



INSTITUT
POLYTECHNIQUE
DE PARIS



NNT : 20XXIPXXXXX

Thèse de doctorat

Méthodes d'assimilation de données pour des simulations lagrangiennes

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°574 École doctorale de mathématiques
Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées

Thèse présentée et soutenue à Ville de soutenance, le Date, par

MARIUS DUVILLARD

Composition du Jury :

Prénom Nom	
Statut, Établissement (Unité de recherche)	Président
Prénom Nom	
Statut, Établissement (Unité de recherche)	Rapporteur
Prénom Nom	
Statut, Établissement (Unité de recherche)	Rapporteur
Prénom Nom	
Statut, Établissement (Unité de recherche)	Examinateur
Olivier Le Maître	
Directeur de Recherche, CMAP-CNRS	Directeur de thèse
Prénom Nom	
Statut, Établissement (Unité de recherche)	Co-directeur de thèse
Prénom Nom	
Statut, Établissement (Unité de recherche)	Invité
Prénom Nom	
Statut, Établissement (Unité de recherche)	Invité

Contents

1	Introduction	11
1.1	Contexte Industriel	11
1.2	Objectifs de la thèse	21
2	Outils et méthodes	23
2.1	Assimilation de données	24
2.1.1	Formulation du problème	24
2.1.2	Approches stochastiques	27
2.1.3	Approches variationnelles	31
2.1.4	Lien avec le filtre EnKF	35
2.1.5	Bilan	37
2.2	Méthodes particulaires pour la simulation	38
2.2.1	Méthodes des éléments discrets, <i>Discrete element method</i> (DEM)	38
2.2.2	Méthodes particulaires discréтиsan un milieu continu	39
2.2.3	Hydrodynamique des particules lissées, <i>Smoothed particle hydrodynamics</i> (SPH)	49
2.2.4	Méthode des points matériels, <i>Material Point Method</i> (MPM) . .	50
2.2.5	Méthodes vortex	52
2.2.6	Bilan	55
3	Méthodes lagrangiennes et assimilation de données	57
3.1	Adaptation du filtre de Kalman pour la correction d'intensité	58
3.1.1	Rappel : mise à jour EnKF	58
3.1.2	Cas eulérien	58
3.1.3	Cas lagrangien	60
3.1.4	Première contribution - Adaptation du filtre EnKF	63
3.2	Alignement pour la correction de la discréétisation particulaire	64
3.2.1	Assimilation de données par déplacement ou alignement	65
3.2.2	Seconde contribution - correction de position et d'intensité	67
3.3	Bilan du chapitre	67
4	Adaptations du filtre EnKF pour des simulations lagrangiennes par correction d'intensité	69
4.1	Résumé	70
4.2	Abstract	70
4.3	Introduction	71
4.4	Background	73
4.4.1	Data assimilation	73
4.4.2	Particle-based methods	76
4.4.3	Particle-based function manipulations	77
4.5	Methods	80

4.5.1	Remesh-EnKF	81
4.5.2	Particles-EnKF	82
4.6	1D density advection-diffusion problem	85
4.6.1	Description of the problem	85
4.6.2	Assimilation parameters and ensemble generation	86
4.6.3	Results	87
4.7	2D vortex-in-cell problem	91
4.7.1	Description of the method	91
4.7.2	Lamb-Chaplygin dipole and simulation parameters	91
4.7.3	Assimilation parameters and ensemble generation	93
4.7.4	Results	95
4.8	Conclusion	100
4.A	Stochastic Ensemble Kalman Filter	101
4.B	Moment conservation of particle discretization	102
4.C	Parameters	103
4.C.1	One dimension problem	103
4.C.2	Two dimension problem	103
5	Méthodes d'assimilation de donnée par correction de position pour des simulations lagrangiennes	105
5.1	Résumé	106
5.2	Abstract	106
5.3	Introduction	106
5.4	Background	110
5.4.1	Model setting	110
5.4.2	Classical ensemble data assimilation method	111
5.4.3	Ensemble Filtering with Displacement Error Adjustment	112
5.5	Methods	114
5.5.1	Position transformation	114
5.5.2	Reduced-order modeling based on ensemble decomposition	115
5.5.3	Gradient computation	116
5.5.4	Selection of the penalization coefficient	117
5.6	Applications	117
5.6.1	Problem description	117
5.6.2	Assimilation of vortex centers	118
5.6.3	Amplitude filters	120
5.6.4	Alignment step	121
5.6.5	Assimilation with position-strength correction	125
5.6.6	Filters performances and parametric studies	127
5.7	Conclusion	129
5.A	Parameters	130
5.B	Gradient computation	131
6	Conclusion générale	133
6.1	Synthèse des résultats	133

Contents 5

6.2 Orientations futures	134
A Conservation des moments particulaires du schéma de remaillage	137
Bibliography	139

List of Figures

1.1	Représentation du cycle du combustible nucléaire. La fermeture du cycle passe par un recyclage en aval du cycle. ©CEA/Com Ci Com Ca	12
1.2	Les principales étapes de la fabrication des pastilles de MOX sont : (a) les matières premières, (b) le mélange des poudres, (c) la mise en forme des pastilles et (d) le frittage.	13
1.3	Image de billes d'acier s'écoulant d'un tas. Trois phases de l'écoulement granulaire, se comportant comme un gaz, un liquide ou un solide, peuvent être identifiées [Forterre and Pouliquen, 2008].	14
1.4	Représentation des différents régimes d'écoulement au sein du tambour en rotation avec la méthode DEM.	15
1.5	Catégorisation des simulations de l'écoulement granulaire au sein d'un tambour en rotation.	17
1.6	Exemples d'acquisition sur la maquette du broyeur à boulets au laboratoire du SA3E.	18
1.7	Variables et représentation graphique du jumeau numérique du broyeur à boulets.	19
1.8	Principales sources d'incertitudes de modèle à l'origine d'erreur de prédiction.	20
1.9	Schéma d'assimilation de données au cours du temps.	20
2.1	Chaîne de Markov cachée.	27
2.2	Résumé des différentes méthodes d'assimilation en fonction de leurs hypothèses.	37
2.3	Une fonction noyau est associée à chaque particule. Sa valeur en un point \mathbf{x}_q est fonction de sa distance avec son centre \mathbf{x}_p	41
2.4	Exemples de distribution particulaire non admissible.	44
2.5	Schéma de remaillage d'une distribution particulaire. Après l'étape d'affectation sur la grille des particules sont régulièrement générées pour interpoler les valeurs de champ 2.5b.	47
3.1	Illustration d'un problème où la définition de l'état uniquement par les positions des particules produit en estimateur loin de tous les membres.	63
3.2	Visualisation de l'effet de double pénalisation. Un décalage de la solution entraîne une mesure de l'erreur plus élevée que celle associée à la fonction nulle.	66
4.1	Particle support before and after the assimilation correction.	81
4.2	4.2a Each particles set is project on a fixed Eulerian grid. 4.2b EnKF matrix operations are apply to nodal values. 4.2c New regular sets of particles are generate to interpolate analyzed solutions.	83
4.3	One dimensional illustration. 4.3a A member forecast function is discretized by a set of three particles. 4.2b The analyzed solution is determined. 4.3c The forecast particle intensities are updated to fit the analyzed solution.	84

4.4	The analytical solution of the convection-diffusion problem evolves over time, with the final snapshot revealing a complete spatial period.	85
4.5	On the left the initial parameters sample, v in abscissa and D in ordinate. On the right is the initial ensemble state.	88
4.6	Data assimilation over assimilation step for the Remesh-EnKF filter. . .	88
4.7	State error with respect to assimilation time step.	89
4.8	Left: Error with respect to particle support size, Middle: Final step for a support of 100 particles, Right: Final step for a support of 60 particles.	89
4.9	The Lamb-Chaplygin dipole vorticity field on a normalized space.	92
4.10	Trajectory of the ground truth. The vorticity is represented on a regularly spaced grid. For $t = [1, 5, 10]s$	93
4.11	Effect of the parameter ε_ω on the particle discretization of the solution for one member. From left to right, results for $\varepsilon_\omega = 0.1, 0.01$, and $1.e^{-6}$.	93
4.12	Six samples from the initial ensemble.	94
4.13	Observed and reference velocity fields. The error on each component is a sample from a centered normal distribution with the nominal value $\sigma_{\text{obs}} = 0.05$	95
4.14	error curves through assimilation steps. Left: L_2 -error of the field, Right: Error for the viscosity parameter. With Part-EnKF in blue and Remesh-EnKF in red.	95
4.15	Box plots of the state error w.r.t. $1/\sigma_{\text{obs}}^2$	96
4.16	Box plots of the state error w.r.t. N_{assim}	97
4.18	Assimilation of one member with a forecast discretization unadapted to the analyses solution. The forecast discretization used by the Part-EnKF does not always support approximation for the analyses and introduces discretization errors.	98
4.19	Error box plots for simulation parameters. The effect of ε_ω on the error is particularly observed on high value. The Part-EnKF error is strongly linked to d_p through the particle approximation error.	98
5.1	Although the union of the particle supports of all ensemble members overlaps with the true solution, individual ensemble members may have disjoint supports that do not fully cover the solution.	108
5.2	Illustration of the double penalty effect: A shift in the solution results in an error measurement that exceeds the error associated with the null function.	109
5.3	Alignment fields and particle distributions with and without regularization in the simulation. Without regularization, the field is noisier, leading to distortions in particle distribution.	116
5.4	Initial conditions and trajectories of the reference three-vortex problem. The figure illustrates the influence of the vortex interactions and boundary conditions on the movement of the vortex centers.	118
5.5	Evolution of the dispersion of the ensemble without assimilation. The figures illustrate the dispersion at different time points: $t = 0$, $t = 30$, and $t = 50$	119

5.6	Position error e_c over time for the amplitude filters. The Remesh-EnKF filter demonstrates convergence for the given configuration, whereas the Part-EnKF filter shows divergence as time progresses.	120
5.7	Update of member 2 using Part-EnKF. The particle configuration does not align with the correction, resulting in partial recovery of the continuous assimilation solution.	121
5.8	Contour plots of the vortex ensemble before and after the alignment process at the first assimilation step. The alignment step significantly improves the match between the dispersed ensemble and the reference configuration.	122
5.9	Particle alignment for two different members ($i = 2$ and $i = 8$). The left column shows the particle configurations before alignment (forward particles), while the right column displays the configurations after alignment, including the velocity fields \mathbf{u}_i used for the alignment. The case of particle 2 (top row) shows a situation where all vortices are shifted from the mean trajectory. The bottom row (particle 8) shows a case where the misalignment is orthogonal to the trajectory.	123
5.10	Singular Value Decomposition (SVD) of alignment fields. The first five subfigures illustrate the principal modes of alignment, showing the predominant correction patterns, while the last subfigure depicts the corresponding singular values. The modes reveal various types of corrections applied to the vortex distributions, including phase corrections and stretching.	124
5.11	Analysis of ensemble size impact on alignment and coefficient distribution. (a) Error after alignment versus ensemble size and (b) Distribution of coefficients a for varying ensemble sizes.	125
5.12	Update of member 2 using the Part-Align-EnKF during the initial assimilation. The particle configuration has been adjusted to effectively accommodate the correction.	126
5.13	Position error e_c for different filtering methods over time. (a) Error for the Align-Filter. (b) Error for the Part-Align-EnKF. The plots show how the position error evolves with each assimilation step for both filtering approaches.	126
5.14	Comparison of vortex contour plots for different assimilation methods and steps. The first row corresponds to the Align-Filter, and the second row to the Part-Align-EnKF filter.	127
5.15	Normalized vorticity error (ε_ω) error for varying forecast time steps and ensemble sizes. Alignment-based filters consistently demonstrate superior performance, achieving convergence with fewer ensemble members compared to the Remesh-EnKF filter.	128
5.16	Normalized vorticity error (ε_ω) as a function of different parameters: number of observations and precision, both plotted on a logarithmic scale. Alignment-based filters consistently show improved performance compared to the Remesh-EnKF filter.	129

Introduction

1.1 Contexte Industriel

Dans le domaine de la production d'énergie électrique, l'énergie nucléaire est une source d'énergie qui s'est imposée à de nombreux pays industrialisés. En 2023, l'industrie électronucléaire a représenté 65% de la production totale d'électricité en France, avec un parc composé de 56 réacteurs à eau pressurisée (REP) répartis sur 18 centrales [Réseau de Transport d'Electricité (RTE), 2023]. Dans le monde, la production nucléaire ne fait qu'augmenter. Fin 2022, la capacité totale des 438 réacteurs nucléaires de puissance en exploitation dans 32 pays s'établissait à 393,8 gigawatts électriques (GWe). Si aujourd'hui, le nucléaire représente près de 9,8% de la production mondiale, elle pourrait attendre 14% du bouquet électrique en 2050. D'ici 2035, le nombre de pays qui exploitent des centrales nucléaires pourrait augmenter de quelque 30% [Agence internationale de l'énergie atomique (AIEA), 2023]. Le secteur est également en constante mutation avec le développement de réacteurs de IV^{ème} génération ou des SMR (*Small Modular Reactor*) [Prospective en Énergie de l'Académie des sciences, 2022]. Si ce secteur attire, c'est en particulier car il offre un très bon rapport qualité prix et est faiblement émettrice en gaz à effet de serre. Son facteur d'émission, c'est à dire la quantité d'émissions de gaz à effet de serre par unité d'énergie produite, est estimé à 12g eqCO₂/kWh [Schlömer et al., 2014]. Il serait même encore plus faible en France avec 4g eqCO₂/kWh [EDF R&D, 2022]. Ainsi elle est aussi émettrice que l'éolien ou bien la production photovoltaïque et est 100 à 1000 fois moins émettrice que les centrales à énergie fossiles.

L'urgence climatique pousse à considérer l'énergie nucléaire comme un levier essentiel dans la transition énergétique, offrant une alternative viable aux énergies fossiles. Cependant, cette option soulève une série de préoccupations. Outre les inquiétudes liées à la sécurité des installations et au risque de prolifération nucléaire [ONU - Assemblée générale, 1968], il est crucial d'aborder la question des déchets hautement radioactifs générés tout au long du fonctionnement des réacteurs. Chaque année, la production d'électricité entraîne la création de près de 2 kg de déchets par habitant. Une infime proportion constitue les déchets à vie longue, mais ils représentent la majorité de l'activité radioactive (0.2% des stocks pour 95% de l'activité). En outre, il est impératif de préserver les réserves de combustible nucléaire. Dans cette optique, les avancées technologiques telles que les réacteurs de quatrième génération visent à optimiser l'utilisation des ressources en transmutant l'uranium 238 en plutonium 239. Ainsi, la question du retraitement et de la fermeture du cycle nucléaire reste cruciale (Figure 1.1).

C'est dans cette optique que le combustible MOX (*Mixed Oxide*), un mélange d'oxyde

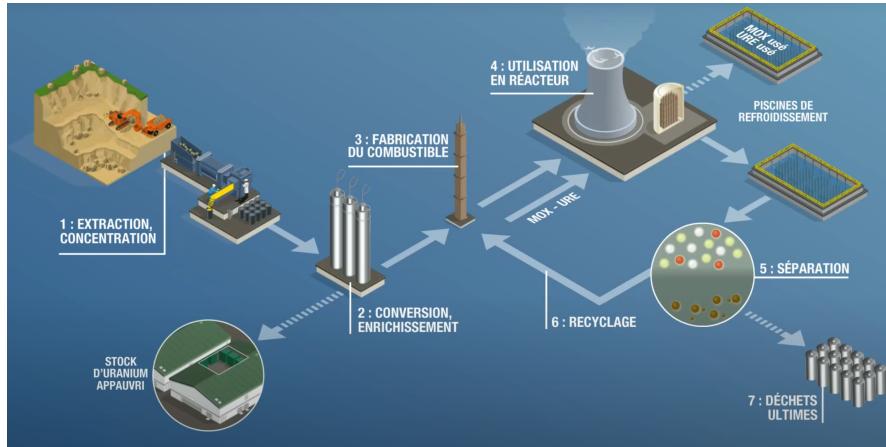


Figure 1.1: Représentation du cycle du combustible nucléaire. La fermeture du cycle passe par un recyclage en aval du cycle. ©CEA/Com Ci Com Ca

de plutonium (PuO_2) et d’oxyde d’uranium (UO_2), a été développé pour recycler une partie des matières nucléaires issues du retraitement des combustibles à uranium naturel enrichi (UNE). La proportion de PuO_2 dans le MOX varie entre 8% et 30%, en fonction des besoins spécifiques du réacteur. Le plutonium utilisé dans le MOX provient du retraitement du combustible irradié, le plutonium étant un sous-produit de la fission de l’uranium 235. Les deux oxydes ont des propriétés différentes : la surface spécifique de l’ UO_2 est de $2 \text{ m}^2/\text{g}$, tandis que celle du PuO_2 atteint $6 \text{ m}^2/\text{g}$. Morphologiquement, les particules d’ UO_2 forment des agglomérats, tandis que le PuO_2 se présente sous forme de plaquettes submicroniques.

Tout comme le combustible à base d’uranium, le combustible MOX est présent dans les réacteurs sous forme de pastille cylindrique, de diamètre et de hauteur d’environ 1 cm, empilées dans des gaines métalliques et constitue un élément de crayon d’environ 4 m de hauteur. Ces éléments sont ensuite réunis dans un assemblage dans une grille de près de 250 éléments. La fabrication de ces pastilles passe par différentes étapes basées sur la métallurgie de poudres, illustrées en Figure 1.2 en particulier une phase de mélange et de broyage qui a lieu dans un broyeur à boulets comme illustré

Ce dispositif cylindrique, rempli de boules de broyage, appelés corps broyants, met en œuvre un processus de rotation pour broyer finement le mélange de poudres d’oxyde. Ce processus dure entre 2 et 4h et qui va permettre de mélanger les deux poudres afin d’avoir un mélange homogène à une granulométrie moyenne de $5\mu\text{m}$. . Cette propriété est déterminante tout au long de la fabrication et particulièrement pour permettre la mise en forme des pastilles [Abe and Asakura, 2012]. D’autre part, cette étape est déterminante pour éviter d’avoir des haute concentrations de Pu, à l’origine de points chauds ou du relâchement de gaz de fission, tous deux facteurs d’accident dans le réacteur [Bouloré et al., 2015; Oudinet et al., 2015].

Cependant, le contrôle de cette étape de mélange et de broyage est déterminé de manière empirique par l’expérimentateur sur une variété de paramètres tels que la vitesse de ro-

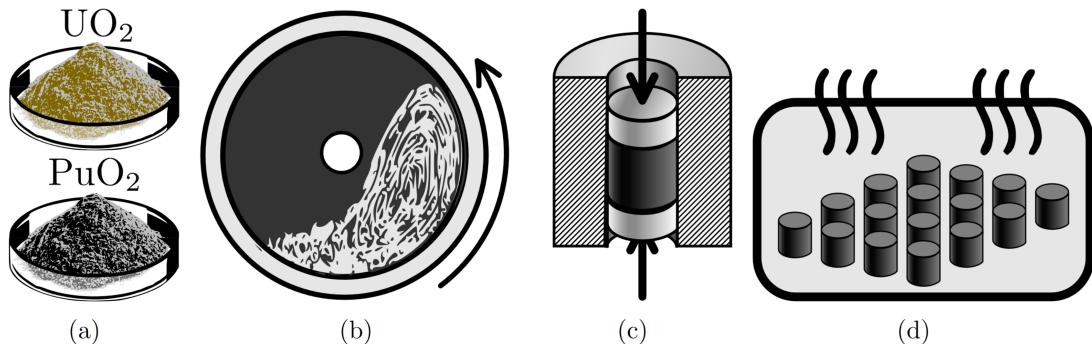


Figure 1.2: Les principales étapes de la fabrication des pastilles de MOX sont : (a) les matières premières, (b) le mélange des poudres, (c) la mise en forme des pastilles et (d) le frittage.

tation, le degré de remplissage, les proportions d'alimentation et de corps broyants. En effet, cette étape critique, assez simple dans son principe, reste encore mal comprise dans ses mécanismes physiques [Austin and Bagga, 1981; Brandao et al., 2020; Mankosa et al., 1986; Datta and Rajamani, 2002; Capece et al., 2014]. C'est dans cette perspective que des outils complémentaires d'analyse sont nécessaires. Ainsi, des études expérimentales ont été mises en place. En particulier, la thèse de Giraud pour pouvoir déterminer des relations empiriques entre les propriétés microscopiques du combustible MOX et du comportement rhéologique macroscopique du mélange [Giraud, 2020]. Ces études reposent sur des campagnes de mesures complexes nécessitant de la manipulation de poudres irradiées. L'environnement doit en particulier être scellé pour éviter toute contamination. Les équipements sont spécialisés et exclusifs et un blindage est mis en place contre les radiations et la production de chaleur. De plus, les mesures effectuées sont soumises à des incertitudes et sont souvent spécifiques au problème. D'autre part, des méthodes de simulation haute-fidélité permettent de représenter le comportement des poudres. En particulier la méthode des éléments discrets (DEM) qui a été utilisée dans les thèses de Orozco [Orozco, 2019] et Vu [Vu, 2023] pour simuler l'écoulement et les mécanismes de fragmentation au sein d'un broyeur à boulets. Ces études ont permis en particulier de déterminer des grandeurs adimensionnelles pour pouvoir réaliser des changements d'échelle. Cependant il existe encore un écart substantiel entre les connaissances actuelles sur les mécanismes physiques régissant le processus de broyage et le besoin de modélisation prédictive et quantitative en vue des applications industrielles.

C'est dans cette optique que la construction d'un jumeau numérique est souhaitée afin d'améliorer la compréhension du procédé de mélange et de broyage ainsi d'optimiser ce procédé. Nous nous focaliserons dans ce travail uniquement sur la compréhension du mécanisme de mélange.

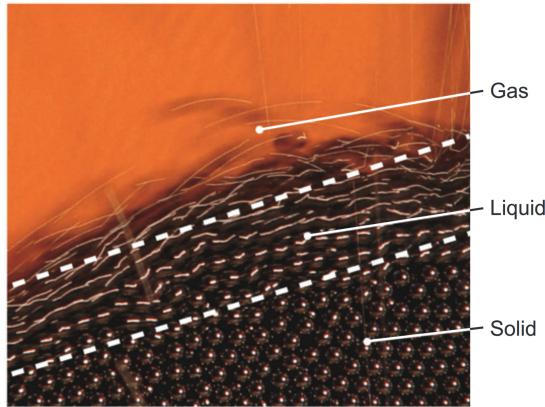


Figure 1.3: Image de billes d’acier s’écoulant d’un tas. Trois phases de l’écoulement granulaire, se comportant comme un gaz, un liquide ou un solide, peuvent être identifiées [Forterre and Pouliquen, 2008].

Écoulement d’un milieu granulaire

La simulation du broyeur repose sur la représentation de l’écoulement d’un milieu granulaire. Si les écoulements granulaires sont très présents dans la nature et dans l’industrie, du fait de la nature discrète du milieu et sa forte dissipation d’énergie, ils sont bien moins compris que l’écoulement d’un liquide qui se base sur les équations de Navier-Stokes. L’écoulement granulaire va se distinguer par trois types de régime que l’on assimile aux trois états de la matière: solide où réarrangement des particules est compact et le comportement est quasi-statique ce qui se traduit par des contacts grain-grain persistants dans le temps, liquide où les grains s’écoulent avec une certaine inertie, et gaz où les particules se déplacent à des vitesses plus élevées de manière chaotique et les contacts deviennent purement collisionnels. Ces trois états peuvent coexister dans le même écoulement comme illustré Figure 1.3, ce qui complexifie sa caractérisation rhéologique.

Le régime solide ou frictionnel intervient en particulier en mécanique des sols pour la prédiction de la défaillance des sols dans les applications de génie civil [Campbell, 2006]. Dans ce cas, le critère de rupture de Mohr-Coulomb [Juvinal and Marshek, 1991], accompagné d’une règle de flux de la plasticité des métaux, est suffisant pour décrire le comportement de l’écoulement granulaire comme un processus continu, sans prendre en compte l’interaction des grains individuels. La loi est alors paramétrée par des quantités interprétables : l’angle de friction interne, la cohésion, ainsi que l’angle de dilatation. La loi de Drucker-Prager [Drucker and Prager, 1952], version lissée du critère de Mohr-Coulomb est également couramment utilisée.

Le régime liquide ou d’écoulement est principalement décrit à l’aide de lois rhéologique. Les travaux récents convergent vers une loi de comportement viscoplastique définie sous le nom de la loi $\mu(I)$ [GDR MiDi, 2004; Jop et al., 2006]. Des simulations du tambour

en rotation [Cortet et al., 2009] ont pu être réalisé et montre une bonne correspondance pour le cas d'écoulement avec surface libre [Chou and Lee, 2009]. Toutefois, cette loi trouve certaines limites dans le cas d'écoulements confinés où le coefficient de tassement change et où le mouvement de chaque grain entraîne des modifications significatives dans les chaînes de force. Si la prédiction est bonne loin des bords, elle reste toutefois insuffisante au niveau des parois [Rognon et al., 2015].

Finalement le régime gaz ou cinétique correspond aux écoulements granulaires dispersés. Ce sont alors des modèles de cinétique des gaz qui sont utilisés pour modéliser le comportement du milieu granulaire [Ng et al., 2008].

Simulation de l'écoulement granulaire au sein d'un tambour en rotation

Dans le tambour tournant, l'ensemble des trois zones d'écoulement sont présentes. En fonction du nombre de Froude, divers régimes d'écoulement apparaissent [Mellmann, 2001]. Entre autres, on retrouve le glissement, l'avalanche, la cascade, le cataracte et la centrifugation illustrés Figure 1.4. Ces différents régimes déterminent la qualité du mélange, du broyage. C'est en particulier le régime en cascade qui est recherché pour la réduction de taille de grain dans le broyeur à boulets. C'est dans ce régime que la surface libre prend la forme caractéristique d'un *S*. Outre la difficulté dans la définition de la loi de comportement décrit précédemment, c'est le choix de la modélisation qui est complexe pour ce type de simulation. En effet, ce problème présente un cas d'écoulement en grande transformation, avec une surface libre et nécessitant de tenir compte des interactions avec une paroi mobile voir des corps broyants.

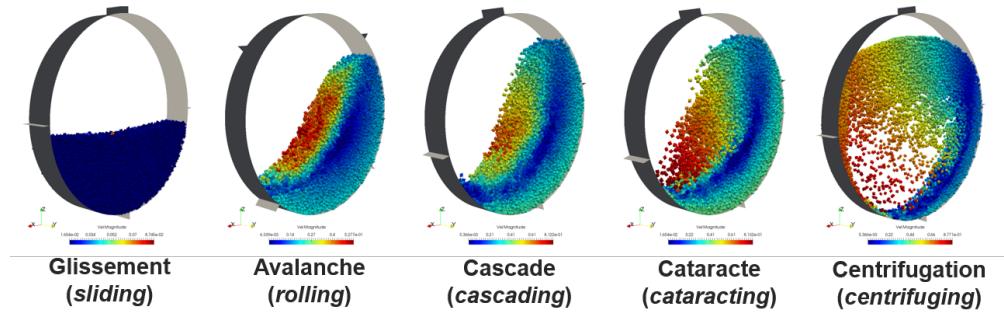


Figure 1.4: Représentation des différents régimes d'écoulement au sein du tambour en rotation avec la méthode DEM.

Ce sont d'abord des lois empiriques ou analytiques qui ont été proposées dans la littérature [Ding et al., 2001; Boateng, 1998; Nicholas and Gray, 2001] mais ceux-ci sont généralement dépendant du problème traité, simplifié et ne permettent donc pas une bonne généralisation. Les méthodes de simulation ce sont à la fois portées sur des représentations continues ou discrète du milieu.

Les méthodes discrètes vont conserver une représentation particulière du milieu en considérant un jeu de particules en équilibre. Introduite en 1979 par Cundall and Strack [Cundall and Strack, 1979], la méthode des éléments discrets (DEM) a été utilisé pour la première fois pour la modélisation du tambour en rotation par Mishra and Rajamani [Mishra and Rajamani, 1992]. Ces méthodes ont l'avantage de pouvoir représenter les différents régimes d'écoulement granulaire, le mélange ou bien les phénomènes de ségrégation. On trouve également des extensions pour prendre en compte la fragmentation [Orozco et al., 2020]. Malgré ces différents avantages, la méthode DEM est très coûteuse en temps de calcul, en particulier lors de la détection des contacts. C'est d'autant plus le cas pour la simulation du tambour qui fait intervenir des grains de l'ordre du *mm* pour décrire un écoulement de l'ordre du mètre. Les grains simulés sont alors agrandis, représentés avec des géométries plus régulières et lisses, pour permettre un temps de calcul acceptable. De plus, la DEM nécessite l'introduction d'un terme dissipatif qu'il est souvent difficile à justifier physiquement.

De l'autre côté, les méthodes continues représentent le milieu granulaire comme un milieu continu. Le système est gouverné par les équations de conservation de masse et de quantité de mouvement. Ce type d'approche est plus à même de représenter des écoulements à grande échelle. Dans cette famille de méthode on retrouve des méthodes qui utilise un maillage pour discréteriser les différents champs approchés. En particulier, on retrouve des méthodes utilisés en dynamique des fluides comme la méthode des volumes finis [Santos et al., 2013; Arseni et al., 2020] ou bien en mécanique des solide avec des extensions de la méthodes des éléments finis comme les éléments finis eulérien [Zheng and Yu, 2015]. Si ces méthodes sont relativement plus efficaces en terme de calcul que la méthode DEM, elles ne sont pas bien adaptées aux problèmes avec des interfaces matérielles mobiles et des surfaces libres en raison de leur nature eulérienne. Finalement, une autre classe de méthode continue a été plus récemment utilisées pour traiter des problèmes impliquant des grandes transformations, des écoulements à surface libre et des problèmes à géométries complexes. Il s'agit des méthodes lagrangiennes continues. Contrairement au méthode eulérienne à maillage fixe, ces méthodes dites lagrangiennes, utilise un ensemble de *particules* comme support de discréterisation qui vont évoluer en suivant l'écoulement. De cette manière, ces méthodes ne peuvent pas souffrir de distorsion de maillage et elles sont beaucoup moins sensibles à la dissipation lors de la phase d'advection. Elles permettent de prendre en compte un mélange en affectant, à chaque particule, une phase ou un matériau différents et traite le problème de surface libre nativement. Une des plus connues est la méthode SPH (*Smoothed Particle hydrodynamics*), où chaque particule transporte des quantités matérielles ainsi qu'une fonction noyau pour interpoler et discréteriser les champs continus et leurs opérateurs différentiels. Cette méthode a été utilisée pour simuler le tambour en rotation à l'aide d'une loi $\mu(I)$ couplée à une surface de charge de Drucker-Prager [Zhu et al., 2022]. Une autre méthode, particulièrement utilisé en mécanique des solides, est la méthode MPM (*Material Point Method*). Cette dernière est une extension de la méthode PIC *Particle In Cell*. Cette méthode hybride a été largement utilisée pour les écoulement granulaire [Kumar et al., 2017] ainsi que pour la simulation du tambour en rotation [Zuo et al., 2020; Chandra et al., 2021] et a pu être comparé à la fois à la méthode DEM à

l'expérimental pour étudier le mélange. Cependant, contrairement aux méthodes avec maillage, on perd en efficacité de calcul soit par des étapes de recherche de plus proches voisins en SPH, ou par le calcul des transferts grille/particules en MPM.

Finalement, bien que le sens physique ne soit pas le même entre méthodes discrètes et continues, il existe une prédominance des méthodes lagrangiennes dites particulières pour traiter le problème de l'écoulement dans un tambour en rotation comme illustré Figure 1.5.

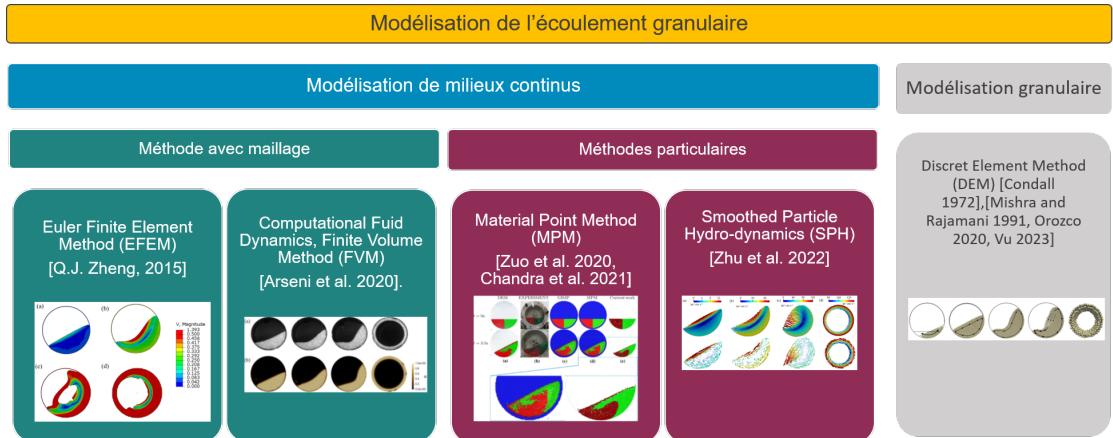


Figure 1.5: Catégorisation des simulations de l'écoulement granulaire au sein d'un tambour en rotation.

Mesures appliquées au tambour en rotation

Outre l'utilisation d'outils de simulation, la validation et la compréhension du procédé se voient renforcées par l'utilisation accrue de méthodes de mesure durant la phase de fonctionnement.

Ces données sont de différents types. D'une part, des données issues de l'imagerie [Jarry et al., 2019; Adepu et al., 2021]. Celles-ci nous permettent de mesurer des mesures de champ de vitesse capturées à travers la face avant d'un hublot transparent grâce à la méthode *Particle Image Velocimetry* (PIV), de post traiter des grandeurs caractéristiques comme l'angle de repos dynamique ou de mesurer des indices de mélange.

D'autre part, des mesures acoustiques du broyeur peuvent révéler des informations sur divers aspects du processus, tels que les conditions de fonctionnement, les anomalies potentielles, ou encore l'usure des éléments du broyeur [Owusu et al., 2021; Almond and Valderrama, 2004].

On trouve également dans la littérature l'usage de mesure vibratoire ou de jauge de contraintes sur la paroi afin de déterminer la position du corps broyant [Davey et al., 2012; Tano et al., 2005], des données issues du moteur [Pedrayes et al., 2017] ou bien

l'instrumentation des boulets [Wang et al., 2021] (ce qui difficilement réalisable dans notre situation où la matière est contaminée).

Le laboratoire expérimental SA3E a instrumenté une maquette de broyeur dont des acquisitions sont représentées sur la Figure 1.6.

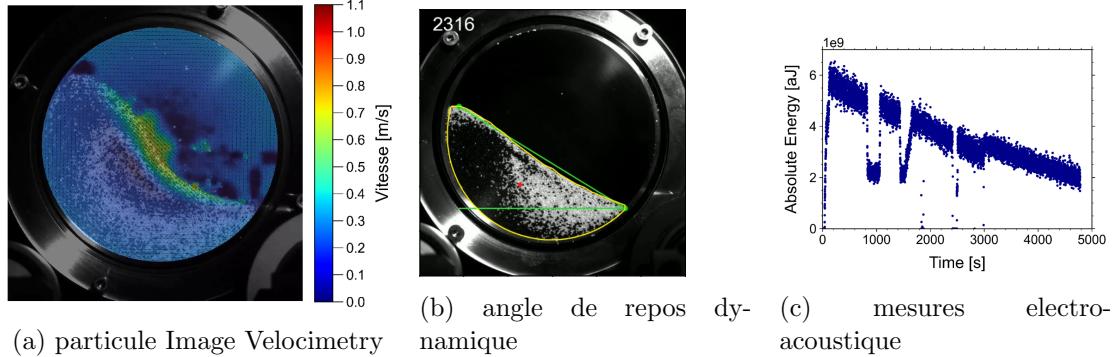


Figure 1.6: Exemples d'acquisition sur la maquette du broyeur à boulets au laboratoire du SA3E.

Une large gamme de mesures sont donc disponibles. Cependant, ces données mesurées ne permettent que d'avoir une représentation partielle et bruitée de l'état réel du broyage et du mélange. De plus, la qualité des résultats analysée reste néanmoins très variable. Outre le bruit de mesure, il peut être difficile d'expliquer la relation entre l'état du système et l'observable. De ce fait, l'interprétabilité de la plupart des mesures se fait au travers de méthode de corrélation et non pas via des approches inductives.

Finalement, les données mesurées que nous possédons ne représentent qu'une vue brute et partielle de l'état réel, tandis que les données simulées de cet état sont sujettes à des erreurs de modélisation. Ainsi, nous souhaitons trouver un moyen d'améliorer notre connaissance des processus en tenant compte à la fois de la simulation et des mesures expérimentales.

Jumeau numérique

La notion de jumeau numérique trouve un essor considérable aujourd'hui à l'air où les données n'ont jamais été aussi présentes. Le potentiel grandissant de la *Big Data*, *l'Internet Of Things*, le calcul haute performance (HPC) mais aussi de l'intelligence artificiel (IA) et en particulier l'apprentissage profond ouvre la voie à de nouveaux outils. Néanmoins, on trouve difficilement une définition univoque de ce paradigme. En effet, le jumeau numérique peut être tour à tour vu comme une réplique haute-fidélité capable d'émuler un système réel, un modèle de substitution rapide à évaluer pour l'optimisation, un outil de visualisation, ou bien un modèle d'intégration de données et d'analyse. Notre définition est inspirée de celle proposée dans [AIAA Digital Engineering Integration Committee, 2020] comme

Definition 1 (Jumeau numérique). Modèle dynamique qui est capable d'être mis à jour

avec les données de son pendant physique au long de son cycle de vie afin d'apporter une aide à la décision.

Le jumeau numérique n'est autre qu'un modèle, et répond au principe d'utilité énoncé par Box [Box, 1979]: *All models are wrong, some are useful.*

Le jumeau numérique pour l'étape de fabrication du combustible peut donc répondre à différents niveaux d'usages comme :

- avoir une meilleure compréhension du procédé ;
- avoir une prédition fidèle de l'état du système ;
- permettre une optimisation des paramètres de contrôle ;
- surveiller l'état du procédé dans une optique de maintenance prédictive ;
- contrôler la qualité du produit.

Finalement, le jumeau numérique peut être formalisé mathématiquement sous la forme d'un modèle graphique probabiliste proposée par Kapteyn et al. [Kapteyn et al., 2021]. Elle permet de relier de bout en bout les flux de données entre capteur, et modèle numérique, mais aussi avec les paramètres de contrôle et de récompense. Le modèle probabiliste permet ainsi d'intégrer dans la construction du jumeau numérique la notion de quantification d'incertitude. La représentation graphique du jumeau numérique du broyeur est présentée en Figure 1.7.

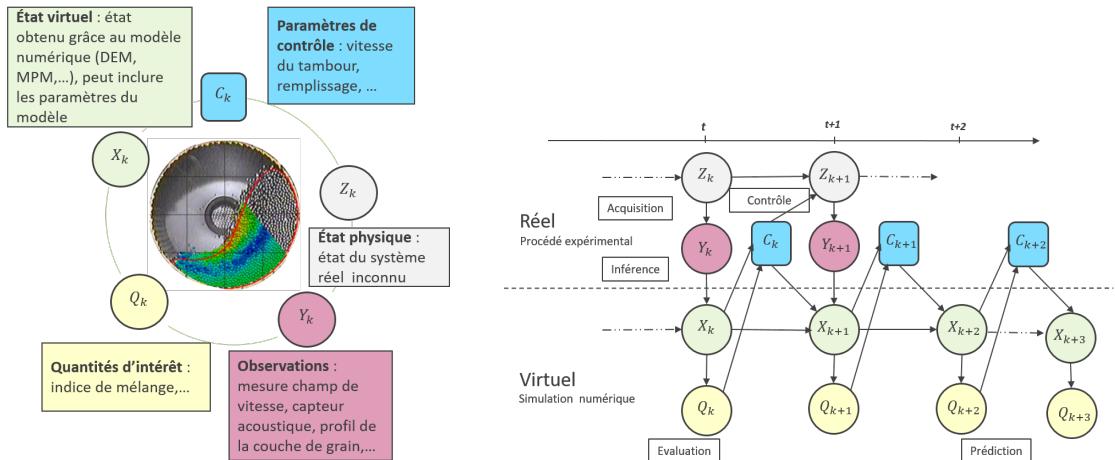


Figure 1.7: Variables et représentation graphique du jumeau numérique du broyeur à boulets.

