les capteurs utilisés sont indiqués sur la Fig. C.3. Ceux-ci reprennent certains capteurs utilisés pour l'optimisation de la Sec. 7.3 situés dans le jet en sortie de la conduite d'admission en additionnant certains capteurs situés dans le plan $x = 0.124$ mm. L'*a priori* 2 est utilisé comme conditions initiales de la description paramétrique des 35 membres de l'ensemble. Le reste de la configuration et des hyper-paramètres est identique à la configuration utilisée pour l'optimisation de l'*a priori* 2. La Fig. C.4 montre le résultat de cette optimisation. Celle-ci présente de très grandes similarités avec les résultats obtenus dans la Sec. 7.3. Cependant, un déphasage de l'onde résultante est ici visible. Ceci suggère que l'espace paramétrique disponible via l'utilisation de l'inflation dans la procédure d'assimilation n'est pas

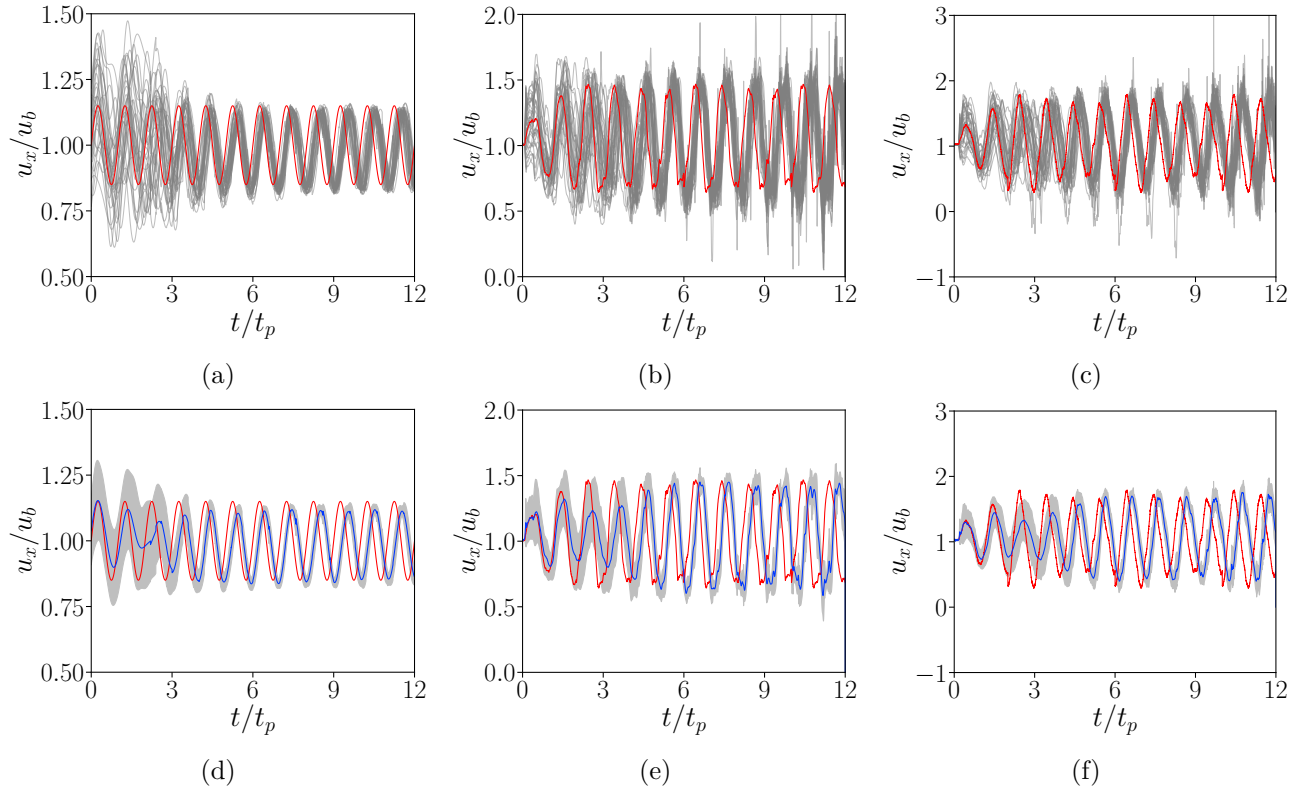


FIGURE C.4 : Évolution de la condition d'entrée de vitesse axiale pour la distribution ***a priori*** 2 avec un jeu d'observations différent. Courbe rouge : condition d'entrée de référence LES-HF. Courbes grises : conditions d'entrée des 35 membres d'ensemble. Courbe bleue : moyenne des conditions des membres d'ensemble. Zone grisée : écart type par rapport à la moyenne. De gauche à droite : Conditions mesurées au début, au milieu et à la fin de la conduite d'admission par les capteurs 7.4.

assez large. Il ne permet pas de cibler l'optimisation optimale de la description paramétrique à partir de ce jeu d'observations. Une solution simple serait d'augmenter le taux d'inflation paramétrique imposé dans la procédure. Cette solution n'est toutefois pas étudiée ici par souci d'économie. En outre, cette solution apporterait des conclusions similaires aux procédures de la Sec. 7.3.

C.2 Reconstruction et synchronisation de structures locales

C.2.1 Synchronisation des composantes u_y et u_z et de la pression p

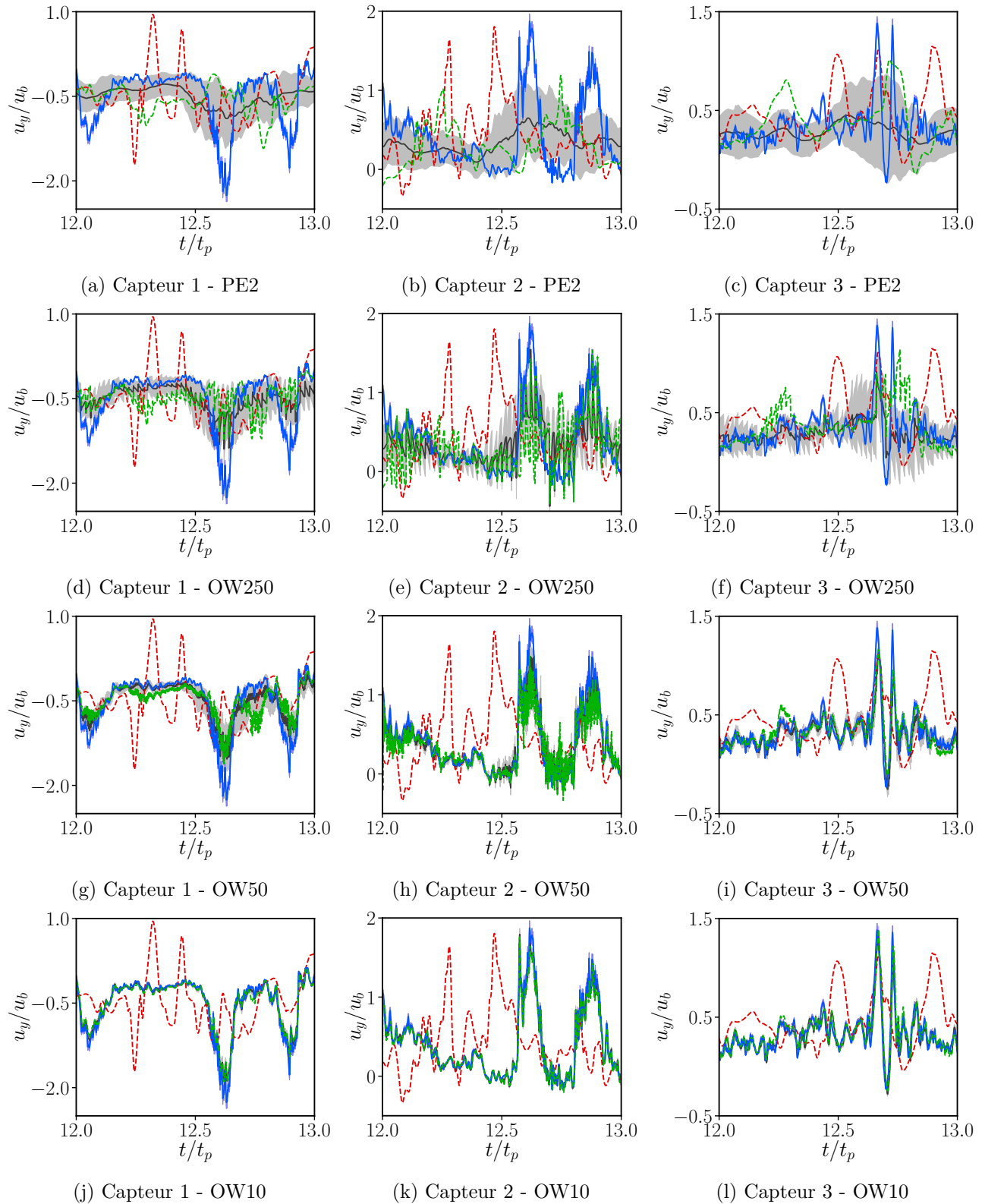


FIGURE C.5 : Synchronisation de la composante de la vitesse u_y adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (*a priori*) et les vertes dans le cas associé à chaque configuration (*a posteriori*).