D'une part, on représente bien l'entité réelle qui n'est connue qu'au travers des observations présentées Section 1.1. Mathématiquement il s'agit d'un modèle de Markov caché. D'autre part, on retrouve le modèle virtuel qui permet d'avoir accès aux variables d'intérêt et de prédire l'évolution de l'état du système à l'aide des modèles physiques comme présenté en Section 1.1 Finalement, on remarque le lien d'inférence qui permet de relier les observations et avec l'état virtuel du système. Fondamentalement, ce sont

les méthodes d'inférence bayésienne et d'assimilation de données qui sont utilisés à cette étape.

Assimilation de données

La construction d'un système expert ou jumeau numérique a besoin de méthodes qui puissent faire le lien entre observations et simulation afin d'améliorer l'estimation de l'état. En effet, les solutions calculées par les modèles numériques comportent des erreurs qu'il est crucial de comprendre, de quantifier et de réduire. Cette incertitude peut se manifester sous diverses formes comme rassemblées Figure 1.8, telles que l'ambiguïté des valeurs des paramètres du modèle, l'imprécision des conditions initiales, ou l'incertitude dans la définition des conditions aux limites ou des forces extérieures. De plus, bien que les modèles numériques intègrent généralement des principes physiques essentiels, ils impliquent certaines simplifications. Finalement l'erreur numérique apparaît en raison de l'algorithme et de la discréétisation. Elle s'étend également à l'incertitude associée aux futures mesures expérimentales calibrant les modèles numériques.

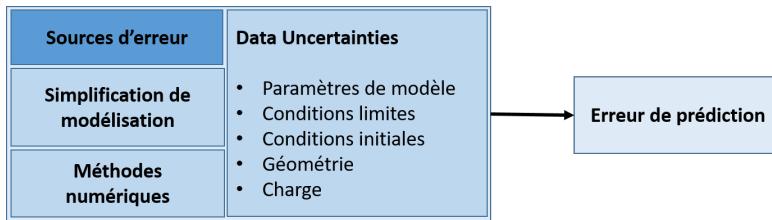


Figure 1.8: Principales sources d'incertitudes de modèle à l'origine d'erreur de prédiction.

Ainsi, les méthodes d'assimilation de données combinent de manière optimale les informations des modèles numériques et des observations en tenant compte de leurs incertitudes respectives. Cette méthode dynamique ajuste la combinaison tout au long du processus, comme le montre la Figure 1.9. On parle parfois de méthode inverse au cours du temps. L'intégration des prévisions des modèles et des données d'observation a

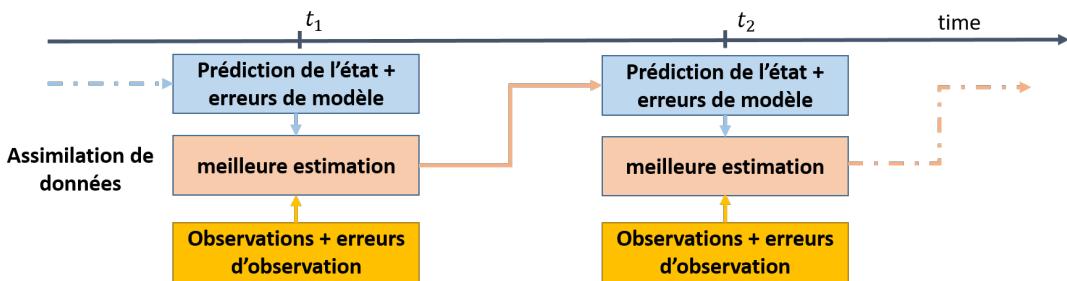


Figure 1.9: Schéma d'assimilation de données au cours du temps.

été largement appliquée dans des disciplines telles que la météorologie, l'océanographie,

et les géosciences [Bocquet, 2014]. Dans le domaine de l'assimilation de données, deux grandes familles d'approches émergent :

- les **méthodes séquentielles stochastiques** cherchent à estimer la distribution à chaque nouvelle acquisition d'une observation à travers le développement de filtre bayésien. On trouve principalement le filtre de Kalman [Kalman, 1960] et ses variantes ainsi que les méthodes de Monte-Carlo séquentielles ;
- les **méthodes variationnelles** cherchent à estimer l'estimateur de maximum a posteriori par la minimisation d'une fonction de coût [Sasaki, 1958]. L'estimation peut être réalisé à chaque nouvelle acquisition comme avec la méthodes 3DVar, ou bien sur une fenêtre d'assimilation avec l'approche 4DVar [Talagrand, 1997].

1.2 Objectifs de la thèse

Les méthodes d'assimilation de données permettent de combiner de manière optimale des données mesurées avec les simulations numériques, permettant ainsi d'obtenir une meilleure estimation des états dynamiques et une quantification des incertitudes. De cette manière, le dialogue entre simulation et expérimentation peut être valorisé pour améliorer l'estimation de l'état du mélange et du broyage lors du procédé de fabrication.

Depuis plusieurs décennies, les méthodes d'assimilation de données sont largement utilisées dans des domaines variés pour mettre à jour des états définis par des champs continus tels que la vitesse, la température ou la pression. Dans ces cas, l'état est généralement représenté sur des maillages ou des grilles fixes. Toutefois, les simulations employées pour modéliser l'écoulement dans un broyeur à boulets reposent sur des méthodes lagrangiennes particulières, où l'état est défini par des particules mobiles suivant l'écoulement du milieu. Malgré quelques tentatives récentes d'application de méthodes d'assimilation à ce type de simulation [Chen et al., 2022; Darakananda et al., 2018], aucune méthode généralisée et adaptée n'a encore été développée.

Cette thèse a pour ambition de combler cette lacune en proposant des méthodes d'assimilation de données adaptées aux simulations lagrangiennes.

Les contributions majeures de cette thèse sont les suivantes :

- Développer des méthodes d'assimilation par correction d'intensité en adaptant le filtre de Kalman d'ensemble ;
- Développer des méthodes d'assimilation par correction successive de position et d'intensité à l'aide d'une approche variationnelle.

Nous commencerons par présenter un état de l'art des méthodes d'assimilation de données et des simulations lagrangiennes dans le Chapitre 2. À la lumière de ces méthodes, le Chapitre 3 est une analyse critique de l'adaptabilité des méthodes d'assimilation à différents types de simulations lagrangiennes. Conscients des spécificités de ce type de simulation, les chapitres suivants détailleront les deux contributions principales de cette thèse. Le Chapitre 4 présente deux nouvelles adaptations du filtre de Kalman

d'ensemble pour les simulations particulières continues, où une mise à jour indépendante de la discréétisation de l'état a été utilisée pour corriger les intensités particulières. Dans un second temps, en complément de la correction d'intensité, le Chapitre 5 introduit une nouvelle méthode d'assimilation de donnée pour corriger la position du support de discréétisation des simulations lagrangiennes. Cette méthode prend en compte une erreur d'alignement pour modifier la position des particules. L'idée est d'introduire une transformation permettant d'aligner les particules dans un cadre variationnel. Chacune de ces filtres a été évalué sur des applications en mécanique des fluides en utilisant la méthode des vortex comme modèle simplifié d'une méthode lagrangienne continue. Enfin, dans le Chapitre 6, nous présentons nos remarques finales et esquissons les perspectives de recherche future, en particulier en ce qui concerne l'adaptation de ces méthodes aux simulations d'écoulements granulaires.

Outils et méthodes

Contents

2.1 Assimilation de données	24
2.1.1 Formulation du problème	24
Définition de l'état	24
Définition des observations	25
Inférence bayésienne récursive	25
Estimation des paramètres du modèle par augmentation de l'état .	27
2.1.2 Approches stochastiques	27
Filtrage bayésien	27
Filtre de Kalman	28
Filtre de Kalman d'Ensemble (EnKF)	29
2.1.3 Approches variationnelles	31
Estimation du maximum a posteriori	32
Méthode 3DVar	33
Méthodes variationnelles d'ensemble	34
2.1.4 Lien avec le filtre EnKF	35
Inflation et localisation	36
2.1.5 Bilan	37
2.2 Méthodes particulières pour la simulation	38
2.2.1 Méthodes des éléments discrets, <i>Discrete element method</i> (DEM) .	38
2.2.2 Méthodes particulières discréétisant un milieu continu	39
Formulation particulière	40
Discrétisation à l'aide de particules	40
Exemples de fonction de noyau	42
Distribution particulière admissible	43
Méthode de remaillage pour obtenir un support de particule commun	45
Approximation d'un champs continu par une discréétisation partic- ulaire	47
2.2.3 Hydrodynamique des particules lissées, <i>Smoothed particle hydrody- namics</i> (SPH)	49
2.2.4 Méthode des points matériels, <i>Material Point Method</i> (MPM) . .	50
2.2.5 Méthodes vortex	52

Méthode vortex classique	52
<i>Vortex-In-Cell</i> (VIC)	54
2.2.6 Bilan	55

2.1 Assimilation de données

Dans cette section, nous présenterons la formulation du problème d’assimilation sous sa forme probabiliste. Nous présenterons ensuite certaines méthodes stochastiques et variationnelles ainsi que leurs extensions d’ensemble que nous utiliserons dans les chapitres suivants. Des explications plus détaillées sont disponibles dans les références suivantes [Law et al., 2015; Asch et al., 2016; Evensen et al., 2022a].

2.1.1 Formulation du problème

Definition 2 (assimilation de données). Méthodes et algorithmes qui permettent d’estimer l’état d’un système, à l’aide d’un **modèle** mathématique décrivant son évolution, des **observations** disponibles, des statistiques décrivant les **erreurs**, et de toute autre information possible sur cet état.

Définition de l’état

Nous définissons un état \mathbf{z}_k comme la variable qui représente complètement la connaissance du système à l’instant $t_k \in \mathbb{R}^+$. La dynamique de l’état du système au cours du temps est obtenue grâce à un modèle \mathcal{M} qui décrit l’évolution du système. Nous noterons $\mathcal{Z}_k = \{\mathbf{z}_0, \dots, \mathbf{z}_k\}$ la trajectoire du modèle jusqu’au pas de temps t_k . Nous supposerons que le modèle admet des incertitudes. Celle-ci sont issues de

- **L’erreur de discréétisation** dans l’espace et le temps. Soit \mathbf{z}^c l’état réel continu. Le modèle numérique ne traite que des représentations discrètes du champ physique. Ainsi, ce n’est non pas l’état \mathbf{z}^c qui est estimé mais une approximation dans l’espace de discréétisation. On estimera $\mathbf{z}^t = \Pi(\mathbf{z}^c)$, où Π est un projecteur sur l’espace de discréétisation. On parle ici d’erreur de représentation.
- **L’erreur de modèle.** Le modèle est une représentation simplifiée et seules certaines physiques sont prises en compte. C’est une erreur qui tient compte de la mauvaise représentation de l’évolution du système, mais également de sa discréétisation.
- **L’erreur de données** Le modèle mathématique doit être complété par des données et des paramètres spécifiant les caractéristiques physiques du système simulé parmi la classe des systèmes représentés par le modèle. Ces données peuvent concerner la géométrie du système, les conditions aux limites et initiales, ainsi que les sollicitations externes. Les paramètres peuvent être des constantes physiques ou des constantes du modèle prescrivant les lois constitutives du système.

Ainsi, nous traiterons l'état comme une variable aléatoire à laquelle est associée une incertitude à la prédiction η_k

$$\mathcal{Z}_k = \mathcal{M}(\mathbf{z}_0, t_k; \boldsymbol{\theta}) + \boldsymbol{\eta}_k.$$

où $\boldsymbol{\theta}$ sont l'ensemble des paramètres du modèle et \mathbf{z}_0 l'état initial.

Définition des observations

A l'équation d'évolution, nous supposons également connu une équation d'observation. Celle-ci relie l'état à l'espace de mesures. On définit $\mathcal{D}_k = \{y_0, \dots, y_k\}$ les mesures prédites par la fenêtre d'état \mathcal{Z}_k . Tout comme l'état, les mesures sont sujettes à des incertitudes issues de plusieurs sources

- **L'erreur de mesure.** L'observable \mathbf{y}^c est issue d'un signal réel fonction de l'état continu \mathbf{z}^c . Or ce signal est mesuré par une capteur sujet à des erreurs instrumentales $\boldsymbol{\varepsilon}^\mu$. C'est une erreur intrinsèque à la méthode d'acquisition et tient compte par exemple d'interférences environnementales, du bruit électronique, ou des biais systématiques des capteurs.
- **L'erreur de représentation.** L'observation est prédite par un opérateur d'observation numérique \mathcal{H} via \mathbf{z}_k . Ainsi, une erreur $\boldsymbol{\varepsilon}^r$ supplémentaire est induite par la représentation de l'opérateur \mathcal{H} et celle de la approximation de l'état continu avec Π . Par exemple, la discréétisation du modèle numérique ne peut pas représenter fidèlement les plus petites échelles spatiales ou temporelles présentes dans le système réel. Cette limitation entraîne une perte d'information et une simplification excessive des dynamiques fines, ce qui peut impacter directement les résultats de prédiction d'observation.

En supposant que ces erreurs sont additives, on définit l'équation d'observation comme

$$\mathcal{D}_k = \mathcal{H}(\mathcal{Z}_k) + \boldsymbol{\varepsilon}_k$$

où $\boldsymbol{\varepsilon}_k = \boldsymbol{\varepsilon}^\mu + \boldsymbol{\varepsilon}^r$ définit l'incertitude sur l'observation \mathcal{D}_k relatif à la prédiction \mathcal{Z}_k .

Inférence bayésienne réursive

Le problème d'assimilation de données peut être formulé comme un problème d'inférence bayésienne réursive. Celle-ci est une méthode statistique pour estimer l'état \mathcal{Z}_k en utilisant à la fois une information a priori, obtenue à partir d'un modèle et des connaissances initiales, et les données observées. Cette méthode repose sur le théorème de Bayes qui décrit la relation entre la distribution a posteriori de l'état étant donné les observations avec $p(\mathcal{Z}_k | \mathcal{D}_k)$, la distribution a priori de l'état $p(\mathcal{Z}_k)$ et la vraisemblance des données $p(\mathcal{D}_k | \mathcal{Z}_k)$ conditionnellement à l'état. Cette formule est la suivante

$$p(\mathcal{Z}_k | \mathcal{D}_k) = \frac{p(\mathcal{D}_k | \mathcal{Z}_k) p(\mathcal{Z}_k)}{p(\mathcal{D}_k)},$$

où $p(\mathcal{D}_k)$ est la distribution marginale des observations. Elle agit comme constante de normalisation afin d'assurer que l'intégrale de la distribution a posteriori soit égale à 1.

Nous souhaitons résoudre le problème d'assimilation de manière séquentielle. C'est-à-dire, mettre à jour l'état à chaque nouvelle observation à l'instant t_k . Pour ce faire, nous utilisons deux approximations :

- Le modèle dynamique est une **chaîne de Markov d'ordre 1**. Cette hypothèse suppose que l'état futur \mathbf{z}_{k+1} est indépendant des états passés \mathcal{Z}_{k-1} conditionnellement à l'état présent \mathbf{z}_k . Le modèle dynamique s'écrit alors

$$\mathbf{z}_k = \mathcal{M}_k(\mathbf{z}_{k-1}; \boldsymbol{\theta}) + \boldsymbol{\eta}_k,$$

ce qui implique que

$$p(\mathbf{z}_{k+1} \mid \mathbf{z}_k, \mathbf{z}_{k-1} \dots \mathbf{z}_0) = p(\mathbf{z}_{k+1} \mid \mathbf{z}_k).$$

Ainsi, la probabilité de l'état $p(\mathcal{Z}_k) = p(\mathbf{z}_0, \dots, \mathbf{z}_k)$ devient

$$\begin{aligned} p(\mathcal{Z}_k) &= p(\mathbf{z}_0)p(\mathbf{z}_1 \mid \mathbf{z}_0)p(\mathbf{z}_2 \mid \mathbf{z}_1) \dots p(\mathbf{z}_k \mid \mathbf{z}_{k-1}) \\ &= p(\mathbf{z}_0) \prod_{\ell=1}^k p(\mathbf{z}_\ell \mid \mathbf{z}_{\ell-1}). \end{aligned}$$

- Les observations sont indépendantes entre chaque assimilation. Cette hypothèse suppose que les observations présentes \mathbf{y}_k soient indépendantes des états et observations passées conditionnellement à \mathbf{z}_k . Ceci correspond à définir une loi d'émission locale

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{z}_k) + \boldsymbol{\varepsilon}_k,$$

ainsi qu'une vraisemblance comme le produit de vraisemblance locale

$$p(\mathcal{D}_k \mid \mathcal{Z}_k) = \prod_{\ell=1}^k p(\mathbf{y}_\ell \mid \mathbf{z}_\ell).$$

Ainsi la trajectoire de l'état et des observations suivent les hypothèses d'un modèle de Markov cachés, ici à temps discret, et qui peut être schématisé par le schéma Figure 2.1.

On reconnaît la partie inférieure du graphique présenté Figure 1.7. C'est cette modélisation qui va être utilisée pour propager et estimer les distributions de l'état en fonction des données.

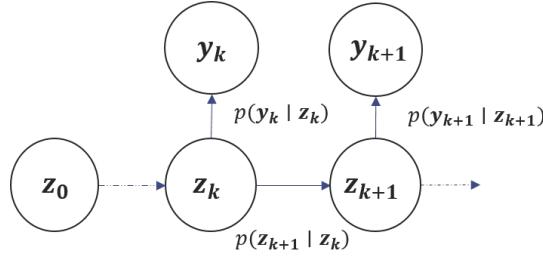


Figure 2.1: Chaîne de Markov cachée.

Estimation des paramètres du modèle par augmentation de l'état

Nous avons supposé que le système était complètement décrit par la variable d'état z que nous souhaitons estimer. Cependant, nous avons aussi supposé que le modèle était imparfait à cause d'erreur sur les données des paramètres de modèle. Ainsi, les paramètres θ ne sont pas connus avec certitude. L'estimation ou calibration de ces paramètres est possible en définissant un état augmenté (z, θ) .

Le modèle d'évolution est toutefois différent car les paramètres du modèle sont supposés constants dans le temps tel que

$$\begin{cases} z_{k+1} = \mathcal{M}(z_k) + \eta_{k+1} \\ y_{k+1} = \mathcal{H}(z_{k+1}) + \varepsilon_{k+1} \\ \theta_{k+1} = \theta_k + \xi_{k+1} \end{cases},$$

L'ajout des paramètres dans la variable d'état a pu être utilisé pour résoudre des problèmes inverses sans calcul de gradient [Iglesias et al., 2013].

2.1.2 Approches stochastiques

Les approches que nous appelons stochastiques sont des méthodes qui vont chercher à approcher la distribution a posteriori de l'état. À partir des éléments vus précédemment, nous présenterons les étapes du filtrage bayésien comme une étape de propagation et d'analyse. Puis, nous présenterons deux adaptations : le filtre de Kalman dans le cas linéaire gaussien, et le filtre de Kalman d'Ensemble dans le cas non-linéaire avec approximation gaussienne.

Filtrage bayésien

Le filtrage bayésien consiste à écrire la récurrence sur les lois de probabilité, pour estimer, en fonction des observations passées et courante \mathcal{D}_k l'état courant z_k et de prédire la distribution de l'état future z_{k+1} .

Puis pour tout $k \geq 0$ les lois de probabilité sont tout d'abord propagées.

L'étape de **propagation** ou *forecast* est obtenue grâce à la loi des probabilités totales

$$p(\mathbf{z}_{k+1} \mid \mathcal{D}_k) = \mathbb{E}_{\mathbf{z}_k} [p(\mathbf{z}_{k+1} \mid \mathbf{z}_k, \mathcal{D}_k)] = \mathbb{E}_{\mathbf{z}_k} [p(\mathbf{z}_{k+1} \mid \mathbf{z}_k)].$$

La loi a priori de la $k + 1$ observation peut être obtenue de nouveau grâce à la loi de probabilité totale

$$p(\mathbf{y}_{k+1} \mid \mathcal{D}_k) = \mathbb{E}_{\mathbf{z}_{k+1}} [p(\mathbf{y}_{k+1} \mid \mathbf{z}_{k+1}, \mathcal{D}_k)] = \mathbb{E}_{\mathbf{z}_{k+1}} [p(\mathbf{y}_{k+1} \mid \mathbf{z}_{k+1})].$$

Après la $k + 1$ observation \mathbf{y}_{k+1} , l'étape d'**analyse**, permet grâce à la loi de Bayes de déterminer la loi a posteriori de l'état

$$\begin{aligned} p(\mathbf{z}_{k+1} \mid \mathcal{D}_{k+1}) &= p(\mathbf{z}_{k+1} \mid \mathbf{y}_{k+1}, \mathcal{D}_k) = \frac{p(\mathbf{y}_{k+1} \mid \mathbf{z}_{k+1}, \mathcal{D}_k)p(\mathbf{z}_{k+1} \mid \mathcal{D}_k)}{p(\mathbf{y}_{k+1} \mid \mathcal{D}_k)}, \\ &= \frac{p(\mathbf{y}_{k+1} \mid \mathbf{z}_{k+1})p(\mathbf{z}_{k+1})}{p(\mathbf{y}_{k+1})}. \end{aligned}$$

Ainsi, les méthodes de filtrage présentent sous diverses formes ces deux étapes de propagation et d'analyse pour mettre à jour la distribution de l'état au cours du temps et après mesure des observations. Dans la suite du manuscrit, on distinguera la variable a priori par la lettre f pour *forecast* et la variable a posteriori par la lettre a pour *analysis*.

Filtre de Kalman

Le filtre de Kalman introduit en 1960 [Kalman, 1960] est une version du filtre bayésien appliquée à un modèle linéaire gaussien

$$\begin{cases} \mathbf{z}_{k+1} = \mathbf{M}(\mathbf{z}_k) + \boldsymbol{\eta}_{k+1}, \\ \mathbf{y}_{k+1} = \mathbf{H}(\mathbf{z}_{k+1}) + \boldsymbol{\varepsilon}_{k+1}. \end{cases}$$

Dans ces conditions, la distribution de l'état a priori de l'état et des observations sont défini par leur deux premiers moments. Soit $(\mathbf{z}^a, \mathbf{P}^a)$ la moyenne et la covariance au pas de temps k , la propagation se résume à

$$\begin{aligned} \mathbf{z}^f &= \mathbb{E}[\mathbf{z}_{k+1} \mid \mathcal{D}_k] = \mathbf{M}\mathbf{z}^a, \\ \mathbf{P}^f &= \mathbb{V}[\mathbf{z}_{k+1} \mid \mathcal{D}_k] = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{Q}, \end{aligned}$$

et le modèle d'observation donne

$$\begin{aligned}\mathbf{y}^f &= \mathbb{E}[\mathbf{y}_{k+1} \mid \mathcal{D}_k] = \mathbf{H}\mathbf{z}^f, \\ \mathbf{C}_{y,y} &= \mathbb{V}[\mathbf{y}_{k+1} \mid \mathcal{D}_k] = \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}, \\ \mathbf{C}_{z,y} &= \text{Cov}[\mathbf{z}_{k+1}, \mathbf{y}_{k+1} \mid \mathcal{D}_k] = \mathbf{P}^f\mathbf{H}^T,\end{aligned}$$

De telle sorte que la distribution a posteriori de la mesure $\mathbf{y}_{k+1} = \mathbf{d}$, si cette dernière est non-dégénérée (ce qui est le cas si \mathbf{R} n'est pas singulière), est gaussienne et définie par ses deux premiers moments qui sont

$$\begin{aligned}\mathbf{z}^a &= \mathbb{E}[\mathbf{z}_{k+1} \mid \mathbf{y}_{k+1}, \mathcal{D}_k] = \mathbf{z}^f + \mathbf{C}_{z,y}\mathbf{C}_{y,y}^{-1}(\mathbf{d} - \mathbf{y}^f), \\ \mathbf{P}^a &= \mathbb{V}[\mathbf{z}_{k+1} \mid \mathbf{y}_{k+1}, \mathcal{D}_k] = \mathbf{P}^f - \mathbf{C}_{z,y}\mathbf{C}_{y,y}^{-1}\mathbf{C}_{z,y}^T.\end{aligned}$$

Ainsi, la distribution a posteriori est définie comme un produit matriciel où l'estimateur a priori \mathbf{z}^f et sa variance \mathbf{P}^f sont mis à jour à partir du **gain de Kalman** $\mathbf{K} = \mathbf{C}_{z,y}\mathbf{C}_{y,y}^{-1}$ et du **terme d'innovation** ($\mathbf{d} - \mathbf{y}^f$) de telle sorte que les précédentes équations s'écrivent

$$\begin{aligned}\mathbf{z}^a &= \mathbf{z}^f + \mathbf{K}(\mathbf{d} - \mathbf{y}^f), \\ \mathbf{P}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f.\end{aligned}$$

Filtre de Kalman d'Ensemble (EnKF)

Le filtre de Kalman est adapté à des modèles linéaires. Il est possible de l'appliquer à des problèmes faiblement linéaire grâce au filtre de Kalman étendu en utilisant l'opérateur tangent pour propager les distributions. Le filtre de Kalman sans parfum peut également être utilisé. Cependant, toutes ces adaptations deviennent inapplicables pour des problèmes de grande dimension ou si le modèle d'évolution est fortement non-linéaire. Dans ce cas, il est préférable de se tourner vers une méthode de Monte-Carlo séquentielle comme le filtre particulaire [Doucet et al., 2000] ou le filtre de Kalman d'ensemble (EnKF). Le filtre EnKF est une méthode d'assimilation de données introduite par Evensen [Evensen, 1994] qui utilise un ensemble de prévisions de même poids pour estimer l'état. Les statistiques de l'état et de l'observation sont estimées pour calculer le gain de Kalman et appliquer la mise à jour de Kalman précédemment défini. Si ce filtre offre de très bons résultats pour des tailles d'ensemble de faible taille (et ainsi ne souffre pas du fléau de la dimension), il reste un estimateur biaisé même lorsque $N \rightarrow \infty$ [Lei and Bickel, 2011].

Le filtre EnKF utilise un ensemble de N membres $\{\mathbf{z}_i^a\}_{i=1}^N$ pour estimer la distribution de l'état

$$p(\mathbf{z}) = \sum_{i=1}^N \frac{1}{N} \delta(\mathbf{z} - \mathbf{z}_i^a)$$

L'étape de **propagation** consiste à appliquer de manière déterministe le modèle d'évolution à chaque membre \mathbf{z}_i . On obtient alors un nouvelle ensemble $\left\{\mathbf{z}_i^f = \mathcal{M}(\mathbf{z}_i^a)\right\}_{i=1}^N$ qui définit notre prior.

Pour l'**analyse**, le filtre EnKF utilise l'ensemble pour approcher le gain de Kalman grâce aux statistiques de l'ensemble. Pour cela, nous définissons la matrice d'état $\mathbf{Z} = [\mathbf{z}_1^f, \dots, \mathbf{z}_N^f]$ et d'anomalies \mathbf{A} dont les colonnes sont les états de chaque membre normalisé et centré ce que l'on peut écrire de la manière suivante

$$\mathbf{A} = \frac{1}{\sqrt{N-1}} (\mathbf{Z}^f - \bar{\mathbf{z}}^f \mathbf{1}^T) \quad \text{avec} \quad \bar{\mathbf{z}}^f = \frac{1}{N} \sum_{j=1}^N \mathbf{z}_j^f.$$

où $\mathbf{1} \in \mathbb{R}^N$ est un vecteur de 1. Respectivement, la matrice d'observation et les anomalies d'observation sont $\mathcal{Y} = [\mathcal{H}(\mathbf{z}_1^f), \dots, \mathcal{H}(\mathbf{z}_N^f)]$ et \mathbf{Y} , où les colonnes sont données par

$$\mathbf{Y} = \frac{1}{\sqrt{N-1}} (\mathcal{Y} - \bar{\mathbf{y}} \mathbf{1}^T) \quad \text{avec} \quad \bar{\mathbf{y}} = \frac{1}{N} \sum_{j=1}^N \mathcal{H}(\mathbf{z}_j^f).$$

Les matrices de covariances d'état et des observations, sont remplacées par leur versions échantillonées

$$\begin{aligned} \bar{\mathbf{C}}_{y,y} &\simeq \frac{1}{N-1} \sum_{i=1}^N \left[\mathcal{H}(\mathbf{z}_i^f) - \bar{\mathbf{y}}^f \right] \left[\mathcal{H}(\mathbf{z}_i^f) - \bar{\mathbf{y}}^f \right]^T + \mathbf{R} = \mathbf{Y} \mathbf{Y}^T + \mathbf{R}, \\ \bar{\mathbf{C}}_{z,y} &\simeq \frac{1}{N-1} \sum_{i=1}^N (\mathbf{z}_i^f - \bar{\mathbf{z}}^f) \left[\mathcal{H}(\mathbf{z}_i^f) - \bar{\mathbf{y}}^f \right]^T = \mathbf{A} \mathbf{Y}^T, \end{aligned}$$

afin d'approcher le gain de Kalman comme

$$\tilde{\mathbf{K}} = \bar{\mathbf{C}}_{z,y} \bar{\mathbf{C}}_{y,y}^{-1} \simeq \mathbf{A} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{R})^{-1}.$$

Cette implémentation sans matrice d'observation repose sur l'approximation par la méthode des sécantes $\mathcal{H}(\mathbf{z}_i^f - \bar{\mathbf{z}}^f) \approx \mathcal{H}(\mathbf{x}_i^f) - \bar{\mathbf{y}}^f$. Ensuite, la prévision est mise à jour vers un ensemble a posteriori $[\mathbf{z}_a^i]_{i=1}^N$ tel que

$$\mathbf{Z}^a = \mathbf{Z}^f + \tilde{\mathbf{K}} (\mathbf{D} - \mathcal{Y}), \tag{2.1}$$

où $[\mathbf{D}]_i = \mathbf{d} + \boldsymbol{\varepsilon}_i$ est l'observation perturbée avec $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, $\tilde{\mathbf{K}}$ est la matrice de gain de Kalman ensembliste et $(\mathbf{D} - \mathcal{Y})$ est le terme d'innovation. L'étape de prévision est ensuite appliquée à l'ensemble analysé jusqu'à l'observation suivante. Sur la base de cette formulation, nous pouvons déduire une formule de correction basée uniquement sur les prédictions des membres et les observations.

Cette version de l'EnKF est parfois appelée EnKF stochastique, car les observations \mathbf{y}_i correspondent aux données mesurées bruitées, i.e. $\mathbf{y}_i = \mathbf{d} + \boldsymbol{\varepsilon}_i$ où $\boldsymbol{\varepsilon}_i$ correspond au bruit de mesure. Ce bruit numérique permet de supprimer un biais statistique sur l'estimation de l'état [Leeuwen, 2020]. On trouve également d'autres implémentations du filtre de Kalman d'ensemble appelé le filtre EnKF déterministe comme le filtre ETKF [Bishop et al., 2001; Hunt et al., 2007]. Dans ce second cas, la solution est supposée appartenir à l'espace généré par les membres.

Cette propriété est aussi vérifiée pour le filtre EnKF stochastique. En effet, on peut réécrire la correction du filtre EnKF en décomposant $\tilde{\mathbf{K}}$ et en remarquant que $\mathbf{1}^T \mathbf{Y}^T = \mathbf{0}$ on a

$$\begin{aligned}\mathbf{Z}^a &= \mathbf{Z}^f + \frac{1}{\sqrt{N-1}} (\mathbf{Z}^f - \bar{\mathbf{z}}^f \mathbf{1}^T) \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{R})^{-1} (\mathbf{D} - \mathcal{Y}), \\ &= \mathbf{Z}^f + \mathbf{Z}^f \frac{1}{\sqrt{N-1}} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{R})^{-1} (\mathbf{D} - \mathcal{Y}), \\ &= \mathbf{Z}^f + \mathbf{Z}^f \mathbf{F}.\end{aligned}$$

Ainsi la correction du filtre EnKF est une combinaison linéaire des membres qui composent les colonnes de \mathbf{Z}^f . De plus, en utilisant la formule de Sherman-Morrison-Woodbury (SMW) [Max, 1950], la matrice de correction \mathbf{F} peut être calculée à l'aide uniquement de l'inversion de la matrice \mathbf{R} et d'une matrice de taille N

$$\mathbf{F} = \frac{1}{\sqrt{N-1}} (\mathbf{I}_N + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{R}^{-1} (\mathbf{D} - \mathcal{Y}), \quad (2.2)$$

ce qui devient avantageux dans le cas où \mathbf{R} est une matrice diagonale et $N < N_{\text{obs}}$. Finalement, la correction \mathbf{F} du filtre EnKF peut être entièrement définie uniquement à l'aide des prédictions d'observation \mathbf{Y} , de la matrice d'erreur d'observation \mathbf{R} et le terme d'innovation $(\mathbf{D} - \mathcal{Y})$. Elle est complètement indépendante de la définition de l'état.

2.1.3 Approches variationnelles

L'objectif des méthodes variationnelles est de déterminer une estimation de l'état ou de sa distribution à partir de la minimisation de fonctions de coût. Pour cela, nous partons de l'estimation du maximum a posteriori.

Estimation du maximum a posteriori

L'estimateur MAP (*Maximum A Posteriori*) est la valeur la plus probable de la distribution a posteriori de l'état connaissant les données mesurées. Il est défini comme

$$\mathbf{z}_{\text{MAP}} = \arg \max_{\mathbf{z}} p(\mathbf{z} \mid \mathbf{y}),$$

Le MAP peut être obtenu par des méthodes d'échantillonnage de distribution. Cependant, lorsque les distributions sont gaussiennes, la distribution a posteriori peut être réécrite comme

$$p(\mathbf{z} \mid \mathbf{y}) \propto \exp(-\mathcal{L}(\mathbf{z})),$$

avec $\mathcal{L}(\mathbf{z})$ défini comme

$$\mathcal{L}(\mathbf{z}) = \frac{1}{2} \left\| \mathbf{z} - \mathbf{z}^f \right\|_{\mathbf{P}}^2 + \frac{1}{2} \left\| \mathcal{H}(\mathbf{z}) - \mathbf{d} \right\|_{\mathbf{R}}^2,$$

où l'on note $\|\mathbf{v}\|_{\mathbf{B}}^2 = \mathbf{v}^T \mathbf{B}^{-1} \mathbf{v}$ avec \mathbf{B} une matrice définie positive. Ceci est équivalent à minimiser l'opposé de la log-vraisemblance et ainsi

$$\mathbf{z}_{\text{MAP}} = \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{z}).$$

Cette définition est à l'origine d'un ensemble de méthodes variationnelles pour l'assimilation de données dont les méthodes 3DVar et 4DVar couramment utilisées en météorologie [Talagrand, 1997].

Dans la méthode 3DVar, \mathbf{z} et $\mathcal{H}(\mathbf{z})$ correspondent à l'état et la prédiction à l'instant k . Dans l'approche 4DVar le vecteur d'état est défini par l'état initial $\mathbf{z} = \mathbf{z}^0$. L'état est propagé sur une fenêtre d'assimilation $0 : k$ donnant ainsi \mathcal{Z}_k , ainsi, l'opérateur d'observation correspond à $\mathcal{D}_k = \mathcal{H}(\mathcal{Z}_k)$.

Le minimum \mathbf{z}^a de cette fonction est obtenue en annulant son gradient qui se trouve être en ce point

$$\nabla_{\mathbf{z}} \mathcal{L}(\mathbf{z}^a) = \mathbf{P}^{-1}(\mathbf{z}^a - \mathbf{z}^f) + \nabla_{\mathbf{z}} \mathcal{H}(\mathbf{z}^a) \mathbf{R}^{-1}(\mathcal{H}(\mathbf{z}^a) - \mathbf{d}) = \mathbf{0}.$$

On peut faire plusieurs remarques :

- L'inverse de la dérivée seconde de la fonction coût \mathcal{L} , Hessienne, est une approximation à l'ordre 1 de la matrice de covariance a posteriori,
- Si l'opérateur d'observation h est non-linéaire alors, le problème n'est pas convexe et une méthode itérative de minimisation est utilisée.

Méthode 3DVar

La méthode 3DVar est un cas particulier de l'équation précédente qui permet de résoudre le problème de minimisation sur un seul instant, sans prendre en compte l'évolution temporelle du système. Elle consiste à supposer fixe et connu les matrices de covariance \mathbf{P} et \mathbf{R} . Dans le cas où la fonction d'observation est linéaire $\mathcal{H}(\mathbf{z}) = \mathbf{Hz}$, la fonction coût devient

$$\mathcal{L}_{3D}(\mathbf{z}) = \frac{1}{2} \|\mathbf{y} - \mathbf{Hz}\|_{\mathbf{R}}^2 + \frac{1}{2} \|\mathbf{z} - \mathbf{z}^f\|_{\mathbf{P}}^2,$$

où \mathbf{R} est la matrice de covariance sur les observations et \mathbf{P} sur l'état. Dans ce cas, l'annulation du gradient de la fonction coût se réduit à l'expression suivante

$$\mathbf{P}^{-1}(\mathbf{z}^a - \mathbf{z}^f) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{Hz}^a - \mathbf{d}) = \mathbf{0},$$

ce qui nous permet d'obtenir une expression de l'estimateur MAP

$$\mathbf{z}^a = \mathbf{z}^f + (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{Hz}^f),$$

qui est l'expression de la mise à jour dans l'espace d'état. Il peut être coûteux d'inverser la matrice $(\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$ si l'espace d'état est de grande dimension. En appliquant deux fois la formule de Sherman-Morrisson-Woodbury, si \mathbf{P}^{-1} et \mathbf{R}^{-1} sont inversibles, la mise à jour peut être réécrite dans l'espace de mesure

$$(\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1}$$

Ainsi la matrice de covariance d'état \mathbf{P} ne nécessite pas d'être inversé. De plus on retrouve le gain de Kalman $\mathbf{K} = \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1}$ précédemment défini. Cette estimateur ainsi obtenu est également le BLUE (*Best Linear Unbiased Estimator*). C'est en effet la combinaison linéaire de \mathbf{z}^f et \mathbf{d} dont l'erreur ε^a est non biaisée ($\mathbb{E}[\varepsilon^a] = \mathbf{0}$), et dont la variance ($\text{Tr}(\mathbf{P}^a)$) est minimale.

Finalement, sachant que la posterior est gaussienne, en prenant l'inverse de dérivée seconde de la fonction coût, la matrice de covariance peut être obtenue

$$(\mathbf{P}^a)^{-1} = \mathbf{P}^{-1} + \mathbf{H} \mathbf{R}^{-1} \mathbf{H}^T.$$

En utilisant de nouveau l'identité de SMW, on retrouve l'expression suivante

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P},$$

qui n'est autre que la mise à jour de la covariance avec le filtre de Kalman. que la mise à jour EnKF était aussi défini sur l'espace

Méthodes variationnelles d'ensemble

Les méthodes variationnelles d'ensemble sont des méthodes hybrides qui combinent les avantages des méthodes variationnelles et d'ensemble. Elles consistent à résoudre différents problèmes variationnels en parallèle. De cette manière il est en particulier possible de définir une matrice d'erreur \mathbf{P} de manière dynamique. D'autre part, il s'agit de pouvoir estimer la distribution a posteriori ainsi que de ne plus avoir besoin de calculer le modèle tangent avec les méthodes 4DEnVar [Hoteit and Köhl, 2006; Liu et al., 2008]. Finalement, il est possible de traiter des problèmes des problèmes très fortement non-gaussien où EnKF ne sera pas performant pour estimer le terme de correction.

Pour cela, nous nous plaçons dans le cas d'un prior d'état gaussien \mathbf{z}^f de matrice de covariance \mathbf{P}^f . On suppose N échantillon i.i.d. de cette distribution $\left\{\mathbf{z}_i^f\right\}_{i=1}^N$. Après introduction de la mesure perturbée $\mathbf{d}_i = \mathbf{d} + \boldsymbol{\varepsilon}_i$ avec $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, la distribution à posteriori peut-être échantillonnée en minimisant un ensemble de fonctions coût définies pour chaque membre $i = 1, \dots, N$

$$\mathcal{L}_i(\mathbf{z}) = \frac{1}{2} \left\| \mathcal{H}(\mathbf{z}_i^f) - \mathbf{d}_i \right\|_{\mathbf{R}}^2 + \frac{1}{2} \left\| \mathbf{z} - \mathbf{z}_i^f \right\|_{\mathbf{P}^f}^2. \quad (2.3)$$

Celles-ci sont indépendantes les unes des autres et offrent une forme similaire au MAP mais autour de la valeur des membres et de la mesure perturbée. Ainsi, les mêmes algorithmes de minimisation peuvent être appliqués pour résoudre ces fonctions coût. La méthode d'échantillonnage *Randomized-Maximum-Likelihood* (RML) introduite par Kitanidis [Kitanidis, 1995] et Olivier et al. [Oliver et al., 1996], permet dans le cas d'un problème faiblement non-linéaire d'approcher les échantillons de la distribution a posteriori en minimisant l'ensemble des fonctions coût définie en (2.3).