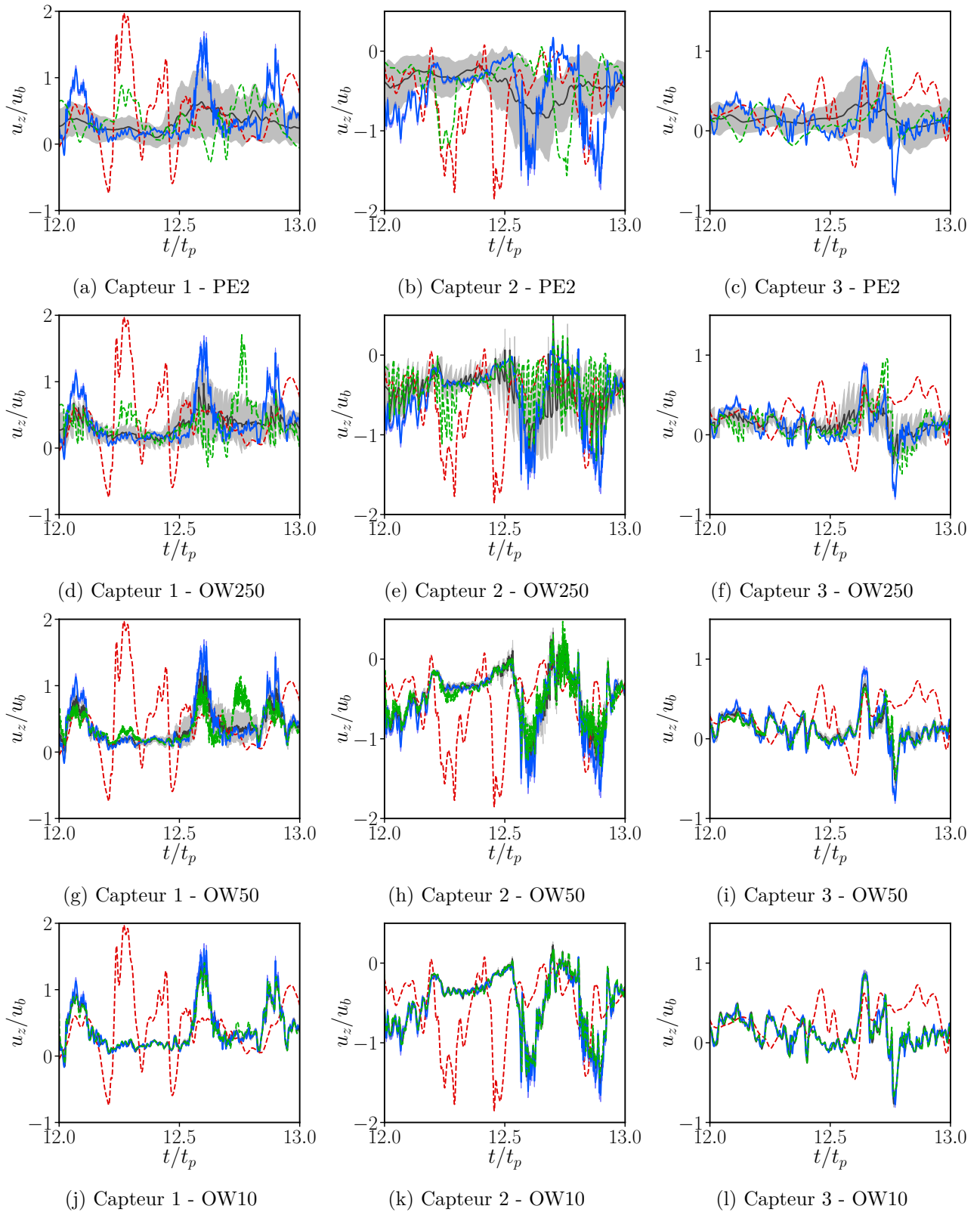


FIGURE C.6 : Synchronisation de la composante de la vitesse u_y adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (*a priori*) et les vertes dans le cas associé à chaque configuration (*a posteriori*).

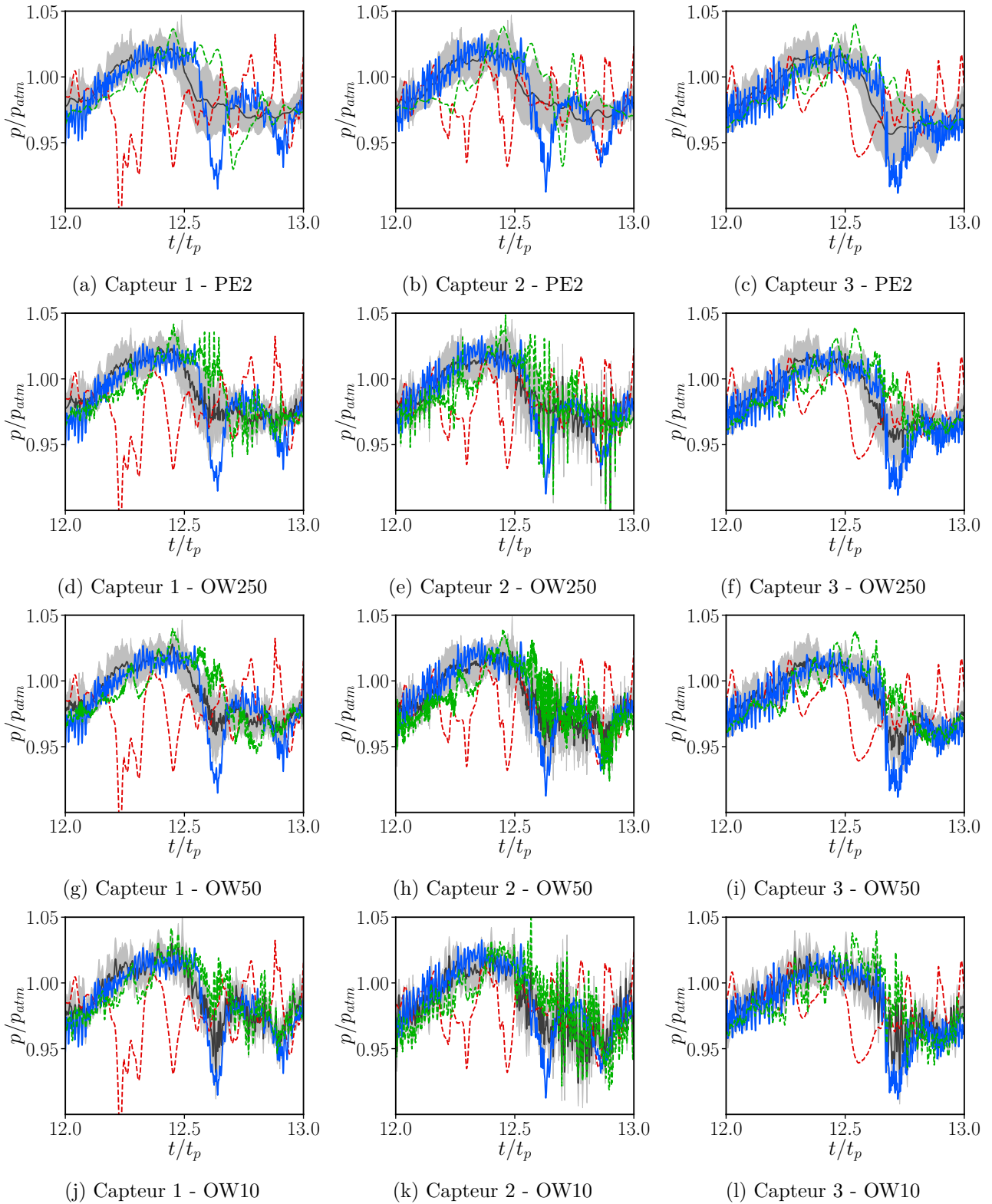


FIGURE C.7 : Synchronisation de la pression p adimensionnée par p_{atm} en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (*a priori*) et les vertes dans le cas associé à chaque configuration (*a posteriori*).

C.2.2 Capacités de synchronisation pour un nombre d'observations plus élevé

Un jeu de capteurs plus dense est utilisé dans cette section annexe pour la synchronisation des champs de vitesse inférés avec les données observées. Ce jeu d'observations, nommé SE-set2 est composé de 1000 capteurs répartis sur la même couronne que le jeu SE-set1. Il est illustré sur la Fig. C.8. La taille des zones d'hyper-localisation est encore diminuée pour réduire le chevauchement des zones. Elle est de 0.000 75 m pour SE-set2. L'échelle de corrélation associée est égale à $l = 0.00025$. Ceci conduit à 24 048 mailles soit $N = 72\,144$ degrés de liberté dans la procédure répartis sur 485 zones hyper-localisées. Tous les autres paramètres de la configuration sont identiques à la configuration LES-OW10 du chapitre 7.

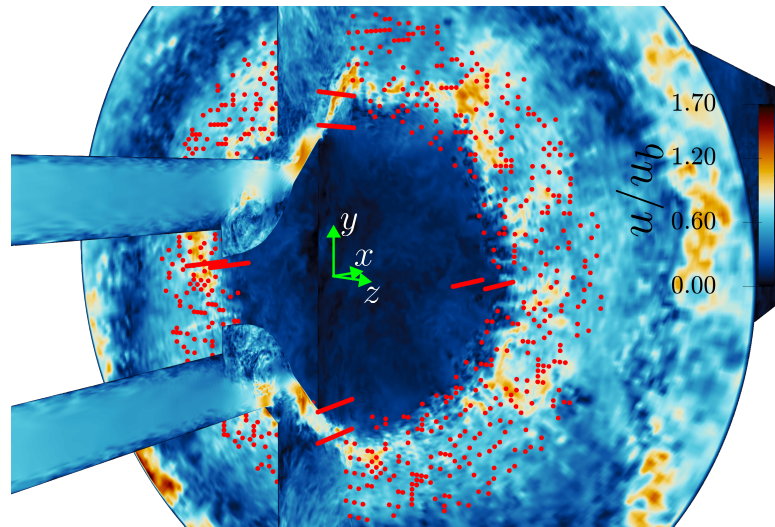


FIGURE C.8 : Le champ de vitesse est tracé sur les sections $z = 0$ et $x = 0.124$ m. Les positions des capteurs du jeu SE-set2 d'observations sont indiquées par les points rouges.