Une première hybridation entre 3DVar et EnKF réside dans la définition de la matrice de covariance de l'erreur. Le filtre 3DVar suppose cette matrice connues fixe et de rang plein inversible, ce qui permet en particulier de précalculer la matrice Hessienne dans le cas non-linéaire. D'un autre côté, le filtre EnKF estime une matrice d'erreur de manière dynamique en utilisant la matrice de covariance échantillonnée $\mathbf{P}^f \simeq \mathbf{A}\mathbf{A}^T$. Cette dernière est néanmoins singulière. Hamill et al. [Hamill and Snyder, 2000] propose que l'ensemble des problèmes 3DVar défini en 2.3 soient simplement résolu en prenant une combinaison linéaire de ces deux matrices $\mathbf{P}^* = \gamma \mathbf{P}_{3D} + (1 - \gamma) \mathbf{P}^f$, où $\gamma \in [0, 1]$ afin que \mathbf{P}^* soit inversible. Une autre manière est de régulariser \mathbf{P}^f sur sa diagonale en ajoutant $\varepsilon \mathbf{I}$ où $\varepsilon \in \mathbb{R}^+$. L'avantage de cette première hybridation est de pouvoir améliorer les performances du filtre EnKF lorsque la taille de l'ensemble est faible (grâce à la composante $\gamma \mathbf{P}_{3D}$), et du filtre 3DVar lorsque la grille d'observation n'est pas assez dense.

Cependant, en conservant la matrice de rang faible \mathbf{P}^f il est possible d'exprimer la solution dans l'espace de perturbations. En effet, \mathbf{P}^f étant symétrique, $\text{Span}(\mathbf{P}^f)$ est bijectif avec l'espace généré par les colonnes de \mathbf{A} . Ainsi, comme suggéré par Hunt et al. [Hunt et al., 2007], la perturbation $\mathbf{z} - \mathbf{z}_i^f$ est exprimée comme une combinaison

linéaire dans l'espace de perturbation $\mathbf{z} = \mathbf{z}_i^f + \mathbf{A}\mathbf{w}$. En utilisant le pseudo-inverse indiqué par \dagger , le terme $\|\mathbf{z} - \mathbf{z}_i^f\|_{\mathbf{P}^f}^2$ devient alors

$$\begin{aligned}\|\mathbf{z} - \mathbf{z}_i^f\|_{\mathbf{P}^f}^2 &= \mathbf{w}^T \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^\dagger \mathbf{A}\mathbf{w}, \\ &= \mathbf{w}^T \mathbf{A}^\dagger \mathbf{A}\mathbf{w}, \\ &= (\mathbf{w}^T \mathbf{A}^\dagger \mathbf{A})(\mathbf{A}^\dagger \mathbf{A}\mathbf{w}), \\ &= \tilde{\mathbf{w}}^T \tilde{\mathbf{w}},\end{aligned}$$

où l'on a utilisé les propriétés du pseudo-inverse [Ben-Israel and Greville, 2003] et $\tilde{\mathbf{w}}$ la projection sur l'espace de perturbation des membres. En constatant que $\mathbf{A}\mathbf{w} = \mathbf{A}\tilde{\mathbf{w}}$, sans perte de généralité, la fonction coût peut être reformulée comme

$$\mathcal{J}_i(\mathbf{w}) = \frac{1}{2} \|\mathbf{d} - \mathcal{H}(\mathbf{z}_i^f + \mathbf{A}\mathbf{w})\|_{\mathbf{R}}^2 + \frac{1}{2} \|\mathbf{w}\|_2^2,$$

retrouvant alors une formulation d'optimisation dans l'espace de perturbation des membres. De cette manière, l'assimilation consiste à résoudre N problèmes d'optimisation de dimension N .

Cette fonction non-linéaire peut être minimisée à l'aide d'un algorithme de descente de gradient de type Gauss-Newton. Le gradient des fonction coût devient

$$\nabla_w \mathcal{J}_i(\mathbf{w}) = \mathbf{w} + (\mathbf{G}_i \mathbf{A})^T \mathbf{R}^{-1} (\mathbf{d} - \mathcal{H}(\mathbf{z}_i^f + \mathbf{A}\mathbf{w})),$$

où $\mathbf{G}_i^T = \nabla_w \mathcal{H}(\mathbf{z}_i^f + \mathbf{A}\mathbf{w})$. La matrice Hessienne peut être approchée au premier ordre par

$$\nabla_w^2 \mathcal{J}_i(\mathbf{w}) \approx \mathbf{I} + (\mathbf{G}_i \mathbf{A})^T \mathbf{R}^{-1} (\mathbf{G}_i \mathbf{A})$$

2.1.4 Lien avec le filtre EnKF

Tout comme le filtre Kalman avec la fonction coût du 3DVar, il est possible de retrouver la mise à jour du filtre EnKF. Pour cela, plusieurs approximations sont nécessaires. Tout d'abord, il s'agit de linéariser l'opérateur d'observation comme pour le filtre de Kalman étendu pour approcher le gradient

$$\nabla_w \mathcal{J}_i(\mathbf{w}) \approx \mathbf{w} + (\mathbf{G}_i \mathbf{A})^T \mathbf{R}^{-1} (\mathbf{d} - \mathcal{H}(\mathbf{z}_i^f) + \mathbf{G}_i \mathbf{A}\mathbf{w}).$$

Puis, en utilisant une représentation ensembliste de la sensibilité, on peut approcher $\mathbf{G}_i \approx \mathbf{G} = \mathbf{Y}\mathbf{A}^\dagger$, où chaque sensibilité individuelle est approchée avec une moyenne sur toutes les réalisations. En remarquant que $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^T$ lorsque $rg(\mathbf{A}) = N - 1$, $\mathbf{A}^\dagger \mathbf{A}$ est un projecteur qui retire la moyenne de l'ensemble. Étant déjà le cas pour \mathbf{Y} ; on obtient $\mathbf{G}\mathbf{A} = \mathbf{Y}$. Pour chaque fonction coût \mathcal{J}_i , un minimum global s'écrit alors comme

$$\mathbf{w}_i = (\mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y} + \mathbf{I})^{-1} \mathbf{Y}^T \mathbf{R}^{-1} \left(\mathbf{d}_i - \mathcal{H}(\mathbf{z}_i^f + \mathbf{A}\mathbf{w}) \right).$$

On retrouve la mise à jour du filtre de Kalman d'ensemble pour chaque membre en remarquant que les colonnes de la matrice \mathbf{F} , comme décrite en (2.2), ne sont rien d'autre que les vecteurs \mathbf{w}_i .

Inflation et localisation

Le filtre EnKF et les filtres variationnels de rang faible définissent la solution analysée dans le sous-espace vectoriel généré par l'ensemble des membres. Cependant, l'utilisation d'une approximation de rang faible conduit à des limitations majeures. En approchant les matrices de covariance d'un système de grande dimension avec un nombre de modes bien inférieur, des erreurs d'échantillonnage importantes peuvent apparaître. En particulier, des corrélations fallacieuses peuvent émerger pour des variables spatialement décorrélées, qui devraient être indépendantes. De plus, ces méthodes limitent la définition de l'état à l'espace engendré par les membres. On ne peut ainsi espérer se rapprocher d'un état qui est en dehors de cet espace. Enfin, en utilisant une représentation de rang faible, le filtre EnKF n'est pas capable de prendre en compte des observations qu'il ne peut représenter. Ceci se vérifie dans (2.2) où l'état est mis à jour à partir du terme $\mathbf{Y}^T \mathbf{R}^{-1} (\mathbf{D} - \mathcal{Y})$. Celui-ci n'est autre que la projection dans l'espace des mesures prédictes du terme d'innovation normalisé par la matrice \mathbf{R} . Les risques majeurs engendrés par cette situation sont le *collapse* de l'ensemble ou la divergence du filtre pour des ensembles de taille trop réduite.

Pour y remédier, deux méthodes sont couramment utilisées. La **localisation** consiste à réduire l'impact des corrélations fallacieuses à grande distance [Sakov and Bertino, 2011]. Une première méthode, appelée **localisation de la matrice de covariance**, cherche à augmenter le rang de la matrice de covariance \mathbf{P}^f . Pour cela, la matrice est régularisée en lissant les effets des corrélations à longue distance. Une matrice de corrélation spatiale $\boldsymbol{\rho}$ est appliquée à \mathbf{P}^f à l'aide du produit de Hadamard, de sorte que les termes de la nouvelle matrice de covariance deviennent $[\mathbf{P}^f \circ \boldsymbol{\rho}]_{ij} = [\mathbf{P}^f]_{ij} [\boldsymbol{\rho}]_{ij}$. Cette méthode a pu être aussi bien appliquée dans l'espace d'état [Hamill, Whitaker, et al., 2001] ou approximé dans l'espace d'observation [Houtekamer and Mitchell, 2001].

Une autre approche consiste à faire une **analyse locale** [Haugen and Evensen, 2002]. Cette méthode considère que les observations induisent une correction uniquement sur la partie de l'état qui leur est proche. L'analyse est alors réalisée en parallèle sur des sous-domaines de l'espace. Enfin, la méthode dite de l'**inflation** est mise en place pour contrebalancer la réduction excessive de la variance induite par une représentation de rang faible [Anderson and Anderson, 1999]. Celle-ci consiste à "gonfler" la matrice de covariance par un facteur empirique λ^2 , tel que $\mathbf{P}^a \leftarrow \lambda^2 \mathbf{P}^a$. Cette correction peut être directement appliquée aux membres de l'ensemble, tel que

$$\mathbf{Z}^a \leftarrow \lambda \mathbf{Z}^a.$$

La valeur de λ^2 est généralement comprise entre 1.05 et 1.1.

2.1.5 Bilan

Après un rappel de la formulation du problème d'assimilation, nous avons présenté deux routes principales pour réaliser l'assimilation de données : les approches stochastiques et variationnelles. Ces méthodes se basent sur la loi de Bayes et un modèle de chaîne de Markov cachée pour réaliser une récurrence sur les lois de probabilité et former le filtre bayésien.

Finalement, on trouve des équivalences entre les différentes approches suivant les approximations réalisées comme résumé en Figure 2.2. En particulier, nous avons introduit des méthodes ensemblistes afin d'approcher la distribution de l'état dans des cas non-linéaires et de grandes dimensions. En particulier, le filtre EnKF peut être pensé comme la minimisation d'un ensemble de fonctions coût où l'opérateur d'observation est linéarisé autour de chaque prédiction et son opérateur tangent approché à l'aide de l'ensemble. Dans tous les cas, nous avons en particulier supposé que l'état pouvait être défini sous forme vectorielle et ainsi décrire sa distribution sur chacune de ses composantes. Il est ainsi supposé que l'état est défini sur une discréttisation univoque pour représenter l'état. Il est également supposé que la mise à jour statistique ou de l'optimisation variationnelle ne respecte pas de contraintes. Or, ce n'est pas le cas pour les états de simulation sans maillage basé sur une discréttisation particulière.

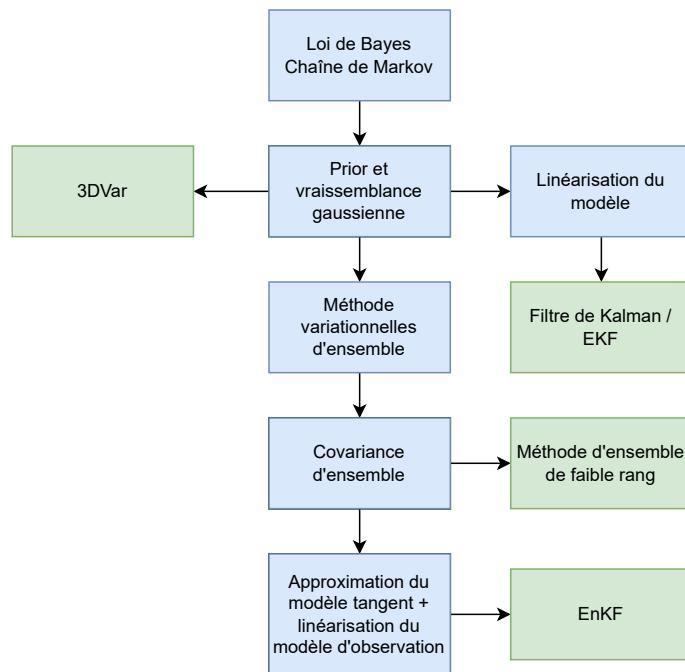


Figure 2.2: Résumé des différentes méthodes d'assimilation en fonction de leurs hypothèses.

2.2 Méthodes particulières pour la simulation

Les méthodes de simulation de l'écoulement au sein du broyeur à boulets reposent principalement sur des discrétisations particulières comme évoqué en Section 1.1.

Ces méthodes lagrangiennes offrent des avantages face à des méthodes reposant sur l'utilisation d'un maillage comme la méthode des éléments finis. En autres, ces méthodes permettent de traiter:

- Des problèmes de grande déformation [Chen et al., 1997; Zhang et al., 2014] ;
- Des géométries complexes obtenues soit par CAD soit à partir de données image [Klaas and Shephard, 2000; Brydon et al., 2005; Bardenhagen, Brydon, et al., 2005] ;
- Des changements de structure topologique du milieu, comme la propagation d'une rupture [Rao and Rahman, 2004; Yang et al., 2014], une bande de cisaillement ou une explosion [Ma et al., 2009; Banerjee, 2004] ;
- Des problèmes à plusieurs phases [Zhang et al., 2008; Wang et al., 2016; Stomakhin et al., 2014] ou à surface libre [Monaghan, 1994] ;
- Des problèmes multi-échelles [Nguyen et al., 2012; Geers et al., 2010] ;

Outre la simulation des milieux granulaires [Kumar et al., 2017; Dunatunga and Kamrin, 2015], ces méthodes sont de plus en plus utilisées aussi bien en mécanique des solides [Sulsky et al., 1994] qu'en mécanique des fluides [Shadloo et al., 2016], en astrophysique [Hultman and Pharasyn, 1999], en biomécanique [Guilkey et al., 2006], en dynamique des sols [Andersen and Andersen, 2010; Solowski and Sloan, 2015], des avalanches [Gaume et al., 2018], la fabrication [Ambati et al., 2012; Russell et al., 2018], en mécanique du contact [Nairn et al., 2018], ou encore en infographie et pour des applications temps réel [Jiang, Schroeder, Teran, et al., 2016; Hu, Liu, et al., 2019].

C'est ce type de modélisation qu'il faut prendre en compte dans nos algorithmes d'assimilation de données. Dans cette section, nous reprenons les principales familles de formulation particulaire en considérant les méthodes discrètes et celles discrétisant un milieu continu. En plus des méthodes particulières pour la modélisation de l'écoulement granulaire, nous présenterons la méthode vortex en Section 2.2.5 comme une méthode modèle pour la suite du manuscrit.

2.2.1 Méthodes des éléments discrets, *Discrete element method (DEM)*

Les méthodes particulières discrètes considèrent les particules comme des objets solides indépendants. Ces entités physiques sont caractérisées par leur géométrie, des propriétés intrinsèques conservées comme la masse ou l'inertie, et des variables cinématiques comme la vitesse ou l'accélération. Les particules interagissent entre elles au travers de lois de

contact, de frottement et de cohésion. L'objectif est de déterminer la trajectoire des particules. Parmi ces méthodes, on peut mentionner la méthode des éléments discrets (DEM) [Radjai and Dubois, 2011]. Sans perte de généralité pour l'applicabilité au problème d'assimilation, nous présenterons la méthode classique DEM.

La DEM [Cundall and Strack, 1979] est une extension du problème à N -corps introduit dans les années 60 aussi bien pour la dynamique moléculaire [Alder and Wainwright, 1959], la dynamique des fluides [Allen and Tildesley, 1987a], ou encore l'astrophysique [Hoerner, 1960]. La méthode consiste à considérer le mouvement d'un ensemble de N grains composant le milieu.

Le principe fondamental de la dynamique s'applique sur cet ensemble, et s'exprime pour chaque particule sous la forme

$$\begin{cases} m_i \frac{d^2 \mathbf{x}_i}{dt^2} = \mathbf{f}_i, \quad i = 1 \dots N \\ I_i \frac{d\boldsymbol{\omega}_i}{dt} = \mathbf{M}_i, \quad i = 1 \dots N \end{cases}$$

où N est le nombre de grains, m_i est la masse, I_i est le moment d'inertie, \mathbf{x}_i est la position, $\boldsymbol{\omega}_i$ est la rotation, \mathbf{f}_i est la force exercée sur le grain considéré et \mathbf{M}_i le moment associé à la force \mathbf{f}_i .

Le couplage entre particules s'exprime au travers de la force \mathbf{f}_i qui se décompose comme

$$\mathbf{f}_i = \sum_{j \neq i} \mathbf{f}_{ij}^c + \mathbf{f}_i^{\text{ext}}$$

où \mathbf{f}_{ij}^c est la force induite par le contact de la particule i avec la particule j , et $\mathbf{f}_i^{\text{ext}}$ la force au centre de la particule i (par exemple la force de gravité).

Les forces de contact sont décrites par des lois de contact entre grains, et entre grains et la paroi à la frontière du domaine. Dans le cas élastique, la modèle de Hertz est adapté pour décrire l'interaction normale. Le modèle Hertz-Mindlin permet de déterminer les interactions tangentielles élastiques. Ces lois sont fonctions de l'interpénétration et de la vitesse relative interparticulaire. Par exemple, dans le cas de particules sphériques, l'interpénétration est définie comme $\delta_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| - (R_i - R_j)$ où \mathbf{x} est la position du centre du grain et R son rayon. De même, on définit des loi de contact pour décrire la composante tangentielle.

La force de contact tend ainsi à pénaliser l'interpénétration via des coefficients de raideur k et d'amortissement γ qui dépendent des propriétés mécaniques des grains de la paroi. Il existe plusieurs algorithmes pour résoudre numériquement les équations de la dynamique. Le plus utilisé est l'algorithme de Verlet en vitesse [Swope et al., 1982].

2.2.2 Méthodes particulières discrétisant un milieu continu

Des méthodes particulières sont également envisagées pour résoudre des problèmes en milieu continu en mécanique des fluides ou des solides. Cela inclut des méthodes telles

que l'hydrodynamique par particules lissées (SPH) [Lucy, 1977; Gingold and Monaghan, 1977] et la méthode des vortex (VM) [Cottet and Koumoutsakos, 2000a], et s'étend à d'autres méthodes comme la méthode des points de matériau (MPM) [Sulsky et al., 1994]. Cette fois, elles partagent toutes de décomposer le domaine en un ensemble \mathcal{P} de particules qui suivent la dynamique du problème. Les particules ne sont donc pas ici des entités physiques distincts. Nous parlerons de grain pour les méthodes discrètes et de particule pour les méthodes continues. Ainsi, la discréétisation suit la transformation appliquée au milieu en transportant des quantités attachées à chaque particule. Elles sont en cela des méthodes lagrangiennes.

Formulation particulaire

Discréétisation à l'aide de particules

La solution représentée par un jeu de « particules » est obtenue grâce à deux éléments: une approximation grâce à un noyau de lissage et l'approximation particulaire d'un opérateur intégral. Si la représentation de la solution peut évoluer suivant la méthode (en particulier avec la méthode MPM), la solution peut toujours être exprimée de cette manière suivant le choix du noyau.

Tout champ régulier \mathbf{u} sur $\Omega \times \mathbb{R}^+$ peut être écrit grâce à la propriété de criblage du Dirac

$$\mathbf{u}(\mathbf{x}, t) = \int_{\Omega} \mathbf{u}(\mathbf{x}', t) \delta(\mathbf{x}' - \mathbf{x}) d\mathbf{x}',$$

avec δ la distribution de Dirac.

Une fonction de noyau ϕ_ε est introduite pour obtenir une estimation moyenne $\langle \mathbf{u} \rangle$ de \mathbf{u} telle que

$$\langle \mathbf{u}(\mathbf{x}, t) \rangle = \int_{\Omega} \mathbf{u}(\mathbf{x}', t) \phi_\varepsilon(\mathbf{x} - \mathbf{x}') d\mathbf{x},$$

où ε est la longueur de lissage. Le noyau lisse doit au moins respecter les propriétés suivantes

$$\begin{cases} \int_{\Omega} \phi_\varepsilon(\mathbf{x}) d\mathbf{x} = 1, \\ \phi_\varepsilon(\mathbf{x}) \rightarrow \delta(\mathbf{x}), \quad \varepsilon \rightarrow 0, \\ \phi_\varepsilon(\mathbf{x}) \in C^k, \quad k \geq 1, \end{cases}$$

où les deux premières propriétés sont des propriétés résiduelles de la distribution de Dirac et la dernière est une exigence de différentiabilité nécessaire pour approcher les opérateurs différentielles. La figure 2.3. La fonction moyenne $\langle \mathbf{u} \rangle$ est ensuite utilisée pour approximer la fonction d'origine. Dans un second temps, le domaine d'origine Ω est subdivisé avec N_p sous-domaines Ω_p associés à une particule lagrangienne à l'emplacement

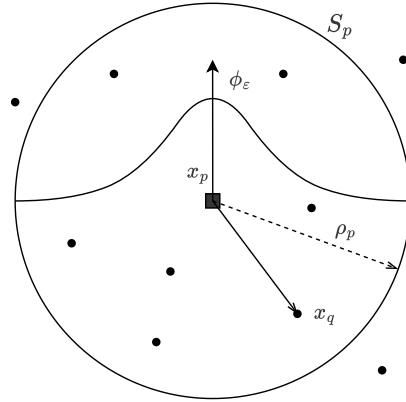


Figure 2.3: Une fonction noyau est associée à chaque particule. Sa valeur en un point \mathbf{x}_q est fonction de sa distance avec son centre \mathbf{x}_p .

$\mathbf{x}_p(t) \in \Omega_p$. Nous notons V_p le volume de Ω_p . Par la suite, nous noterons \mathcal{P} = l'ensemble des particules de cette partition spatiale. Cette discréétisation est ensuite utilisée pour approximer la fonction moyenne par une méthode de quadrature sur les particules telles que

$$\begin{aligned} \langle \mathbf{u}(\mathbf{x}, t) \rangle &= \sum_{p \in \mathcal{P}} \int_{\Omega_p} \mathbf{u}(\mathbf{x}', t) \phi_\varepsilon(\mathbf{x} - \mathbf{x}') d\mathbf{x}' \\ &\simeq \sum_{p \in \mathcal{P}} \mathbf{u}(\mathbf{x}_p(t), t) V_p \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p(t)) \\ &\simeq \sum_{p \in \mathcal{P}} \mathbf{\Gamma}_p(t) \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p(t)). \end{aligned}$$

Ainsi, toute fonction définie sur une discréétisation par particule est caractérisée par un ensemble de positions de particules \mathbf{x}_p associées à une valeur de particule $\mathbf{\Gamma}_p(t) = \mathbf{u}(\mathbf{x}_p, t) V_p$ et un noyau lisse ϕ_ε . Sur la base de cette discréétisation, l'opérateur différentiel peut être dérivé à travers cette formulation. Tout comme le champ \mathbf{u} , la même interpolation peut être appliquée pour obtenir

$$\nabla \mathbf{u}(\mathbf{x}, t) = \sum_{p \in \mathcal{P}} \mathbf{\Gamma}_p(t) \nabla \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p(t)).$$

Il existe toutefois une grande variété de formule pour approximer l'opérateur gradient. Dans ce cas, ce sont les propriétés de conservation associées au champ qui sont privilégiées. Généralement, un terme additionnel est introduit lorsque le champ est évalué à la position d'une particule $p \in \mathcal{P}$ tel que

$$\nabla \mathbf{u}(\mathbf{x}_p(t), t) = \sum_{q \in \mathcal{P}} (\mathbf{\Gamma}_q(t) - \mathbf{\Gamma}_p(t)) \nabla \phi_\varepsilon(\mathbf{x}_q(t) - \mathbf{x}_p(t)),$$

où $\sum_q \mathbf{\Gamma}_p(t) \nabla \phi_\varepsilon(\mathbf{x}_q(t) - \mathbf{x}_p(t))$ est nul par propriété de localisation du noyau ϕ_ε .

Finalement, l'ensemble de particules \mathcal{P} va permettre de résoudre les équations de la dynamique, et ce, soit par discrétisation des opérateurs de la formulation forte comme avec la méthode SPH, soit par discrétisation de la formulation faible comme avec la méthode MPM. Ainsi, un modèle d'évolution est associé aux positions $\mathbf{x}_p(t)$ et intensités $\mathbf{\Gamma}_p(t)$ tel que

$$\begin{cases} \frac{d\mathbf{\Gamma}_p(t)}{dt} = M_\Gamma(\mathbf{\Gamma}_p(t); \mathcal{P}(t)), \\ \frac{d\mathbf{x}_p(t)}{dt} = M_x(\mathbf{x}_p(t); \mathcal{P}(t)). \end{cases} \quad (2.4)$$

C'est en particulier grâce à la mise à jour de la position \mathbf{x}_p qui le terme d'advection est pris en compte dans la résolution du problème dynamique.

Exemples de fonction de noyau

Plusieurs noyaux ont été utilisés en fonction de la méthode. La formulation originale de la MPM n'utilise pas de noyau de substitution et écrit la densité comme suit

$$\mathbf{u}(\mathbf{z}, t) = \sum_{p \in \mathcal{P}} \mathbf{\Gamma}_p(t) \delta(\mathbf{z} - \mathbf{z}_p(t)).$$

Ce n'est pas tant cette fonction que la projection sur une grille de fond associée à certaines fonctions de forme qui est utilisée pour discrétiser le problème [Sulsky et al., 1994]. La méthode GIMP [Bardenhagen and Kober, 2004] est une formulation différente qui utilise la fonction de Heaviside et associe donc bien volume autour de chaque particule

$$M_1(r) = \frac{\alpha}{\varepsilon} \begin{cases} 1; & r \leq \varepsilon, \\ 0; & \text{sinon,} \end{cases}$$

où $r = \|\mathbf{z}\|_2$ et α un facteur de normalisation. Cette méthode a été introduite pour éviter le problème de *cell crossing* lorsque une particule se déplace d'une cellule à une autre à travers la grille de fond. Dans la méthode SPH, comme son nom l'indique, un noyau lisse est associé pour approximer la solution. Théoriquement, il pourrait s'agir de la fonction de noyau gaussien

$$\phi_g(r) = \frac{1}{(\pi \varepsilon^2)^{d/2}} \exp(-r^2/\varepsilon^2).$$

Ce noyau est infiniment différentiable mais défini sur un support non-compact. En pratique, nous utilisons un seuillage pour supprimer les valeurs négligeables pour une

grande distance par rapport à une particule. D'autres noyaux, basés sur des fonctions B-Spline, permettent de travailler sur un support compact. Ces fonctions sont également positives, ce qui est une exigence pour certains champs comme la densité. Par exemple, le B-spline quadratique M_3 est défini par

$$M_3(r) = \frac{\alpha}{\varepsilon^d} \begin{cases} \frac{3}{4} - |q|^2 & 0 \leq |q| < \frac{1}{2} \\ \frac{1}{2} \left(\frac{3}{2} - |q| \right)^2 & \frac{1}{2} \leq |q| < \frac{3}{2} \\ 0 & \frac{3}{2} \leq |q| \end{cases}$$

avec $r = \|z\|_2$, $q = r/\varepsilon$ et α la condition de normalisation et d la dimension spatiale. Ce noyau garantit la continuité C^1 . Le noyau cubique est un autre noyau B-Spline qui est

$$M_4(r) = \frac{\alpha}{\varepsilon^d} \begin{cases} \frac{1}{6}(-|q|+2)^3 - \frac{4}{6}(-|q|+1)^3 & 0 \leq |q| \leq 1 \\ \frac{1}{6}(-|q|+2)^3 & 1 \leq |q| \leq 2 \\ 0 & 2 \leq |q| \end{cases}$$

Dans ce dernier cas, le facteur de normalisation α est

$$\alpha = \begin{cases} 1; & 1 \text{ d} \\ 30/14\pi; & 2 \text{ d} \\ 3/2\pi; & 3 \text{ d} \end{cases}$$

Notons que ces noyaux ont été définis avec la coordonnée radiale r . Une autre possibilité serait de définir le noyau multidimensionnel comme le produit tensoriel du noyau 1d. C'est ce qui est généralement fait pour définir les fonctions de forme associées à la grille de la méthode MPM ou bien les noyaux de redistribution utilisés pour remailler une discrétisation particulière comme en Section 4.4.3.

Distribution particulière admissible

Les méthodes particulières sont ici des méthodes d'approximation de champs continus. Ces discrétisations doivent ainsi vérifier un certain nombre de critères pour être conforme à la solution. Il s'agit en particulier de donner ici des critères de régularité pour définir si une méthode d'interpolation particulière est valide. On trouvera des détails supplémentaires dans le livre [S. Li and W. K. Liu, 2004] Chapitre 4. On rappelle tout d'abord ce qu'est le support d'une particule p . Chaque particule est associée à une fonction d'interpolation $\phi_\varepsilon(\mathbf{x})$. Le support compact est donc

$$S_p = \{\mathbf{x} \mid \|\phi_\varepsilon(\mathbf{x} - \mathbf{x}_p)\| > 0\} \cap \Omega,$$

On définit également ρ_p le rayon du support compact comme

$$\rho_p = \max\{\|\mathbf{x} - \mathbf{x}_p\| > 0 \mid \mathbf{x} \in S_p\}.$$

C'est l'ensemble de ces sous-domaines qui forment la distribution particulaire. Nous supposons un nombre fini de particules comme décrit précédemment.

Definition 3 (distribution particulaire admissible). On définit une distribution admissible selon plusieurs conditions

1. L'union des supports de particules

$$S := \bigcup_{p \in \mathcal{P}} S_p$$

est inclus dans le domaine $\bar{\Omega}$ dans lequel réside les particules de tel sorte que $\bar{\Omega} \subseteq S$.

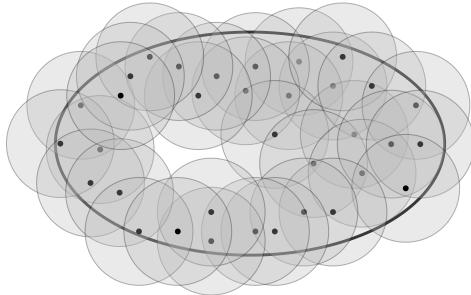
2. $\forall \mathbf{x} \in \Omega$, il existe une boule

$$\mathcal{B}(x) = \{x \mid \|\mathbf{x} - \bar{\mathbf{x}}\| < \rho\}$$

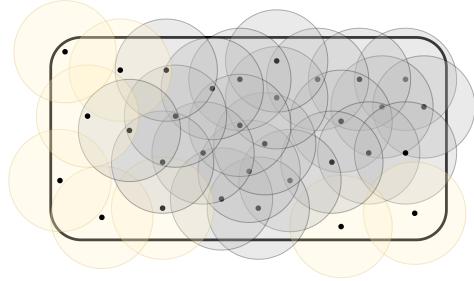
tel que le nombre de particule dans \mathcal{B} satisfait deux bornes $0 < N_{min} < N_{max} < \inf$

$$N_{min} \leq N \leq N_{max}$$

3. Pour $\Omega \in \mathbb{R}^d$, il est nécessaire que $B(x)$ admette au moins $d + 1$ particules dont les vecteurs positions forment un simplexe non-dégénéré de rang d .



(a) La zone centrale du domaine n'est couverte par aucun support de particule. La condition 1 n'est pas respectée.



(b) Les particules de support de couleur jaune. La condition 3 n'est pas respectée.

Figure 2.4: Exemples de distribution particulaire non admissible.

La condition 1 est une condition de recouvrement. Cette dernière est une condition nécessaire de la condition 2. Elle peut être vue comme une condition de chevauchement, nécessaire afin de « lier » les particules entre elles. Finalement, la condition 3 est une condition nécessaire pour que la matrice des moments soit inversible. La figure 2.4

présente deux distributions particulières non admissibles. La définition d'admissibilité a été développée en termes de (ρ, p) -régularité par Han et Meng [Han and Meng, 2001] et par Duarte et Oden [Duarte and Oden, 1996]. Cette définition est en particulier déterminante pour pouvoir appliquer les schémas d'interpolation définis dans la Section 4.4.3.

Méthode de remaillage pour obtenir un support de particule commun

La méthode de remaillage est une technique essentielle dans notre approche pour maintenir un support de particule uniforme. À l'origine développée pour atténuer les distorsions dans la distribution des particules, cette méthode se fonde sur un schéma de redistribution sur une grille régulière de particules, utilisant des opérateurs de projection et d'interpolation. Ce processus permet de passer d'une représentation lagrangienne à une représentation eulérienne, puis ensuite revenir à une représentation lagrangienne.

L'essence de cette méthode réside dans l'utilisation d'opérateurs similaires à ceux employés par les méthodes *Particles-in-Cell* (PIC) comme la méthode *Vortex-In-Cell* (VIC) ou la *Material Point Method* (MPM).

Le remaillage est défini en deux étapes.

Consiste en un schéma de redistribution sur une grille régulière de particules à l'aide d'opérateur de projection et d'interpolation. Elle consiste à passer d'une représentation lagrangienne à eulérienne, puis de revenir sur une grille de particules lagrangienne. En fait même type d'opérateur qu'utilisé par les méthodes PIC.

Dans notre méthodologie, nous proposons une approche en deux étapes. Tout d'abord, nous effectuons une étape d'assignation ((a)) pour transférer la discréétisation des particules à une discréétisation sur une grille. Ensuite, une étape d'interpolation ((b)) est réalisée pour obtenir un nouvel ensemble de particules régulièrement espacées. Notre analyse se rapporte au scénario unidimensionnel, où $\Omega \in \mathbb{R}$. L'extension au cas n -dimensionnel peut être réalisée par la tensorisation de l'approche unidimensionnelle.

(a) Affectation sur une grille eulérienne

Nous désignons par z_I et z_p respectivement les emplacements sur la grille et les anciens emplacements des particules. Les nouvelles particules sont définies sur une grille de n_g éléments avec un espace régulier $\ell_I = 2d_p$, où d_p est la taille caractéristique des particules. Nous définissons les intensités des particules comme Γ_p et les valeurs du champ nodal comme \mathbf{u}_I . En utilisant une fonction de forme W , l'étape d'affectation des particules à chaque nœud $I \in \Lambda$ peut être écrite comme

$$\mathbf{u}_I = \frac{1}{V_I} \sum_{p \in \mathcal{P}} \Gamma_p W \left(\frac{z_I - z_p}{\ell_I} \right).$$

Où W détermine une redistribution de l'intensité sur la grille, la nouvelle discréétisation

peut ensuite être utilisée pour approximer le champ \mathbf{u}_p , défini par la discréétisation des particules par interpolation

$$\mathbf{u}_p(z) \approx \mathbf{u}_g(z) = \sum_{I \in \Lambda} \mathbf{u}_I W\left(\frac{z - z_I}{\ell_I}\right) \quad \forall z \in \Omega.$$

(b) *Interpolation sur une nouvelle discréétisation régulière des particules*

Un nouvel ensemble de particules est défini au quart de chaque cellule de sorte que la nouvelle position est définie à $z_{p'} = d_p/2 + i dp$, $i = 0, \dots, 2n_g$. La valeur du champ est alors interpolée à cette nouvelle position et multipliée par le volume de la particule $\Gamma_{p'} = \mathbf{u}_g(z_{p'})V_{p'}$ afin de donner une nouvelle approximation particulaire de ce champ

$$\mathbf{u}_g(z) \approx \mathbf{u}_{p'}(z) = \sum_{p' \in \mathcal{P}'} \mathbf{u}_g(z_{p'})V_p,$$

La combinaison de ces deux étapes peut être initialement utilisée pour générer une nouvelle distribution de particules non déformées. La fonction de forme W détermine le type et la qualité du transfert. Le critère de la qualité de la méthode réside dans la conservation des premiers moments des distributions de particules, comme détaillé en Annexe A.

Pour W , on peut utiliser une fonction d'interpolation affine, qui garantit la conservation du moment 0. Pour une conservation des moments supérieurs, la fonction B-spline fournit une fonction de lissage d'ordre plus élevé.

Monaghan [Monaghan, 1985] développe une approche systématique pour améliorer la précision et maintenir la régularité par extrapolation. Le concept implique la construction d'une nouvelle fonction de forme basée sur une coupure et sa dérivée radiale. Pour $m = 4$, la B-spline cubique est améliorée par le noyau d'interpolation suivant

$$M'_4(z) = \begin{cases} 1 - \frac{5}{2}z^2 + \frac{3}{2}|z|^3 & 0 \leq |z| \leq 1 \\ \frac{1}{2}(2 - |z|)^2(1 - |z|) & 1 \leq |z| \leq 2 \\ 0 & 2 \leq |z|. \end{cases}$$

que nous utiliserons dans les applications.

Enfin, dans l'espace multidimensionnel, le noyau de redistribution W peut être obtenu comme le produit du noyau unidimensionnel appliqué à chaque coordonnée, comme suit

$$\begin{aligned}\boldsymbol{\Gamma}_p &= \sum_{I \in \Lambda} \boldsymbol{\Gamma}_I W(\boldsymbol{z}_p - \boldsymbol{z}_I, \ell_I) \\ &= \sum_{I \in \Lambda} \boldsymbol{\Gamma}_I \prod_{i=1}^d W_{1D} \left(\frac{\boldsymbol{z}_{I,i} - \boldsymbol{z}_{p,i}}{\ell_I} \right)\end{aligned}$$

Le schéma est illustré dans la Figure 2.5

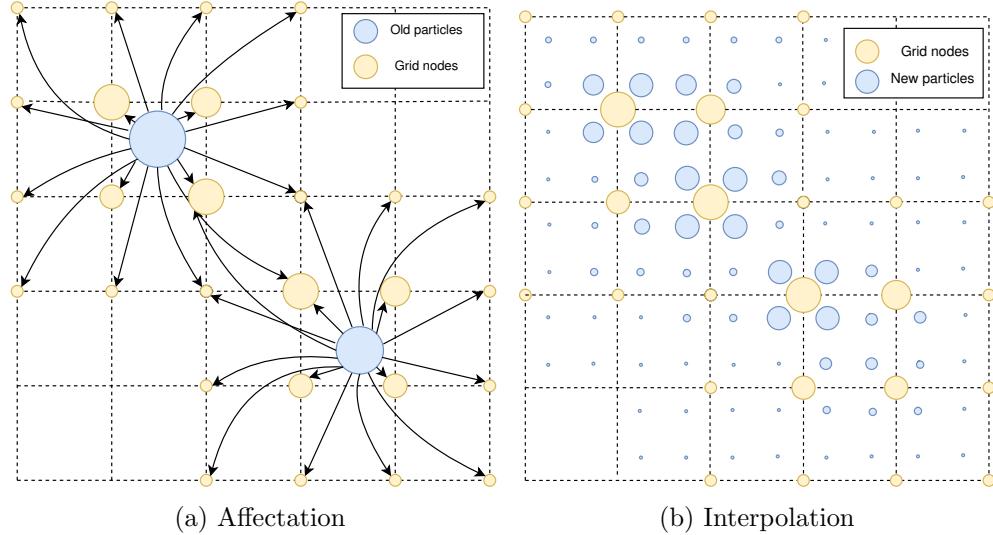


Figure 2.5: Schéma de remaillage d'une distribution particulaire. Après l'étape d'affectation sur la grille des particules sont régulièrement générées pour interpoler les valeurs de champ 2.5b.

Approximation d'un champs continu par une discréétisation particulière

Différentes méthodes ont pu être introduites pour approcher une fonction \mathbf{u} par une discréétisation particulière. Ceci est particulièrement utile lors de l'initialisation ou bien pour réduire les effets de distorsion de la distribution particulière.

Approximation particulière Cette première méthode consiste à utiliser les particules comme point de quadrature. Comme dans les hypothèses initiales des méthodes particulières (2.4), il convient alors d'évaluer le champ à la position des particules x_p afin d'obtenir une approximation de l'intensité comme intégration locale du champ

$$\boldsymbol{\Gamma}_p^a = \int_{\Omega_p} \mathbf{u}^a(\mathbf{x}) d\mathbf{x} = \mathbf{u}(\mathbf{x}_p) V_p, \quad p \in \mathcal{P}_i^f$$

Si cette approximation est très simple à évaluer mais ne garantit pas d'obtenir une bonne approximation du champ si la qualité de la quadrature est de mauvaise qualité. Ceci est particulièrement le cas si la distribution particulière n'est pas uniforme ou ne recouvre pas entièrement le domaine Ω .