La Fig. C.9 montre les figures de synchronisations pour les capteurs explicités dans la Sec. 7.4. La première et deuxième ligne de figures correspondent aux configurations LES-PE2 et LES-OW10 déjà commentés dans le corps du chapitre 7. La dernière ligne correspond aux figures de la configuration utilisant le jeu d'observations SE-set2. Pour cette configuration, le troisième capteur (colonne de droite) est un capteur qui n'est pas utilisé pour l'assimilation de données. Pour ce dernier, la vitesse moyenne indiquée par la courbe noire est similaire au cas sans estimation d'état LES-PE2. L'écart type est en revanche largement réduit, ce qui implique que les valeurs de vitesses éloignées des capteurs d'observations sont affectées par de légères corrections permettant une diminution de l'écart entre les simulations inférées et la référence. Les capteurs 1 et 2 (colonne de gauche et du milieu) présentent des résultats très similaires au cas LES-OW10. Leurs écarts types légèrement plus élevés indiquent que la longueur de localisation du jeu SE-set2 et le possible chevauchement de quelques zones d'observations est responsable d'une légère diminution de la précision de la correction pour certains capteurs.

La Fig. C.10 illustre les cartographies des modes principaux de la transformée de Fourier azimutale de la vitesse axiale $u_x - \bar{u}_x$ au temps $t = 13t_p$ pour les 11 916 capteurs de la section $x = 0.124$ tel qu'explicité dans la Sec. 7.4 du chapitre 7. En particulier, les figures de gauche et du milieu présentent les résultats dans le cas de référence LES-HF ainsi que dans le cas LES-

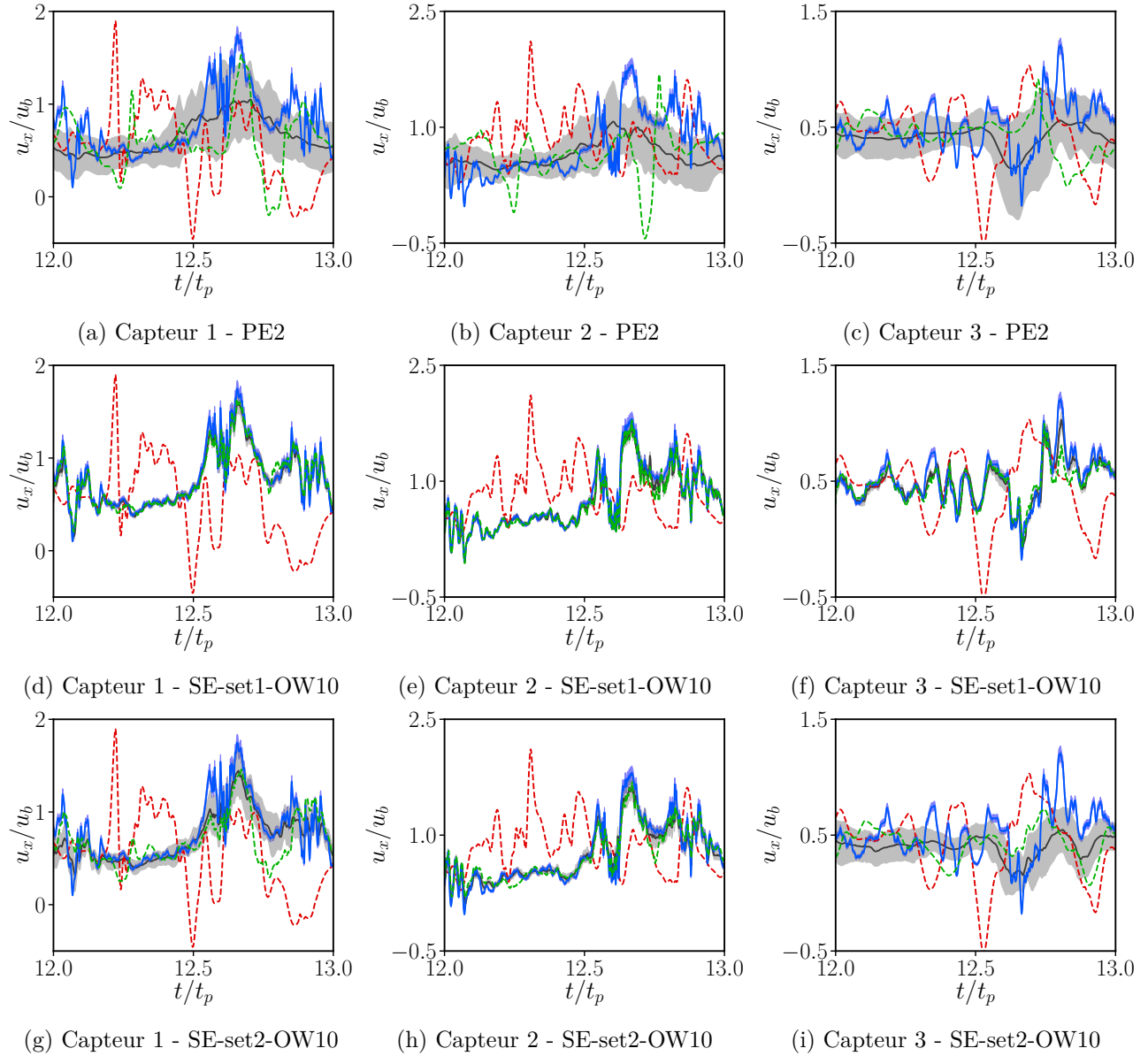


FIGURE C.9 : Synchronisation de la composante axiale de vitesse u_x adimensionnée par u_b en trois points utilisés par le HLEnKF et pour différentes configurations. Les données en bleu représentent l'observation. Les lignes noires indiquent la prédiction moyenne de l'ensemble. Les lignes rouges correspondent aux évolutions pour le membre 27 sans aucune assimilation de données (*a priori*) et les vertes dans le cas associé à chaque configuration (*a posteriori*).

OW10 pour le SE-set1 d'observations déjà commentés dans le corps du manuscrit. La figure de droite correspond à la cartographie dans le cas du jeu SE-set2 d'observations. Cette dernière affiche une très grande similarité dans les distributions d'énergie cinétique des 10 premiers modes de Fourier par rapport au cas SE-set1. Bien que les niveaux d'énergie semblent différer selon les modes, toutes les positions d'énergies maximales pour chaque mode selon r/R semblent identiques.

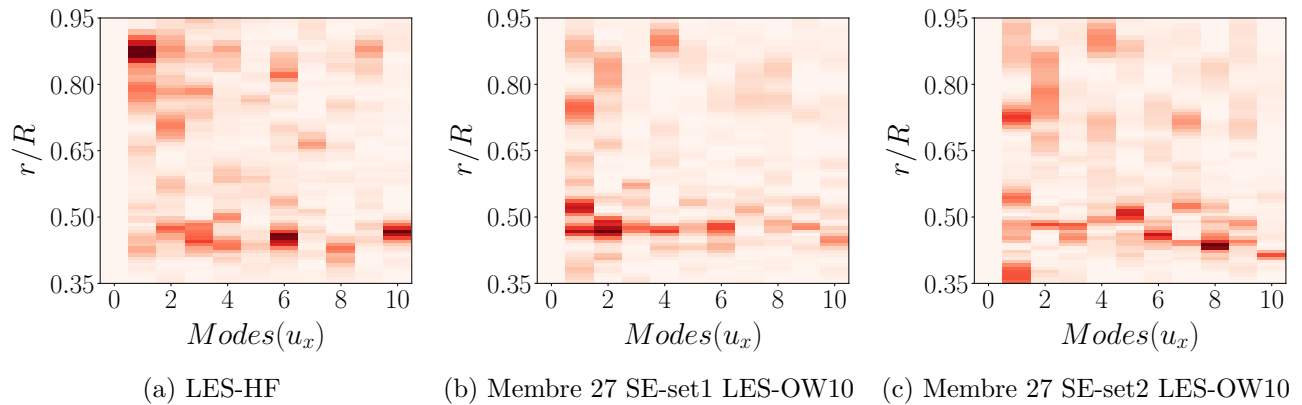


FIGURE C.10 : Cartographie des modes principaux de la transformée de Fourier azimutale de la vitesse axiale $u_x - \overline{u_x}$ au temps $t = 13t_p$. Une transformée est réalisée pour chaque r/R disponible. Seule la région du jet et de la recirculation entre $r/R = 0.35$ et $r/R = 0.95$ est illustrée. Les figures sont colorées en fonction de la puissance spectrale P_s normalisée par la valeur maximale de puissance dans chaque cas.

La légère diminution dans la précision de la synchronisation des champs inférés avec l'observation ainsi que la très grande similarité de la répartition énergétique de l'écoulement pour le plan $x = 0.124$ avec la configuration utilisant le jeu d'observations SE-set1 permettent de conclure que le jeu plus dense SE-set2 n'entraîne pas d'amélioration de la prédiction du HLEnKF. De plus, le temps de calcul pour cette configuration s'est avéré être presque 2 fois plus élevé que pour la configuration utilisant le SE-set1 (environ 139 000 contre 73 000 heures scalaires). Dans ces conditions, la configuration SE-set1 apparaît suffisante et à privilégier au regard de l'économie réalisée sur les ressources de calcul nécessaire à l'inférence d'une période complète.

C.3 Temps de calcul et consommation

Pour la configuration SE-set1, le temps de calcul d'une phase d'assimilation de données est d'environ 22s auxquels s'ajoutent les temps de manipulation et de transfert de l'état complet vers les membres d'ensembles d'environ 30s. Pour la configuration SE-set2, la phase d'assimilation de données dure 150s. Environ 600 000 heures scalaires ont été nécessaires pour le calcul et la mesure des données de référence de la LES-HF. La calibration paramétrique a nécessité environ 440 000 heures scalaires par configuration testée, pour le calcul de 12 périodes de l'onde sinusoïdale de référence. L'étape de synchronisation du champ de vitesse via l'estimation d'état était réalisée uniquement sur une période. Dans le cas LES-OW10 utilisant le jeu SE-set1, le calcul était deux fois plus long que la LES-PE2 sur une période (environ 40 heures sur 1821 cœurs) soit approximativement 73 000 heures scalaires. Dans le cas SE-set2 le temps monte à environ 139 000 heures scalaires.

Bibliographie

- [1] A. H. AFAILAL. “Numerical Simulation of Non-Reactive Aerodynamics in Internal Combustion Engines Using a Hybrid RANS/LES Approach”. Thèse de doct. Université de Pau et des pays de l’Adour, 2021 (cf. p. [20](#), [134](#)).
- [2] A. H. AFAILAL et al. “Development and Validation of a Hybrid Temporal LES Model in the Perspective of Applications to Internal Combustion Engines”. In : *Oil & Gas Science and Technology – Revue d’IFP Energies nouvelles* 74 (2019). Sous la dir. de C. ANGELBERGER, p. 56. ISSN : 1294-4475, 1953-8189. DOI : [10.2516/ogst/2019031](#) (cf. p. [135](#)).
- [3] S. AHMED et al. “A nudged hybrid analysis and modeling approach for realtime wake-vortex transport and decay prediction”. In : *Computers & Fluids* 221 (2021), p. 104895 (cf. p. [110](#)).
- [4] M. ASCH, M. BOCQUET et M. NODET. *Data Assimilation : Methods, Algorithms, and Applications*. Philadelphia, PA : Society for Industrial et Applied Mathematics, 2016. DOI : [10.1137/1.9781611974546](#). (Visité le 13/01/2022) (cf. p. [24](#), [43](#), [46](#), [51](#), [57](#), [59](#), [62](#), [146](#), [183](#)).
- [5] C. BAI et al. “Industry 4.0 technologies assessment : A sustainability perspective”. In : *International Journal of Production Economics* 229 (2020), p. 107776 (cf. p. [23](#)).
- [6] P. BAUWERAERTS et J. MEYERS. “Reconstruction of turbulent flow fields from lidar measurements using large-eddy simulation”. In : *Journal of Fluid Mechanics* 906 (2021), A17 (cf. p. [52](#)).
- [7] M. BEN ALI et al. “An Adjoint Approach for the Analysis of RANS Closure Using Pressure Measurements on a High-Rise Building”. In : (2022) (cf. p. [52](#)).
- [8] M. BEN ALI et al. “Mean wind flow reconstruction of a high-rise building based on variational data assimilation using sparse pressure measurements”. In : *Journal of Wind Engineering and Industrial Aerodynamics* 231 (2022), p. 105204 (cf. p. [32](#), [52](#), [88](#), [100](#)).
- [9] B. BLOCKEN. “Computational Fluid Dynamics for urban physics : Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations”. In : *Building and Environment* 91 (2015), p. 219-245 (cf. p. [22](#)).
- [10] L. BRICTEUX, M. DUPONCHEEL et G. WINCKELMANS. “A Multiscale Subgrid Model for Both Free Vortex Flows and Wall-Bounded Flows”. In : *Physics of Fluids* 21.10 (oct. 2009), p. 105102. ISSN : 1070-6631, 1089-7666. DOI : [10.1063/1.3241991](#) (cf. p. [41](#)).
- [11] S. BRUNTON, B. NOACK et P. KOUMOUTSAKOS. “Machine Learning for Fluid Mechanics”. In : (2019) (cf. p. [24](#)).

- [12] A. CARRASSI et al. “Data assimilation in the geosciences : An overview of methods, issues, and perspectives”. In : *WIREs Climate Change* 9 (2018). DOI : [10.1002/wcc.535](https://doi.org/10.1002/wcc.535) (cf. p. 59).
- [13] A. S. CATO et al. “Comparison of Different Data-Assimilation Approaches to Augment RANS Turbulence Models”. In : *Computers & Fluids* 266 (nov. 2023), p. 106054. ISSN : 00457930. DOI : [10.1016/j.compfluid.2023.106054](https://doi.org/10.1016/j.compfluid.2023.106054) (cf. p. 53).
- [14] P. CHANDRAMOULI, E. MEMIN et D. HEITZ. “4D large scale variational data assimilation of a turbulent flow with a dynamics error model”. In : *Journal of Computational Physics* 412 (2020), p. 109446 (cf. p. 24, 52).
- [15] P. CHASSAING et É. CÉPADUÈS. *Mécanique Des Fluides 3E Édition*. Polytech (Institut National Polytechnique de Toulouse). Cépaduès-éditions, 2010. ISBN : 9782854289299. URL : <https://books.google.fr/books?id=NVfkZwEACAAJ> (cf. p. 29).
- [16] S. CHERROUD et al. *Space-dependent aggregation of data-driven turbulence models*. 2023 (cf. p. 102).
- [17] M. CHETAN, S. YAO et D. GRIFFITH. “Multi-fidelity digital twin structural model for a sub-scale downwind wind turbine rotor blade”. In : *Wind Energy* 24 (2021), p. 1368-1387 (cf. p. 110).
- [18] M. CHEVALIER et al. “State Estimation in Wall-Bounded Flow Systems. Part 2. Turbulent Flows”. In : *Journal of Fluid Mechanics* 552.-1 (mars 2006), p. 167. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/S0022112005008578](https://doi.org/10.1017/S0022112005008578) (cf. p. 66).
- [19] C. H. COLBURN, J. B. CESSNA et T. R. BEWLEY. “State Estimation in Wall-Bounded Flow Systems. Part 3. The Ensemble Kalman Filter”. In : *Journal of Fluid Mechanics* 682 (sept. 2011), p. 289-303. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/jfm.2011.222](https://doi.org/10.1017/jfm.2011.222) (cf. p. 66).
- [20] COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT. *Émissions de gaz à effet de serre du transport*. <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-transports-2023/20-emissions-de-gaz-a-effet-de-serre-du-transport-.php> (cf. p. 19).
- [21] E. CONSTANT et al. “An Immersed Boundary Method in OpenFOAM : verification and validation”. In : *Computers and Fluids* 157.3 (2017), p. 55 -72 (cf. p. 32).
- [22] R. V. CRUZ et E. LAMBALLAIS. “Physical/numerical duality of explicit/implicit subgrid-scale modelling”. In : *Journal of Turbulence* 24.6-7 (2023), p. 235-279 (cf. p. 126).
- [23] A. DA SILVA et T. COLONIUS. “Flow state estimation in the presence of discretization errors”. In : *Journal of Fluid Mechanics* 890 (2020), A10 (cf. p. 68).
- [24] S. B. DALEY. *Atmospheric Data Analysis*. Cambridge University Press, 1991 (cf. p. 43).
- [25] J. C. DEL ÁLAMO et J. JIMÉNEZ. “Spectra of the Very Large Anisotropic Scales in Turbulent Channels”. In : *Physics of Fluids* 15.6 (2003), p. L41 (cf. p. 112).
- [26] P. A. DELLENBACK, D. E. METZGER et G. P. NEITZEL. “Measurements in Turbulent Swirling Flow through an Abrupt Axisymmetric Expansion”. In : *AIAA Journal* 26.6 (juin 1988), p. 669-681. ISSN : 0001-1452, 1533-385X. DOI : [10.2514/3.9952](https://doi.org/10.2514/3.9952) (cf. p. 135).

- [27] Z. DING, K. TRUFFIN et S. JAY. “Cause-and-Effect Chain Analysis of Combustion Cyclic Variability in a Spark-Ignition Engine Using Large-Eddy Simulation, Part I : From Tumble Compression to Flame Initiation”. In : *Combustion and Flame* 267 (sept. 2024), p. 113566. ISSN : 00102180. DOI : [10.1016/j.combustflame.2024.113566](https://doi.org/10.1016/j.combustflame.2024.113566) (cf. p. 20).
- [28] Z. DING, K. TRUFFIN et S. JAY. “Cause-and-Effect Chain Analysis of Combustion Cyclic Variability in a Spark-Ignition Engine Using Large-Eddy Simulation, Part II : Origins of Flow Variations from Intake”. In : *Combustion and Flame* 267 (sept. 2024), p. 113565. ISSN : 00102180. DOI : [10.1016/j.combustflame.2024.113565](https://doi.org/10.1016/j.combustflame.2024.113565) (cf. p. 20).
- [29] Z. DING, K. TRUFFIN et S. JAY. “Cause-and-effect chain analysis of combustion cyclic variability in a spark-ignition engine using large-eddy simulation, Part I : From tumble compression to flame initiation”. In : *Combustion and Flame* (2024), in press . DOI : <https://doi.org/10.1016/j.combustflame.2024.113566> (cf. p. 110, 134).
- [30] Z. DING, K. TRUFFIN et S. JAY. “Cause-and-effect chain analysis of combustion cyclic variability in a spark-ignition engine using large-eddy simulation, Part II : Origins of flow variations from intake”. In : *Combustion and Flame* (2024), in press . DOI : <https://doi.org/10.1016/j.combustflame.2024.113565> (cf. p. 110, 134).
- [31] Z. DING et al. “On the Use of LES and 3D Empirical Mode Decomposition for Analyzing Cycle-to-Cycle Variations of In-Cylinder Tumbling Flow”. In : *Flow, Turbulence and Combustion* (2023), p. 1-50 (cf. p. 110).
- [32] K. DURASAMY, H. XIAO et G. IACCARINO. “Turbulence modeling in the age of data”. In : *Annual Review of Fluid Mechanics* 51 (2019), p. - (cf. p. 24, 38).
- [33] .-. EN. *Eurocode 1 : Actions on structures-Part 1-4 : General actions - Wind actions*. Eur. Comm. Stand., 2005 (cf. p. 89).
- [34] G. EVENSEN. *Data Assimilation : The Ensemble Kalman Filter*. Springer-Verlag/Berlin/Heidelberg, 2009 (cf. p. 25).
- [35] G. EVENSEN. “The ensemble Kalman Filter for combined state and parameter estimation - Monte Carlo techniques for data assimilation in large systems”. In : *IEEE Control Systems* 29 (2009), p. 83-104 (cf. p. 57).
- [36] G. EVENSEN et P. J. VAN LEEUWEN. “An Ensemble Kalman Smoother for Nonlinear Dynamics”. In : *Monthly Weather Review* 128.6 (juin 2000), p. 1852-1867. ISSN : 0027-0644, 1520-0493. DOI : [10.1175/1520-0493\(2000\)128<1852:AEKSFN>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2) (cf. p. 144, 152).
- [37] G. EVENSEN, F. C. VOSSEPOEL et P. J. VAN LEEUWEN. *Data Assimilation Fundamentals : A Unified Formulation of the State and Parameter Estimation Problem*. Springer Textbooks in Earth Sciences, Geography and Environment. Springer International Publishing, 2022. DOI : [10.1007/978-3-030-96709-3](https://doi.org/10.1007/978-3-030-96709-3) (cf. p. 24, 45, 46, 48, 51, 63).
- [38] J. H. FERZIGER, M. PERIĆ et R. L. STREET. *Computational Methods for Fluid Dynamics*. Cham : Springer International Publishing, 2020. ISBN : 978-3-319-99691-2 978-3-319-99693-6. DOI : [10.1007/978-3-319-99693-6](https://doi.org/10.1007/978-3-319-99693-6) (cf. p. 35, 36).
- [39] J. FERZIGER et M. PERIC. *Computational Methods in Fluid Dynamics*. New-York : Springer-Verlag, 1996 (cf. p. 32, 92).