Régression RBF Afin d'avoir une meilleure évaluation du champ en chaque position de particules \mathbf{x}_p , Beale [Beale, 1988] introduit une méthode itérative afin de déterminer les intensités Γ_p tel que

$$\sum_{q \in P^f} \Gamma_p \phi_\varepsilon(\mathbf{x}_p - \mathbf{x}_q) = \mathbf{u}^a(\mathbf{x}_p), \quad p \in \mathcal{P}^f.$$

En réalité, cette méthode revient à résoudre un problème de régression. C'est cette dernière approche qui a été développé par Barba et al. [Barba, 2004] pour modifier les intensités particulières. Ce sont ces types de méthodes qui sont couramment utilisées pour la résolution d'EDP avec des fonctions à base radiales [Fornberg and Flyer, 2015; Sperotto et al., 2022].

On cherche à déterminer les intensités de particule définies sous la forme d'un vecteur $\boldsymbol{\Gamma} = [\Gamma_1, \dots, \Gamma_p]^T$. On suppose la valeur du champ à approcher \mathbf{u}^a connu en position des particules et dont les évaluations sont consignés dans un vecteur $\mathbf{u} = [\mathbf{u}_1(z_1), \dots, \mathbf{u}_p(z_p)]^T$. L'approximation particulière peut être évalué en chaque position \mathbf{x}_p et doit permettre d'approcher en ces points le vecteur \mathbf{u} comme

$$\mathbf{u} \simeq \tilde{\mathbf{u}} = \boldsymbol{\Phi} \boldsymbol{\Gamma},$$

où $\boldsymbol{\Phi}_{ij} = \phi_\varepsilon(z_i - z_j)$.

Minimiser l'erreur quadratique entre la prédiction $\tilde{\mathbf{u}}$ et \mathbf{u} revient à résoudre le problème suivant

$$\boldsymbol{\Gamma}^* = \arg \min_{\boldsymbol{\Gamma}} \|\mathbf{u} - \boldsymbol{\Phi} \boldsymbol{\Gamma}\|_2^2.$$

L'opérateur étant linéaire, la solution de ce problème n'est autre que $\boldsymbol{\Gamma}^* = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi} \mathbf{u}$. Cependant, ce problème peut être mal conditionné. En particulier dans le cas où les particules sont mal distribuées (cas dégénéré lorsque deux particules se superposent). Pour éviter ces cas, un terme de pénalisation sur l'amplitude du vecteur d'intensité $\boldsymbol{\Gamma}$ est introduit. En introduisant un régularisation à la Thikhonov de la forme $\lambda \|\boldsymbol{\Gamma}\|_2^2$, on obtient un problème dit de régression *Ridge*, où le coefficient λ est le coefficient de pénalisation. Le problème devient alors

$$\boldsymbol{\Gamma}_{\text{ridge}}^* = \arg \min_{\boldsymbol{\Gamma}} \|\mathbf{u} - \boldsymbol{\Phi} \boldsymbol{\Gamma}\|_2^2 + \lambda \|\boldsymbol{\Gamma}\|_2^2,$$

donnant la solution suivante $\boldsymbol{\Gamma}_{\text{ridge}}^* = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \lambda \mathbf{I})^{-1} \boldsymbol{\Phi} \mathbf{u}$.

Finalement, cette adaptation du filtre EnKF permet d'approcher les solutions analysées pour chaque membre, en modifiant uniquement les intensités de particule des membres en question. De cette manière, nous évitons l'accroissement excessif du nombre de particules. De plus, elle permet d'être entièrement définie sans étape de remaillage.

2.2.3 Hydrodynamique des particules lissées, *Smoothed particle hydrodynamics* (SPH)

La méthode SPH a été développée indépendamment par Lucy [Lucy, 1977], et Gingold et Monaghan [Gingold and Monaghan, 1977]. Elle a été formulée initialement pour des problème de formation et d'évolution des systèmes stellaires [Hultman and Pharasyn, 1999] mais elle trouve un gain d'intérêt considérable en mécanique des fluides [Kum et al., 1995; Shadloo et al., 2016]. Toutefois, la méthode n'est pas interprétée de la même manière. D'une part, comme en mécanique quantique elle est utilisée pour lisser un système discret afin d'obtenir un milieu continu discrétisé par un ensemble de particules. En mécanique, cette méthode est vu comme une méthode de discrétisation particulaire d'un milieu continu. C'est cette représentation que nous utilisons ici.

Les particules sont des interpolants qui transportent toutes les quantités nécessaires comme la densité, la vitesse, le tenseur de contrainte ou l'énergie. Elle consiste à résoudre la forme forte des équations de la dynamique. Les équations résolues sont l'équation de continuité et l'équation de conservation de la quantité de mouvement à travers l'équation d'Euler

$$\begin{cases} \frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0, \\ \frac{d\mathbf{v}}{dt} = \frac{1}{\rho} \nabla \cdot \boldsymbol{\sigma}, \end{cases}$$

où ρ est la densité, \mathbf{v} la vitesse, $\boldsymbol{\sigma}$ la contrainte de Cauchy.

En utilisant la règle d'approximation du gradient le terme $\rho \nabla \cdot \mathbf{v}$ peut être approximé pour chaque particule $p \in \mathcal{P}$, ce qui donne pour l'équation de continuité

$$\frac{d\rho_p}{dt} = \sum_{q \in \mathcal{P}} m_j (\mathbf{v}_q - \mathbf{v}_p) \cdot \nabla \phi_\varepsilon(\mathbf{x}_p - \mathbf{x}_q),$$

où m_p est la masse de la particule assignée initialement comme $m_p = \rho_p V_p$, et \mathbf{v}_p sa vitesse. L'équation d'équilibre des quantités de mouvement est discrétisée de la même manière. La forme suivante est généralement utilisée

$$m_p \frac{d\mathbf{v}_p}{dt} = \sum_q m_p m_q \left(\frac{\boldsymbol{\sigma}_p}{\rho_p^2} + \frac{\boldsymbol{\sigma}_q}{\rho_q^2} \right) \cdot \nabla \phi_\varepsilon(\mathbf{x}_p - \mathbf{x}_q).$$

Cette version est symétrique par rapport aux indices p et q ce qui favorise les propriétés de conservation. Comme défini en (2.4), on obtient des équations d'évolution pour la

densité ρ_p et la vitesse \mathbf{v}_p , utilisées pour mettre à jour la position des particules \mathbf{x}_p via par exemple un algorithme *leap-frog*. La variable de contrainte $\boldsymbol{\sigma}_p$ est mise à jour au travers d'une loi de comportement. En pratique, pour des problèmes de choc ou dominé par la convection, un terme de viscosité artificiel est ajouté dans l'équation d'évolution de la vitesse [Monaghan and Gingold, 1983].

2.2.4 Méthode des points matériels, *Material Point Method* (MPM)

La méthode des points matériels (MPM) est une autre méthode particulaire qui utilise une double description: lagrangienne au travers de particules et eulérienne à l'aide d'une grille fixe. Tout comme la méthode SPH, elle utilise une discréétisation particulaire, les points matériels, qui en se déplaçant transportent toutes les quantités nécessaires pour le problème : masse, vitesse, tenseur de déformation, ou variables internes. D'autre part, l'interpolation du champ de déplacement est réalisée à l'aide de coordonnées spatiales de la grille fixe et souvent régulière. Les deux discréétisations interagissent au travers d'opérateurs de projection et interpolation entre particules et les noeuds et de la grille qui les entourent. Cette méthode fait partie de la famille *particle-in-cell* (PIC). La méthode MPM est un adaptation en mécanique des solides développé par Sulsky et al. [Sulsky et al., 1994] de la méthode FLIP de Brackbill [Brackbill et al., 1988] en mécanique des fluides.

Comme en éléments finis, la méthode MPM consiste à résoudre le problème aux valeurs sous sa forme faible

$$\int_{\Omega} \rho \mathbf{a} \cdot \mathbf{q} d\Omega = - \int_{\Omega} \rho \boldsymbol{\sigma} : \nabla \cdot \mathbf{q} d\Omega + \int_{\Omega} \rho \mathbf{b} \cdot \mathbf{q} d\Omega + \int_{\partial\Omega_t} \mathbf{t} \cdot \mathbf{q} dS.$$

où \mathbf{a} est l'accélération, \mathbf{v} la vitesse, $\boldsymbol{\sigma}$ le tenseur de contrainte, \mathbf{q} une fonction test, \mathbf{t} le vecteur de traction et \mathbf{b} la force volumique.

Le schéma MPM peut être obtenu en concentrant la masse sur N_p particules. La densité est alors représentée comme une somme de Dirac

$$\rho(\mathbf{x}) = \sum_{p \in \mathcal{P}} V_p \rho_p \delta(\mathbf{x} - \mathbf{x}_p) = \sum_{p \in \mathcal{P}} m_p \delta(\mathbf{x} - \mathbf{x}_p).$$

De même, on discréétise la contrainte $\boldsymbol{\sigma}$

$$\boldsymbol{\sigma}(\mathbf{x}) = \sum_{p \in \mathcal{P}} \boldsymbol{\sigma}_p \delta(\mathbf{x} - \mathbf{x}_p).$$

Finalement, en injectant cette discréétisation de la densité dans (2.2.4)

$$\begin{aligned} \sum_{p \in \mathcal{P}} m_p \mathbf{q}(\mathbf{z}_p(t), t) \cdot \mathbf{a}(\mathbf{z}_p(t), t) &= - \sum_{p \in \mathcal{P}} m_p \boldsymbol{\sigma}(\mathbf{z}(t), t) : \nabla \mathbf{u}(\mathbf{z}(t), t) \\ &\quad + \int_{\partial \Omega_t} \mathbf{t} \mathbf{q} dS + \sum_p m_p \mathbf{q}(X(t), t) \cdot \mathbf{b}(X(t), t). \end{aligned}$$

On trouve un certain nombre de généralisation dans la littérature comme la méthode GIMP [Bardenhagen and Kober, 2004] en définissant une représentation particulière de la densité où les Dirac sont remplacés par des fonction caractéristique χ_p tel que

$$\rho(\mathbf{z}) = \sum_{p \in \mathcal{P}} m_p \chi(\mathbf{z} - \mathbf{z}_p)$$

Les variables cinématiques sont ensuite discrétisées sur la grille eulérienne. La précédente équation est résolue sur la grille et l'accélération est déterminée sur chaque nœud $I \in \Lambda$ de la grille comme

$$\sum_{J \in \Lambda} m_{IJ} \mathbf{a}_J = \mathbf{f}_I^{\text{int}} + \mathbf{f}_I^{\text{ext}},$$

où m_{IJ} la matrice de masse, $\mathbf{f}_I^{\text{int}}$ sont les forces internes, et $\mathbf{f}_I^{\text{ext}}$ les forces externes.

En pratique, afin d'obtenir $m_{IJ}, \mathbf{a}_J, \mathbf{f}_I^{\text{int}}$ à partir des quantités particulières, les deux niveaux de discrétisation communiquent selon des étapes de projection (*particles to grid*) puis d'interpolation (*grid to particles*).

Projection Les valeurs de la grille sont initialisées à 0. La masse m_p , la quantité de mouvement $m_p \mathbf{v}_p$ et les forces \mathbf{f}_p sont transférées à la grille à l'aide des fonctions de forme ϕ_I associées à chaque nœud

$$\begin{cases} m_I = \sum_p \varphi_I(\mathbf{x}_p) m_p, \\ m_I \mathbf{v}_I = \sum_p \varphi_I(\mathbf{x}_p) m_p \mathbf{v}_p, \\ \mathbf{f}_I = \sum_p \varphi_I(\mathbf{x}_p) \mathbf{f}_p. \end{cases}$$

Mise à jour sur la grille L'équation 2.2.4 est résolue sur les nœuds de la grille avec un schéma explicite

$$\begin{cases} m_I \mathbf{a}_I = \mathbf{f}_I + \mathbf{f}_g, \\ m_I \mathbf{v}^{n+1} = \mathbf{v}^n + \Delta t (\mathbf{f}_I + \mathbf{f}_g)/m_I, \\ \mathbf{x}_I^{n+1} = \mathbf{x}_I^n + \Delta t \mathbf{v}^{n+1}. \end{cases}$$

Interpolation Les quantités particulières sont mises à jour par interpolation des quantités nodales afin de suivre la déformation de la grille mettant à jour positions \mathbf{x}_p et vitesses \mathbf{v}_p . C'est ainsi que l'étape d'advection est prise en compte.

Celle-ci est réalisée directement par interpolation de la vitesse, de l'accélération, ou une combinaison des deux [Stomakhin et al., 2013] tel que

$$\mathbf{v}_p^{n+1} = \alpha \left(\mathbf{v}_p^n + \sum_I \varphi_{Ip} (\mathbf{v}_I^{n+1} - \mathbf{v}_I^n) \right) + (1 - \alpha) \sum_I \varphi_{Ip} \mathbf{v}_I^{n+1}$$

avec $\alpha \in [0, 1]$.

La position est mise jour en interpolant la déformation de la grille de telle sorte que

$$\mathbf{x}_p^{n+1} = \mathbf{x}_p^n + \Delta t \sum_I \varphi_I(\mathbf{x}_p^n) \mathbf{v}_p^{n+1}$$

Des transferts plus plus complexes ont été développés afin d'offrir des schémas plus stables et conservatifs [Jiang, Schroeder, Selle, et al., 2015; Fu et al., 2017; Hu, Fang, et al., 2018]. Finalement, on retrouve également comme en (2.4) des équations d'évolution pour les quantités particulières en particulier pour la position et la vitesse

$$\begin{cases} \frac{d\mathbf{v}_p(t)}{dt} = \mathbf{I}_v \circ \mathbf{P}_v(\mathbf{v}_p(t); \mathcal{P}), \\ \frac{d\mathbf{x}_p(t)}{dt} = \mathbf{I}_x \circ \mathbf{P}_x(\mathbf{x}(t); \mathcal{P}), \end{cases} \quad (2.5)$$

où \mathbf{P} et \mathbf{I} sont respectivement des opérateurs de projection et d'interpolation.

Tout comme la méthode SPH, des lois de comportement sont introduites pour pouvoir calculer \mathbf{f}_p , généralement grâce au tenseur de déformation, le tenseur de contrainte et de variable interne comme fait classiquement en mécanique des solides.

2.2.5 Méthodes vortex

Méthode vortex classique

La méthode vortex est une méthode particulaire utilisée pour résoudre les équations de Navier-Stokes pour des écoulements incompressibles [Cottet and Koumoutsakos, 2000b]. Elle a été pour la première fois développée indépendamment par Prager [Prager, 1928] et Rosenhead [Rosenhead, 1931]. Elle se base sur la discrétisation du champ de tourbillon par un ensemble de particules, et résout la formulation vorticité-vitesse des équations de Navier-Stokes

$$\begin{cases} \frac{\partial \boldsymbol{\omega}}{\partial t} + (\mathbf{u} \cdot \nabla) \boldsymbol{\omega} = (\boldsymbol{\omega} \cdot \nabla) \mathbf{u} + \nu \Delta \boldsymbol{\omega}, \\ \Delta \mathbf{u} = -\nabla \times \boldsymbol{\omega}, \end{cases}$$

où \mathbf{u} la vitesse, $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ le champ de tourbillon, et ν la viscosité.

Les formes lagrangiennes des équations précédentes sont

$$\begin{cases} \frac{d\mathbf{x}_p}{dt} = \mathbf{u}(\mathbf{x}_p, t), \\ \frac{d\boldsymbol{\omega}_p}{dt} = -[\nabla \times \mathbf{u}(\mathbf{x}_p, t)] \boldsymbol{\omega}_p + \nu \Delta \boldsymbol{\omega}(\mathbf{x}_p, t). \end{cases}$$

Dans le cas d'un écoulement bi-dimensionnelle, le champ tourbillon est un champ scalaire porté par la troisième dimension. En particulier dans un repère cartésien $\omega = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}$. De plus le terme d'étiirement disparaît $(\boldsymbol{\omega} \cdot \nabla) \mathbf{u}$, ainsi les équations lagrangiennes se simplifient

$$\begin{cases} \frac{d\mathbf{x}_p(t)}{dt} = \mathbf{u}(\mathbf{x}_p, t) \\ \frac{d\omega(\mathbf{x}_p, t)}{dt} = \nu \Delta \omega(\mathbf{x}_p, t) \end{cases}$$

Le champ de vorticité est discréteisé à l'aide d'un ensemble de particules p définies à une position \mathbf{x}_p , une quantité de circulation locale Γ_p qui est par définition la circulation autour de la particule $\Gamma_p = \oint_{\partial\Omega_p} \mathbf{v} = \int_{\Omega_p} \omega dS$. Ainsi, pour tout point $\mathbf{x} \in \Omega \subset \mathbb{R}^2$ la vorticité peut être exprimée comme

$$\omega(\mathbf{x}, t) = \sum_{i=1}^{N_p} \Gamma_p(t) \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p(t)),$$

où ϕ_ε est le noyau de lissage.

La vitesse \mathbf{u} peut être obtenue en résolvant l'équation de Poisson suivante

$$\Delta \mathbf{u} = -\nabla \times \boldsymbol{\omega}.$$

Finalement, par une représentation intégrale et en choisissant $\phi_\varepsilon = \delta$, on obtient dans le cas 2D l'équation de Biot-Savart suivante

$$\mathbf{u}(\mathbf{x}) = \sum_{p \in \mathcal{P}} \frac{\Gamma_p}{2\pi} \frac{(\mathbf{x} - \mathbf{x}_p) \times \mathbf{k}}{\|\mathbf{x} - \mathbf{x}_p\|_2^2},$$

où \mathbf{k} est le vecteur unitaire normal au plan.

En pratique, le choix d'un noyau en Dirac rend impossible le calcul de la vitesse sur la discrétisation particulière à cause du dénominateur en $\|\boldsymbol{x} - \boldsymbol{x}_p\|^2$. En choisissant un noyau de type gaussien de taille ε on obtient alors

$$\boldsymbol{u}(\boldsymbol{x}) = \sum_{p \in \mathcal{P}} \frac{\Gamma_p(1 - \exp(-r^2/\varepsilon^2))}{2\pi r^2} (\boldsymbol{x} - \boldsymbol{x}_p) \times \boldsymbol{k}, \quad r = \|\boldsymbol{x} - \boldsymbol{x}_p\|_2.$$

Afin de tenir compte de la diffusion, une approche par fractionnement est généralement utilisée. Introduite pour la première fois par Chorin [Chorin and Bernard, 1973], elle permet dans le cas de problèmes où le terme de transport est dominant, de traiter séparément et successivement les termes d'advection et de diffusion. Ainsi après avoir mis à jour la position des particules sans tenir compte de la viscosité, l'équation suivante est résolue

$$\frac{d\omega(\boldsymbol{x}_p, t)}{dt} = \nu \Delta \omega(\boldsymbol{x}_p).$$

Pour se faire, deux méthodes sont principalement utilisées : soit la méthode par marche aléatoire [Chorin and Bernard, 1973], par étalement du noyau ou *core spreading* [Leonard, 1980], ou par échange d'intensité [Degond and Mas-Gallic, 1989a]. Dans le dernier cas, l'opérateur différentiel est approximé à l'aide de la discrétisation comme il est fait dans la méthode SPH. Dans ce dernier cas l'intensité évolue comme

$$\frac{d\omega_p}{dt} = \nu \varepsilon^{-2} \sum_q V_q [\omega_q - \omega_p] \phi_\varepsilon(\boldsymbol{x}_p - \boldsymbol{x}_q).$$

Finalement, on obtient également les équations d'évolution des quantités particulières

$$\begin{cases} \frac{d\boldsymbol{x}_p(t)}{dt} = \boldsymbol{u}(\boldsymbol{x}_p(t), t), \\ \frac{d\Gamma_p(t)}{dt} = \nu \varepsilon^{-2} \sum_{q \in \mathcal{P}} [V_p \Gamma_q(t) - V_q \Gamma_p(t)] \phi_\varepsilon(\boldsymbol{x}_p(t) - \boldsymbol{x}_q(t)) \end{cases}$$

Vortex-In-Cell (VIC)

La méthode *Vortex-In-Cell* [Christiansen, 1973], tout comme la méthode MPM 2.2.4, est une version *Particle-In-Cell* [Birdsall and Fuss, 1969] de la méthode Vortex précédemment décrite. Elle a été développé pour tenir compte de ses faiblesses. Comme la méthode MPM, celle-ci repose sur la représentation particulaire pour tenir compte du terme d'advection mais également d'une grille de calcul pour résoudre l'équation de Poisson en utilisant des méthodes eulériennes sur grille.

Le schéma de transfert est similaire à celui de la méthode MPM. D'abord une projection du champ de vorticité sur la grille de noeuds Λ pour obtenir les valeurs nodales ω_I à l'aide d'une fonction de forme W_I

$$\omega_I = \frac{1}{V_I} \sum_{p \in \mathcal{P}} \Gamma_p W_I(\mathbf{x}_p).$$

L'équation de Poisson 2.2.5 est résolue sur la grille, soit par différences finies, soit par une méthode FFT pour obtenir des vitesses au nœud \mathbf{u}_I . La vitesse est ensuite interpolée sur les particules pour mettre à jour leurs positions.

$$\frac{d\mathbf{x}_p}{dt} = \sum_{I \in \Lambda} \mathbf{u}_I W_I(\mathbf{x}_p).$$

Le terme de diffusion peut également être pris en compte sur la grille de calcul. De plus, de cette manière l'étape de recherche de plus proche voisin n'est pas nécessaire, la résolution des équations se faisant directement sur la grille.

2.2.6 Bilan

Les méthodes lagrangiennes sont une classe de méthode de simulation qui se basent sur des ensembles particulaires pour résoudre un problème discret ou bien pour discréteriser un champ continu. Plusieurs méthodes lagrangiennes utilisées pour la simulation ont été présentées. Elles ont toutes en commun de définir chaque particule par une position et une intensité, qui évoluent au cours du temps par l'intégration des équations d'évolution sous leur forme lagrangienne. Ces grandeurs évoluent au cours du temps par intégration des équations d'évolution sous leur forme lagrangienne. En conséquence, le support de discréétisation défini par les positions particulières n'est jamais fixe. Nous avons également présenté les méthodes vortex pour la résolution des équations de Navier-Stokes dans le cas d'un écoulement incompressible. Ce modèle, qui ne dépend que d'un champ scalaire, sera utilisé lors du développement et de l'évaluation des Chapitres 4 et 5.

Méthodes lagrangiennes et assimilation de données

Contents

3.1 Adaptation du filtre de Kalman pour la correction d'intensité	58
3.1.1 Rappel : mise à jour EnKF	58
3.1.2 Cas eulérien	58
3.1.3 Cas lagrangien	60
3.1.4 Première contribution - Adaptation du filtre EnKF	63
3.2 Alignement pour la correction de la discréétisation particulière	64
3.2.1 Assimilation de données par déplacement ou alignement	65
3.2.2 Seconde contribution - correction de position et d'intensité	67
3.3 Bilan du chapitre	67

L'assimilation de données est une technique essentielle en modélisation numérique, visant à ajuster l'état d'un système en intégrant les observations disponibles avec un modèle numérique, tout en prenant en compte les erreurs propres aux deux sources d'information. Cette méthode repose sur une combinaison optimale entre le modèle a priori et les observations, chacune étant associée à ses propres incertitudes. Les méthodes d'ensemble, telles que le filtre EnKF détaillé en Section 2.1.2, sont particulièrement efficaces pour traiter des problèmes non-linéaires de grande dimension. Elles permettent de résoudre le problème en se concentrant sur un sous-espace vectoriel formé par les membres de l'ensemble. Ainsi, l'analyse implique l'estimation des statistiques des variables d'état ou d'observation, puis la déduction d'une combinaison linéaire des vecteurs de l'état discréétisé pour obtenir l'état corrigé.

Néanmoins, l'application de ces méthodes aux simulations lagrangiennes introduit des défis spécifiques. Contrairement aux approches traditionnelles, les méthodes lagrangiennes impliquent que les particules de discréétisation suivent l'écoulement qu'elles génèrent, entraînant une évolution dynamique de la discréétisation de la solution au fil du temps ou d'une simulation à l'autre.

De ce fait, il devient complexe de pouvoir mettre à jour au cours du temps des membres qui ne partage pas le même espace d'approximation fonctionnel. C'est pourquoi, les méthodes de calibration basées sur des méthodes d'assimilation de données avec

des méthodes lagrangienne (DEM [Cheng et al., 2019] ou MPM [Levy et al., 2010] et SPH [Brezzi et al., 2016]) ne mettent jamais mis à jour l'état de la simulation. La procédure prévoit alors de relancer la simulation à chaque itération de la procédure de calibration.

Cependant, certaines approches dans la littérature [Chen et al., 2022; Darakananda et al., 2018; Le Provost and Eldredge, 2021] traitent les particules discrétilisant le champ comme des entités indépendantes, caractérisées uniquement par leur position et leur intensité. Nous montrerons que cette définition de l'état, bien que séduisante, présente des problèmes significatifs. Dans ce chapitre, nous présenterons donc la démarche qui nous a amenés à proposer différentes méthodes d'assimilation adaptées aux méthodes lagrangiennes discrétilisant un milieu continu.

3.1 Adaptation du filtre de Kalman pour la correction d'intensité

3.1.1 Rappel : mise à jour EnKF

Dans la Section 2.1.2, nous avons présenté le filtre EnKF comme une approximation Monte-Carlo du filtre de Kalman. Nous rappelons différentes formes équivalentes du filtre en utilisant les notations introduites précédemment.

La formulation classique du filtre EnKF résulte de l'approximation du gain de Kalman, permettant une combinaison optimale entre l'état prédit et le terme d'innovation. Les matrices de covariance, $\bar{C}y, y$ et $\bar{C}z, y$, sont approximées pour obtenir l'ensemble mis à jour selon l'expression suivante

$$\mathbf{Z}^a = \mathbf{Z}^f + \tilde{\mathbf{K}}(\mathbf{D} - \mathcal{Y}).$$

La correction dans $\tilde{\mathbf{K}}$ dépend à la fois des statistiques sur l'état et les observations.

Nous avons mis en évidence à deux reprises en 2.1.2 et 2.1.3 l'approximation de rang faible de la mise à jour EnKF, définie dans l'espace des membres comme suit

$$\mathbf{Z}^a = \mathbf{Z}^f + \mathbf{Z}^f \mathbf{F}, \quad (3.1)$$

où \mathbf{F} dépend de l'ensemble des prédictions $\mathcal{H}(\mathbf{z}_i^f) \in \mathbb{R}^{d_y}$, de l'observation perturbée \mathbf{d}_i et de l'erreur associée \mathbf{R} . Ainsi, on se retrouve avec une combinaison linéaire des états, avec un terme de correction qui est indépendant de la discrétilisation de la variable d'état \mathbf{z} .

3.1.2 Cas eulérien

Jusqu'à présent, nous avons supposé que \mathbf{z} était une variable vectorielle pour laquelle il est possible de déterminer des statistiques. Cependant ce sont généralement des états continus qui nous intéressent, par exemple, des champs de densité ou de vitesse défini spatialement sur un domaine $\Omega \in \mathbb{R}^d$.

Ces champs sont généralement représentés à l'aide d'une discréétisation. Pour cela, on introduit une base $\{\phi_i\}_{i=1}^n$ qui permet d'approximer le champ u dans l'espace discréétisé.

Dans le cas eulérien, la discréétisation utilisée pour représenter l'espace d'approximation fonctionnelle est fixe au cours du temps et est indépendante des événements tel que

$$u_h(\mathbf{x}, t) = \sum_{i=1}^n U_i \phi_i(\mathbf{x}),$$

où $\{U_i\}_{i=1}^N$ représente le vecteur des intensités dans cette base, noté \mathbf{z} , et ϕ_i sont les fonctions de base définissant la discréétisation. Le vecteur $\mathbf{z} \in \mathbb{R}^N$ contient alors les coefficients de cette approximation, et constitue la variable d'état discréétisée.

Ce principe est notamment appliqué dans les méthodes de discréétisation comme les éléments finis ou les volumes finis. Considérons un domaine $\Omega \subset \mathbb{R}^d$, avec un maillage $\mathcal{T} = (\mathcal{N}, \mathcal{E})$, où $\mathcal{N} = \{\mathbf{x}_i\}_{i=1}^n$ est l'ensemble des nœuds, et $\mathcal{E} = \{K_i\}_{i=1}^m$ l'ensemble des éléments.

- **La méthode des éléments finis (FEM)** [Allaire and Craig, 2007] associe à chaque nœud i une fonction de forme ϕ_i . L'état de la solution discréétisée u_h peut alors être défini comme

$$u_h(\mathbf{x}, t) = \sum_{i=1}^n U_i(t) \phi_i(\mathbf{x}),$$

où U_i sont les valeurs nodales à déterminer. Le vecteur d'état est donc défini sur les nœuds du maillage par $\mathbf{z} = \{U_1, \dots, U_n\}$.

- **La méthode des volumes finis (FVM)** [Eymard et al., 2000] approche le champ u par une valeur moyenne sur chaque élément K_i comme suit :

$$U_i(t) = \frac{1}{|K_i|} \sum_{i=1}^m \int_{K_i} u \, dK.$$

Ainsi, la solution approchée est une solution constante par élément

$$u_h(\mathbf{x}, t) = \sum_{i=1}^m U_i(t) \mathbf{1}_{K_i}(\mathbf{x}),$$

où $\mathbf{1}_{K_i}$ est la fonction indicatrice de l'élément K_i . L'état est donc défini par $\mathbf{z} = \{U_1, \dots, U_m\}$.

On peut réécrire la formule de mise à jour pour chaque membre l à partir de la formule dans l'espace des membres 3.1. Pour tout $\mathbf{x} \in \Omega$, la mise à jour est donnée par

$$u_l^a(\mathbf{x}) = u_l^f(\mathbf{x}) + \sum_{k=1}^N F_{kl} u_k^f(\mathbf{x}).$$

Cette formule peut être décomposée sur une base commune, ce qui permet d'exprimer la mise à jour directement sur les variables d'état $\mathbf{z}_{ii} = 1^N$

$$\mathbf{z}_l^a = \mathbf{z}_l^f + \sum_{k=1}^N F_{kl} \mathbf{z}_k^f.$$

Lorsque le vecteur d'état est aléatoire, il définit complètement le champ aléatoire \mathbf{u}_h dans l'espace discrétilisé. Cette propriété est facilitée par la linéarité entre les valeurs du champ et les valeurs nodales. Par exemple, pour la méthode FEM, pour $\mathbf{x}, \mathbf{y} \in \Omega$, on a

$$\begin{aligned} \mathbb{E}[u(\mathbf{x})] &= \sum_{i=1}^n \mathbb{E}[u_i] \phi_i(\mathbf{x}), \\ \text{Cov}(u(\mathbf{x}), u(\mathbf{y})) &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_i, u_j) \phi_i(\mathbf{x}) \phi_j(\mathbf{y}). \end{aligned}$$

Ainsi, travailler sur les statistiques des valeurs discrétilisées est équivalent à travailler sur celles du champ. Les méthodes décrites en Section 2.1 peuvent donc être appliquées directement aux valeurs nodales ou au centre des éléments.

3.1.3 Cas lagrangien

Dans la Section 2.2.2, nous avons étudié les méthodes lagrangiennes continues qui discrétilisent le champ de manière similaire comme

$$u(\mathbf{x}, t) \approx \sum_{p \in \mathcal{P}} \Gamma_p(t) \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p(t))$$

mais où les fonctions de forme évoluent au cours du temps. En particulier, elles sont paramétrées par la position des particules $\mathbf{x}_p(t)$ de \mathcal{P} qui comme décrit en 2.4 varie au cours du temps.

Cas où les positions sont indépendante de l'événement

Lorsque la position des particules $\mathbf{x}_p(t)$ est indépendante de l'événement, c'est-à-dire identique pour tous les membres, nous obtenons une méthode d'interpolation utilisant des fonctions de base radiale (RBF)[Fornberg and Flyer, 2015]. Dans ce cas, si ϕ est une fonction strictement positive et si les points \mathbf{x}_i sont distincts, la solution discrétilisée u_h est définie de manière unique par les valeurs nodales Γ_p . De façon similaire aux méthodes eulériennes, l'équation 3.2 pour $\mathbf{z} = \{\Gamma_p\}_{p=1}^{N_p}$.

Expression générale Dans le cas générale chaque membre dispose de son propre ensemble particulaire.

Cependant, grâce à la formulation 3.1, il est possible de réécrire la mise à jour à partir des champs prédit u_l et de les décomposer de la manière suivante

$$\begin{aligned} u_l^a &= \sum_{p \in \mathcal{P}_l^f} \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) + \sum_{k=1}^N F_{kl} \sum_{p' \in \mathcal{P}_k^f} \Gamma_{p'}^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_{p'}) \\ &= \sum_{p \in \mathcal{P}_l^f} \left[(1 + F_l) \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) \right] + \sum_{j \neq i} \sum_{p' \in \mathcal{P}_k^f} F_{kl} \Gamma_{p'}^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_{p'}) \\ &= \sum_{k=1}^N \sum_{p \in \mathcal{P}_k^f} \Gamma_p^a \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p). \end{aligned}$$

où u_l^a est bien une solution particulière.

Sans perte de généralité, en supposant que chaque particule a des positions \mathbf{x}_p distinctes, l'espace d'approximation est de dimension $\sum_{i=1}^N \text{Card}(\mathcal{P}_i^f)$.

Ainsi, à chaque étape d'assimilation de données, le nombre de particules nécessaires pour approcher la solution analysée augmente exponentiellement avec le nombre total de particules. Par conséquent, cette mise à jour ne peut être mise en pratique qu'en utilisant des méthodes de réduction du nombre de particules.

Mise à jour sur les variables de discréétisation Il est néanmoins toujours possible de définir complètement l'état du champ en utilisant

$$\mathbf{z}(t) = \{(\mathbf{x}_p(t), \boldsymbol{\Gamma}_p(t))\}_{p=1}^{N_p}$$

Lorsque le nombre de particules $\text{Card}(\mathcal{P}) = N_p$ est indépendant de l'événement, de manière analogue on est tenté d'écrire

$$\begin{cases} \boldsymbol{\Gamma}_l^a = \boldsymbol{\Gamma}_l^f + \sum_{k=1}^N F_{kl} \boldsymbol{\Gamma}_k^f \\ \mathbf{X}_l^a = \mathbf{X}_l^f + \sum_{k=1}^N F_{kl} \mathbf{X}_k^f. \end{cases} \quad (3.2)$$

où \mathbf{X}_l est le vecteur des positions pour le membre l . Ce choix de définition de l'état a notamment été utilisé avec la méthode vortex [Darakananda et al., 2018; Le Provost and Eldredge, 2021]. De la même manière on retrouve des approches semblables avec la méthode DEM [Chen et al., 2022].

Agir ainsi présente cependant certaines limites. Premièrement, cela suppose la possibilité d'ordonner de manière unique des ensembles de particules de même taille. La mise à jour

devient alors sensible à la permutation. Pour un grand nombre de particules, l'affectation entre elles peut être arbitraire et influencer les résultats de l'analyse. De plus, deux particules proches à l'instant t peuvent se retrouver très éloignées en un temps fini. Une solution pour surmonter les limitations des méthodes d'assimilation de données dans les méthodes lagrangiennes continues consiste à agglomérer les particules proches entre elles. Cette stratégie a été explorée dans des travaux tels que ceux de Chen et al.[[Chen et al., 2022](#)] et de Darakananda et al.[[Darakananda et al., 2018](#)], où chaque nouveau cluster de particules est traité comme une entité distincte. Cependant, cette approche se fait au détriment du détail de la discréétisation, car elle simplifie la représentation du champ en regroupant les particules plutôt qu'en les traitant individuellement. Enfin, réaliser une combinaison linéaire sur des positions pour définir la mise à jour est en soi problématique. En effet, les hypothèses du filtre de Kalman d'Ensemble rappelées en Section 2.1.4 d'où découle la formule de mise à jour ne sont pas vérifiées. En particulier, l'hypothèse d'approximation du modèle tangent par l'ensemble ni la linéarisation du modèle d'observation par rapport à la variable de position. En effet, contrairement à l'amplitude, les variables d'observation sont généralement non-linéaire par rapport à \mathbf{x}_p à cause de la fonction de lissage.

Pour illustrer cet effet, nous considérons des champs u_i défini sur le cercle unité par des ensembles de particules de même intensité régulièrement espacées comme illustré dans la Figure 3.1a. Chaque particule est numérotée avec une incertitude sur l'angle de position par rapport à la première particule $\theta \sim \mathcal{U}([0, 2\pi])$. Un ensemble de champs $\{u_i\}_{i=1}^N$ sont générés, et nous notons les vecteurs d'état associés $\left\{z_i = \{\mathbf{x}_{ip}, \Gamma_{ip}\}_{ip \in \mathcal{P}}\right\}_{i=1}^N$. Nous estimons la moyenne du champ soit directement à partir de l'ensemble $\{u_i\}_{i=1}^N$ avec l'estimateur e_1 , ou bien à partir de la distribution de \mathbf{z} par substitution avec e_2 tel que

$$\begin{aligned} e_1(\mathbf{x}) &= \frac{1}{N} \sum_{i=1}^N u_i(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \sum_{p \in \mathcal{P}} \Gamma_{ip} \phi_\varepsilon(\mathbf{x} - \mathbf{x}_{ip}), \\ e_2(\mathbf{x}) &= \sum_{p \in \mathcal{P}} \mathbb{E}(\Gamma_p) \phi_\varepsilon(\mathbf{x} - \mathbb{E}(\mathbf{x}_p)) = \frac{1}{N} \sum_{i=1}^N \sum_{p \in \mathcal{P}} \Gamma_{ip} \phi_\varepsilon(\mathbf{x} - \bar{\mathbf{x}}_p), \end{aligned}$$

où $\bar{\mathbf{x}}_p = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_{jp}$ est la moyenne pondérée pour la position p .

L'estimateur e_1 , illustré Figure 3.1c préserve les propriétés de l'ensemble, tandis que l'estimateur e_2 représenté Figure 3.1b, tend à concentrer les particules vers le centre du cercle. Lorsque $N \rightarrow \infty$, la position moyenne $\mathbb{E}(\mathbf{z}) = \mathbf{0}$.

En conclusion, l'application directe de la mise à jour du filtre EnKF par combinaison linéaire des positions de discréétisation peut entraîner des résultats biaisés. Cela est principalement dû au fait que les hypothèses du filtre EnKF ne sont pas respectées, en raison de la non-linéarité entre le champ u et \mathbf{z} induite par le noyau de lissage ϕ_ε . Par conséquent, pour mettre à jour les positions des particules de manière plus appropriée, une approche variationnelle apparaît comme une solution plus adaptée.