- [40] FORUM INTERNATIONAL DES TRANSPORTS. *Perspectives Des Transports FIT 2023*. Perspectives Des Transports FIT. OECD, avr. 2024. ISBN : 978-92-821-1630-2 978-92-821-1507-7 978-92-821-1823-8. DOI : [10.1787/32bbcca5-fr](https://doi.org/10.1787/32bbcca5-fr) (cf. p. 19).
- [41] D. P. G. FOURES et al. “A data-assimilation method for Reynolds-averaged Navier-Stokes-driven mean flow reconstruction”. In : *Journal of Fluid Mechanics* 759 (2014), p. 404-431 (cf. p. 51).
- [42] L. FRANCESCHINI, D. SIPP et O. MARQUET. “Mean-flow data assimilation based on minimal correction of turbulence models : Application to turbulent high Reynolds number backward-facing step”. In : *Physical Review Fluids* 5 (2020), p. 094603 (cf. p. 52).
- [43] B. GALMICHE et al. “Turbulence characterization of a high-pressure high-temperature fan-stirred combustion vessel using LDV, PIV and TR-PIV measurements”. In : *Experiments in fluids* 55.1 (2014), p. 1636 (cf. p. 154).
- [44] S. GHOSH et al. “A Robust Computational Framework for Variational Data Assimilation of Mean Flows with Sparse Measurements Corrupted by Strong Outliers”. In : *Journal of Computational Physics* 508 (juill. 2024), p. 113008. ISSN : 00219991. DOI : [10.1016/j.jcp.2024.113008](https://doi.org/10.1016/j.jcp.2024.113008) (cf. p. 52).
- [45] C. GORLÉ et G. IACCARINO. “A framework for epistemic uncertainty of turbulent scalar flux models for Reynolds-averaged Navier-Stokes simulations”. In : *Physics of Fluids* 25 (2013), p. 055105 (cf. p. 24).
- [46] L. GRAFTIEAUX, M. MICHARD et N. GROSJEAN. “Combining PIV, POD and Vortex Identification Algorithms for the Study of Unsteady Turbulent Swirling Flows”. In : *Measurement Science and Technology* 12.9 (sept. 2001), p. 1422-1429. ISSN : 0957-0233, 1361-6501. DOI : [10.1088/0957-0233/12/9/307](https://doi.org/10.1088/0957-0233/12/9/307) (cf. p. 135).
- [47] C. GREENSHIELDS et H. WELLER. *Notes on Computational Fluid Dynamics : General Principles*. Reading, UK : CFD Direct Ltd, 2022 (cf. p. 32).
- [48] A. GRONSKIS, D. HEITZ et E. MÉMIN. “Inflow and Initial Conditions for Direct Numerical Simulation Based on Adjoint Data Assimilation”. In : *Journal of Computational Physics* 242 (juin 2013), p. 480-497. ISSN : 00219991. DOI : [10.1016/j.jcp.2013.01.051](https://doi.org/10.1016/j.jcp.2013.01.051) (cf. p. 53, 54).
- [49] F. HAAN. “The effect of turbulence on the aerodynamics of the long-span bridges”. Thèse de doct. Department of Aerospace et Mechanical Engineering, University of Notre Dame, 2000 (cf. p. 105).
- [50] H. HEMIDA et al. “On the Flow over High-rise Building for Wind Energy Harvesting : An Experimental Investigation of Wind Speed and Surface Pressure”. In : *Applied Sciences* 10 (2020), p. 5283 (cf. p. 87, 91).
- [51] J. HÖPFFNER et al. “State Estimation in Wall-Bounded Flow Systems. Part 1. Perturbed Laminar Flows”. In : *Journal of Fluid Mechanics* 534 (juin 2005), p. 263-294. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/S0022112005004210](https://doi.org/10.1017/S0022112005004210) (cf. p. 66).
- [52] I. HOTEIT et al. “Mitigating Observation Perturbation Sampling Errors in the Stochastic EnKF”. In : *Monthly Weather Review* 143.7 (juill. 2015), p. 2918-2936. ISSN : 0027-0644, 1520-0493. DOI : [10.1175/MWR-D-14-00088.1](https://doi.org/10.1175/MWR-D-14-00088.1) (cf. p. 59).

- [53] B. R. HUNT, E. J. KOSTELICH et I. SZUNYOGH. “Efficient Data Assimilation for Spatiotemporal Chaos : A Local Ensemble Transform Kalman Filter”. In : *Physica D : Nonlinear Phenomena* 230.1-2 (juin 2007), p. 112-126. ISSN : 01672789. DOI : [10.1016/j.physd.2006.11.008](https://doi.org/10.1016/j.physd.2006.11.008) (cf. p. 63, 69).
- [54] IEA. *Transport - Energy System*. <https://www.iea.org/energy-system/transport> (cf. p. 19, 20).
- [55] F. JORGENSEN. *How to Measure Turbulence with Hot-wire Anemometers – a Practical Guide*. Rapp. tech. Dantec Dynamics, 2004 (cf. p. 88).
- [56] R. E. KALMAN. “A new approach to linear filtering and prediction problems”. In : *Journal of Basic Engineering* 82 (1960), p. 35-45 (cf. p. 25, 54).
- [57] H. KATO et al. “A data assimilation methodology for reconstructing turbulent flows around aircraft”. In : *Journal of Computational Physics* 283 (2015), p. 559-581 (cf. p. 69).
- [58] H. KATO et S. OBAYASHI. “Approach for Uncertainty of Turbulence Modeling Based on Data Assimilation Technique”. In : *Computers & Fluids* 85 (oct. 2013), p. 2-7. ISSN : 00457930. DOI : [10.1016/j.compfluid.2012.09.002](https://doi.org/10.1016/j.compfluid.2012.09.002) (cf. p. 69).
- [59] M. KATZFUSS, J. R. STROUD et C. K. WIKLE. “Understanding the Ensemble Kalman Filter”. In : *The American Statistician* 70 (2016). DOI : [10.1080/00031305.2016.1141709](https://doi.org/10.1080/00031305.2016.1141709) (cf. p. 93).
- [60] M. KHALIL et al. “Uncertainty quantification in LES of a turbulent bluff-body stabilized flame”. In : *Proceedings of the Combustion Institute* 35.2 (2015), p. 1147-1156 (cf. p. 24).
- [61] K. KOSTADINOVIĆ-VRANEŠEVIĆ et al. “Furthering knowledge on the flow pattern around high-rise buildings : LES investigation of the wind energy potential”. In : *J. Wind Eng. Ind. Aerodyn* 226 (2022), p. 105029 (cf. p. 100).
- [62] J. W. LABAHN et al. “Data assimilation using high-speed measurements and LES to examine local extinction events in turbulent flames”. In : *Proceedings of the Combustion Institute* 37 (2019), p. 2259-2266 (cf. p. 68).
- [63] J. W. LABAHN et al. “Ensemble Kalman Filter for Assimilating Experimental Data into Large-Eddy Simulations of Turbulent Flows”. In : *Flow, Turbulence and Combustion* 104.4 (avr. 2020), p. 861-893. ISSN : 1386-6184, 1573-1987. DOI : [10.1007/s10494-019-00093-1](https://doi.org/10.1007/s10494-019-00093-1) (cf. p. 68).
- [64] E. LAMBALLAIS, R. VICENTE CRUZ et R. PERRIN. “Viscous and hyperviscous filtering for direct and large-eddy simulation”. In : *Journal of Computational Physics* 431 (2021), p. 110115. ISSN : 0021-9991. DOI : <https://doi.org/10.1016/j.jcp.2021.110115>. URL : <https://www.sciencedirect.com/science/article/pii/S0021999121000073> (cf. p. 110).
- [65] G. LAMBERTI et al. “Optimizing Turbulent Inflow Conditions for Large-Eddy Simulations of the Atmospheric Boundary Layer”. In : *Journal of Wind Engineering and Industrial Aerodynamics* 177 (juin 2018), p. 32-44. ISSN : 01676105. DOI : [10.1016/j.jweia.2018.04.004](https://doi.org/10.1016/j.jweia.2018.04.004) (cf. p. 134).
- [66] B. LAUNDER et B. SHARMA. “Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc”. In : *Letters in Heat and Mass Transfer* 1 (1974), p. 131-137 (cf. p. 37, 93).
- [67] Y. LI et al. “Small-Scale Reconstruction in Three-Dimensional Kolmogorov Flows Using Four-Dimensional Variational Data Assimilation”. In : *Journal of Fluid Mechanics* 885 (fév. 2020), A9. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/jfm.2019.960](https://doi.org/10.1017/jfm.2019.960) (cf. p. 52).

- [68] Y. LI et al. “Deep learning based real-time energy extraction system modeling for flapping foil”. In : *Energy* 246 (2022), p. 123390 (cf. p. 110).
- [69] L. MARGHERI et al. “Epistemic uncertainties in RANS model free coefficients”. In : *Computers & Fluids* 102 (2014), p. 315-335 (cf. p. 24, 93, 95).
- [70] L. MARGHERI et al. “Epistemic uncertainties in RANS model free coefficients”. In : *Computers & Fluids* 102 (2014), p. 315-335. DOI : <http://dx.doi.org/10.1016/j.compfluid.2014.06.029> (cf. p. 38).
- [71] A. MARIOTTI et al. “Large-Eddy Simulations of the Accelerating Flow Around a Square Cylinder”. In : *Direct and Large Eddy Simulation XIII*. Sous la dir. de C. MARCHIOLI et al. Cham : Springer Nature Switzerland, 2024, p. 23-28 (cf. p. 110).
- [72] M. MARTÍNEZ et M. MELDI. “A physics-infused Immersed Boundary Method using online sequential Data Assimilation”. In : *submitted, available on ArXiv* 2310.09087 (2023), p. 1-27 (cf. p. 111).
- [73] M. MELDI. “Augmented Prediction of Turbulent Flows via Sequential Estimators : Sensitivity of State Estimation to Density of Time Sampling for Available Observation”. In : *Flow, Turbulence and Combustion* 101 (2018), p. 389-412 (cf. p. 68, 127).
- [74] M. MELDI, D. LUCOR et P. SAGAUT. “Is the Smagorinsky coefficient sensitive to uncertainty in the form of the energy spectrum?” In : *Physics of Fluids* 23 (2011), p. 125109 (cf. p. 24, 117).
- [75] M. MELDI et A. POUX. “A reduced order model based on Kalman Filtering for sequential Data Assimilation of turbulent flows”. In : *Journal of Computational Physics* 347 (2017), p. 207-234 (cf. p. 66, 116).
- [76] M. MELDI, M. V. SALVETTI et P. SAGAUT. “Quantification of errors in large-eddy simulations of a spatially evolving mixing layer using polynomial chaos”. In : *Physics of Fluids* 24 (2012), p. 035101 (cf. p. 24, 32, 117).
- [77] F. R. MENTER, M. KUNTZ et R. LANGTRY. “Ten Years of Industrial Experience with the SST Turbulence Model”. In : *Turbulence, Heat and Mass Transfer* 4 (2003) (cf. p. 37).
- [78] F. R. MENTER, M. KUNTZ et R. LANGTRY. “Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications”. In : *AIAA Journal* 32.8 (1993), p. 1598-1605 (cf. p. 38).
- [79] J. MEYERS et P. SAGAUT. “Is Plane-Channel Flow a Friendly Case for the Testing of Large-Eddy Simulation Subgrid-Scale Models?” In : *Physics of Fluids* 19 (2007), p. 048105 (cf. p. 111).
- [80] P. MOIN et al. “A Dynamic Subgrid-Scale Model for Compressible Turbulence and Scalar Transport”. In : *Physics of Fluids A : Fluid Dynamics* 3.11 (nov. 1991), p. 2746-2757. ISSN : 0899-8213. DOI : [10.1063/1.858164](https://doi.org/10.1063/1.858164) (cf. p. 39).
- [81] P. MOIN et J. KIM. “Numerical Investigation of Turbulent Channel Flow”. In : *Journal of Fluid Mechanics* 118 (1982), p. 341 (cf. p. 40).
- [82] G. MOLDOVAN et al. “A multigrid/ensemble Kalman filter strategy for assimilation of unsteady flows”. In : *Journal of Computational Physics* 443 (2021), p. 110481 (cf. p. 69, 72).
- [83] G. MOLDOVAN et al. “Data-driven augmented LES for the analysis of the BARC test case”. In : *ArXiv* - (2022), p. 1-26 (cf. p. 24, 61, 69, 95).

-
- [84] G. MOLDOVAN et al. “Optimized parametric inference for the inner loop of the Multigrid Ensemble Kalman Filter”. In : *Journal of Computational Physics* 443 (2022), p. 110481 (cf. p. 69, 93).
- [85] R. MOLINARO et al. “Embedding data analytics and CFD into the digital twin concept”. In : *Computers & Fluids* 214 (2021), p. 104759 (cf. p. 110).
- [86] V. MONS, J. C. CHASSAING et P. SAGAUT. “Optimal sensor placement for variational data assimilation of unsteady flows past a rotationally oscillating cylinder”. In : *Journal of Fluid Mechanics* 823 (2021), p. 230-277 (cf. p. 51, 53).
- [87] V. MONS, Y. DU et T. ZAKI. “Ensemble-variational assimilation of statistical data in large-eddy simulation”. In : *Physical Review Fluids* 6 (2021), p. 104607 (cf. p. 24, 53, 61, 93, 110-112, 115, 118, 120, 125, 132).
- [88] V. MONS et O. MARQUET. “Linear and nonlinear sensor placement strategies for mean flow reconstruction via data assimilation”. In : *Journal of Fluid Mechanics* 923 (2021), A1 (cf. p. 51, 53).
- [89] V. MONS, Q. WANG et T. A. ZAKI. “Kriging-enhanced ensemble variational data assimilation for scalar-source identification in turbulent environments”. In : *Journal of Computational Physics* 398 (2019), p. 108856 (cf. p. 53).
- [90] V. MONS et al. “Data assimilation-based reconstruction of urban pollutant release characteristics”. In : *Journal of Wind Engineering & Industrial Aerodynamics* 169 (2017), p. 232-250 (cf. p. 53).
- [91] V. MONS et al. “Reconstruction of unsteady viscous flows using data assimilation schemes”. In : *Journal of Computational Physics* 316 (2016), p. 255-280 (cf. p. 53, 54, 152).
- [92] T. MOUSSIE, P. ERRANTE et M. MELDI. *Statistical Inference of Upstream Turbulence Intensity for the Flow around a Bluff Body with Massive Separation*. Déc. 2023. arXiv : [2312.14526 \[physics\]](https://arxiv.org/abs/2312.14526) (cf. p. 70, 134).
- [93] C. NEUHAUS. *Numerische Frequenzabhaengige Kalibrierung Langer Druckmessschlauchsysteme, Internal Report, Wind Engineering and Flow Mechanics. Ruhr-Universität Bochum, Germany*. 2010 (cf. p. 91).
- [94] E. NICOUD. “Quantifying Combustion Robustness in GDI Engines by Large-Eddy Simulation”. Thèse de doct. Université de Paris-Saclay, 2018 (cf. p. 20, 135).
- [95] F. NICOUD et F. DUCROS. “Subgrid-scale stress modelling based on the square of the velocity gradient tensor”. In : *Flow, turbulence and Combustion* 62.3 (1999), p. 183-200 (cf. p. 39-41).
- [96] F. NICOUD et al. “Using singular values to build a subgrid-scale model for large eddy simulations”. In : *Physics of fluids* 23.8 (2011) (cf. p. 39-41).
- [97] Y. NISHI et P. V. DOAN. “Hybrid Boundary Condition Combined with Data Assimilation for Simulations of Free Surface Flows Using Lattice Boltzmann Method”. In : *Computers & Fluids* 88 (déc. 2013), p. 108-114. ISSN : 00457930. DOI : [10.1016/j.compfluid.2013.08.010](https://doi.org/10.1016/j.compfluid.2013.08.010) (cf. p. 134).
- [98] T. NORTON et al. “Optimizing the ventilation configuration of naturally ventilated livestock buildings for improved indoor environmental homogeneity”. In : *Build Environ.* 45 (2010), 983e95 (cf. p. 92).