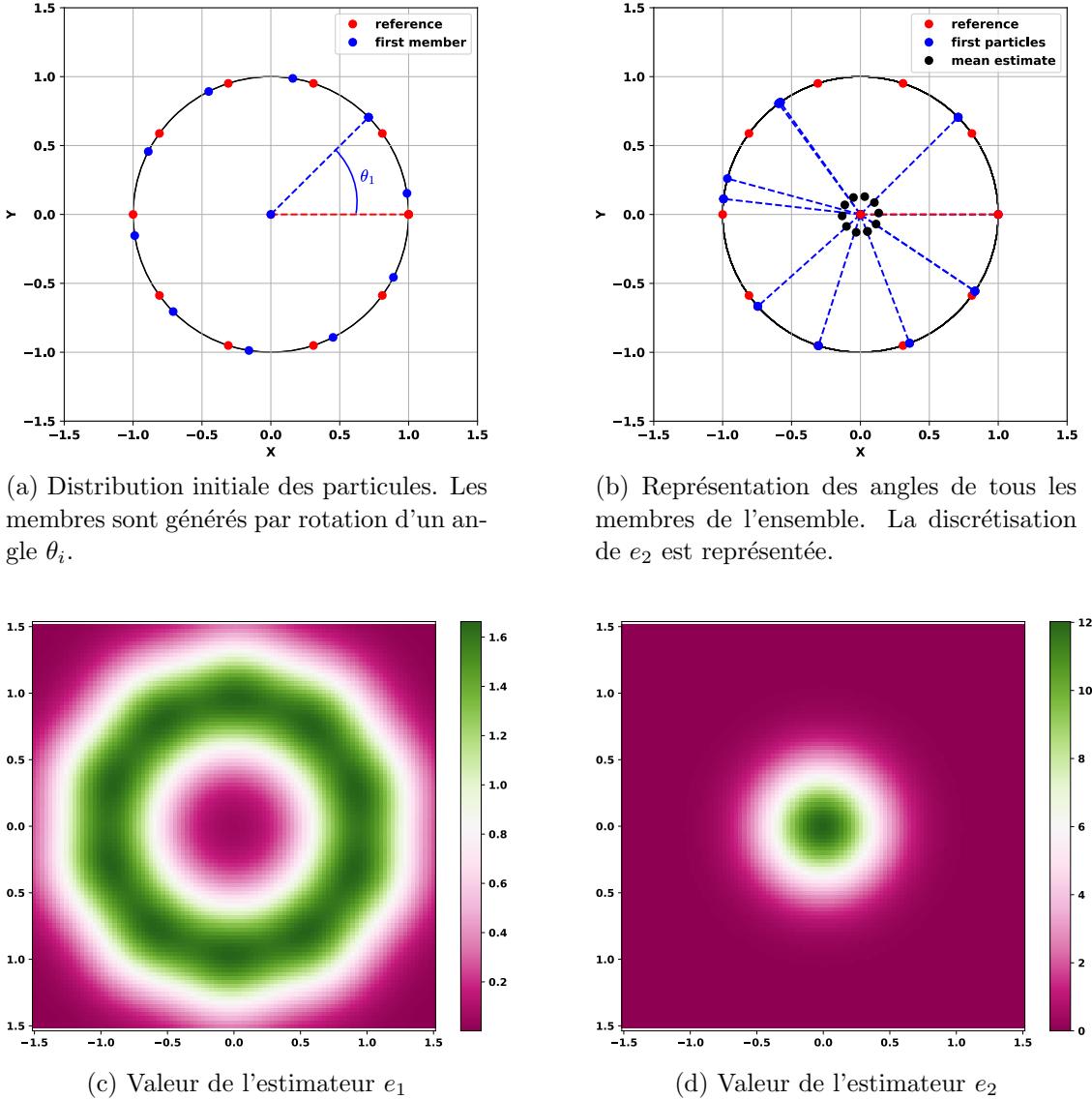


Figure 3.1: Illustration d'un problème où la définition de l'état uniquement par les positions des particules produit un estimateur loin de tous les membres.

3.1.4 Première contribution - Adaptation du filtre EnKF

Notre première contribution vise à répondre à la question suivante : comment définir la mise à jour de Kalman avec des membres discréétisés sur des ensembles particulaires différents, sans pour autant accroître exponentiellement le nombre de particules ?

En réalité, les méthodes de simulation lagrangiennes ne sont pas les seules à présenter des défis liés à une discréétisation évoluant au cours du temps. Ce problème se retrouve également dans les méthodes utilisant des maillages adaptatifs. Une solution souvent employée consiste à définir une discréétisation de référence commune à tous les membres

de l'ensemble. Que ce soit à travers des techniques de multirésolution (MRA) ou avec des méthodes de maillage mobile, les états sont projetés sur une grille de référence servant à la mise à jour des états [Siripatana et al., 2019; Bonan et al., 2017]. Une fois la mise à jour effectuée, la solution est réinterpolée sur le maillage initial ou sur une nouvelle discréétisation adaptée.

Remesh-EnKF En nous inspirant des méthodes de maillage adaptatif présentées précédemment, nous proposons une approche permettant de définir les membres sur une discréétisation commune. Les opérateurs d'affectation, tels que ceux définis en Section 2.2.2, sont utilisés pour remailler la discréétisation particulière. Ainsi, nous pouvons nous ramener au cas décrit en Section 3.1.3.

Le schéma de mise à jour devient alors

$$u_i^f(x) = \sum_{p \in \mathcal{P}_i^f} \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) \rightarrow u_i^a(x) = \sum_{q \in \mathcal{P}^{\text{reg}}} \Gamma_q^a \phi_\varepsilon(\mathbf{x} - \mathbf{x}_q),$$

où P^{reg} est une nouvelle configuration de particules commune à tous les membres après remaillage.

Part-EnKF Une seconde approche consiste à approcher le champ analysé u_l^a de manière continue comme défini en 3.2 en mettant à jour uniquement les intensités de la discréétisation du membre associé. Cette méthode conserve autant que possible la représentation lagrangienne de la solution à la fin de l'étape de propagation. Le schéma de mise à jour devient alors

$$u_i^f(x) = \sum_{p \in \mathcal{P}_i^f} \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) \rightarrow u_i^a(x) = \sum_{p \in \mathcal{P}_i^f} \Gamma_p^a \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p).$$

Dans ces deux premières formulations, nous travaillons avec un support de particules donné, qu'il soit défini sur une grille régulière ou bien issu de l'étape de propagation, en privilégiant une correction sur les intensités lors de l'analyse.

3.2 Alignement pour la correction de la discréétisation particulière

Les approches précédentes permettent d'adapter le filtre EnKF aux simulations lagrangiennes. L'assimilation s'appuie principalement sur des corrections d'intensité, soit en générant une correction d'intensité sur la configuration avant l'assimilation, soit en définissant une nouvelle discréétisation indépendante pour chaque membre. Cependant, dans ces approches, les positions des particules qui discréétisent la solution u_i ne sont pas directement mises à jour au cours de la procédure d'assimilation de données.

Cela soulève une question importante : est-il nécessaire de corriger la discréétisation particulière elle-même, c'est-à-dire les positions des particules, en plus des intensités ?

Dans la Section 2.2.2, nous avons vu dans les différents exemple que les position étaient mise à jour par l'intégration du champ de vitesse. Pour la méthode Vortex elle est définie pour une particule p comme $\frac{d\mathbf{x}_p(t)}{dt} = \mathbf{u}(\mathbf{x}_p, t; \mathcal{P}(t))$ où $\mathcal{P}(t)$ est la discréétisation particulaire à l'instant t . Le champ de vitesse étant connu avec incertitude, il impacte directement les trajectoires des particules.

En corrigent les positions des particules, l'objectif est de réduire l'effet des erreurs liées à l'intégration du champ de vitesse sur la trajectoire des particules. Cela permet d'améliorer la précision globale de la simulation en limitant la propagation d'erreurs dans les étapes suivantes de la modélisation.

3.2.1 Assimilation de données par déplacement ou alignement

Si la prise en compte des corrections pour les positions de la discréétisation pose des difficultés, c'est également en raison du fait que les méthodes d'assimilation sont fondamentalement des techniques de correction d'intensité. Cela découle de la nature des mesures d'erreur qui reposent sur des normes L_2 . Cette approche présente de nombreuses limitations, particulièrement lorsque des erreurs de position sont présentes. En effet, lorsque l'erreur est mesurée à l'aide d'une norme quadratique, les erreurs dues à un mauvais alignement sont susceptibles d'être surévaluées.

Pour illustrer ce phénomène, la Figure 3.2 montre que l'erreur due à un mauvais alignement est excessivement pénalisée par rapport à l'erreur associée à une fonction nulle. Ce phénomène est qualifié d'effet de *double pénalisation*, car il se manifeste tant dans l'évaluation de l'erreur sur le modèle que dans celle des observations [Amodei and Stein, 2009]. Cette double pénalisation contribue de manière significative à l'erreur de représentativité, comme l'ont souligné Jancić et al. [Janjić et al., 2018].

Une seconde limitation concerne la nécessité pour les champs d'état et d'observation a priori de recouvrir la solution à analyser. En effet, les méthodes d'assimilation de données classiques reposent principalement sur une interpolation dans l'espace des valeurs des champs, ce qui limite l'analyse au support de l'état de fond et à celui de l'observation. Cette contrainte est particulièrement évidente dans le cadre du filtre EnKF, qui se définit précisément comme une combinaison des membres de l'ensemble, comme décrit en Section 2.1.2. Ainsi, cette limitation est directement liée à notre problème de support de discréétisation particulaire.

Pour surmonter cette difficulté, plusieurs approches ont été proposées. Une méthode élégante consiste à s'inspirer des techniques de transport optimal pour définir une correction dans un espace d'interpolation plus riche, prenant en compte à la fois les intensités et les déplacements [Villani et al., 2009; Benamou and Brenier, 2000]. Cela implique de remplacer la norme quadratique par une norme de Wasserstein. Par exemple, les travaux de thèse de Feyeux [Feyeux, 2016] ont exploré l'utilisation de la distance de Wasserstein pour l'assimilation de données issues d'images. D'autre part, Bocquet et al. [Bocquet et al., 2024] ont adapté la méthode 3DVar au transport optimal, en appliquant cette approche aux distributions d'état et d'observation avec des masses

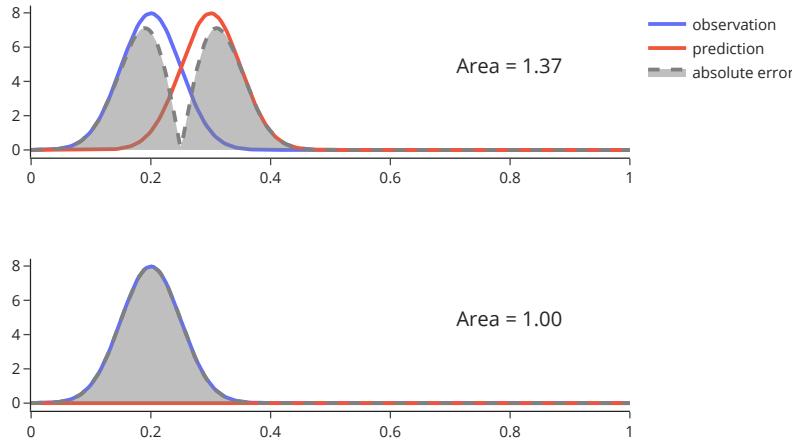


Figure 3.2: Visualisation de l’effet de double pénalisation. Un décalage de la solution entraîne une mesure de l’erreur plus élevée que celle associée à la fonction nulle.

potentiellement différentes. Bien que les algorithmes de transport optimal aient gagné en efficacité computationnelle [Cuturi and Doucet, 2014; Peyré and Cuturi, 2019; Simsekli et al., 2018], leur application aux méthodes d’assimilation de données est encore à ses débuts.

Cependant, plusieurs développements récents introduisent des transformations spatiales pour améliorer l’ajustement des modèles. Par exemple, Percival et al. [Percival, 2008] utilisent des techniques de réarrangement pour appliquer des transformations de coordonnées. Pour des problèmes avec des caractéristiques cohérentes telles que les ouragans, les orages, les fronts de feu, les lignes de grains, et les fronts de pluie, il est crucial d’ajuster l’état du modèle numérique en déformant la grille ainsi qu’en corrigeant les amplitudes de manière additive. En 2007, Ravela et al. [Ravela et al., 2007] ont proposé un modèle d’ajustement conjoint de position et d’amplitude basé sur des ensembles, dérivant une approximation séquentielle applicable tant à l’EnKF qu’à d’autres formulations. Leur approche ne repose pas sur l’hypothèse d’indépendance ou de normalité conjointe des erreurs d’amplitude et de position, contrairement à d’autres méthodes. Cette méthode a également été adaptée avec succès dans des formulations multi-échelles [Ying, 2019; Ying et al., 2023]. Enfin, Rosenthal et al. [Rosenthal et al., 2017] ont développé une méthode séquentielle en deux étapes pour aligner et corriger les intensités successivement. Leur objectif est de préserver les propriétés morphologiques des tourbillons en appliquant d’abord une transformation cinématiquement admissible pour corriger la position, suivie d’une correction des intensités. Cette approche offre également l’avantage de pouvoir être couplée aux méthodes classiques d’assimilation par correction d’intensité.

3.2.2 Seconde contribution - correction de position et d'intensité

L'erreur liée à l'intégration du champ de vitesse entraîne une erreur de position dans les discrétisations particulières, pouvant rendre le support de particules inadapté pour représenter précisément la solution analysée. En définissant la correction uniquement dans l'espace des intensités des particules, comme proposé dans les approches précédentes, il est possible que l'espace des solutions ne soit pas suffisamment riche pour représenter correctement la solution analysée. Cette configuration peut ne pas être admissible au sens de la Section 2.2.2, notamment en l'absence de recouvrement entre tous les membres. Pour y remédier, la contribution détaillée au Chapitre 5 permet de mettre à jour le support des particules lors de l'étape d'analyse pour prendre en compte cette source d'erreur. Nous faisons l'hypothèse que cette erreur peut être corrigée par une procédure d'alignement. Cette idée s'inspire des travaux existants sur l'assimilation des données avec ajustement spatial, tels que ceux de [Ravela et al., 2007] et [Rosenthal et al., 2017], et repose sur une méthode de correction particulaire basée sur une approche variationnelle, capable de mettre à jour successivement la position et l'intensité des particules. Le schéma de mise à jour devient alors

$$u_i^f(x) = \sum_{p \in \mathcal{P}_i^f} \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) \rightarrow \sum_{p \in \mathcal{P}_i^f} \Gamma_p^f \phi_\varepsilon(\mathbf{x} - \Psi(\mathbf{x}_p)) \rightarrow \sum_{p \in \mathcal{P}_i^f} \Gamma_p^a \phi_\varepsilon(\mathbf{x} - \Psi(\mathbf{x}_p))$$

où Ψ est une transformation spatiale à définir et Γ_p^a l'intensité corrigée.

En particulier, nous cherchons à ce que cette approche respecte les hypothèses physiques du modèle sous-jacent. Par exemple, dans le cas de la méthode vortex, il est essentiel que le schéma de mise à jour tienne compte des principes d'un écoulement incompressible. Cette formulation consiste en une méthode d'ajustement complet simultanément sur les positions et les intensités des particules, permettant ainsi une meilleure représentation des solutions analysées, même dans des contextes où l'erreur sur les positions peut affecter de manière significative les résultats de l'assimilation.

3.3 Bilan du chapitre

Les méthodes lagrangiennes présentent la particularité d'utiliser des discrétisations qui évoluent au cours du temps, rendant l'application directe des méthodes d'ensemble, comme l'EnKF, plus complexe sur les configurations particulières. En effet, les positions et les intensités des particules ne peuvent être traitées de manière uniforme lors de l'assimilation de données.

Cette observation nous a conduit à proposer plusieurs contributions pour aborder ce problème. Nous avons ainsi développé des approches distinctes pour corriger soit les intensités, soit les positions des particules. Ces contributions sont traitées en détail dans les chapitres 4, qui se concentre sur la correction des intensités, et 5, qui explore la mise à jour des positions à l'aide de méthodes variationnelles d'alignement.

Adaptations du filtre EnKF pour des simulations lagrangiennes par correction d'intensité

Contents

4.1 Résumé	70
4.2 Abstract	70
4.3 Introduction	71
4.4 Background	73
4.4.1 Data assimilation	73
Data assimilation setting	73
Bayesian filtering	74
Ensemble Kalman Filter	75
4.4.2 Particle-based methods	76
Particle discretization	76
4.4.3 Particle-based function manipulations	77
Approximation operator	77
Regression operator	77
Remeshing operator	78
4.5 Methods	80
4.5.1 Remesh-EnKF	81
4.5.2 Particles-EnKF	82
4.6 1D density advection-diffusion problem	85
4.6.1 Description of the problem	85
4.6.2 Assimilation parameters and ensemble generation	86
4.6.3 Results	87
4.7 2D vortex-in-cell problem	91
4.7.1 Description of the method	91
4.7.2 Lamb-Chaplygin dipole and simulation parameters	91
4.7.3 Assimilation parameters and ensemble generation	93
4.7.4 Results	95
Error through time	95

Error with respect to assimilation parameters	96
Error with respect to simulation parameters	96
4.8 Conclusion	100
4.A Stochastic Ensemble Kalman Filter	101
4.B Moment conservation of particle discretization	102
4.C Parameters	103
4.C.1 One dimension problem	103
4.C.2 Two dimension problem	103

4.1 Résumé

Cette étude propose de nouvelles approches pour adapter la méthode d'assimilation de données par filtre de Kalman d'ensemble(EnKF) pour les simulations lagrangiennes continues. Bien que les méthodes d'assimilation de données aient été appliquées aux simulations eulériennes depuis longtemps, elles sont très peu utilisées dans le contexte d'une discréttisation de solution lagrangienne. Une des limitations majeures réside dans le fait que chaque membre est défini sur une discréttisation qui lui est propre et évolue au cours du temps. Deux méthodologies spécifiques sont introduites pour adapter l'analyse du filtre EnKF. La première repose sur l'utilisation d'une transformation eulérienne intermédiaire combinant un processus de projection et de remaillage d'une grille régulièrre de particules. La seconde est un schéma purement lagrangien applicable lorsque la génération d'une grille de particules n'est pas souhaitée. Ces méthodes sont évaluées à l'aide d'un modèle unidimensionnel d'advection-diffusion avec des conditions aux limites périodiques. Par la suite, les schémas d'assimilation sont appliqués à un problème d'écoulement incompressible non-linéaire bidimensionnel, résolu par la méthode *Vortex-In-Cell*. Dans le scénario unidimensionnel, la performance de ces filtres est comparée à celle d'un filtre d'assimilation basé sur grille. Dans le cas bidimensionnel, l'étude démontre la faisabilité de l'application de ces méthodes dans des scénarios plus complexes.

4.2 Abstract

This study presents a novel approach for integrating data assimilation techniques into particle-based simulations using the Ensemble Kalman Filter. If data assimilation methods have been applied to Eulerian simulations for a long, they have never been properly used in the context of a Lagrangian solution discretization. Two specific methodologies are introduced to complete the analysis. The first one is based on the use of an intermediary Eulerian transformation combining a projection and a remeshing process. The second is a purely Lagrangian scheme that is applicable when remeshing is not adapted. These methods are evaluated using a one-dimensional advection-diffusion model with periodic boundaries. Subsequently, assimilation schemes are applied to a non-linear two-dimensional incompressible flow problem, solved via the Vortex-In-Cell method. In

the one-dimensional scenario, the performance of these filters is benchmarked against a grid-based assimilation filter. In the two-dimensional case, the study demonstrates the feasibility of applying these methods in more intricate scenarios.

4.3 Introduction

Numerical simulation enables the assessment of complex real-world systems, for instance, to facilitate the optimization of complex systems and perform risk analysis, all while reducing experimental costs. Thanks to the increasing computational resources, they help in understanding and designing processes, particularly in the mechanical field. The solid and fluid mechanics historically leaned on grid-based or mesh-based methods. These techniques necessitate the use of structured meshes. The shift towards meshless methods offers significant promises for complex physics or large deformations (moving interfaces, material disintegration, or distortion) to avoid computing complex geometries.

Meshless methods, specifically particle-based methods, describe geometry as a collection of particles that move with the deformation flow in a Lagrangian fashion. Each particle transports material properties and internal variables. Particles can discretize a continuum medium and are associated with a kernel to reconstruct continuous fields and differential operators. In this article, we will mainly focus on the Vortex Method [Cottet and Koumoutsakos, 2000a; Mimeau and Mortazavi, 2021], which discretizes the vorticity field and solves the incompressible fluid flow equation with the vorticity-stream function formulation.

The computed solution may contain errors that need to be understood, quantified, and minimized. If observations are available, integrating this information can lead to a more accurate estimation of the simulation state. In this context, data assimilation techniques offer an optimal way to combine various sources of information, resulting in a more precise estimation of the system state. Integrating model predictions and observational data has been widely applied in disciplines such as meteorology, oceanography, hydrology, and geosciences [Bocquet, 2014].

In the domain of data assimilation, two prominent families of approaches have emerged: variational and stochastic methods. Variational approaches [Sasaki, 1958] focus on minimizing a cost function that measures the misfit between model predictions and observations, seeking the optimal system state. The most common formulations derive from 3DVar, 4DVar [Talagrand, 1997].

On the other hand, stochastic approaches go beyond mere state estimation; they delve into the quantification of uncertainty associated with the estimated states. Uncertainty quantification is a critical aspect, especially in dynamic and uncertain systems, where acknowledging and characterizing becomes paramount for reliable decision-making and model improvement. In this case, the estimate is sequentially updated based on previous and current observations. The assimilation process is performed through a Bayesian

framework with a forecast and an analysis step. The Kalman filter [Kalman, 1960] is an example of a sequential formulation considering a linear model and Gaussian distribution assumptions. However, more advanced filters have been introduced to be adapted to nonlinear and arbitrary distributions. One of the most popular Bayesian filters is undoubtedly the Ensemble Kalman Filter, introduced by Evensen [Evensen, 1994], primarily because of its adaptability to high-dimensional problems with any evolution model and its remarkable resilience to deviations from the initial Gaussian assumptions. It consists of approximating the probability distribution of a state thanks to an ensemble of simulations called particles or members.

This paper introduces new approaches to applying ensemble data assimilation techniques to meshless simulations that discretize a continuum domain. The hypothesis considers several members of different particle distributions. The Ensemble Kalman Filter (EnKF) has been extensively employed for Eulerian discretization frameworks. However, its application in the Lagrangian approach presents unique challenges. These issues primarily revolve around defining a unified state representation across all ensemble members and effectively updating this state during the analysis phase.

When particle-based methods discretize a field on a continuum discretization, the particles are point entities and thus allow a certain flexibility to the update. Operations such as agglomeration, splitting, or resampling are utilized to update particle configurations, primarily to mitigate issues like distortion, excessive deformation or to manage particle count [Y. et al., 2015; Cottet, Ould Salihi, et al., 1999].

Nevertheless, the crux of the challenge lies in the inherent disparity in discretization across different ensemble members. The first solution is to consider a reference discretization for all members. In fixed-grid methodologies with Multi-Resolution Analysis (MRA) and moving mesh scenarios, the state definition on varied grids with assimilation is managed through projection and interpolation techniques to establish a reference grid for state updates [Siripatana et al., 2019; Bonan et al., 2017]. The selection of the reference and updated grids provides a spectrum of implementation possibilities. Furthermore, Siripatana et al. [Siripatana et al., 2019] highlight that the EnKF correction is contingent solely upon the predictions and observations, thereby rendering it independent of the state definition.

Another solution consists of defining the state with the union of the particles, considering the position and associated intensities of each particle [Darakananda et al., 2018]. Complex filters have been developed to estimate correctly the posterior discretization with a nonlinear observation model or with a deficient number of sensors of pressure [Le Provost et al., 2021]. However, these methods grapple with scenarios involving markedly divergent particle discretizations or highly variant model flows. In this general case, using a particle state for all particles for the update is unfeasible. Indeed, the update implies a linear combination of all members, leading to an exponential increase of particles. On the other hand, the state could be associated with the spatial field defined in a functional space. The updated fields could be evaluated on the entire domain. Finally, using approximation or regression, a new particle discretization could be approximated.

These final modifications have already been introduced in the Vortex Method to better approximate the vorticity field by changing the particle intensities regroup under the label Meshless Rezoning Methods in [Mimeau and Mortazavi, 2021]. It mainly involved iterative methods [Beale, 1988], triangulation [Russo and Strain, 1994] or Radial Basis Function (RBF) interpolation [Barba, 2004; Sperotto et al., 2022]. RBF offers to introduce new particles or introduce penalization to regularize optimization problems.

Based on those different tools, we propose two novel EnKF-based filters. First, the Remesh-EnKF uses an Eulerian intermediate update. This way, the analysis is defined on an Eulerian framework, and the method is reduced to the previous development of data assimilation. The projection step is something familiar for Particle-In-Cell methods [Sulsky et al., 1994; Brackbill et al., 1988]. Nevertheless, it involves remesh the discretization entirely on a regular grid of particles. This method is based on the regridding of the particle discretization as described by [Cottet, Ould Salihi, et al., 1999]. Then, in a case where the particle discretization would be preserved, the Particle-EnKF is introduced. In this case, the analyzed field is approximated on the previous particle discretization of each members. The particle positions are unchanged; only the strengths are modified by regression or approximation of the analyzed solution. In the next part, background on sequential filtering and EnKF algorithm will be introduced 4.4.1, then on particle-based methods 4.4.2. Then, the two categories of method will be described in section 4.5. Afterward, those filters will be compared with a grid-based filter in a 1D Advection-Diffusion problem in section 4.6, and an incompressible viscous flow is solved using a Vortex Method 4.7 where the filters are quantitatively analyses.

4.4 Background

4.4.1 Data assimilation

Data assimilation could be generally formulated with a probabilistic framework. It allows us to rigorously deal with measurement and model error in order to not only deduce an estimate of the real state but also associate uncertainty. Thus, state and observation are modeled as random variables. A filtering approach is then applied to estimate the current state based on past observations sequentially.

The goal is to establish the recurrence in probability distributions that, through Bayesian estimation, will enable us to estimate the current state and predict the future state.

Data assimilation setting

A hidden Markov chain is used to model this recurrence. We position ourselves within a finite-dimensional context. The forecast and observation are introduced, such as for $k \geq 0$,

$$\begin{cases} \mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \boldsymbol{\eta}_{k+1}, \\ \mathbf{y}_{k+1} = \mathcal{H}_{k+1}(\mathbf{x}_{k+1}) + \boldsymbol{\varepsilon}_{k+1}, \end{cases}$$

where \mathcal{M}_{k+1} is the model operator describing the time evolution of the state from time k to time $k + 1$, and \mathcal{H}_k is the observation operator. The term $\mathbf{x}_k \in \mathbb{R}^n$ is the vector state at time k and $\mathbf{y}_k \in \mathbb{R}^m$ the observation vector, $\boldsymbol{\eta}_k$ is the model error that accounts for error in the numerical model and the errors due to discretization, and $\boldsymbol{\varepsilon}_k$ is the observation error which combine measurement error and representativeness error. We assume that $\boldsymbol{\eta}_k$, $\boldsymbol{\varepsilon}_k$ are random variables following Gaussian distributions with zero mean and covariance matrices \mathbf{Q}_k and \mathbf{R}_k respectively. Finally, we assume that the observation and the model errors are independent though the time and that initial error on \mathbf{x}_0 , $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are mutually independent. Let $\mathcal{D}_k = \{\mathbf{y}_l\}_{l=1}^k$ represent the accumulated data up to time k . Thus, \mathbf{x}_{k+1} and \mathcal{D}_k are conditionally independent with respect to \mathbf{x}_k , as well as \mathbf{y}_{k+1} and \mathbf{x}_{k+1} , leading to simplifications in the next paragraph.

Bayesian filtering

The filtering problem consists of assessing the current state of the signal by utilizing data observation up to the present moment. Filtering involves the determination of $p(\mathbf{x}_k | \mathcal{D}_k)$, the probability density function associated with the probability measure on the random variable $\mathbf{x}_k | \mathcal{D}_k$. Specifically, filtering focuses on the sequential updating of this probability density function as the index k is incremented. The state density is initialized by the a priori density of the initial state p_{x_0} . Then, for all $k \geq 0$, probability distributions are propagated. The forecast step is obtained through the law of total probability

$$p(\mathbf{x}_{k+1} | \mathcal{D}_k) = \mathbb{E}_{\mathbf{x}_k} [p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathcal{D}_k)] = \mathbb{E}_{\mathbf{x}_k} [p(\mathbf{x}_{k+1} | \mathbf{x}_k)].$$

The a priori law of the $k + 1$ observations can be obtained again through the law of total probability

$$p(\mathbf{y}_{k+1} | \mathcal{D}_k) = \mathbb{E}_{\mathbf{x}_{k+1}} [p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) | \mathcal{D}_k].$$

After the $k + 1$ observation of the random variable \mathbf{y}_{k+1} , the analysis step determines the a posteriori law of the state using Bayes law

$$p(\mathbf{x}_{k+1} | \mathcal{D}_{k+1}) = p(\mathbf{x}_{k+1} | \mathbf{y}_{k+1}, \mathcal{D}_k) = \frac{p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}, \mathcal{D}_k)p(\mathbf{x}_{k+1} | \mathcal{D}_k)}{p(\mathbf{y}_{k+1} | \mathcal{D}_k)}.$$

Due to the independence hypothesis the formula reduced to

$$p(\mathbf{x}_{k+1} | \mathcal{D}_{k+1}) = \frac{p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1})p(\mathbf{x}_{k+1})}{p(\mathbf{y}_{k+1})} \propto p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1})p(\mathbf{x}_{k+1}).$$

This finally lead to a mapping from the prior $p(\mathbf{x}_{k+1} | \mathcal{D}_k)$ to the posterior $p(\mathbf{x}_{k+1} | \mathcal{D}_{k+1})$. We remove the time subscript k in the rest of the section for simplicity and present the forecast and analysis step for a time increment.

Ensemble Kalman Filter

The Kalman filter [Kalman, 1960] is a Bayesian filter that, in addition to the previously mentioned assumptions, requires \mathcal{M}_k and \mathcal{H}_k to be linear operators. In this case, the posterior distribution of the state is still Gaussian, so only the mean and the variance are updated. The Kalman estimator is thus a recursive version of the Minimum Mean Square Error applied to the Gaussian Linear model.

The ensemble Kalman Filter (EnKF) is a data assimilation method adapted to high dimensional non-linear problems introduced by Evensen [Evensen, 1994]. The formulation uses an ensemble of discrete samples based on the assumptions of a multivariate Gaussian distribution, such as the Kalman filter. We present the stochastic EnKF, where the observations are perturbed to account for observation errors.

Assuming we have an ensemble of N states $\{\mathbf{x}_i\}_{i=1}^N$, we forecast the ensemble by propagating each state with the dynamic model and obtain a forecast ensemble. The two first moments of the error are given by

$$\begin{aligned}\bar{\mathbf{x}}^f &= \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^f, \\ \mathbf{P}^f &= \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)^T,\end{aligned}$$

where $\bar{\mathbf{x}}^f$ and \mathbf{P}^f are respectively the empirical estimates of the mean and the covariance matrix of the state distribution obtained from the ensemble members.

Likewise, the mean and covariance of the observation $\{\mathcal{H}(\mathbf{x}_i^f)\}_{i=1}^N$ could be estimated as well as the covariance between state and observation.

We develop the general formulation of the EnKF filter in the Appendix 4.A.

Finally, our formulation of EnKF takes advantage of a correction of the state defined in the member space. We define \mathbf{F} , the correction matrix that gives the update in terms of linear combinations of the forward states

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{X}^f \mathbf{F}, \quad (4.1)$$

where the matrix \mathbf{F} only depends on the ensemble members through the predicted observations ensemble $\{\mathcal{H}(\mathbf{x}_i^f)\}_{i=1}^N$, the observation \mathbf{y} and the associate perturbations $\{\boldsymbol{\varepsilon}_i\}_{i=1}^N$

This independence is made possible by the linearization of the observation operator and the low-rank approach. The same property is verified for the ensemble transform Kalman filter [Bishop et al., 2001]. Indeed, the filter is explicitly defined in the perturbation space.

4.4.2 Particle-based methods

We consider particle methods for solving continuous problems in fluid or solid mechanics. The Lagrangian methods decompose the domain on a set $\mathcal{P} = \{\mathbf{z}_p, \mathbf{U}_p\}_{p=1}^{n_p}$, where $\mathbf{z}_p \in \Omega$, the spatial coordinates a particle p and $\mathbf{U}_p \in \mathbb{R}^n$ its intensity.

We focus our work on methods that discretize a solution \mathbf{u} on a continuous domain $\Omega \subset \mathbb{R}^d$ with Ω the spatial domain. This includes methods like Smoothed particle hydrodynamics (SPH) [Gingold and Monaghan, 1977; Lucy, 1977], or the Vortex Method (VM) [Cottet and Koumoutsakos, 2000a] and is extended to other methods like the Material Point Method (MPM) [Sulsky et al., 1994].

Particle discretization

Any smooth field \mathbf{u} on Ω could be written

$$\mathbf{u}(\mathbf{z}) = \int_{\Omega} \mathbf{u}(\mathbf{z}') \delta(\mathbf{z}' - \mathbf{z}) d\mathbf{z}',$$

with δ the Dirac distribution.

A kernel function ϕ_ε is introduced to obtain an average estimate $\langle \mathbf{u} \rangle$ of \mathbf{u} such that

$$\langle \mathbf{u}(\mathbf{z}) \rangle = \int_{\Omega} \mathbf{u}(\mathbf{z}') \phi_\varepsilon(\mathbf{z} - \mathbf{z}') d\mathbf{z},$$

where ε is the smoothing length. The smooth kernel should at least respect the following properties

$$\begin{aligned} \int_{\Omega} \phi_\varepsilon(\mathbf{z}) d\mathbf{z} &= 1, \\ \phi_\varepsilon(\mathbf{z}) &\rightarrow \delta(\mathbf{z}), \quad \varepsilon \rightarrow 0, \\ \phi_\varepsilon(\mathbf{z}) &\in C^k, \quad k \geq 1, \end{aligned}$$

where the two first properties are remanent properties of the Dirac distribution and the last is a differentiability requirement.

The average function $\langle \mathbf{u} \rangle$ is then used to approximate the original function.

Finally, the original domain Ω is subdivided with N_p subdomain Ω_p associated with a Lagrangian particle in the location $\mathbf{z}_p \in \Omega_p$. We denote by V_p the volume of Ω_p . This discretization is then used to approximate the average function such that

$$u(z) \approx \sum_{p \in \mathcal{P}} \mathbf{U}_p \phi_\varepsilon(\mathbf{z} - \mathbf{z}_p). \tag{4.2}$$

Thus, any function defined on a particle discretization is defined by a particle discretization \mathcal{P} and a smoothing kernel ϕ_ε . Based on this discretization, the differential operator could be derived through this formulation.

Several kernels have been used depending on the method. Theoretically, it could be the Gaussian kernel function

$$\phi_g(\mathbf{r}) = \frac{1}{(\pi\varepsilon^2)^{d/2}} \exp(-\|\mathbf{r}\|^2/\varepsilon^2).$$

This kernel is infinitely differentiable but defined on non-compact support. In practice, we use a cutoff to remove negligible value for a large distance from a particle.

4.4.3 Particle-based function manipulations

Based on particle discretization, we present several particle manipulations that will be used in our methods. Initially, those manipulations are either dedicated to improving the quality of the approximation, avoiding high distortion by creating a new particle discretization or projecting the solution on an Eulerian configuration. The different operators will be used in the assimilation process in order to update the particle solution of each member in Section 4.5.

Approximation operator

The first category of manipulations aims to improve the approximation of the field by modifying particle strength. A first approximation could be to use the particle approximation to reevaluate the particle intensities like in Equation 4.2 such as

$$\mathbf{U}_p = \int_{\Omega_p} \mathbf{u}(z) dz = \mathbf{u}(\mathbf{z}_p) V_p,$$

where \mathbf{z}_p is the particle location.

This approximation is easily computable but does not ensure the conservation of all the moments of the field. A better approximation could be obtained using the iterative Beale's formula [Beale, 1988], which corrected circulation values in order to recover the vorticity field at the particle locations.

Regression operator

Based on regression methods, the new intensities of the particles defined $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_p]^T$ could be obtained by minimizing the quadratic error. Assume that we have some vector $\mathbf{u} = [\mathbf{u}_1(z_1), \dots, \mathbf{u}_p(z_p)]^T$ of the continuous field evaluations. The particle approximation could be computed on each particle position \mathbf{z}_p given

$$\mathbf{u} \simeq \Phi \mathbf{U},$$

where $\Phi_{ij} = \phi_\varepsilon(z_i - z_j)$.

Finding the new intensities \mathbf{U}^* corresponds to solving the previous system in the least square sense. It corresponds to the classical problem of finding the minimizer of the following quadratic function

$$\mathbf{U}^* = \arg \min_{\mathbf{U}} \|\mathbf{u} - \Phi \mathbf{U}\|_2^2.$$

In this case, the solution is $\mathbf{U}^* = (\Phi^T \Phi)^{-1} \Phi \mathbf{u}$. This problem may be ill-posed, particularly in the case of a large set of non-well-distributed particles. We choose to regularize the solution by introducing a penalization term. The Ridge regression introduces a penalization on of the form $\lambda \|\mathbf{U}\|_2^2$, where λ is a penalization coefficient, such as the new problem is

$$\mathbf{U}_{\text{ridge}}^* = \arg \min_{\mathbf{U}} \|\mathbf{u} - \Phi \mathbf{U}\|_2^2 + \lambda \|\mathbf{U}\|_2^2,$$

given the following solution $\mathbf{U}_{\text{ridge}}^* = (\Phi^T \Phi + \lambda \mathbf{I})^{-1} \Phi \mathbf{u}$.

Remeshing operator

A second type of manipulation is based this time on a complete projection of the solution on a new regular grid of particles [Cottet and Koumoutsakos, 2000a; Cottet, Ould Salihi, et al., 1999]. This method allows us to switch from a Lagrangian discretization \mathcal{P} to an Eulerian one Λ and then switch back to an entirely new regular particles discretization \mathcal{P}' that conserves as much as possible the moment of the particle discretization.

In our methodology, we propose a two-step approach. First, we execute an assignment step ((a)) to transfer the particle discretization to the grid discretization. Subsequently, an interpolation step ((b)) is performed to yield a new set of regularly spaced particles.

Our analysis pertains to the one-dimensional spatial scenario, where $\Omega \in \mathbf{R}$. The extension to the n -dimensional case can be achieved through the tensorization of the one-dimensional approach.

(a) Assignment on an Eulerian grid

We denote by z_I and z_p respectively the grid and the old particle locations. The new particles are defined on a grid of n_g elements with regular spacing $\ell_I = 2d_p$ where d_p is the characteristic size of the particles. We define the particle intensities as \mathbf{U}_p and the nodal field values as \mathbf{u}_I . By using some shape function W , the assignment step from particles to each node $I \in \Lambda$ can be written as

$$\mathbf{u}_I = \frac{1}{V_I} \sum_{p \in \mathcal{P}} \mathbf{U}_p W \left(\frac{z_I - z_p}{\ell_I} \right).$$

Where W determines a redistribution of the intensity on the grid, the new discretization can then be used to approximate the field \mathbf{u}_p , defined by the particle discretization by interpolation given.

$$\mathbf{u}_p(z) \approx \mathbf{u}_g(z) = \sum_{I \in \Lambda} \mathbf{u}_I W\left(\frac{z - z_I}{\ell_I}\right) \quad \forall z \in \Omega.$$

(b) *Interpolation on a new regular particle discretization*

A new set of particles is defined at the quarter of each cell such that the new position is defined at $z_{p'} = d_p/2 + i dp$, $i = 0, \dots, 2n_g$. The value of the field is then interpolated at that new location and multiplied with the volume of the particle $\mathbf{U}_{p'} = \mathbf{u}_g(z_{p'})V_{p'}$ in order to give a new particle approximation of the field.

$$\mathbf{u}_g(z) \approx \mathbf{u}_{p'}(z) = \sum_{p' \in \mathcal{P}'} \mathbf{u}_g(z_{p'})V_{p'},$$

The combination of these two steps can initially be utilized to generate a new undistorted particle distribution. The shape function W determines the type and quality of the transfer. The method effectiveness is evaluated by assessing the conservation of the first moments of the particle distributions, as detailed in the Appendix 4.B.

For W , one may employ the piecewise linear interpolation function, which ensures the conservation of moment 0. For higher moment conservation, the B-spline function provides a smoothing function for higher order.

However, while higher-order B-splines improve the smoothness of the solution, their accuracy is limited to the second order, allowing only exact interpolation of linear functions.

Monaghan [Monaghan, 1985] proposes a systematic approach to enhance accuracy and maintain smoothness through extrapolation. The concept involves constructing a new shape function based on a cutoff and its radial derivative. For $m = 4$, the cubic B-spline is improved by the following new interpolating kernel

$$M'_4(z) = \begin{cases} 1 - \frac{5}{2}z^2 + \frac{3}{2}|z|^3 & 0 \leq |z| \leq 1 \\ \frac{1}{2}(2 - |z|)^2(1 - |z|) & 1 \leq |z| \leq 2 \\ 0 & 2 \leq |z|. \end{cases}$$

The drawback of this method is that it does not ensure positivity. Therefore, we opt to utilize the M'_4 kernel for our implementation.

Finally, in multidimensional space, the redistribution kernel W can be obtained as the product of the one-dimensional kernel applied to each coordinate, as follows

$$\begin{aligned} \mathbf{U}_p &= \sum_{I \in \Lambda} \mathbf{U}_I W(\mathbf{z}_p - \mathbf{z}_I, \ell_I) \\ &= \sum_{I \in \Lambda} \mathbf{U}_I \prod_{i=1}^d W_{1D} \left(\frac{\mathbf{z}_{I,i} - \mathbf{z}_{p,i}}{\ell_I} \right) \end{aligned}$$

4.5 Methods

This section outlines the development of ensemble data assimilation techniques tailored for particle-based simulations. While the forward step aligns with the traditional Ensemble Kalman Filter, the primary challenge lies in the update during the analysis step. In order to be state independent, we define the analysis thanks to (4.1). This step benefits from an observation matrix-free implementation, rendering the analysis independent of state discretization. Hence, the correction matrix in the analysis \mathbf{F} relies solely on the observation \mathbf{y} , predictive observations $\{\mathbf{h}_i\}_{i=1}^{N_{\text{ens}}}$, and noise samples $\{\boldsymbol{\varepsilon}_i\}_{i=1}^{N_{\text{ens}}}$. The analysis fields are obtained at any spacial coordinates thanks to the particle approximation of each member field \mathbf{u}_i^f such as for all $\mathbf{z} \in \Omega$

$$\mathbf{u}_i^a(\mathbf{z}) = \mathbf{u}_i^f(\mathbf{z}) + \sum_j F_{ji} \mathbf{u}_j^f(\mathbf{z}) \quad i = 1, \dots, N$$

Thus, solutions are also described on a particle discretization $\mathcal{P}_i^a = \bigcup_k \mathcal{P}_k^f$. Such as

$$\mathbf{u}_i^a(\mathbf{x}) = \sum_{p \in \mathcal{P}_i^f} \mathbf{\Gamma}_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) + \sum_{j=1}^N F_{ji} \sum_{p' \in \mathcal{P}_j^f} \mathbf{\Gamma}_{p'}^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_{p'}) \quad (4.3)$$

$$= \sum_{p \in \mathcal{P}_i^f} \left[(1 + F_{ii}) \mathbf{\Gamma}_p^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p) \right] + \sum_{j \neq i} F_{ji} \mathbf{\Gamma}_{p'}^f \phi_\varepsilon(\mathbf{x} - \mathbf{x}_{p'}) \quad (4.4)$$

$$= \sum_{j=1}^N \sum_{p \in \mathcal{P}_j^f} \mathbf{\Gamma}_p^a \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p), \quad (4.5)$$

where \mathbf{u}_i^a is still defined on a particle discretization. Without loss of generality, assuming that each particle has a distinct position x_p , the approximation space has a dimension of $\sum_{i=1}^N \text{Card}(\mathcal{P}_i^f)$. Thus, at each data assimilation step, the number of particles required to approximate the analyzed solution increases exponentially with the union of the particles from all members, as shown in Figure 4.1.

To overcome this effect, we introduce two distinct EnKF-adapted filters

- The Remesh-EnKF filter 4.5.1 employs an intermediate Eulerian discretization of the field, consistent across all members. Consequently, classical EnKF can be

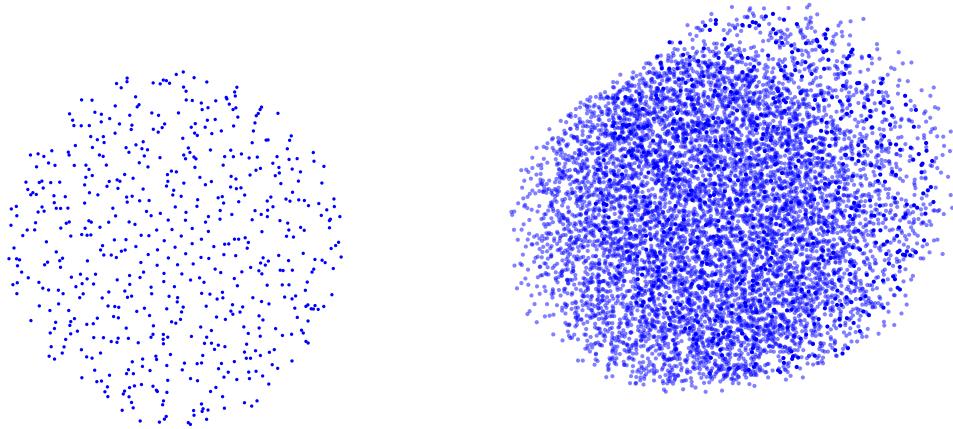


Figure 4.1: Particle support before and after the assimilation correction.

applied. Then a regular grid of particles is set to allow a constant number of particles.

- The Particles-EnKF filter 4.5.2 executes data assimilation with (4.3). Analyzed fields are approximated on each forward member discretization.

The choice of filter depends on the context, particularly regarding the feasibility of a remeshing process.

4.5.1 Remesh-EnKF

The first method consists by defining a scheme that combine an intermediate projection on a grid, to perform the assimilation, with a remeshing process to generate a new set of particles. The global scheme is build to conserve as much as possible the property of the particle discretization of the members. By regularly remeshing, we can control and limit the number of particles, thereby ensuring a more homogeneous and manageable representation of the solution. Remeshing allows for maintaining a balanced distribution of particles while preserving the essential characteristics of the original solution.

The assimilation is performed with the following step:

- **Propagation** The members are forward, given new particle set $\mathcal{P}_i^f = \{(\mathbf{z}_{ip}^f, \mathbf{\Gamma}_{ip}^f)\}_{ip=1}^{N_{ip}}$,
- **Projection** The associate field is project on a regular grid of n_g elements of characteristic length $\ell_{iI} = 2dp$. Using the assignent operator, we obtain for each node $iI \in \Lambda_i$

$$\mathbf{u}_{iI}^f = \frac{1}{V_{iI}} \sum_{ip \in \mathcal{P}_i^f} \mathbf{\Gamma}_{ip}^f \mathbf{W} \left(\frac{\mathbf{z}_{iI} - \mathbf{z}_{ip}^f}{\ell_{iI}} \right)$$

- **Analyse** Based on this new discretization, an Eulerian-based data assimilation could be apply on the nodal state values \mathbf{u}_{iI}^f such as the analysis state \mathbf{u}_{iI}^a is

$$\mathbf{u}_{iI}^a = \mathbf{u}_{iI}^f + \sum_{j=1}^{N_{\text{ens}}} F_{ji} \mathbf{u}_{jI}^f,$$

- **Interpolation** A new regular particle discretization is initialized. Two particles by directions are placed inside each cell of the grid. The new particle intensities could be evaluate thanks to the interpolation operator, such as for $ip' \in \mathcal{P}_i^a$

$$\boldsymbol{\Gamma}_{ip'}^a = \sum_{iI \in \Lambda} \mathbf{u}_{iI}^a \left(\frac{\mathbf{z}_{iI} - \mathbf{z}_{ip'}}{\ell_{iI}} \right).$$

The Remesh-Filter update scheme is sum-up in the algorithm 1 and illustrate in the Figure 4.2

Algorithm 1: Remesh Filter analysis update

```

Data:  $\mathbf{G} \in \mathbb{R}^{n_g \times d}, \mathbf{z}^a \in \mathbb{R}^{2n_g \times d}$ ; // grid discretization
Data:  $\mathbf{R} \in \mathbb{R}^m$ ; // observation covariance
Input:  $\mathcal{P}_i^f = \{(\mathbf{z}_{ip}^f, \boldsymbol{\Gamma}_{ip}^f)\}_{ip=1}^{N_{ip}}$ ,  $i = 1, \dots, N_{\text{ens}}$ ; // forward
discretizations
Input:  $\mathbf{Y}_f \in \mathbb{R}^{m \times N_{\text{ens}}}$ ; // the associate observation anomalies
Input:  $\mathbf{D} \in \mathbb{R}^{m \times N_{\text{ens}}}$ ; // the perturbed observations
 $\mathbf{F} = \mathbf{Y}_f^T (\mathbf{Y}_f \mathbf{Y}_f^T + \mathbf{R})^{-1} (\mathbf{D} - \mathcal{Y})$ ; // correction matrix
foreach  $i = 1, \dots, N$  do
     $\lfloor \mathbf{u}[:, i] = \text{Projection}(\mathcal{P}_i^f, \mathbf{G})$ 
     $\mathbf{u} = \mathbf{u} + \mathbf{uF}$ ; // analysis update
foreach  $i = 1, \dots, N$  do
     $\lfloor \mathbf{z}_{ip}^a, \boldsymbol{\Gamma}_{ip}^a = \text{Assign}(\mathbf{u}[:, i])$ 
return  $\mathcal{P}_i^a = \{\mathbf{z}_{ip}^a, U_{ip}^a\}_{ip=1}^{N_a}$ ,  $i = 1, \dots, N_{\text{ens}}$ ; // analyse discretizations

```

We observe that all the operators we have defined are linear. Therefore, performing the update on the grid as in the algorithm or on the new particles is equivalent because the update during the analysis, the interpolation operation, and the projection are all linear with respect to the intensities. However, the current algorithm's order allows for the minimum number of operations to be performed in lower-dimensional spaces, namely on the projection grid.

4.5.2 Particles-EnKF

The previous adaptation of the EnKF filter allows for controlling the number of particles during the analysis by regenerating a new particle distribution. In a way, it involves

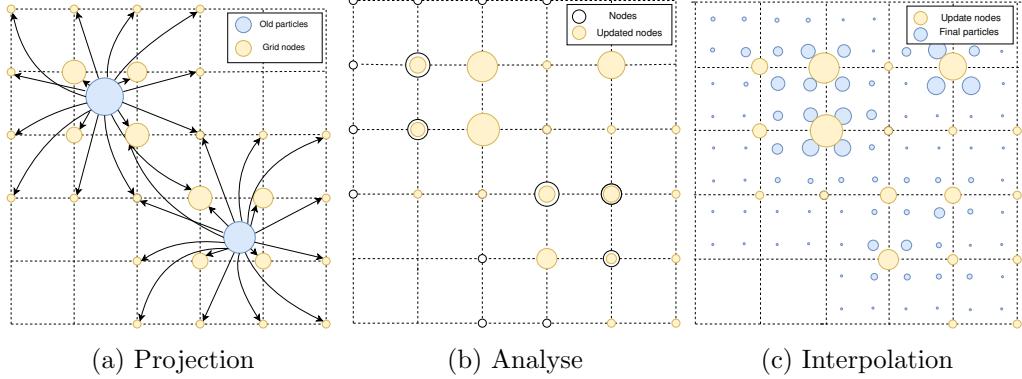


Figure 4.2: 4.2a Each particles set is project on a fixed Eulerian grid. 4.2b EnKF matrix operations are apply to nodal values. 4.2c New regular sets of particles are generate to interpolate analyzed solutions.

projecting the solution onto an Eulerian discretization before applying the analysis in the traditional manner. The goal of the Particles-EnKF formulation is to define the analysis on the particle discretization as much as possible. In this scheme, we keep the particle positions after the forward is unchanged. The change concern the particle intensities. This way, the Lagrangian representation of the solution at the end of the forward step is kept the same as much as possible. The fields, defined in (4.3), are approximated on the previous discretization to avoid exponential growth of the number of particles.

By this way, the analyzed field is approximated by $\hat{\mathbf{u}}_i^a$ such as

$$\mathbf{u}_i^a(\mathbf{z}) \simeq \hat{\mathbf{u}}_i^a(\mathbf{z}) = \sum_{p \in \mathcal{P}_i^f} \boldsymbol{\Gamma}_{ip}^a \varphi_p(\mathbf{z})$$

where $\boldsymbol{\Gamma}_{ip}^a$ have been determined by approximation or regression.

However, because this regression is only performed on the current forecast discretization \mathcal{P}_i^f , additional particles could be introduced at the support border to allow a better estimate.

The Part-EnKF algorithm is expressed in Algo 2

Algorithm 2: Part-EnKF Filter analysis update

```

Data:  $\mathbf{R} \in \mathbb{R}^m$  ;                                // observation covariance
Input:  $\mathcal{P}_i^f = \{(z_{ip}^f, \Gamma_{ip}^f)\}_{ip=1}^{N_{ip}}$ ,    $i = 1, \dots, N_{ens}$  ;           // forward
         discretizations
Input:  $\mathbf{Y}_f \in \mathbb{R}^{m \times N_{ens}}$  ;          // the associate observation anomalies
Input:  $\mathbf{D} \in \mathbb{R}^{m \times N_{ens}}$  ;          // the perturbed observations
 $\mathbf{F} = \mathbf{Y}_f^T (\mathbf{Y}_f \mathbf{Y}_f^T + \mathbf{R})^{-1} (\mathbf{D} - \mathcal{Y})$  ;          // correction matrix
foreach  $i = 1, \dots, N$  do
     $u_{ip}^a = \text{AnalysisFieldValues}(\mathcal{P}_i^f, \mathbf{F})$  ; // evaluate the analysis field
     $\Gamma_{ip}^a = \text{Approx}(u_{ip}^a)$  ;                      // approximate the analysis field
return  $\mathcal{P}_i^a = \{z_{ip}^a, \Gamma_{ip}^a\}_{ip=1}^{N_{ip}}$ ,    $i = 1, \dots, N_{ens}$  ; // analyse discretizations

```

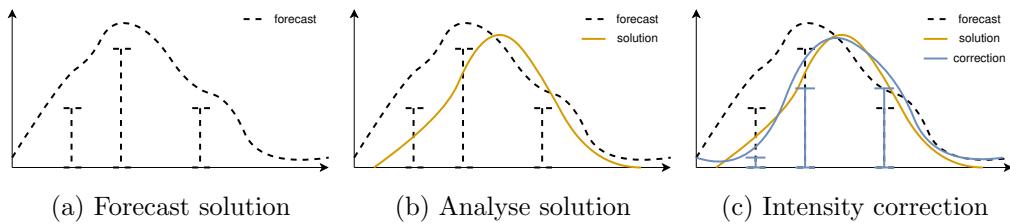


Figure 4.3: One dimensional illustration. 4.3a A member forecast function is discretized by a set of three particles. 4.3b The analyzed solution is determined. 4.3c The forecast particle intensities are updated to fit the analyzed solution.

4.6 1D density advection-diffusion problem

4.6.1 Description of the problem

An initial exploration is conducted on a one-dimensional application to assess the filter performance. We define the following one-dimensional 2π -periodic convection-diffusion problem such as

$$\frac{\partial u}{\partial t}(z, t) + v \frac{\partial u}{\partial z}(z, t) = D \frac{\partial^2 u}{\partial z^2}(z, t),$$

with z the spatial coordinate, v a constant velocity and D a constant diffusion coefficient. For the following application, the reference solution will use $v = 1.0$ and $D = 0.05$ as parameters. We define the 2π -periodic heat kernel in one dimension, such as

$$\phi(u, s) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{4\pi s}} \exp\left(-\frac{(u - 2\pi k)^2}{4s}\right).$$

Considering an initial condition characterized by a Gaussian shape expressed as $u^{gt}(z, 0) = \phi(z - z_0, Dt_0)$, where $z_0 = 0.02$, $t_0 = \frac{\sigma_0^2}{2D}$, and $\sigma_0^2 = 0.5$, we derive the comprehensive analytical solution utilizing the Green equation solution

$$u^{gt}(z, t) = \phi(z - vt - z_0, D(t + t_0)).$$

The analytical solution is succinctly described as a Gaussian function, characterized by a mean that moves at the advection velocity and a standard deviation proportional to t and D . This solution is visually depicted in Figure 4.4 across various assimilation time frames.

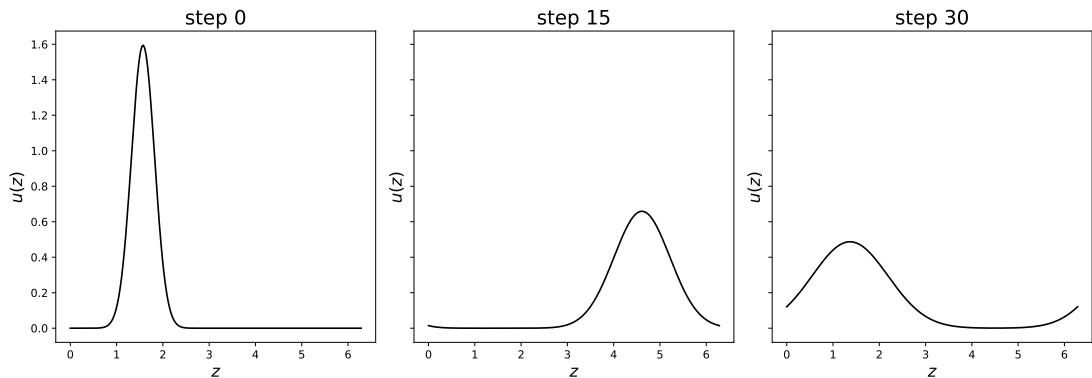


Figure 4.4: The analytical solution of the convection-diffusion problem evolves over time, with the final snapshot revealing a complete spatial period.

Following a Lagrangian perspective by tracking a fluid particle of position z_p and intensity U_p , the equation becomes

$$\frac{dz_p}{dt} = v(z_p, t), \quad \frac{dU_p}{dt} = D \frac{d^2 U_p}{dz^2}$$

To solve the convection-diffusion scheme, we employ the two steps of the viscous splitting algorithm. The advection is taken into account by updating the position of the particle with an Euler explicit scheme. On the other hand, we use a redistribution method called the Particle Strength Exchange Method (PSE) [Degond and Mas-Gallic, 1989b; Cottet and Mas-Gallic, 1990] to approximate the Laplacian term $\frac{d^2 U_p}{dz^2}$.

$$D \frac{d^2 U_p}{dz^2} = D \frac{d^2 u_p}{dz^2} V_p \approx D \varepsilon^{-d} \int [u(z) - u(y)] \phi_\varepsilon(y - z) V_p dz,$$

It deals with the particle approximation by a redistribution of the particle intensities in their previous locations, such as

$$\frac{dU_p}{dt} = D \varepsilon^{-d} \sum_q (V_p U_q - V_q U_p) \phi_\varepsilon(z_q - z_p),$$

where V_p the volume of the particle p and d the dimension of Ω . For further details on the computation, please refer to [Cottet and Mas-Gallic, 1990].

For the periodic boundary problem described in section 4.6, we define an equivalent kernel function $\phi_\varepsilon = \phi_g^P = \sum_{n=-\infty}^{+\infty} \phi_g(r - 2\pi n)$.

The particle-based model employs a discretization of $N_{part} = 100$ particles with a size of $h = \frac{L}{N_{part}}$ and a smoothing length of $\varepsilon = 1.3h$. For the sake of comparison, we solve the convection-diffusion equation with an explicit central finite difference scheme discretized on a regular grid with $N_{grid} = 100$ nodes.

4.6.2 Assimilation parameters and ensemble generation

Ensemble distribution

All filters undergo testing on an identical initial prior ensemble of size $N = 25$ members, characterized by Gaussian shapes that are shifted and scaled. The mean of the function and its standard deviation are sampled. The total mass is set equal to 1. Parameters of velocity v and the diffusion D are also sampled. The different distributions are defined in Appendix 4.1. Moreover, the parameters samples and initial state are illustrated in Figure 4.5.

The observational data is subject to additive noise, denoted as $\eta \sim \mathcal{N}(0, \sigma_y \mathbf{I})$, where $\sigma_y = 0.05$ and \mathbf{I} represents the identity matrix.

Error definition

We define the error as the following relative ratio

$$e_{L_2} = \frac{\left[\frac{1}{N} \sum_{i=1}^N \int_{\Omega} (u_i(z) - u^{gt}(z))^2 dz \right]^{1/2}}{\|u^{gt}\|_{L_2}} \quad (4.6)$$

where u_i denote the i -th member of the ensemble and $\|u\|_{L_2}$ denote the L_2 norm of u^{gt} .

The L_2 norm is computed using a quadrature over a regular grid of an ensemble of cells \mathcal{C} such as for any $f \in L_2$

$$\|f\|_{L_2} = \int_{\Omega} f^2 d\Omega \approx \sum_{c \in \mathcal{C}} f(z_c) V_c$$

where z_c is the center of the cell c and V_c the volume of the cell. The grid is still the same for all the simulations.

Numerical parameters

We conduct $N_{\text{assim}} = 30$ assimilation steps at evenly spaced intervals until the final time $t_f = 2\frac{L}{v}$. During each assimilation step, the field u^{gt} is observed at six regularly spaced positions x_{obs} .

In the particle-based simulation, fields are discretized using regularly spaced particles that are shifted. Intensity values are obtained by fitting an interpolation operator like in Section 4.4.3 to the particle intensity. The parameter $\varepsilon_{\text{mass}}$ is introduced as a cutoff for particle selection, allowing for the definition of varying numbers of particles for each simulation. The particle support poses challenges in the Part-EnKF as described in Section 4.

Simultaneously, a standard Ensemble Kalman Filter (EnKF) update is applied to the nodal variables to construct the reference filter Grid-EnKF, which uses a grid-based model. For grid-based simulation, the fields of each member are interpolated at the node locations. In this way, the ensemble generated is still the same for the sake of comparison.

4.6.3 Results

We compare the different filters on the assimilation of the state. We first compare the Grid-EnKF, the Remesh EnKF, and two Part-EnKF filters with 100 and 60 particles. The parameter sample is still unchanged. We take unknown parameters into account as model uncertainties. The two filters outlined in the Method section 4.5 and the Eulerian filter are compared with the reference filter based on a grid discretization.

In figure 4.7, we appreciate different assimilation steps for the Remesh-EnKF filter.

The result is quite similar for all the different filters except the Part-EnKF with 60 particles.

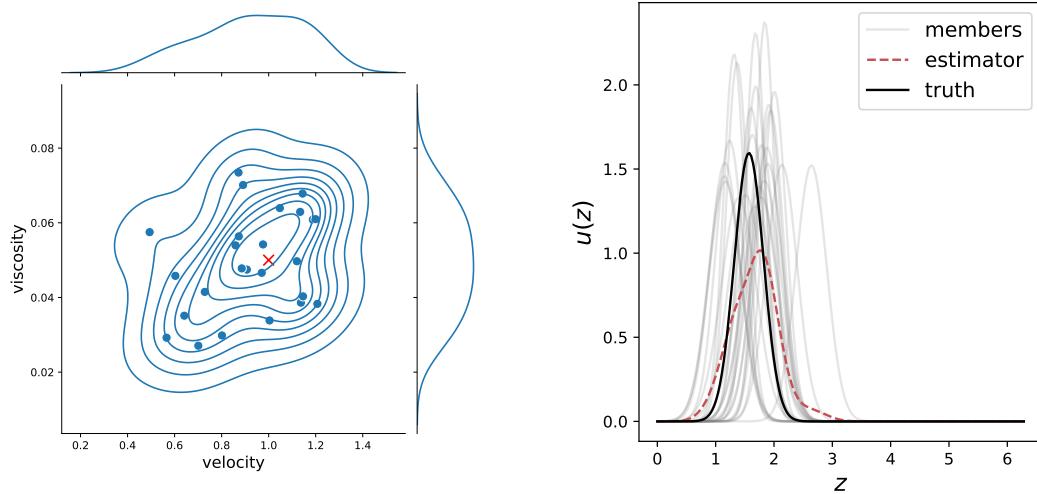


Figure 4.5: On the left the initial parameters sample, v in abscissa and D in ordinate. On the right is the initial ensemble state.

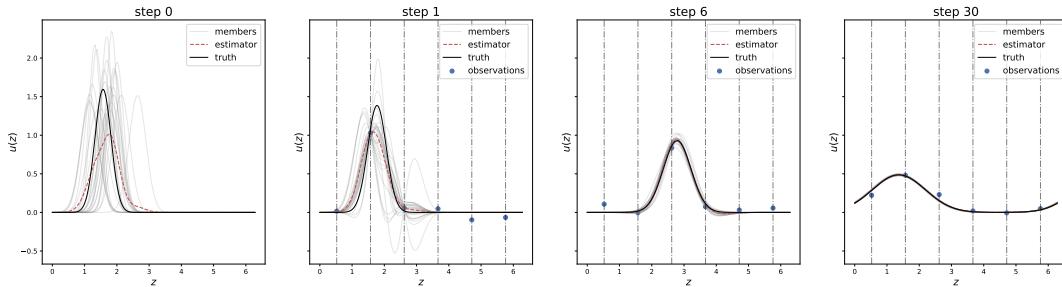


Figure 4.6: Data assimilation over assimilation step for the Remesh-EnKF filter.

The primary issue arises from the regression on non-overlapping support, where the regression struggles to fit the analysis solution defined on a more considerable space support. This leads to heightened variability, particularly in the tail of the distribution. Addressing this common challenge in RBF Regression [Fornberg and Flyer, 2015] involves increasing the Ridge penalization coefficient, a parameter we choose through cross-validation Ridge regression. Even with a more stable regression, it remains a projection of the analysis solution onto the forecast support. It is imperative to increase the number of particles to achieve a better approximation of the analysis solution using the particle approximation operator in Section 4.4.3 or the regression operator in Section 4.4.3.

We validate this assumption by varying the initial support of particles. Quantitatively, as observed in Figure 4.8, the error estimate and dispersion decrease with an increase in the number of particles. At the level of 70 particles support, a threshold is reached. Moreover, qualitatively examining the snapshot on the right reveals that the solution closely aligns with the reference.

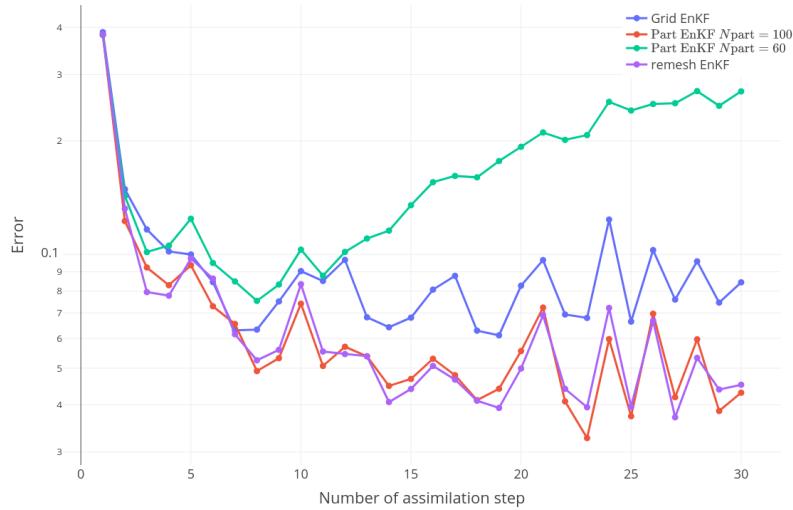


Figure 4.7: State error with respect to assimilation time step.

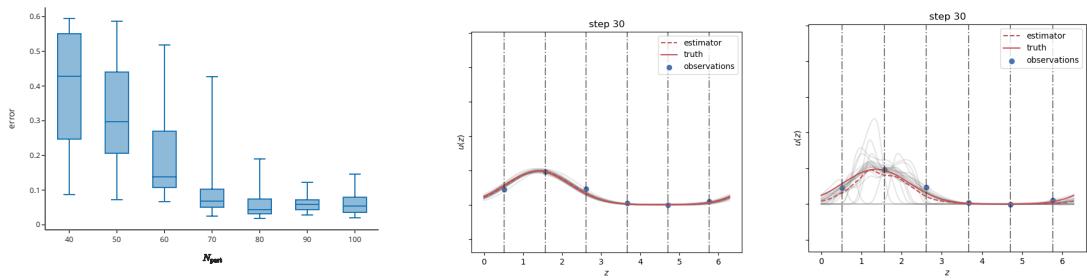


Figure 4.8: Left: Error with respect to particle support size, Middle: Final step for a support of 100 particles, Right: Final step for a support of 60 particles.

However, adding particles to a more complex solution is challenging. Indeed, good spacing between particles and particle density has to be preserved. In this case, we advise defining criteria for the reconstruction' error. Instead of adding particles, we advise generating a new, regularly spaced grid of particles to reconstruct the solution.

In conclusion, this example underscores the Remesh-EnKF filter capability to yield results comparable to the classical EnKF applied to a grid model. Additionally, it highlights the Part-EnKF capability in assimilating on a particle discretization while also emphasizing the importance of addressing spatial discrepancies between members, which can pose challenges in solution reconstruction. The computation of solution error reconstruction provides a straightforward criterion for remeshing a member and applying the analysis solution approximation.

4.7 2D vortex-in-cell problem

4.7.1 Description of the method

In this section, we apply the Vortex Method in a two-dimensional scenario, as outlined by Cottet et al. [Cottet and Koumoutsakos, 2000a]. The Vortex Method is a Lagrangian approach utilizing a particle ensemble to discretize the vorticity field, allowing for the solution of the Navier-Stokes equation for viscous incompressible flow. The method is grounded in the vorticity-velocity formulation of the Euler equation, where $\omega = \nabla \times \mathbf{v}$ satisfies

$$\frac{\partial \omega}{\partial t} + (\mathbf{v} \cdot \nabla) \omega - \nu \Delta \omega = 0, \\ \nabla \cdot \mathbf{v} = 0,$$

where ω denotes vorticity, \mathbf{v} represents velocity, and ν stands for viscosity.

In the context of 2D flow, vorticity is perpendicular to the flow plane, forming a scalar field denoted as ω . In Cartesian coordinates, it is expressed as $\omega = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}$.

The vorticity field is discretized using a collection of discrete vortices, each characterized by a position \mathbf{z}_p , an associated kernel ϕ_ε , and a circulation Γ_p . For all points \mathbf{z} within the domain Ω , the vorticity is expressed as

$$\omega(\mathbf{z}) = \sum_{i=1}^{N_p} \Gamma_p \phi_\varepsilon(\mathbf{z} - \mathbf{z}_p).$$

To address the Navier-Stokes equation, we employ a viscous splitting scheme, following the methodology outlined in [Cottet and Mas-Gallic, 1990], acknowledging the predominance of the convection term over viscosity. We use the Vortex-In-Cell algorithm [Christiansen, 1973; Birdsall and Fuss, 1969], coupled with an FFT solver to compute the advection velocity. The subsequent steps involve assigning particle vorticity values to the grid using a particle-to-grid formula and computing the velocity field by solving the Poisson equation on the grid verified by the stream function. Finally, the velocity is interpolated back onto the particles using the grid-to-particles formula. A Runge-Kutta 3 time-stepping scheme is employed to update the particle positions through a time integration scheme. The final phase involves solving the heat equation and updating the particle intensities thanks to the PSE method previously described in Section 4.6.

4.7.2 Lamb-Chaplygin dipole and simulation parameters

We define the reference as the advection of the Lamb-Chaplygin dipole inside a close domain with stress-free walls. Lamb-Chaplygin dipole is a popular choice for numerical studies [Orlandi, 1990]. The model represents a specific steady, inviscid dipolar vortex flow and offers a non-trivial solution to the two-dimensional Euler equations. The dipole

is characterized by a translation velocity U , a mean position z_0 , a radius R , and an orientation α .

The dipole vorticity field ω could be expressed as

$$\omega(r) = \begin{cases} \frac{-2kUJ_1(kr)}{J_0(kR)} \sin \alpha & \text{for } r < R, \\ 0 & \text{otherwise,} \end{cases}$$

where (r, α) represent the polar coordinates in the dipole reference frame. Here, J_0 and J_1 denote the zeroth and first-order Bessel functions of the first kind, respectively, and k is determined such that kR corresponds to the first non-trivial zero of the first Bessel function. The dipole vorticity field is depicted in Figure 4.9.

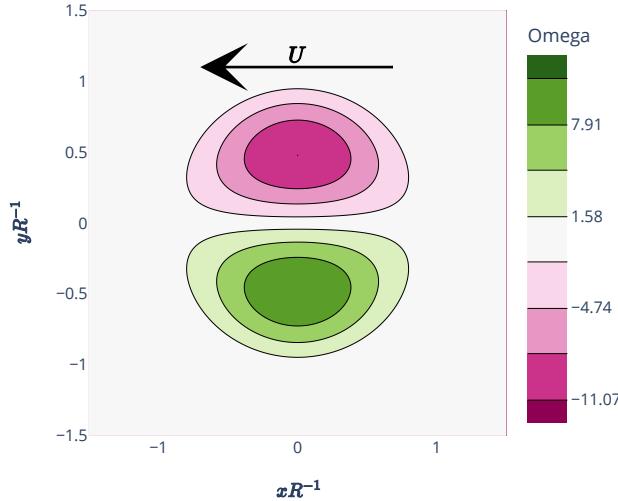


Figure 4.9: The Lamb-Chaplygin dipole vorticity field on a normalized space.

The dipole is positioned at the center of a box with dimensions $[0, \pi] \times [0, \pi]$, featuring an orientation of $\frac{7\pi}{8}$ rad., a radius of 0.5 meters, and a velocity U of 0.25 m.s^{-1} . The complete reference setting is listed in Table 4.2.

The boundary box features stress-free walls, meaning fluid cannot pass through them. The velocity perpendicular to the walls is zero, while tangential velocity remains undetermined. When a vortex, such as a dipole, reaches this boundary, it walks along the wall, sensing its reflection and interacting with it.

Because this problem does not have an explicit solution on a closed domain, we simulate the ground truth with the vortex method for a fined discretization and fixed set of parameters also described in Table 4.2. The trajectory of the ground truth is illustrated in Figure 4.10 on a regularly spaced grid.

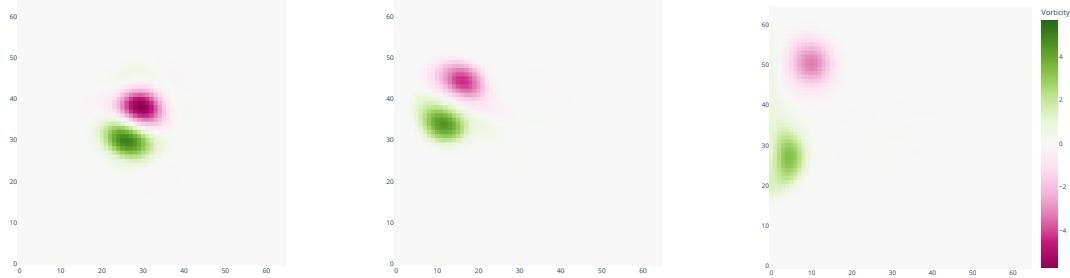


Figure 4.10: Trajectory of the ground truth. The vorticity is represented on a regularly spaced grid. For $t = [1, 5, 10]s$.

Several parameters in the simulation influence the particle distribution and can lead to different results. The first one is the particle size defined by d_p . Another significant parameter is ε_ω , associated with the remeshing process occurring either during the forecast (to prevent high distortion of the particle distribution) or during the Remesh-EnKF filter. ε_ω serves as a threshold, determining whether a particle is retained after the remeshing process based on the condition $V_p \Gamma_p > \varepsilon_\omega$. The impact of this parameter is illustrated for one member after the first forward in Figure 4.11.

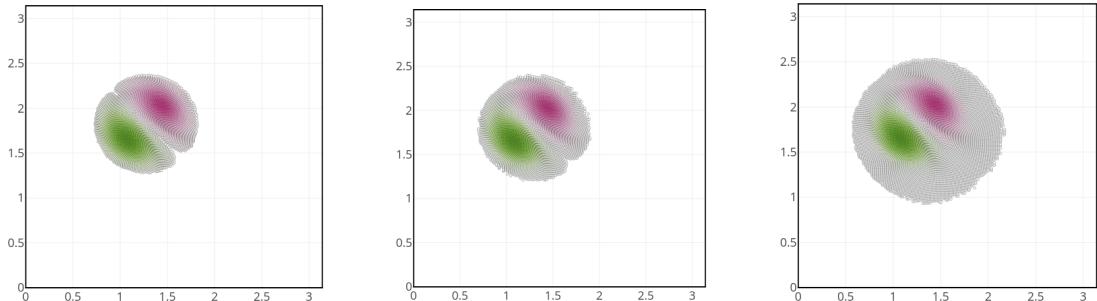


Figure 4.11: Effect of the parameter ε_ω on the particle discretization of the solution for one member. From left to right, results for $\varepsilon_\omega = 0.1, 0.01$, and $1.e^{-6}$.

For the following paragraphs, if the value is not explicitly changed, we use the nominal parameters described in Table 4.3 for the simulation.

4.7.3 Assimilation parameters and ensemble generation

Ensemble distribution

An ensemble of 32 members is created by sampling distributions over the dipole parameters. We sample the radius R , the prescribed velocity U , the orientation α , and the barycenter \mathbf{z}_{mean} . Additionally, the model viscosity ν is also sampled. All the distributions are summarized in Table 4.4. The first six members are plotted in Figure 4.12.

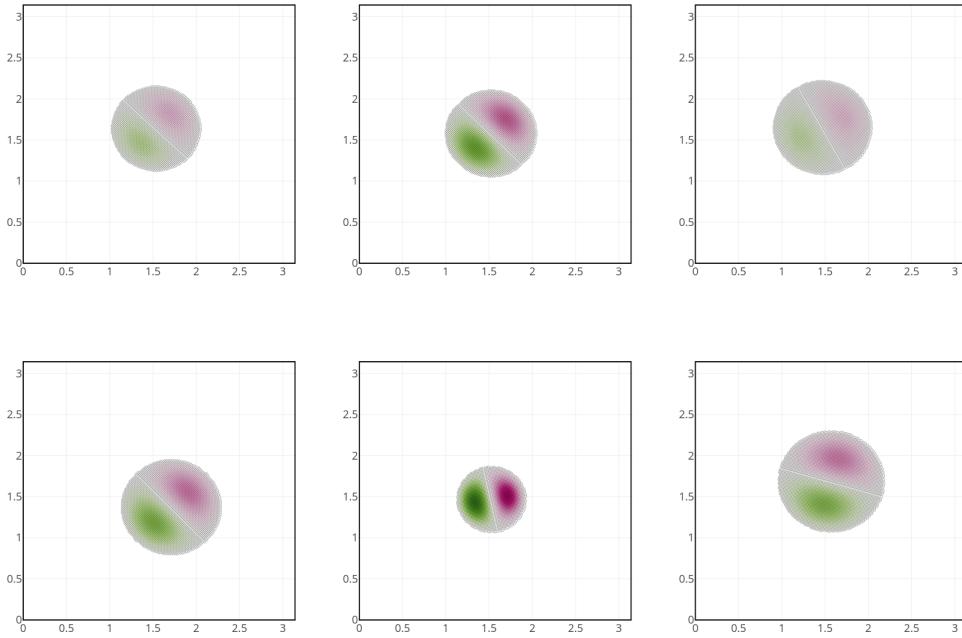


Figure 4.12: Six samples from the initial ensemble.

The initial vorticity field is first discretized on a regular grid of particles with a characteristic length d_p , where each particle receives the circulation $\Gamma_p = \omega(\mathbf{z}_p)V_p$ and $V_p = d_p^2$ represents the volume of the particle.

Error definition

We use an absolute error absolute L_2 -error defined as $\frac{1}{N} \sum_{i=1}^N \int_{\Omega} (\omega_i(\mathbf{z}) - \omega^{gt}(\mathbf{z}))^2 d\mathbf{z}$. We also use the member errors to evaluate the dispersion of the error estimate.

Numerical parameters

The assimilation frequency is defined by the assimilation step dt_a . The simulation is performed over a duration of t_f . All simulation parameters are summarized in Table 4.3.

Observations are collected on a regular grid of size N_{obs} , measuring both components of the velocity. The observations follow a normal distribution $\mathcal{N}(0, \sigma_{\text{obs}}^2 \mathbf{I})$, indicating an ensemble of independent measurements, each characterized by a standard distribution of σ_{obs} . An example of observed velocity with and without noise is illustrated in Figure 4.13.

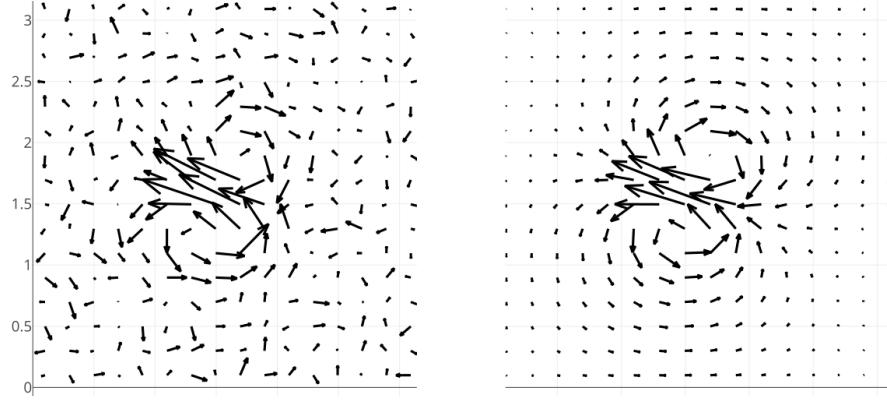


Figure 4.13: Observed and reference velocity fields. The error on each component is a sample from a centered normal distribution with the nominal value $\sigma_{\text{obs}} = 0.05$.

4.7.4 Results

Error through time

We start by analyzing the assimilation error over time. Figure 4.14 illustrates the error throughout the assimilation process for the nominal set of assimilation parameters, demonstrating comparable results for both filters. At each assimilation step, the error decreases and prevents the solution from diverging elsewhere.