- [99] *OpenFOAM - Field Operation And Manipulation*. <https://www.openfoam.com> (cf. p. 32).
- [100] C. PERA et C. ANGELBERGER. “Large Eddy Simulation of a Motored Single-Cylinder Engine Using System Simulation to Define Boundary Conditions : Methodology and Validation”. In : *SAE International Journal of Engines* 4.1 (avr. 2011), p. 948-963. ISSN : 1946-3944. DOI : [10.4271/2011-01-0834](https://doi.org/10.4271/2011-01-0834) (cf. p. 138).
- [101] S. B. POPE. *Turbulent flows*. Cambridge University Press, 2000 (cf. p. 20-22, 35, 36, 109, 111).
- [102] A. POUBEAU et al. *Large-Eddy Simulations of a Speed Transient Performed on a Motored Gasoline Engine*. Rapp. tech. SAE Technical Paper, 2017 (cf. p. 134).
- [103] R. RAMPONI et B. BLOCKEN. “CFD simulation of cross-ventilation for a generic isolated building : impact of computational parameters”. In : *Building and Environment* 53 (2012), p. 34-48 (cf. p. 92).
- [104] A. RASHEED, O. SAN et T. KVAMSDAL. “Digital Twin : Values, Challenges and Enablers From a Modeling Perspective”. In : *IEEE Access* 8 (2020), p. 21980-22012 (cf. p. 23).
- [105] S. S. RAVINDRAN. “A Reduced-Order Approach for Optimal Control of Fluids Using Proper Orthogonal Decomposition”. In : *International Journal for Numerical Methods in Fluids* 34.5 (nov. 2000), p. 425-448. ISSN : 0271-2091, 1097-0363. DOI : [10.1002/1097-0363\(20001115\)34:5<425::AID-FLD67>3.0.CO;2-W](https://doi.org/10.1002/1097-0363(20001115)34:5<425::AID-FLD67>3.0.CO;2-W) (cf. p. 24).
- [106] A. REFLOX et al. “CEDRE Software”. In : *Aerospace Lab* 2 (2011), p. 1-10. URL : <https://hal.archives-ouvertes.fr/hal-01182463> (cf. p. 71).
- [107] S. RENGANATHAN, K. HARADA et D. MAVRIS. “Aerodynamic Data Fusion Toward the Digital Twin Paradigm”. In : *AIAA Journal* 58 (2020), p. 3902-3918 (cf. p. 110).
- [108] A. ROBERT et al. “LES prediction and analysis of knocking combustion in a spark ignition engine”. In : *Proceedings of the Combustion Institute* 35.3 (2015), p. 2941-2948 (cf. p. 134).
- [109] A. ROBERT et al. “Large-eddy simulation analysis of knock in a direct injection spark ignition engine”. In : *International Journal of Engine Research* 20.7 (2019), p. 765-776 (cf. p. 134).
- [110] C. W. ROWLEY. “MODEL REDUCTION FOR FLUIDS, USING BALANCED PROPER ORTHOGONAL DECOMPOSITION”. In : *International Journal of Bifurcation and Chaos* (2005) (cf. p. 24).
- [111] P. RUHNAU et al. “Variational Optical Flow Estimation for Particle Image Velocimetry”. In : *Experiments in Fluids* 38.1 (jan. 2005), p. 21-32. ISSN : 0723-4864, 1432-1114. DOI : [10.1007/s00348-004-0880-5](https://doi.org/10.1007/s00348-004-0880-5) (cf. p. 51).
- [112] P. RUHNAU, A. STAHL et C. SCHNÖRR. “On-Line Variational Estimation of Dynamical Fluid Flows with Physics-Based Spatio-temporal Regularization”. In : *Pattern Recognition*. T. 4174. Berlin, Heidelberg : Springer Berlin Heidelberg, 2006, p. 444-454. ISBN : 978-3-540-44412-1 978-3-540-44414-5. DOI : [10.1007/11861898_45](https://doi.org/10.1007/11861898_45) (cf. p. 51).
- [113] C. SAFTA et al. “Uncertainty quantification in LES of channel flow”. In : *International Journal for Numerical Methods in Fluids* 83.4 (2017), p. 376-401 (cf. p. 24).
- [114] P. SAGAUT. *Large-eddy simulation for incompressible flows. An introduction*. Springer-Verlag, 2005 (cf. p. 39, 109).

-
- [115] P. SAGAUT et C. CAMBON. *Homogeneous Turbulence Dynamics*. Cham : Springer International Publishing, 2018. ISBN : 978-3-319-73161-2 978-3-319-73162-9. DOI : [10.1007/978-3-319-73162-9](https://doi.org/10.1007/978-3-319-73162-9) (cf. p. [35](#)).
- [116] R. D. SANDBERG et Y. ZHAO. “Machine-learning for turbulence and heat-flux model development : A review of challenges associated with distinct physical phenomena and progress to date”. In : *International Journal of Heat and Fluid Flow* 95 (2022), p. 108983 (cf. p. [24](#)).
- [117] M. SCHMELZER, R. DWIGHT et C. P. “Discovery of algebraic Reynolds-stress models using sparse symbolic regression”. In : *Flow, Turbulence and Combustion* 104 (2020), p. 579-603 (cf. p. [24](#)).
- [118] C. SEMERARO et al. “Digital twin paradigm : A systematic literature review”. In : *Computers in Industry* 130 (2021), p. 103469 (cf. p. [23](#)).
- [119] J. SMAGORINSKY. “GENERAL CIRCULATION EXPERIMENTS WITH THE PRIMITIVE EQUATIONS I THE BASIC EXPERIMENT*”. In : *Monthly Weather Review* 91 (1963), p. 99-164 (cf. p. [22](#), [35](#), [39](#), [110](#), [111](#)).
- [120] G. SOLARI. “Thunderstorm Downbursts and Wind Loading of Structures : Progress and Prospect”. In : *Frontiers in Built Environment* 6 (2020), p. 1-24 (cf. p. [110](#)).
- [121] G. SOLARI, M. BURLANDO et M. P. REPETTO. “Detection, simulation, modelling and loading of thunderstorm outflows to design wind-safer and cost-efficient structures”. In : *Journal of Wind Engineering and Industrial Aerodynamics* 200 (2020), p. 104142 (cf. p. [110](#)).
- [122] J. SOUSA et C. GORLÉ. “Computational Urban Flow Predictions with Bayesian Inference : Validation with Field Data”. In : *Building and Environment* 154 (mai 2019), p. 13-22. ISSN : 03601323. DOI : [10.1016/j.buildenv.2019.02.028](https://doi.org/10.1016/j.buildenv.2019.02.028) (cf. p. [69](#), [134](#)).
- [123] C. A. M. STRÖFER et al. “Enforcing Boundary Conditions on Physical Fields in Bayesian Inversion”. In : *Computer Methods in Applied Mechanics and Engineering* 367 (août 2020), p. 113097. ISSN : 00457825. DOI : [10.1016/j.cma.2020.113097](https://doi.org/10.1016/j.cma.2020.113097). arXiv : [1911.06683](https://arxiv.org/abs/1911.06683) [[physics](#), [stat](#)] (cf. p. [69](#)).
- [124] T. SUZUKI. “Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation”. In : *Journal of Fluid Mechanics* 709 (2012), p. 249-288 (cf. p. [66](#)).
- [125] T. SUZUKI et Y. HASEGAWA. “Estimation of Turbulent Channel Flow at Based on the Wall Measurement Using a Simple Sequential Approach”. In : *Journal of Fluid Mechanics* 830 (nov. 2017), p. 760-796. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/jfm.2017.580](https://doi.org/10.1017/jfm.2017.580) (cf. p. [66](#)).
- [126] S. SYMON et al. “Data Assimilation of Mean Velocity from 2D PIV Measurements of Flow over an Idealized Airfoil”. In : *Experiments in Fluids* 58.5 (mai 2017), p. 61. ISSN : 0723-4864, 1432-1114. DOI : [10.1007/s00348-017-2336-8](https://doi.org/10.1007/s00348-017-2336-8) (cf. p. [52](#)).
- [127] T. S. T. POTSI Y. Tominaga. “Computational wind engineering : 30 years of research progress in building structures and environment”. In : *J. Wind Eng. Ind. Aerodyn* 234 (2023), p. 105346 (cf. p. [100](#)).

- [128] P. TANDEO et al. “Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation : a Review”. In : *Monthly Weather Review* 148 (10 2020), 3973–3994. DOI : <https://doi.org/10.1175/MWR-D-19-0240.1> (cf. p. 129, 175).
- [129] L. THOBOIS. “Intérêt et faisabilité de la simulation aux grandes échelles dans les moteurs automobiles.” Thèse de doct. CERFACS, 2006 (cf. p. 134-136).
- [130] L. THOBOIS et al. “Large-Eddy Simulation in IC Engine Geometries”. In : *2004 SAE Fuels & Lubricants Meeting & Exhibition*. Juin 2004, p. 2004-01-1854. DOI : [10.4271/2004-01-1854](https://doi.org/10.4271/2004-01-1854) (cf. p. 135, 138).
- [131] J. THOMAS et al. “A CFD Digital Twin to Understand Miscible Fluid Blending”. In : *AAPS PharmSciTech* 22 (2021), p. 91 (cf. p. 110).
- [132] H. B. TODA, K. TRUFFIN et F. NICLOUD. “Is the dynamic procedure appropriate for all SGS models”. In : *V european conference on computational fluid dynamics, ECCOMAS, Lisbon, Portugal*. 2010, p. 14-17 (cf. p. 39).
- [133] H. B. TODA et al. “A dynamic procedure for advanced subgrid-scale models and wall-bounded flows”. In : *7th International Symposium On Turbulence and Shear Flow Phenomena*. 2011 (cf. p. 39).
- [134] Y. TOMINAGA et al. “AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings”. In : *Journal of Wind Engineering and Industrial Aerodynamics* 96 (2008), 1749–1761 (cf. p. 88, 91, 92).
- [135] C. TORRENCE et G. P. COMPO. “A Practical Guide to Wavelet Analysis”. In : *Bulletin of the American Meteorological Society* 79.1 (jan. 1998), p. 61-78 (cf. p. 123).
- [136] K. TRUFFIN et al. “Using large-eddy simulation and multivariate analysis to understand the sources of combustion cyclic variability in a spark-ignition engine”. In : *Combustion and Flame* 162.12 (2015), p. 4371-4390 (cf. p. 110).
- [137] M. M. VALERO et M. MELDI. *A Physics-Infused Immersed Boundary Method Using Online Sequential Data Assimilation*. Oct. 2023. arXiv : [2310.09087](https://arxiv.org/abs/2310.09087) [physics] (cf. p. 70).
- [138] J. VASSILICOS. “Dissipation in turbulent flows”. In : *Annual Review of Fluid Mechanics* 47 (2015), p. 95-114 (cf. p. 39).
- [139] P. VÉRAS et al. “Reconstruction of Proper Numerical Inlet Boundary Conditions for Draft Tube Flow Simulations Using Machine Learning”. In : *Computers & Fluids* 254 (mars 2023), p. 105792. ISSN : 00457930. DOI : [10.1016/j.compfluid.2023.105792](https://doi.org/10.1016/j.compfluid.2023.105792) (cf. p. 134).
- [140] L. VILLANUEVA, K. TRUFFIN et M. MELDI. “Synchronization and Optimization of Large Eddy Simulation Using an Online Ensemble Kalman Filter”. In : *International Journal of Heat and Fluid Flow* 110 (déc. 2024), p. 109597. ISSN : 0142727X. DOI : [10.1016/j.ijheatfluidflow.2024.109597](https://doi.org/10.1016/j.ijheatfluidflow.2024.109597) (cf. p. 177).
- [141] L. VILLANUEVA et al. “Augmented State Estimation of Urban Settings Using On-the-Fly Sequential Data Assimilation”. In : *Computers & Fluids* (2023), p. 106118. DOI : <https://doi.org/10.1016/j.compfluid.2023.106118> (cf. p. 63, 88).