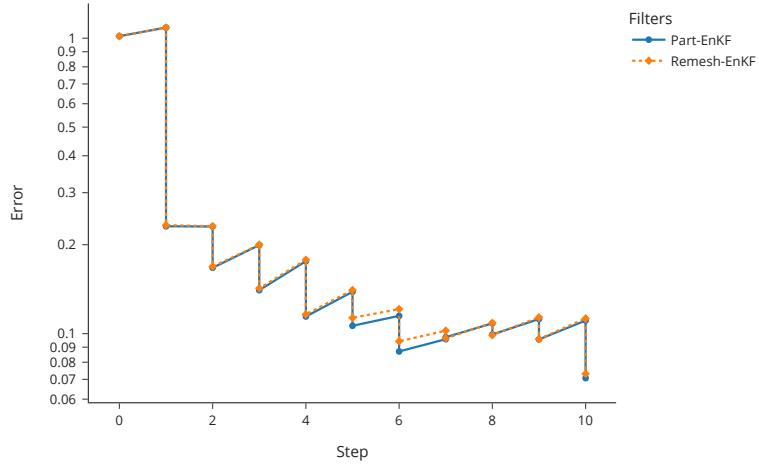


Figure 4.14: error curves through assimilation steps. Left: L_2 -error of the field, Right: Error for the viscosity parameter. With Part-EnKF in blue and Remesh-EnKF in red.

Error with respect to assimilation parameters

We also assess the performances of the different filters by evaluating the convergence of the error with respect to the assimilation parameters.

We observe the convergence rate concerning data assimilation parameters: The observation precision, which is $1/\sigma_{\text{obs}}^2$, the number of observations N_{obs} , the number of assimilation step N_{assim} .

Figure 4.15a illustrates a decreasing error bias and variances with respect to observation precision similarly for both filters. What is striking in the same Figure 4.15b but in the log-scale is the regular convergence rate for both filters with respect to the observation precision. The order of convergence is about 0.68 for Part-EnKF and 0.75 for Remesh-EnKF.

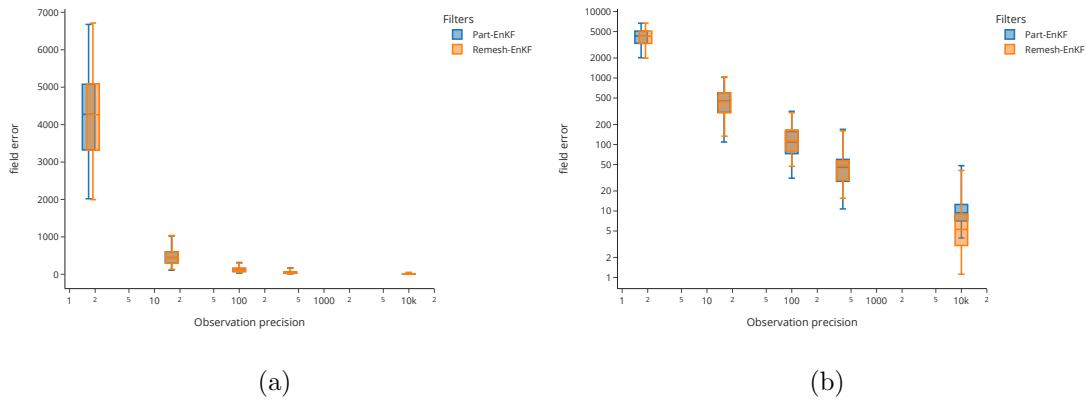


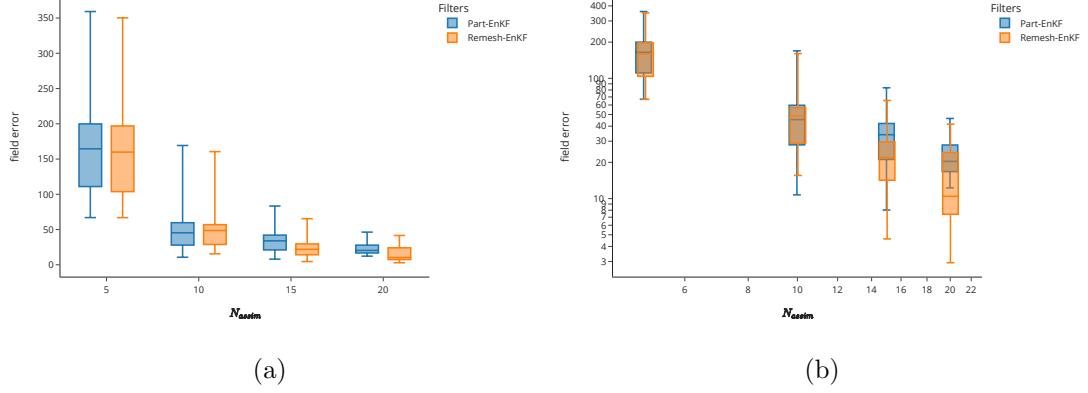
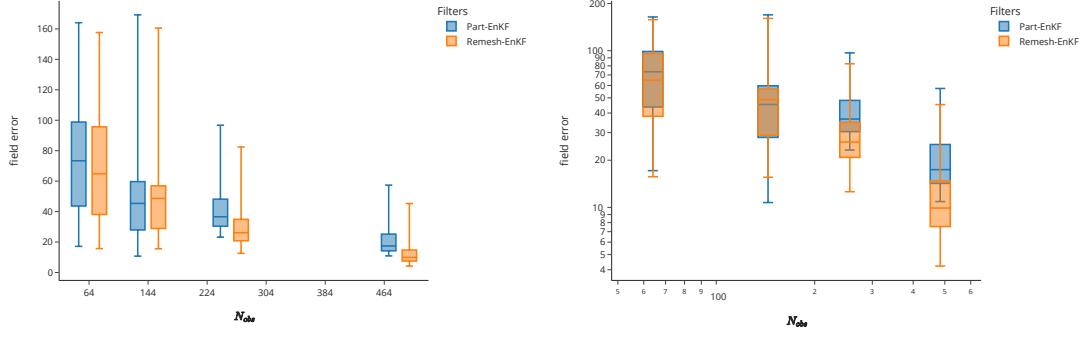
Figure 4.15: Box plots of the state error w.r.t. $1/\sigma_{\text{obs}}^2$.

In Figure 4.16a, the reduction of error is still prominent and shows a reduction of variance as the number of observations increases. In the log-scale Figure 4.16b, the error decrease also at a constant rate for both filters. We notice a more substantial order of 1.8 for the Remesh-EnKF compared to an order of 1.4 for Part-EnKF.

Finally, we analyze the error convergence with respect to the number of observations. Observation locations increase regularly on both axes. In Figure 4.17a, the error estimate and variances decrease. For a relatively small number of observations, the two filters offer similar results when, in contrast, the Remesh-EnKF has relatively better results when the number of observations is increased. Moreover, the convergence rate seems to change around 200 observation points, as illustrated in the log-scale Figure 4.17b. Nevertheless, it illustrates adequate performances for both filters.

Error with respect to simulation parameters

To better understand the differences, let us now turn to the evolution of the error with respect to particle discretization parameters. For the Part-EnKF, remember that each member has its own particle discretization that flows according to the dipole direction

Figure 4.16: Box plots of the state error w.r.t. N_{assim} .(a) Box plots of the state error w.r.t. N_{obs} . (b) Box plots of the state error w.r.t. N_{obs} .

and velocity. Each analyzed member's solutions are then respectively projected on their member discretization. However, this scheme could introduce different sources of error. First, due to particle irregularity in the particle distribution, severe approximation was introduced, which led to errors between the analyzed and the approximated solution. Even more seriously, certain parts of the solution may vanish as no particle in the support can interpolate it. This effect could be appreciated on several samples of the ensemble where the analysis is projected on a non-conforming particle discretization. For instance, we analyzed the first assimilation step of one member for the different filters. If the analyzed field is known over the space domain, we observed in Figure 4.18 that the Remesh-Filter is only able to interpolate the entire solution. The Part-EnKF is not entirely able to interpolate the solution with the forecast member discretization. Moreover, some distortions observed in the particle distribution are not in line with the analyzed field flow. These remarks are more critical when the forecast step is longer, leading to high errors, or when the size of the support is lower. Moreover, the approximation of Section 4.4.3 will introduce the approximation error function of the size of the particles. For instance, the particle will not conserve the total circulation due to a quadrature error, which is the opposite case for the Remesh-EnKF filter.

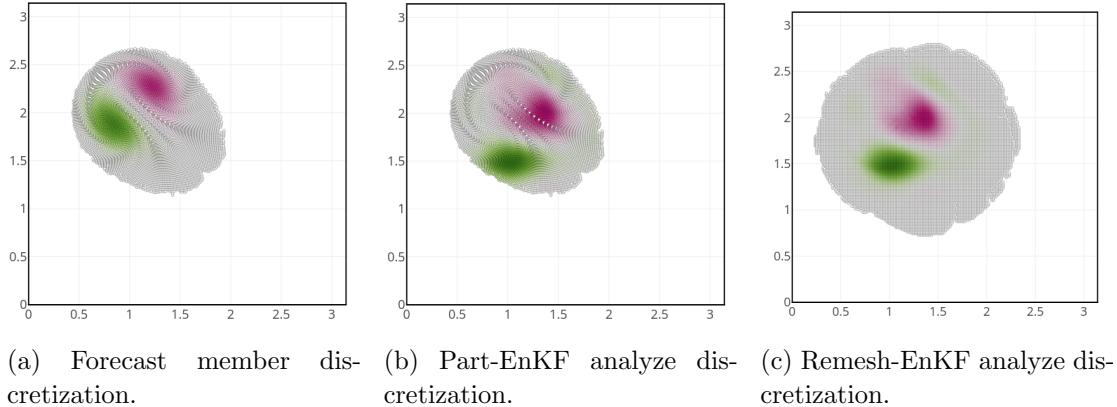
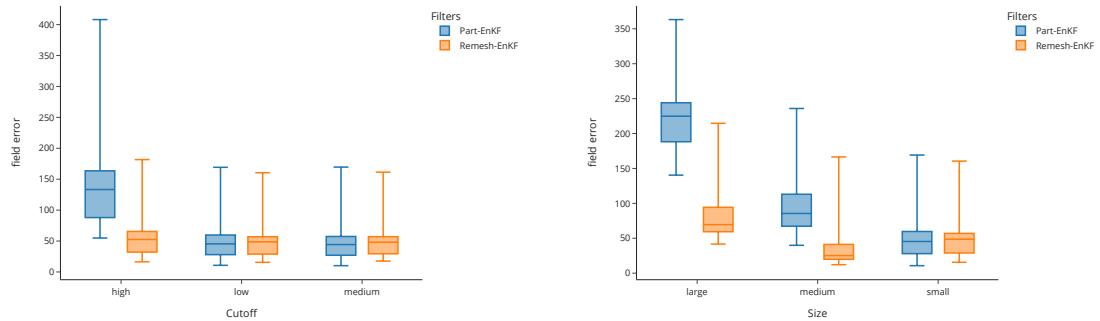


Figure 4.18: Assimilation of one member with a forecast discretization unadapted to the analyses solution. The forecast discretization used by the Part-EnKF does not always support approximation for the analyses and introduces discretization errors.

To evaluate the effect of the size of the support, we varied the value of ε_ω . We have seen in Figure 4.11 that this parameter affects the number of particles and, thus, the size of the support. In Figure 4.19a, we observe high disparities between the two filters. However, the error stabilizes rapidly by decreasing the threshold. These findings suggest an impact on the threshold and, thus, the particle support for the Part-EnKF. By contrast, Remesh-EnKF is less sensible. As a consequence, it only thresholds low values in the analyzed solutions.



(a) state error w.r.t. ε_ω . The High, low, and medium cutoff correspond respectively to $\varepsilon_\omega = 0.1, 1.e^{-6}$ and 0.01.

(b) state error w.r.t. d_p . The large, medium, and low sizes correspond to $d_p = 0.0327, 0.0245$, and 0.0123.

Figure 4.19: Error box plots for simulation parameters. The effect of ε_ω on the error is particularly observed on high value. The Part-EnKF error is strongly linked to d_p through the particle approximation error.

In terms of particle size d_p , we evaluate its effect on the assimilation. In this section, we have used the regression operator defined in Section 4.4.3 for Part-EnKF. This choice has been motivated to accelerate the update process. in Figure 4.19b. We observe that

the error for Part-EnKF increases with d_p . In fact, it increases proportionally as for the approximation error. This high error confirmed, in that case, the high effect of the particle approximation for Part-EnKF. In contrast, the Remesh-EnKF error is relatively low, taking advantage of a high-order projection interpolation scheme. Using a regression operator to approximate the analyzed solution should alleviate this effect provided other particle discretization considerations (distortion, support size) as illustrated in Part 4.6 and the choice of an adequate penalty coefficient to succeed in approximating the solution.

This discussion pointed out the high dependency of the Part-EnKF on particle discretization. The particle discretization of a member may be too far from the solution support of particles. This opens the question of the choice or modification of the particle discretization. As suggested in Part 4.6, an error estimation could be introduced to choose between the different filters. On the other hand, other approaches could be to select the member with the maximum likelihood estimate to approximate the solution. This proposition has to be evaluated because it could considerably reduce the variance of the ensemble. Finally, an alignment of particles might be introduced to better fit the analyzed solution.

4.8 Conclusion

In this study, we introduce a novel framework that combines a sequential ensemble data assimilation approach with particle-based models. Specifically, we have developed two novel Ensemble Kalman Filter schemes dedicated to meshfree Lagrangian simulations. These formulations depend either on updating the particle quantities or on remeshing the particle discretization on a common ensemble grid.

We proposed two update strategies relying on an update matrix that does not directly depend on the state discretization. The first strategy, Remesh-EnKF, is based on the projection of every member onto a new common discretization of particles. The second strategy, Part-EnKF, evaluates the analysis field at particle locations to update the particle quantities.

The different classes have been initially tested on a one-dimensional example and compared with an Eulerian representation of the solution. The results demonstrate comparable performance across the various filters, with the exception of certain configurations of the Part-EnKF. It has been demonstrated that provided the support of the particles is consistent with the analysis solution, the filters yield similar results. However, in cases where the support deviates from the analysis field, members diverge. Increasing the support for the solution is necessary.

A two-dimensional case has also been tested, particularly to assess nonlinear advection schemes with various configurations. We observe good agreement among the different filters. However, this time, the particle approximation is predominant in the Part-EnKF. Both strategies offer several derivations. Remesh-EnKF is mainly dependent on the redistribution kernel to obtain the new regularly spaced particle set. Part-EnKF could be extended because it is highly flexible when defining the new set of particles for each member. This tuning is essential, as we have highlighted the issue of the non-conforming support of the forward particle position with the analysis.

Several methodologies to introduce new particles or change the previous ones could be explored, particularly at the edge of the distribution. For the sake of simplicity, we advocate generating a new set of particles with a regular spacing for any member encountering difficulties in accurately reconstructing the analysis field. Another alternative is to update the positions of the particles in addition to the intensities instead of only changing the intensities. In this way, the error due to a misfit of alignment or approximation could be avoided. These type of adaptation could be derived from optimal transport scheme [Bocquet et al., 2024] or correction of the background ensemble alignment with the observation [Ravela et al., 2007; Rosenthal et al., 2017].

4.A Stochastic Ensemble Kalman Filter

We define the matrix of states and the matrix of anomalies $\mathbf{Z}^f = [\mathbf{x}^1, \dots, \mathbf{x}^N]$, \mathbf{A} whose columns are the member states and the normalized anomalies.

$$\mathbf{A} = \frac{1}{\sqrt{N-1}}(\mathbf{Z}^f - \bar{\mathbf{x}}^f \mathbf{1}^T),$$

where $\mathbf{1} \in \mathbb{R}^N$ is a vector of one.

Respectively the matrix of observation and observation anomalies are $\mathcal{Y}^f = [\mathcal{H}(\mathbf{x}_1^f), \dots, \mathcal{H}(\mathbf{x}_N^f)]$, \mathbf{Y} where columns are

$$\mathbf{Y} = \frac{1}{\sqrt{N-1}}(\mathcal{Y} - \bar{\mathbf{y}} \mathbf{1}^T) \quad \text{with} \quad \bar{\mathbf{y}} = \frac{1}{N} \sum_{j=1}^N \mathcal{H}(\mathbf{x}_j^f).$$

The ensemble defines the covariance between states and observations $\mathbf{P}\mathbf{H}^T$, the covariance between observations $\mathbf{P}\mathbf{H}^T$, and $\tilde{\mathbf{K}}$

$$\begin{aligned} \mathbf{P}\mathbf{H}^T &\simeq \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i^f - \bar{\mathbf{x}}^f) [\mathcal{H}_k(\mathbf{x}_i^f) - \bar{\mathbf{y}}]^T = \mathbf{A}\mathbf{Y}^T, \\ \mathbf{H}\mathbf{P}\mathbf{H}^T &\simeq \frac{1}{N-1} \sum_{i=1}^N [\mathcal{H}_k(\mathbf{x}_i^f) - \bar{\mathbf{y}}] [\mathcal{H}_k(\mathbf{x}_i^f) - \bar{\mathbf{y}}]^T = \mathbf{Y}\mathbf{Y}^T, \\ \tilde{\mathbf{K}} &= \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} = \mathbf{A}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{R})^{-1}. \end{aligned}$$

This observation matrix-free implementation rely on the secant method approximation $\mathcal{H}(\mathbf{x}_i^f - \bar{\mathbf{x}}^f) \approx \mathcal{H}(\mathbf{x}_i^f) - \bar{\mathbf{y}}$. The forecast is then update to a posterior ensemble $\{\mathbf{x}_i^a\}_{i=1}^N$ such as

$$\mathbf{Z}^a = \mathbf{Z}^f + \tilde{\mathbf{K}}(\mathbf{D} - \mathcal{Y}), \quad (4.7)$$

where $[\mathbf{D}]_i = \mathbf{y} + \boldsymbol{\varepsilon}_i$ is the perturbed observation with $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, $\tilde{\mathbf{K}}$ the ensemble Kalman gain matrix and $(\mathbf{D} - \mathcal{Y})$ the innovation term. The forecast step is then applied to the analyzed ensemble until the following observation. Based on this formulation, we can deduce a correction formula only based on the member's predictions and observations.

We can rewrite the classical update formula using the previous anomaly matrices.

$$\mathbf{Z}^a = \mathbf{Z}^f + \mathbf{A}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \mathbf{R})^{-1}(\mathbf{D} - \mathcal{Y})$$

We reformulate the correction term by remarking that $\mathbf{1}^T \mathbf{Y}^T = \mathbf{0}$. We define \mathbf{F} , the correction matrix that gives the update in terms of linear combinations of the forward states

$$\mathbf{Z}^a a = \mathbf{Z}^f + \mathbf{Z}^f \mathbf{F}, \quad \mathbf{F} = \frac{1}{\sqrt{N-1}} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{R})^{-1} (\mathbf{D} - \mathcal{Y}).$$

using the Sherman-Morrison-Woodbury formula we obtain

$$\mathbf{F} = \frac{1}{\sqrt{N-1}} (\mathbf{I}_N + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{R}^{-1} (\mathbf{D} - \mathcal{Y}).$$

4.B Moment conservation of particle discretization

The m -th moment of a particle distribution is defined as the quantity $\sum_p z_p^\alpha \mathbf{U}_p$.

First, we see that the partition of unity is required

$$\sum_{I \in \Lambda} W \left(\frac{z - z_I}{\ell_I} \right) = 1, \quad z \in \Omega \quad (4.8)$$

due to the final particle arrangement \mathcal{P}' on a grid of size d_p , it leads to the following property

$$\sum_{p' \in \mathcal{P}'} W \left(\frac{z - z_{p'}}{\ell_I} \right) = \frac{V_I}{V_p}, \quad z \in \Omega. \quad (4.9)$$

Attention should be focused on the border. Extending the domain with "ghost" particles or nodes allows for verification of properties within Ω .

This property is the necessary condition for the conservation of the first moment. Primarily for the assignment

$$\begin{aligned} \sum_{I \in \Lambda} \mathbf{u}_I V_I &= \sum_{p \in \Lambda} \mathbf{U}_p W \left(\frac{z_I - z_p}{\ell_I} \right) \\ &= \sum_{p \in \mathcal{P}} \mathbf{U}_p \sum_{I \in \Lambda} W \left(\frac{z_I - z_p}{\ell_I} \right) = \sum_{p \in \mathcal{P}} \mathbf{U}_p. \end{aligned}$$

using the property (4.8). Secondary, for the interpolation process

$$\begin{aligned}
\sum_{p' \in \mathcal{P}'} \mathbf{U}_{p'} &= \sum_{p' \in \mathcal{P}'} \mathbf{u}_g(z_{p'}) V_{p'} = \sum_{p' \in \mathcal{P}'} V_{p'} \sum_{I \in \Lambda} \mathbf{u}_I W \left(\frac{z_{p'} - z_I}{\ell_I} \right) \\
&= \sum_{I \in \Lambda} \mathbf{u}_I V_{p'} \sum_{p' \in \mathcal{P}'} W \left(\frac{z_{p'} - z_I}{\ell_I} \right) \\
&= \sum_{I \in \Lambda} \frac{V_I}{V_p} V_{p'} \mathbf{u}_I = \sum_{I \in \Lambda} \mathbf{u}_I V_I = \sum_{p \in \mathcal{P}} \mathbf{U}_p
\end{aligned}$$

using (4.9).

It can be shown moreover that if for $1 \leq |\alpha| \leq m - 1$, W satisfies,

$$\sum_{I \in \Lambda} (\mathbf{z} - \mathbf{z}_I)^\alpha W \left(\frac{\mathbf{z} - \mathbf{z}_I}{\ell_I} \right) = 0, \quad (4.10)$$

The regridding procedure will be ordered at m . Equivalently, the previous equality lead, for $0 \leq |\alpha| \leq m - 1$, to

$$\sum_{I \in \Lambda} \mathbf{z}_I^\alpha W \left(\frac{\mathbf{z}_p - \mathbf{z}_I}{\ell_I} \right) = \mathbf{z}^\alpha,$$

obtained by developing $(\mathbf{z} - \mathbf{z}_q)^\alpha$ and using a recurrence on previous orders. This means that the interpolation is exact for polynomials of degrees less or equal to $m - 1$ or that the moment of order $m - 1$ is conserved.

4.C Parameters

4.C.1 One dimension problem

Table 4.1: Ensemble generation variables

Variables	Distributions
Gaussian mean	$Z_m \sim \mathcal{N}(\pi/2 + 0.6, 0.5)$
Gaussian standard deviation	$S_m \sim \mathcal{U}(0.8, 1.2)$
velocity	$v \sim \mathcal{N}(0.9, 1.2)$
diffusion	$D \sim \mathcal{U}(0.02, 0.08)$

4.C.2 Two dimension problem

Table 4.2: Reference parameters

Parameters	Values
reference viscosity	$v_{\text{ref}} = 0.001$
reference orientation	$\theta_{\text{ref}} = \frac{7\pi}{8} (\text{rad.})$
barycenter position	$\mathbf{z}_{\text{ref}} = \left[\frac{\pi}{2}, \frac{\pi}{2} \right]^T$
translation velocity	$U_{\text{ref}} = 0.25$

Table 4.3: Nominal assimilation and simulation parameters

Parameters	Values
time step	$dt = 0.005$
final time	$t_f = 10$
std. observation	$\sigma_{\text{obs}} = 5.0e^{-2}$
vorticity threshold	$\varepsilon_{\omega} = 1.0e^{-4}$
particle characteristic length	$dp = \frac{pi}{256} \approx 0.01227$
smoothing length	$h = 2.0dp$
number of assimilation	$N_{\text{assim}} = 10$
ensemble size	$N_{\text{ens}} = 32$
number of observation	$N_{\text{obs}} = 12^2 = 144$
grid discretization	$N_{\text{grid}} = 65^2 = 4225$
number of remeshing by forecast	$N_{\text{remesh}} = 2$

Table 4.4: Ensemble generation variables

Variables	Distributions
radius	$R \sim \mathcal{N}(1.0, 0.05^2)$
orientation	$\theta \sim \mathcal{U}\left(\pi, \frac{\pi}{2}\right) (\text{rad.})$
barycenter	$z_{\text{mean},x} \sim \mathcal{N}\left(\frac{\pi}{2}, 0.1^2\right), \quad z_{\text{mean},y} \sim \mathcal{N}\left(\frac{\pi}{2}, 0.1^2\right)$
velocity	$U \sim \mathcal{U}(0, 0.5^2)$
viscosity	$v \sim \mathcal{N}(0.0015, 0.0005^2)$

Méthodes d'assimilation de donnée par correction de position pour des simulations lagrangiennes

Contents

5.1	Résumé	106
5.2	Abstract	106
5.3	Introduction	106
5.4	Background	110
5.4.1	Model setting	110
5.4.2	Classical ensemble data assimilation method	111
5.4.3	Ensemble Filtering with Displacement Error Adjustment	112
5.5	Methods	114
5.5.1	Position transformation	114
5.5.2	Reduced-order modeling based on ensemble decomposition	115
5.5.3	Gradient computation	116
5.5.4	Selection of the penalization coefficient	117
5.6	Applications	117
5.6.1	Problem description	117
5.6.2	Assimilation of vortex centers	118
5.6.3	Amplitude filters	120
5.6.4	Alignment step	121
5.6.5	Assimilation with position-strength correction	125
5.6.6	Filters performances and parametric studies	127
5.7	Conclusion	129
5.A	Parameters	130
5.B	Gradient computation	131

5.1 Résumé

Le filtre de Kalman présenté en Section 2.1.2 est un filtre séquentiel adéquat pour appliquer des méthodes d’assimilation pour des modèles de grande dimension et non-linéaire. Il consiste à mettre à jour un ensemble d’état en déterminant un gain de Kalman pour combiner l’état prédit et un terme d’innovation fonction de l’erreur de prédiction.

Toutefois, la mise à jour du filtre de Kalman d’Ensemble est dépendante de la discréétisation de l’état de ses membres pour le calcul du gain de Kalman d’ensemble. De plus, la mise à jour du filtre de Kalman d’Ensemble est basée sur une combinaison linéaire des membres au travers du gain de Kalman d’ensemble, celle-ci entraîne une explosion du nombre de particules dans la définition de la solution analysée.

L’objectif de ce chapitre est donc de présenter un certain nombre d’adaptation du filtre de Kalman d’Ensemble qui puisse être appliquées à des simulations lagrangiennes.

Nous reformulons tout d’abord l’expression de la mise à jour afin qu’elle soit indépendante de la définition de l’état. Enfin, les solutions analysées étant des combinaisons linéaires des particules de l’ensemble des membres, nous proposons de développer des méthodes pour réduire l’augmentation exponentielle des particules par le biais de deux développement.

5.2 Abstract

We introduce a novel data assimilation formulation for Lagrangian simulations. Specifically, our method addresses problems requiring the modification of the position of particle discretized the solution. We treat the correction of intensity and position separately, employing distinct methodologies for each. To achieve this, we developed a sequential formulation that includes an alignment correction step through the integration of a velocity field, followed by an intensity correction using the EnKF as previously described. Our ensemble-based approach minimizes a set of independent, low-rank cost functions for cost efficiency. Importantly, this method provides a suitable framework for techniques that cannot remesh a new particle discretization and require position updates consistent with physical principles. We test the algorithm to solve the Navier-Stokes equations with the Vortex method on a 2D problem.

5.3 Introduction

Data assimilation is a crucial process employed across various scientific disciplines, including meteorology, oceanography, and hydrology, to integrate observed data with the outputs of numerical models. The goal of this integration is to provide an optimal and consistent representation of a system state by continuously refining model predictions based on newly acquired observations [Bocquet, 2014]. Significant progress has been made in applying data assimilation techniques to high-dimensional and non-linear problems. One notable example is the Ensemble Kalman Filter (EnKF), introduced by Evensen [Evensen, 1994], which utilizes an ensemble of realizations to perform sequential

updates. This approach is particularly well-suited for complex systems with inherent uncertainties. Alternatively, variational data assimilation methods, such as 3DVar and 4DVar, address the assimilation challenge by minimizing a cost function that quantifies the discrepancy between model predictions and observations [Talagrand, 1997]. Hybrid methods have also emerged, aiming to combine the strengths of ensemble and variational approaches by minimizing an ensemble of cost functions. Ensemble methods are particularly valued for their efficiency in estimating the posterior distribution through a low-dimensional representation, making them increasingly popular in high-dimensional assimilation problems [Hoteit and Köhl, 2006; Hunt et al., 2007; Liu et al., 2008]. Although these methods have predominantly been applied to problems defined on fixed Eulerian meshes, their application to Lagrangian (or meshless) simulations requires further development. Unlike Eulerian approaches, where the mesh is fixed, Lagrangian methods discretize continuous fields and operators using particles that move with the flow [S. Li and W. K. Liu, 2004]. Lagrangian methods are particularly advantageous in situations where traditional mesh-based approaches fail to represent complex physical phenomena accurately. This is especially true for problems involving large deformations [Chen et al., 1997; Zhang et al., 2014], complex geometries [Klaas and Shephard, 2000; Brydon et al., 2005; Bardenhagen, Brydon, et al., 2005], fragmentation [Rao and Rahman, 2004; Ma et al., 2009], free-surface flows [Monaghan, 1994], multi-phase flows [Stomakhin et al., 2014], granular flows [Kumar et al., 2017; Dunatunga and Kamrin, 2015], and various applications in geotechnical engineering [Andersen and Andersen, 2010; Sołowski and Sloan, 2015], biomechanics [Guilkey et al., 2006], manufacturing [Ambati et al., 2012; Russell et al., 2018], and computer graphics [Jiang, Schroeder, Teran, et al., 2016].

Applying ensemble methods to Lagrangian frameworks presents significant challenges when handling an ensemble of realizations discretized on different particle configurations. In ensemble filtering, the analysis depends on evaluating ensemble statistics, making it challenging to define a consistent state representation and appropriately update ensemble members. Although some approaches attempt to incorporate the positions and strengths of particles within the state definition [Chen et al., 2022; Darakananda et al., 2018], these methods face difficulties when dealing with ensembles that have varying numbers of particles. Furthermore, the field is non-linear with respect to particle positions, which can significantly violate the Gaussian assumptions typically used in ensemble methods. Recent work [Duvillard et al., 2024] has introduced modifications to the classical Ensemble Kalman Filter to address these issues by using a discretization-free update formula and proposing a strength correction approach. Two formulations have been proposed: the Remesh-EnKF, which generates an identical discretization of particle positions, and the Part-EnKF, which focuses solely on updating particle strengths on the given particle discretization. The first approach effectively reverts to an Eulerian representation of the field, while the second strives to maintain a consistent Lagrangian framework.

However, it is crucial that the prior state and observation fields sufficiently cover the solution being analyzed. Traditional data assimilation methods primarily rely on interpolation within the space of field values, restricting the analysis to the support of the background state and the observations. This limitation is particularly evident in

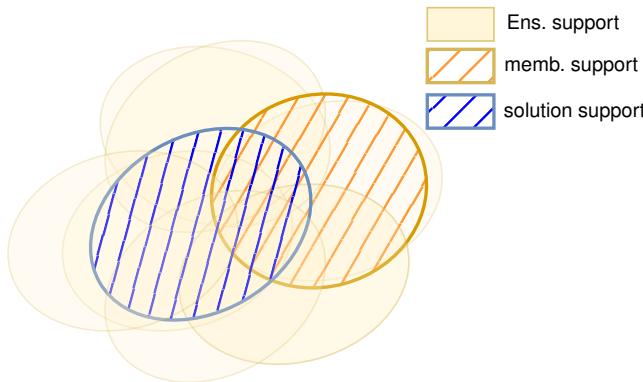


Figure 5.1: Although the union of the particle supports of all ensemble members overlaps with the true solution, individual ensemble members may have disjoint supports that do not fully cover the solution.

the context of the Ensemble Kalman Filter (EnKF), which explicitly combines ensemble members [Hunt et al., 2007]. As a result, this constraint directly relates to our problem of particle discretization support, making the positions of particles a critical factor during the assimilation step. Specifically, the Part-EnKF filter proposed in [Duvillard et al., 2024], which corrects intensities without altering particle positions, projects an analyzed solution onto a particle support that may not be suitable for that solution as illustrated in Figure 5.1. Although projecting onto an Eulerian representation (as with the Remesh-EnKF filter) remains a viable option, our goal is to propose a formulation that updates the particle positions for each ensemble member, thereby preserving the particle-based representation.

Our objective is to develop an assimilation scheme that enables modifications to both the position and intensity of particles. We assume that the position correction must adhere to kinematically admissible steps, as it is designed to capture the effects of unresolved or unmodeled phenomena that are physically constrained within the system. To address this, we assume that the issue of inadequate support arises from a misalignment error between the ensemble members and the solution. In addition to adjusting the interpolation space of the intensities, it is also necessary to identify a transformation to correct this positional error. The literature addresses several developments related to positional errors, particularly in the analysis of coherent structures such as hurricanes, storms, fire fronts, squall lines, and rain fronts [Plu, 2013; Michel, 2011; Ravela, 2014; Chen and Snyder, 2007a]. Nevertheless, current assimilation methods still fall short in adequately addressing these errors. These methods predominantly focus on correcting field intensities, with errors assessed using an L_2 norm. However, this approach tends to overestimate errors arising from misalignment, as illustrated in Figure 5.2. This issue is known as the *double penalty* effect, which adversely affects both the evaluation of model errors and the assessment of observational errors [Amodei and Stein, 2009]. The double penalty effect significantly contributes to representation error, as highlighted by Janjić

et al. [Janjić et al., 2018].

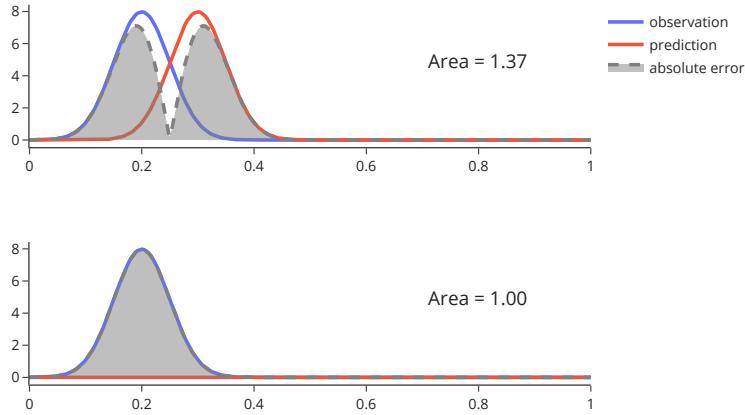


Figure 5.2: Illustration of the double penalty effect: A shift in the solution results in an error measurement that exceeds the error associated with the null function.

Several approaches have been proposed to address this difficulty. An elegant method draws inspiration from optimal transport techniques to define a correction in an interpolation space that considers both intensities and displacements [Villani et al., 2009; Benamou and Brenier, 2000]. This method involves substituting the quadratic norm with the Wasserstein distance. For instance, Feyeux’s thesis [Feyeux, 2016] investigated the use of Wasserstein distance for data assimilation in images. From there, Bocquet et al. [Bocquet et al., 2024] adapted the 3DVar method to optimal transport by applying this approach to state and observation distributions with potentially differing masses. Although these methods are increasingly computationally efficient [Cuturi and Doucet, 2014; Peyré and Cuturi, 2019; Simsekli et al., 2018], they remain in their early stages of data assimilation.

Other approaches incorporate an alignment variable determined through displacement assimilation methods. For instance, Percival et al. [Percival, 2008] solve a rearrangement problem in order to apply a coordinate transformation. In the case the structure features are directly described, Chen et al. [Chen and Snyder, 2007b] propose directly assimilating the position of a tornado using the Ensemble Kalman Filter (EnKF). When the features are implicit, Ravela et al. [Ravela et al., 2007] proposed a joint adjustment model for position and amplitude based on ensembles by introducing grid deformation. From an ensemble variational approach, a two-step sequential approach is proposed. Their method has the advantage of not relying on the assumption of independence or joint normality of the amplitude and position errors. This method has also been successfully adapted to multi-scale formulations [Ying, 2019; Ying et al., 2023]. Finally, [Rosenthal et al., 2017] developed a two-step sequential method to align and correct intensities

successively. Their goal is to preserve the morphological properties of vortices by first applying a kinematically admissible transformation to correct the position, followed by an intensity correction. This approach has the advantage of being compatible with classical assimilation methods for the final step.

Building specifically upon the works of [Ravela et al., 2007; Rosenthal et al., 2017], we propose a sequential two-step scheme to correct successively the position and the intensity of particle discretizations. We utilize the vortex method as described 5.4.1 to solve the Navier-Stokes equations for an inviscid, incompressible flow. In Section 5.4.2, we provide background on the intensity correction algorithm for Lagrangian simulations, followed by the field alignment paradigm introduced by Ravela [Ravela et al., 2007]. In Section 5.5, our method involves introducing a kinematically admissible alignment transformation by integrating a velocity field over an arbitrary time interval. In addition to applying for the first time alignment assimilation with Lagrangian simulation, the novelty of our approach lies in the introduction of a search space and regularization that leverages ensemble velocity fields. This allows the alignment step, which involves a non-linear optimization problem, to be solved within a low-dimensional space. For the intensity correction step, we apply the methods developed in [Duvillard et al., 2024]. In Section 5.6, we implement this method for a problem involving three vortices discretized using multiple particles. Our primary metric is the error in the positions of the vortex centers. We will demonstrate that our alignment approach enables filter convergence even when the discretization supports of the ensemble members are first disjoint.

5.4 Background

5.4.1 Model setting

In this article, we employ the vortex method to solve the Navier-Stokes equation for inviscid incompressible flow. The Vortex Method [Cottet and Koumoutsakos, 2000a] is a Lagrangian method that uses an ensemble of particles to discretize the vorticity field ω . The Euler equation in the velocity-vorticity formulation is

$$\begin{cases} \frac{\partial \omega}{\partial t} + (\mathbf{v} \cdot \nabla) \omega = \mathbf{0}, \\ \Delta \mathbf{v} = -\nabla \times \omega, \end{cases} \quad (5.1)$$

where \mathbf{u} is the velocity field such as $\omega = \nabla \times \mathbf{v}$. In the 2d case, the vorticity field ω is carried by the unit vector normal to the plane \mathbf{k} such as $\omega = \omega \mathbf{k}$.

We express the particle discretization of the vorticity field thanks to a particle set $\mathcal{P} = \{(\mathbf{x}_p, \Gamma_p)\}_{p=1}^{N_p}$ with position \mathbf{x}_p and strength Γ_p such as

$$\omega(\mathbf{x}, t) = \sum_{p \in \mathcal{P}} \Gamma_p \phi_\epsilon(\mathbf{x} - \mathbf{x}_p(t)),$$

where ϕ_ϵ is a smoothing kernel that particularly verifies the unit integral property.

We use the Vortex-In-Cell formulation that uses operator projection/interpolation on the grid to solve the Poisson equation $\Delta \boldsymbol{v} = -\nabla \times \boldsymbol{\omega}$ and determined \boldsymbol{v} . The equation of evolution is defined by using the Lagrangian version of (5.1) such as

$$\begin{cases} \frac{d\boldsymbol{x}_p(t)}{dt} = \boldsymbol{v}(\boldsymbol{x}_p(t), t), \\ \frac{d\Gamma_p(t)}{dt} = 0. \end{cases}$$

Through this integration scheme, we can assume that a positional error may occur during the determination of the velocity field \boldsymbol{v} due to uncertainties in the intensities of the particles and their positions.

During the assimilation process, we measure as observations the velocity field \boldsymbol{v} on a coarse, regular grid. Each coordinate is supposed to be normally distributed with a standard deviation of σ_{obs} .

5.4.2 Classical ensemble data assimilation method

Data assimilation methods are usually defined as the estimation of the posterior distribution of the state $\boldsymbol{\omega}$ given the observations \boldsymbol{y} .