-
- [142] K. K. VRANEŠEVIĆ et al. “Furthering knowledge on the flow pattern around high-rise buildings : LES investigation of the wind energy potential”. en. In : *Journal of Wind Engineering and Industrial Aerodynamics* 226 (juill. 2022), p. 105029. DOI : [10.1016/j.jweia.2022.105029](https://doi.org/10.1016/j.jweia.2022.105029). (Visité le 08/08/2022) (cf. p. [103](#), [105](#)).
- [143] M. WANG, Q. WANG et T. A. ZAKI. “Discrete Adjoint of Fractional-Step Incompressible Navier-Stokes Solver in Curvilinear Coordinates and Application to Data Assimilation”. In : *Journal of Computational Physics* 396 (nov. 2019), p. 427-450. ISSN : 00219991. DOI : [10.1016/j.jcp.2019.06.065](https://doi.org/10.1016/j.jcp.2019.06.065) (cf. p. [52](#)).
- [144] M. WANG et T. A. ZAKI. “State Estimation in Turbulent Channel Flow from Limited Observations”. In : *Journal of Fluid Mechanics* 917 (juin 2021), A9. ISSN : 0022-1120, 1469-7645. DOI : [10.1017/jfm.2021.268](https://doi.org/10.1017/jfm.2021.268) (cf. p. [52](#)).
- [145] M. WANG et T. ZAKI. “Synchronization of turbulence in channel flow”. In : *Journal of Fluid Mechanics* 943 (2022), A4 (cf. p. [110](#), [127](#)).
- [146] Y. WANG et al. “Accelerating Inverse Inference of Ensemble Kalman Filter via Reduced-Order Model Trained Using Adaptive Sparse Observations”. In : *Journal of Computational Physics* 496 (jan. 2024), p. 112600. ISSN : 00219991. DOI : [10.1016/j.jcp.2023.112600](https://doi.org/10.1016/j.jcp.2023.112600) (cf. p. [68](#)).
- [147] C. K. WIKLE et L. M. BERLINER. “A Bayesian Tutorial for Data Assimilation”. In : *physD* 230.1-2 (2007), p. 1-16. DOI : [10.1016/j.physd.2006.09.017](https://doi.org/10.1016/j.physd.2006.09.017) (cf. p. [46](#)).
- [148] D. C. WILCOX. *Turbulence Modeling for CFD*. 3nd. La Canada CA : DCW Industries Inc., 2006 (cf. p. [22](#), [37](#), [93](#)).
- [149] H. XIAO et P. CINNELLA. “Quantification of Model Uncertainty in RANS Simulations : A Review”. In : *Progress in Aerospace Sciences* 108 (2019), p. 1-31 (cf. p. [24](#), [38](#)).
- [150] H. XIAO et al. “Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations : A data-driven, physics informed Bayesian approach”. In : *Journal of Computational Physics* 324 (2016), p. 115-136 (cf. p. [24](#), [68](#)).
- [151] Y. YANG et al. “Enhanced Ensemble-Based 4DVar Scheme for Data Assimilation”. In : *Computers & Fluids* 115 (juill. 2015), p. 201-210. ISSN : 00457930. DOI : [10.1016/j.compfluid.2015.03.025](https://doi.org/10.1016/j.compfluid.2015.03.025) (cf. p. [53](#)).
- [152] X. ZHANG, L. ZHANG et G. HE. “Parallel Ensemble Kalman Method with Total Variation Regularization for Large-Scale Field Inversion”. In : *Journal of Computational Physics* 509 (juill. 2024), p. 113059. ISSN : 00219991. DOI : [10.1016/j.jcp.2024.113059](https://doi.org/10.1016/j.jcp.2024.113059) (cf. p. [69](#)).
- [153] X. ZHANG et al. “Evaluation of ensemble methods for quantifying uncertainties in steady-state CFD applications with small ensemble sizes”. In : *Computers & Fluids* 203 (2020), p. 104530 (cf. p. [24](#), [68](#)).
- [154] R. ZHAO et al. “Generalizability evaluation of k-epsilon models calibrated by using ensemble Kalman filtering for urban airflow and airborne contaminant dispersion”. In : *Building and Environment* 212 (2022), p. 108823 (cf. p. [69](#), [88](#), [100](#), [101](#), [103](#), [105](#)).

Résumé : Dans ce travail de thèse, des outils d'assimilation de données sont utilisés pour augmenter les performances de solveurs de mécanique des fluides dédiés à la simulation des grandes échelles. L'objectif est d'améliorer la prédiction et l'étude d'évènements marginaux nuisibles à l'intégrité des systèmes physiques. Bien que difficile à caractériser et à modéliser, la compréhension détaillée de ces phénomènes physiques complexes est essentielle pour le développement d'applications plus durables. Ces objectifs s'intègrent aux activités de recherche du projet ANR ALEKCIA dans lequel ce travail de thèse s'inscrit. Plus précisément, il s'agit de répondre au besoin de couplage de calculs de mécanique des fluides numérique avec un algorithme d'assimilation de données séquentiel. L'outil CONES (Coupling OpenFOAM with Numerical EnvironmentS), qui a été développé, permet d'y apporter une réponse en s'appuyant sur le logiciel OpenFOAM et le filtre de Kalman d'ensemble. Ce dernier permet à la fois le calibrage des paramètres physiques de la simulation numérique ainsi que l'inférence de champs physiques tels que le champ de vitesse.

CONES est mis à contribution pour l'inférence de trois cas d'étude à la complexité grandissante. Le premier optimise les coefficients de fermeture de modèles de turbulence de type RANS pour un écoulement incompressible via l'assimilation de données expérimentales. La calibration de ces paramètres entraîne notamment une amélioration topologique des structures de recirculation de la géométrie. Le cas démontre également l'importance de la qualité des informations de sources hétérogènes observées plutôt que leur quantité. Dans une deuxième étude, la simulation des grandes échelles est utilisée pour fournir une prédiction des caractéristiques tridimensionnelles instationnaires d'un écoulement incompressible turbulent en canal. Outre l'optimisation du modèle Smagorinsky, le champ de vitesse est partiellement synchronisé avec les données observées pour favoriser la reconstruction des structures instationnaires. L'influence de certains hyper-paramètres tels que l'inflation est mise en lumière. Enfin, une variante de l'algorithme de Kalman, le filtre de Kalman d'ensemble hyper-localisé, est développée pour le dernier cas d'étude. Cette méthode permet notamment une diminution du coût de calcul. Elle est utilisée pour l'inférence d'une LES de l'écoulement compressible d'une géométrie simplifiée de moteur. La condition d'entrée de référence pulsée est correctement calibrée et le champ de vitesse est localement synchronisé sur les simulations inférées. La correction apportée par l'algorithme montre également une amélioration de la répartition énergétique de la région inférée en adéquation avec la répartition de référence. En conclusion, le potentiel du filtre de Kalman d'ensemble pour la calibration de paramètres physiques et la reconstruction de structures locales grâce à l'observation de données haute-fidélité d'un système réel a été démontré. Ceci permettrait l'étude d'évènements extrêmes pouvant nuire à l'intégrité du système physique grâce à la simulation numérique augmentée de ces phénomènes.

Mots-clés : Assimilation de données, Filtrage de Kalman, Écoulement instationnaire, Turbulence, Simulation des grandes échelles

Development of data assimilation tools for augmented estimation of internal flows

Abstract : In this thesis, data assimilation tools are used to increase the performance of fluid mechanics solvers dedicated to large eddy simulations. The aim is to improve the prediction and study of marginal events harmful to the integrity of physical systems. Although difficult to characterize and model, a detailed understanding of these complex physical phenomena is essential for the development of more sustainable applications. These objectives are in line with the research activities of the ANR ALEKCIA project, of which this thesis is a part. More specifically, the aim is to meet the need to couple numerical fluid mechanics calculations with a sequential data assimilation algorithm. CONES (Coupling OpenFOAM with Numerical EnvironmentS) tool has been developed to provide an answer to this need by using the OpenFOAM software and the ensemble Kalman filter. The latter can be used both to calibrate the physical parameters of the numerical simulation and to infer physical fields, such as the velocity field.

CONES is used to infer three case studies of increasing complexity. The first optimizes the closure coefficients of RANS-type turbulence models for an incompressible flow by assimilating experimental data. The calibration of these parameters leads in particular to a topological improvement in the recirculation structures of the geometry. The case also demonstrates the importance of the quality of the heterogeneous source of information observed rather than its quantity. In a second study, large eddy simulation is used to provide a prediction of the unsteady three-dimensional characteristics of a turbulent incompressible flow in a channel. In addition to optimizing the Smagorinsky model, the velocity field is partially synchronized with the observed data to facilitate the reconstruction of the unsteady structures. The influence of hyperparameters such as inflation is highlighted. Finally, a variant of the Kalman algorithm, the hyper-localized ensemble Kalman filter, is developed for the last case study. In particular, this method reduces the computational cost. It is used to infer a LES of the compressible flow of a simplified engine geometry. The pulsed reference input condition is correctly calibrated and the velocity field of the inferred simulations is locally synchronized. The correction provided by the algorithm also shows an improvement in the energy distribution of the inferred region in line with the reference distribution. In conclusion, the potential of the ensemble Kalman filter for calibrating physical parameters and reconstructing local structures by observing high-fidelity data from the real system has been demonstrated. This could enable the study of extreme events that could damage the integrity of the physical system, thanks to the augmented numerical simulation of these phenomena.

Keywords : Data assimilation, Kalman Filtering, Unsteady flow, Turbulence, Large eddy simulation