After each acquisition of observations, the posterior distribution is obtained using Bayes formula

$$p(\boldsymbol{\omega} | \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{\omega})p(\boldsymbol{\omega}), \quad (5.2)$$

where $p(\boldsymbol{\omega}) = \mathcal{N}(\boldsymbol{\omega}^f, \boldsymbol{P}^f)$ is the prior state distribution and $p(\boldsymbol{y} | \boldsymbol{\omega}) = \mathcal{N}(\mathcal{H}(\boldsymbol{\omega}^f), \boldsymbol{R})$ is the likelihood, both assumed to be Gaussian. The Maximum A Posteriori (MAP) is obtained by minimizing the negative log-likelihood

$$\mathcal{L}_{\boldsymbol{\omega}}(\boldsymbol{\omega}) = \frac{1}{2} \|\boldsymbol{u} - \boldsymbol{H}(\boldsymbol{\omega})\|_{\boldsymbol{R}}^2 + \frac{1}{2} \|\boldsymbol{\omega} - \boldsymbol{\omega}^f\|_{\boldsymbol{P}^f}^2,$$

where $\|\cdot\|_{\boldsymbol{A}}^2$ denote the weight norm defined by $\|\boldsymbol{v}\|_{\boldsymbol{A}}^2 = \boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{v}$ for any positive-definite \boldsymbol{A} . To estimate not only the MAP but also to sample from the posterior distribution, an ensemble method based on Randomized Maximum Likelihood (RML) sampling can be employed, as proposed by [Kitanidis, 1995] and [Oliver et al., 1996]. In the weakly non-linear case, the posterior distribution is approximately sampled by minimizing the following ensemble of cost functions

$$\mathcal{L}_{\boldsymbol{\omega},i}(\boldsymbol{\omega}) = \frac{1}{2} \|\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\omega})\|_{\boldsymbol{R}}^2 + \frac{1}{2} \|\boldsymbol{\omega} - \boldsymbol{\omega}_i^f\|_{\boldsymbol{P}^f}^2, \quad \text{for } i = 1, \dots, N, \quad (5.3)$$

which differ from the cost function $\mathcal{L}_{\boldsymbol{\omega}}$ by substituting \boldsymbol{y} with $\boldsymbol{d}_i \sim \mathcal{N}(\boldsymbol{y}, \boldsymbol{R})$, representing the perturbed measurements, and the mean $\boldsymbol{\omega}^f$ with random samples $\boldsymbol{\omega}_i^f \sim \mathcal{N}(\boldsymbol{\omega}^f, \boldsymbol{P}^f)$.

The solutions ω_i^a of the minimization problem form a new ensemble that is independently propagated with the physical model until the next observation.

Ensemble filtering method for Lagrangian simulation

The main drawback of Lagrangian simulation is that each ensemble member is discretized on its own grid, which prevents directly solving equation (5.3). This limitation has been addressed in the work [Duvillard et al., 2024] by utilizing the Ensemble Kalman Filter (EnKF) update. The EnKF, a widely used filtering method introduced by Evensen [Evensen, 1994], provides a solution to (5.3) in the form of the Kalman update [Kalman, 1960]. The key assumptions include using the sample covariance matrix \mathbf{P}^f , linearizing the observation operator \mathcal{H} , and approximating the sensitivity $\nabla \mathcal{H}$ with an ensemble average, as developed in Chapter 8 [Evensen et al., 2022b]. Through this approach, the updated states ω_i^a are directly determined as a linear combination of the ensemble members, such that for $\mathbf{x} \in \Omega$

$$\omega_i^a(\mathbf{x}) = \omega_i^f(\mathbf{x}) + \sum_j \mathbf{F}_{ji} \omega_j^f(\mathbf{x}) \quad i = 1, \dots, N ,$$

where \mathbf{F} is defined as

$$\mathbf{F} = \frac{1}{\sqrt{N-1}} (\mathbf{I}_N + \mathbf{Y}^T \mathbf{R}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{R}^{-1} (\mathbf{D} - \mathcal{Y}),$$

where \mathbf{Y} is the matrix where each column contains the centered prediction of each realization.

Thus, the update relies solely on the observation ω , the predictions $\{\mathbf{y}_i = \mathcal{H}(\omega_i^f)\}_{i=1}^N$, and noise samples $\{\varepsilon_i\}_{i=1}^N$ such as $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. In [Duvillard et al., 2024], two formulations of the EnKF filter are introduced, depending on the choice of discretization for ω_i^a while a particle discretization. On one hand, the Remesh-EnKF approximates ω_i^a by generating a new regular set of particles. On the other hand, the Part-EnKF retains the particle positions of each member and applies linear regression or particle approximation. While the Part-EnKF offers a method to address the assimilation problem by preserving the original discretization, it may need to introduce adaptations to approximate ω_i^a accurately. This issue primarily arises when the support of each ensemble member is disjoint. Consequently, it is necessary to consider the correction of particle positions within the assimilation scheme.

5.4.3 Ensemble Filtering with Displacement Error Adjustment

We first address the alignment error by decomposing the state error as $\boldsymbol{\eta} = \boldsymbol{\eta}^d + \boldsymbol{\eta}^a$, where $\boldsymbol{\eta}^d$ represents the displacement or position error and $\boldsymbol{\eta}^a$ represents the amplitude error.

A common approach to address the displacement error is to introduce an alignment control variable Ψ . We define Ψ as an invertible mapping from the deformed space to the aligned space. Consequently, the aligned field is defined as $\omega_\Psi(x) = \omega(\Psi^{-1}(x))$.

Similarly to (5.2), the posterior distribution is given by

$$p(\boldsymbol{\omega}, \boldsymbol{\Psi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\omega}, \boldsymbol{\Psi}) p(\boldsymbol{\omega} | \boldsymbol{\Psi}) p(\boldsymbol{\Psi}),$$

where the terms are respectively the likelihood, the amplitude prior conditioned by the alignment operator, and the hyper prior of the alignment operator.

The term $p(\boldsymbol{\Psi})$ that describes the prior distribution of the mapping is usually formulated on some physical constraint to regularize the function $\boldsymbol{\Psi}$. Usually, it is supposed to be written as an exponential distribution

$$p(\boldsymbol{\Psi}) \propto \exp(-L(\boldsymbol{\Psi})),$$

where L is a regularization function. For instance, in [Ravela et al., 2007], the transformation is penalized for smoothness, while in [Rosenthal et al., 2017], the deformation is penalized for avoiding high distortion. Therefore, there is no unique way, and the choice of regularization depends on the problem.

We get from the negative log-likelihood the following cost function

$$\mathcal{L}(\boldsymbol{\omega}, \boldsymbol{\Psi}) = \frac{1}{2} \|y - h(\boldsymbol{\omega}_{\boldsymbol{\Psi}})\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\boldsymbol{\omega}_{\boldsymbol{\Psi}} - \boldsymbol{\omega}_{\boldsymbol{\Psi}}^f\|_{\mathbf{P}^f(\boldsymbol{\Psi})^{-1}}^2 + L(\boldsymbol{\Psi}) - \frac{1}{2} \ln(|\mathbf{P}^f(\boldsymbol{\Psi})|).$$

Finally, the analyzed field is $\boldsymbol{\omega}^a = \boldsymbol{\omega}' \circ \boldsymbol{\Psi}'^{-1}$, where $(\boldsymbol{\omega}', \boldsymbol{\Psi}')$ is the minimizer of the previous cost function.

A two-step approach is generally used [Ravela et al., 2007; Rosenthal et al., 2017; Ying, 2019] to determine $\boldsymbol{\omega}'$ and $\boldsymbol{\Psi}'$ sequentially:

- **alignment step:** An approximation of the stationary point is achieved by neglecting the dependency of \mathbf{P}^f on $\boldsymbol{\Psi}$ and assuming that $\boldsymbol{\omega} = \boldsymbol{\omega}^f$ such as

$$\mathcal{L}_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}) = \frac{1}{2} \|y - h(\boldsymbol{\omega}_{\boldsymbol{\Psi}}^f)\|_{\mathbf{R}^{-1}}^2 + L(\boldsymbol{\Psi})$$

- **amplitude step:** $\boldsymbol{\Psi}$ is fixed to $\boldsymbol{\Psi}'$ the solution from the alignment step. $\mathbf{P}^f(\boldsymbol{\Psi}')$ is evaluated and $\boldsymbol{\omega}$ varied given

$$\mathcal{L}_{\boldsymbol{\omega}}(\boldsymbol{\omega}_{\boldsymbol{\Psi}'}) = \frac{1}{2} \|y - h(\boldsymbol{\omega}_{\boldsymbol{\Psi}'})\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\boldsymbol{\omega}_{\boldsymbol{\Psi}'} - \boldsymbol{\omega}_{\boldsymbol{\Psi}'}^f\|_{\mathbf{P}^f(\boldsymbol{\Psi}')^{-1}}^2.$$

In this way, the two errors are supposed to be decomposed such as $\boldsymbol{\eta}^a = \boldsymbol{\omega}_{\boldsymbol{\Psi}'}^f - \boldsymbol{\omega}_{\boldsymbol{\Psi}'}$ and $\boldsymbol{\eta}^p = \boldsymbol{\omega}_{\boldsymbol{\Psi}'} - \boldsymbol{\omega}^t$ where $\boldsymbol{\omega}^t$ is the true solution.

As in equation 5.3, the posterior distribution can be sampled by applying the alignment process to an ensemble of realizations. Additionally, the amplitude step corresponds to the classical variational cost function, as previously defined in Section 5.4.2, but is applied within the aligned coordinate space. This means that the amplitude step may be performed by using classical filters on the state $\boldsymbol{\omega}_{\boldsymbol{\Psi}'}$.

5.5 Methods

Our contribution focuses on defining an alignment process to correct both particle positions and amplitudes in Lagrangian simulations. We build on the two-stage formulation previously introduced, which allows for sequential correction of position and amplitude. Our approach leverages the methods discussed in Section 5.4.2 for the amplitude correction.

Here, we present only the alignment step. We correct for particle positions by specifying the alignment operator Ψ and outlining the procedure for minimizing the associated cost function.

5.5.1 Position transformation

To define the position correction, we introduce a specific mapping Ψ that represents a coordinate transformation. The new field obtained can be obtained by evaluating $\omega \circ \Phi^{-1}$. Alternatively, the aligned field can be derived by pushing forward the discretization supports by the particle positions. Specifically, if $\omega(x) = \sum_{p \in \mathcal{P}} \Gamma_p \phi_\varepsilon(\mathbf{x} - \mathbf{x}_p)$, the updated state is given by $\omega_\Psi(\mathbf{X}) = \sum_{p \in \mathcal{P}} \Gamma_p \phi_\varepsilon(\mathbf{X} - \Psi(\mathbf{x}_p))$.

We choose this transformation to respect the physical constraints of our problem. One of our assumptions is that the alignment error of the particles arises from inaccuracies in the integration of their positions due to perturbations in the velocity field calculation.

In the case of the vortex method, which assumes incompressible flow we define the mapping Ψ by integrating a constant divergence-free velocity $\mathbf{u}(\cdot; \mathbf{a})$ over an arbitrary time interval. Therefore, Ψ is the solution to the following ordinary differential equation (ODE)

$$\begin{cases} \mathbf{x}'(\tau = 0) = \mathbf{x}, \\ \frac{d\mathbf{x}'}{d\tau}(\tau) = \mathbf{u}(\mathbf{x}'(\tau); \mathbf{a}), \quad \Psi(\mathbf{x}) = \mathbf{x}'(\tau = 1). \end{cases} \quad (5.4)$$

By applying this type of transformation for position correction, we ensure that the corrected positions remain physically consistent. More precisely, this approach guarantees that the local volume is preserved during the correction process. Consequently, the problem of position correction involves determining the field $\mathbf{u}(\cdot; \mathbf{a})$, which is parametrized by the vector \mathbf{a} and depends on the discretization of this field.

Alignment step

The cost function defined in (5.4.3) can be reformulated using the previously defined transformation. For each member $\left\{ \omega_i^f \right\}_{i=1}^N$ we generate the following cost functions based on the RML sampling

$$\mathbf{u}_i = \arg \min_{\mathbf{u} \in \mathcal{U}} \left\| \mathcal{H}(\omega_i^f | \mathbf{u}) - \mathbf{d}_i \right\|_R^2 + \mathcal{R}(\mathbf{u}),$$

where \mathcal{U} is the velocity search space, $\mathcal{H}(\omega_i^f \mid \mathbf{u})$ is the observation operator conditioned by \mathbf{u} . Specifically, $\mathcal{H}(\omega^f \mid \mathbf{u}) = \mathcal{H}(\omega_i^f \circ \Psi^{-1}(\cdot; \mathbf{u}))$ represents the observation of ω^f after applying the position correction. The regularization based on Φ is replaced by a regularization on the integrated field $\mathcal{R}(\mathbf{u})$. This term became essential during optimization to avoid the problem from being ill-posed. In our case, because streamlines form closed loops in our finite domain, the value of \mathbf{u} will not be unique. Moreover, regularization helps to be less sensitive to noise and ill-conditioning.

Moreover, in order to address the nonlinear nature of the observation operator with respect to the coordinate transformation, the optimization approach must differ from the amplitude correction strategy described in Section 5.4.2. Unlike the linear approximation leveraged by the EnKF, we employ an iterative gradient descent algorithm for minimization. To effectively solve the problem, we still need to define a suitable search space \mathcal{U} that satisfies the previously stated assumptions, along with an appropriate regularization term to ensure stable and meaningful solutions.

This approach ensures that the optimization process can effectively navigate the complexities introduced by the nonlinear relationship between the state variables and the observations, thereby improving the robustness and accuracy of the position correction mechanism.

5.5.2 Reduced-order modeling based on ensemble decomposition

The choice of the search space and the regularization is directly tied to our prior assumption that alignment errors originate from perturbations in the integration of the velocity field during the forecast. Therefore, as a first approximation, we confine our search for the velocity correction field \mathbf{u} to the space spanned by the member velocity fields at the end of the forecast. This space is defined as $V = \text{Span}(\{\mathbf{v}_j\}_{j=1}^N)$, where \mathbf{v}_j is the velocity field induced by the vorticity of member j . By construction, the fields \mathbf{v}_j are divergence-free, ensuring that the search space also maintains this property due to the linearity of the gradient operator. To express this formally, we introduce vectors $\mathbf{a} \in \mathbb{R}^N$, such that the velocity correction field can be represented as

$$\mathbf{u} = \sum_{j=1}^N a_j \frac{\mathbf{v}_j}{\sqrt{N}}. \quad (5.5)$$

The regularization is implemented by introducing a prior on the coefficients of this decomposition. As a first approximation, we assume that the alignment is unbiased and that equal weight is assigned to each field \mathbf{v}_i in the decomposition. Consequently, we assume that the coefficients a_j are independent and identically distributed such as $a_j \sim \mathcal{N}(0, 1/\lambda^2)$. This leads to a probability distribution of \mathbf{a} given by

$$p(\mathbf{a}) \propto \exp\left(-\frac{\lambda}{2} \|\mathbf{a}\|_2^2\right).$$

This assumption results in a Tikhonov regularization, also known as Ridge regression,

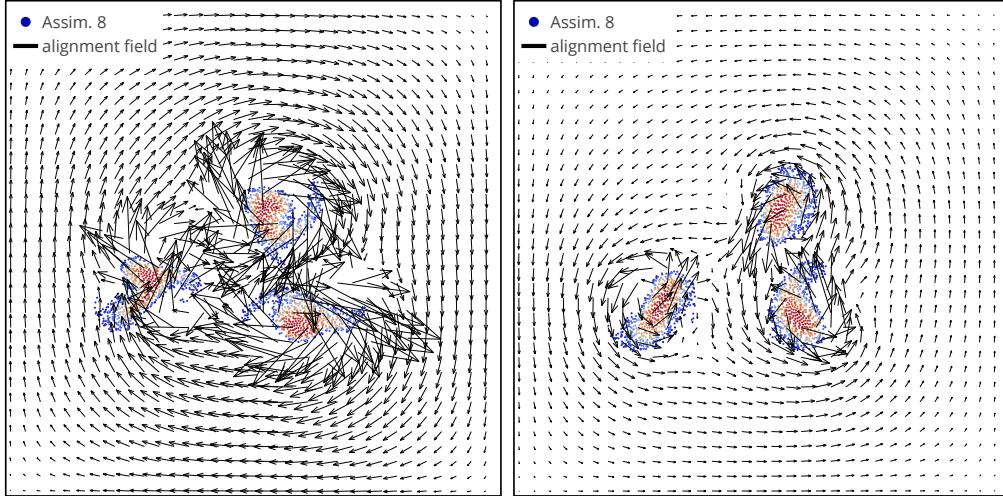


Figure 5.3: Alignment fields and particle distributions with and without regularization in the simulation. Without regularization, the field is noisier, leading to distortions in particle distribution.

where the regularization term $\mathcal{R}(\mathbf{a})$ is defined as

$$\mathcal{R}(\mathbf{a}) = \frac{\lambda}{2} \|\mathbf{a}\|_2^2,$$

Ridge regression is a common choice for avoiding overfitting by penalizing large coefficient values, thereby stabilizing the solution, mainly when the \mathbf{v}_i fields are highly correlated. In fact, this form of regularization corresponds to regularizing the field \mathbf{u} with $\mathcal{R}(\mathbf{u}) = \lambda \|\mathbf{u}\|_{\mathbf{B}}^2$, where the metric $\mathbf{B} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_j \otimes \mathbf{v}_j$. This reflects how much each direction in the space is amplified or attenuated by the transformation given by \mathbf{B} . Figure 5.3 shows that a non-regularized solution is usually unrealistic.

These choices lead to the formulation of N independent minimizing problems with $\mathbf{a} \in \mathbb{R}^N$,

$$\mathcal{L}_i(\mathbf{a}) = \|h(\mathbf{a}) - \mathbf{d}_i\|_{\mathbf{R}}^2 + \frac{\lambda}{2} \|\mathbf{a}\|_2^2,$$

where $h(\mathbf{a})$ is the observation operator conditioned by (\mathbf{a}, V) such as

$$h(\mathbf{a}) = \mathcal{H}(\boldsymbol{\omega}_i^f \mid \mathbf{a}, V) = \mathcal{H} \left(\boldsymbol{\omega}_i^f \circ \Psi^{-1} \left(\cdot; \mathbf{u} = \sum_{j=1}^N a_j \mathbf{v}_j \right) \right).$$

5.5.3 Gradient computation

The cost function is highly nonlinear with respect to the coefficient \mathbf{a} . Therefore, we decided to use a local descent gradient algorithm to minimize it. The gradient of the

cost function with respect to \mathbf{a} is given as

$$\nabla_{\mathbf{a}} \mathcal{L}_i(\mathbf{a}) = 2\nabla_{\mathbf{a}} h(\mathbf{a}) \mathbf{R}^{-1}(h(\mathbf{a}) - \mathbf{y}_i) + \lambda \mathbf{a}.$$

A straightforward way to compute $\nabla_{\mathbf{a}} h(\mathbf{a})$ is by using finite differences

$$\frac{\partial h}{\partial a_j}(\mathbf{a}) = D_{\mathbf{v}_j} \mathbf{h}(\mathbf{a}) \approx \frac{\mathbf{h}(\mathbf{a} + \delta a_j \mathbf{e}_j) - \mathbf{h}(\mathbf{a})}{\delta a_j},$$

where \mathbf{e}_i is the unit vector along the direction of \mathbf{v}_i . This approach requires evaluating the integration of the velocity field $N + 1$ times. An alternative method involves evaluating the gradient during a single model evaluation by propagating N perturbations around a main trajectory, as detailed in Appendix 5.B. This method efficiently computes the gradient by leveraging information from the perturbations, thereby reducing computational costs.

5.5.4 Selection of the penalization coefficient

The coefficient λ is a hyperparameter that controls the variability of the field \mathbf{u} . It has been introduced to penalize the decomposition coefficients and represent the relative weight of the regularization term compared to the discrepancy measure between observation and prediction. However, it depends on both the time integration, which is arbitrarily set to 1 in (5.4), and the norm of the member velocity fields \mathbf{v}_i . The normalization of the decomposition by N ensures that it is independent of the size of the ensemble. Therefore, this hyperparameter must be carefully calibrated to avoid overfitting. Several methods can be used for this purpose, including L-curve analysis [Hansen, 2000], cross-validation [Dubrule, 1983], or the discrepancy principle [Blanchard and Mathé, 2012]. In our study, we qualitatively selected the coefficient that facilitated correction while minimizing significant distortion to the problem under investigation.

5.6 Applications

5.6.1 Problem description

We assess the performance of the filters by using the three-vortex problem [Aref, 1979; Yim et al., 2022]. This problem is analogous to the N-body problem in celestial mechanics. Notably, Henri Poincaré demonstrated that problems involving three bodies exhibit sensitivity to initial conditions, which laid the groundwork for the development of modern chaos theory [Poincaré, 1890; Diacu and Holmes, 1996].

In our case, three vortices are initially placed within a domain of size $[0, \pi]^2$. Each vortex follows the distribution of a Bessel vortex [Van Geffen et al., 1996]. A Bessel vortex is defined by a continuous vorticity field on a circle of radius R , such as

$$\omega(r) = \begin{cases} \Gamma J_0\left(\frac{kr}{R}\right), & r < R, \\ 0 & \text{else,} \end{cases}$$

where J_0 represents the first Bessel function, k is its first non-trivial zero of J_0 , r denotes the radial distance from the vortex center, and Γ is the vortex strength.

In an unbounded domain, a single Bessel vortex is a stationary solution to the Euler equations. This vortex rotates about its center with a constant angular velocity of $\frac{\Gamma}{2\pi}$. In a system of vortices, each vortex generates a local velocity field that causes it to rotate rapidly around itself. In addition to this self-induced rotation, the vortex experiences a background velocity field generated by the influence of other vortices and boundary conditions. This leads each vortex to move around each other. The Figures 5.4a and 5.4c show the initial condition respectively with the induced vorticity field and the resulting trajectories of three vortices center over time in a case the vortex has the configuration described in Table 5.2.

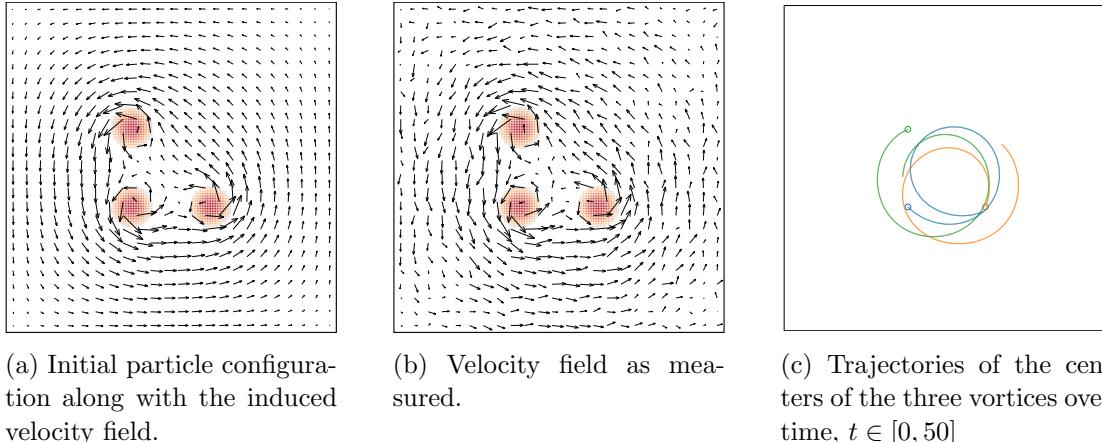


Figure 5.4: Initial conditions and trajectories of the reference three-vortex problem. The figure illustrates the influence of the vortex interactions and boundary conditions on the movement of the vortex centers.

5.6.2 Assimilation of vortex centers

Due to the inherent chaotic nature of the problem, even minor perturbations in the initial conditions lead to a significant amplification of errors in the vortex trajectories. As an illustration, we generate an ensemble of perturbed members based on the distribution provided in Table 5.2. Even if the positions of the vortex centers, the core size R , as well as their amplitudes Γ , are not too far from the reference configuration, the Figure 5.5 shows that at the end of the forecast, the vortices are extensively dispersed. They are no longer closely aligned with the reference three-vortex configuration. The primary goal of this analysis is to track the positions of the vortices over time and reduce errors in the vorticity field.

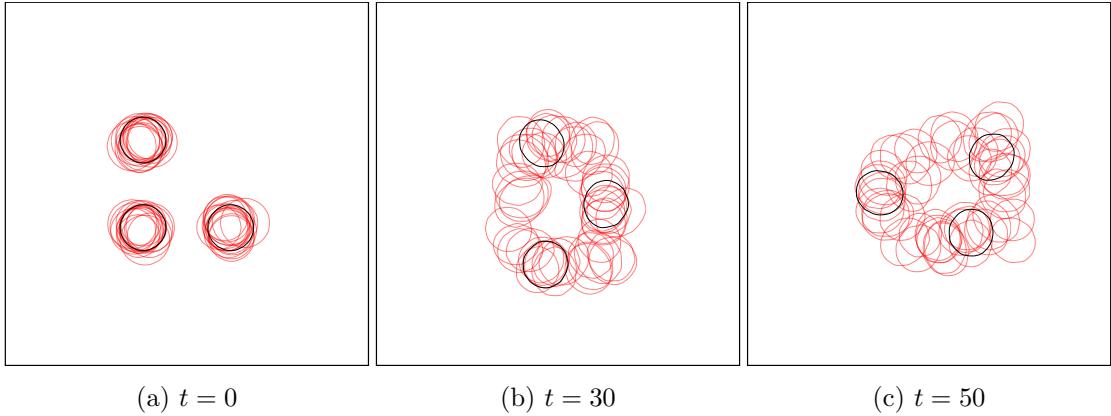


Figure 5.5: Evolution of the dispersion of the ensemble without assimilation. The figures illustrate the dispersion at different time points: $t = 0$, $t = 30$, and $t = 50$.

Assimilation parameters

We observe a noisy coarse grid with dimensions $N_{\text{obs.}} = 24^2 = 576$ points to measure velocity. These components are assumed to be independent and are affected by observational noise, characterized by a standard deviation σ_{obs} as shown in Figure 5.4b.

The number of assimilation steps is fixed at $N_{\text{assim}} = 5$, with each assimilation occurring after a forecast interval $\Delta t = 10$. This parameter will be varied to analyze how overly dispersed vortices influence the assimilation process.

We evaluate the effectiveness of the new alignment step in correcting misalignment errors through particle position adjustments. The two-step sequential method enables a distinct combination of alignment and amplitude corrections, as summarized in Table 5.1.

	\emptyset	Part-EnKF	Remesh-EnKF
\emptyset	\emptyset	Part-EnKF	Remesh-EnKF
Align	Align-Filter	Part-Align-EnKF	Remesh-Align-EnKF

Table 5.1: Summary of different filter combinations.

Error definitions

The errors of alignment of the center are assessed from each vortex $v = 1, 2, 3$ of member $i = 1, \dots, N$. A particle p of the particle set \mathcal{P}^i is assigned to a vortex v at the initialisation and is included in the subset \mathcal{P}_v^i . It induces partition such as $\mathcal{P}^i = \mathcal{P}_1^i \cup \mathcal{P}_2^i \cup \mathcal{P}_3^i$. The vortex number is treated as a particle variable, and is consequently added to the state vector during assimilation as well as remeshing process.

The error is then defined as the normalized value $e_{c,i}^v = \frac{1}{R} \left\| \mathbf{x}_{c,\text{gt}}^v - \mathbf{x}_{c,i}^v \right\|_2$, where R is the core size of the ground truth vortices and $\mathbf{x}_{c,v}^i$ the normalized first moment of the vorticity of member i . It is obtained thanks to the particle discretization such as $\mathbf{x}_{c,v}^i = \frac{M_1^v}{M_0^v} = \frac{\sum_{p \in \mathcal{P}_v^i} \Gamma_p \mathbf{x}_p}{\sum_{p \in \mathcal{P}_v^i} \Gamma_p}$.

On the other hand, the vorticity errors are defined on the entire field thanks to the normalized L_2 norm such as $e_{\omega,i} = \frac{\|\omega^i - \omega^{gt}\|_{L_2}}{\|\omega^{gt}\|_{L_2}}$ where $\|\mathbf{u}\|_{L_2} = \sqrt{\int_{\Omega} \|\mathbf{u}_i(\mathbf{x})\|^2 \, d\mathbf{x}}$. The distribution of errors over the ensemble will be represented thanks to box plots or bias-variance plots.

5.6.3 Amplitude filters

We begin by performing data assimilation using two previously developed EnKF-based filters: Part-EnKF and Remesh-EnKF [Duvillard et al., 2024]. The Part-EnKF filter maintains the same particle support, while the Remesh-EnKF generates a new particle distribution over the entire domain. Although the Remesh-EnKF filter successfully reduces position errors over time, the Part-EnKF eventually diverges, as illustrated in Figure 5.6. After a finite time, the majority of the ensemble under Part-EnKF exhibits a position error e_c exceeding 5, whereas the Remesh-EnKF maintains a median error below 0.05. For the given forecast step and particle support, increasing the ensemble size does not mitigate this effect.

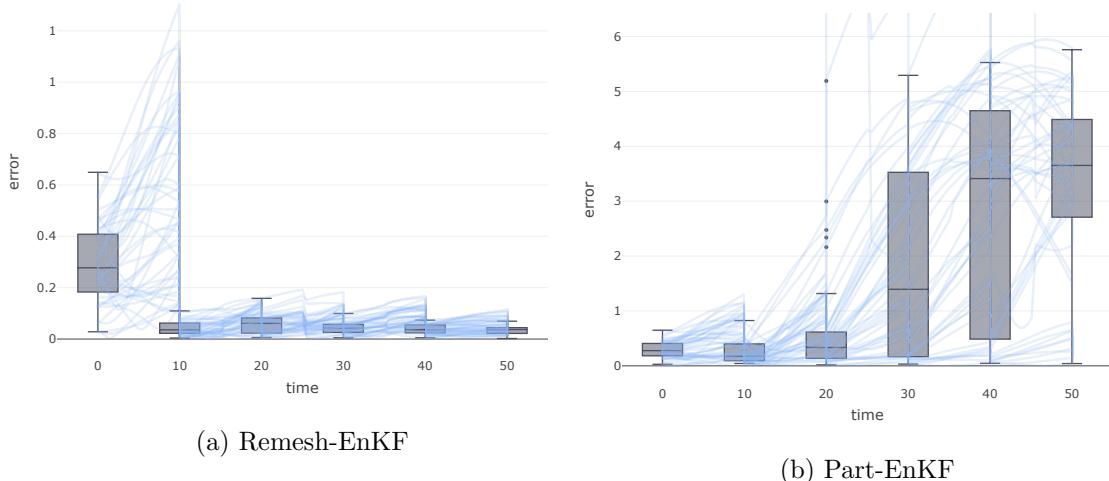


Figure 5.6: Position error e_c over time for the amplitude filters. The Remesh-EnKF filter demonstrates convergence for the given configuration, whereas the Part-EnKF filter shows divergence as time progresses.

Even if the continuous assimilation field for a given ensemble member is accurately computed, this does not ensure that the particle support of that member constitutes an

admissible particle distribution for discretization. This issue is illustrated by the ensemble member shown in Figure 5.7, where the solutions are only partially recovered. In this instance, the vortices become significantly dispersed even after the initial assimilation step. Consequently, the particle support of a member may not be present in the principal regions of the solution, capturing only minor perturbations in amplitude, which ultimately leads to the divergence of the member solution. This situation illustrates how solutions with misaligned vortices can prevent the Part-EnKF filter from functioning effectively. Our goal is to determine whether the misalignment in particle support can be mitigated through the alignment step, thereby improving the performance of the Part-EnKF compared to the Remesh-EnKF.

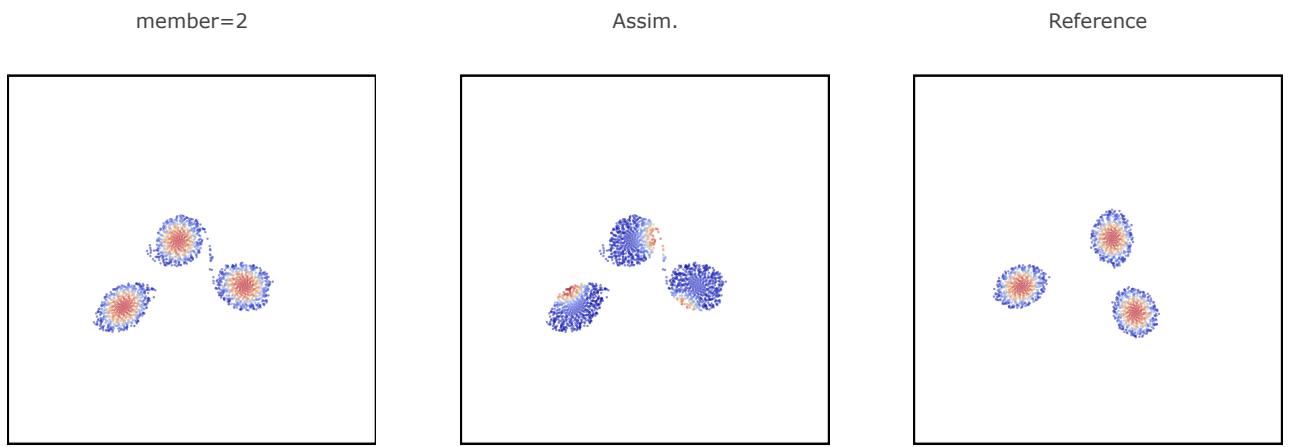


Figure 5.7: Update of member 2 using Part-EnKF. The particle configuration does not align with the correction, resulting in partial recovery of the continuous assimilation solution.

5.6.4 Alignment step

In this section, we focus on the alignment step performed during the first assimilation. Following the acquisition of data, the alignment is achieved by minimizing independent cost functions of (5.5.2) over an ensemble of 16 members. In this way, we determined for each member i the alignment velocity field \mathbf{u}_i , which is then used to adjust the particle positions.

In Figure 5.8, the previously dispersed ensemble of vortices is shown to be significantly closer to the reference configuration after a single alignment step. This result demonstrates that the alignment method is effective in adjusting the discretization support where it is most needed, even when the vortices differ in core size and amplitude. Although the alignment step may introduce changes in the vortex shape, the subsequent amplitude correction step can project necessary adjustments to refine the vortex geometry and enhance the alignment.

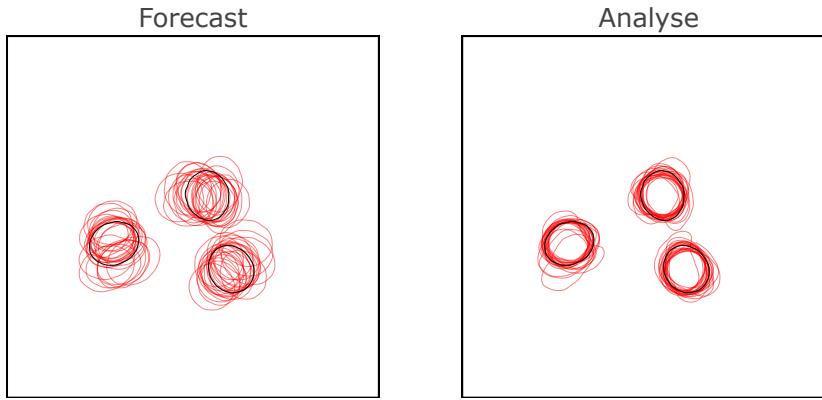


Figure 5.8: Contour plots of the vortex ensemble before and after the alignment process at the first assimilation step. The alignment step significantly improves the match between the dispersed ensemble and the reference configuration.

The alignment step is illustrated explicitly in Figure 5.9 for members 2 and 8 (the worst-case scenario). The left column displays the particle configurations before alignment (forward particles), while the right column shows the configurations after alignment, including the associated velocity fields \mathbf{u}_i used for the alignment. Member 2 exemplifies a case where the vortices are misaligned along the mean trajectory, which is effectively corrected by the alignment process. In contrast, member 8 demonstrates a scenario where the misalignment is orthogonal to the trajectory of the vortices. In this case, the correction is applied to the perturbation velocity space, which may lead to an elongation of the particle distribution, as seen with the vortex in the bottom left. Although this distortion tends to dissipate due to the vortex structure, it highlights a potential limitation of using a fixed velocity field \mathbf{u}_i . Additionally, we note that while the particle configuration may be slightly distorted in regions of low vortex strength, high-strength regions generally remain well-aligned with the reference configuration.

Overall, the alignment fields across the ensemble reveal distinct patterns of positional correction. This is primarily due to the choice of a low-rank decomposition. but we want to understand the predominant modes of alignment and their weights. To better understand the predominant modes of alignment and their associated weights, we perform a Singular Value Decomposition (SVD) to identify the principal orthogonal directions and their corresponding singular values. The first modes are illustrated on 5.10a. We get that the first singular mode 5.10a corresponds to a phase correction and looks close to the mean component. In contrast, the mode depicted in Figure 5.10b represents an outward/inward stretching mode, where the three vortices undergo a uniform perpendicular translation. This translation is achieved by coupling two vortices with opposite signs to form a dipole structure. Secondary modes provide more detailed corrections, allowing for finer-scale adjustments in vortex positions. These modes enable more independent displacement corrections for individual vortices.

However, these detailed modes are residual ones. In fact, it is observed that very few

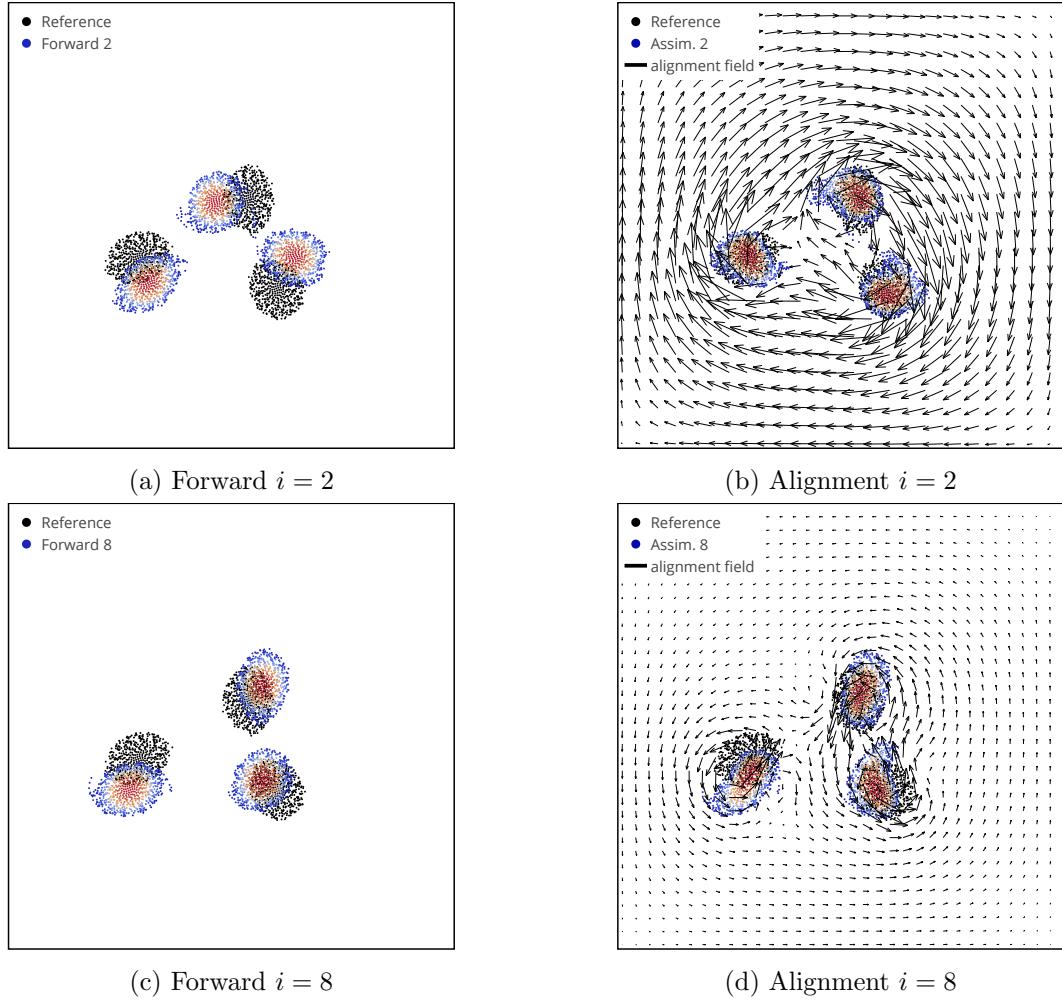


Figure 5.9: Particle alignment for two different members ($i = 2$ and $i = 8$). The left column shows the particle configurations before alignment (forward particles), while the right column displays the configurations after alignment, including the velocity fields \mathbf{u}_i used for the alignment. The case of particle 2 (top row) shows a situation where all vortices are shifted from the mean trajectory. The bottom row (particle 8) shows a case where the misalignment is orthogonal to the trajectory.

corrections are needed beyond a certain number of modes, as indicated by the decreasing distribution of singular values on the log scale, as shown in Figure 5.10f. For instance, the mode 23 only capture noise in the velocity field. This suggests that the variations generated by the ensemble have fewer dominant modes than the size of the ensemble and that last modes do not induce support modification. In other words, the velocity field does not require as much detail to reconstruct the correction field on a given basis. Specifically, it is noted that increasing the ensemble size does not improve the error after alignment, as represented in Figure 5.11a. This observation is corroborated by examining the distribution of coefficients relative to ensemble size in Figure 5.11b. It is

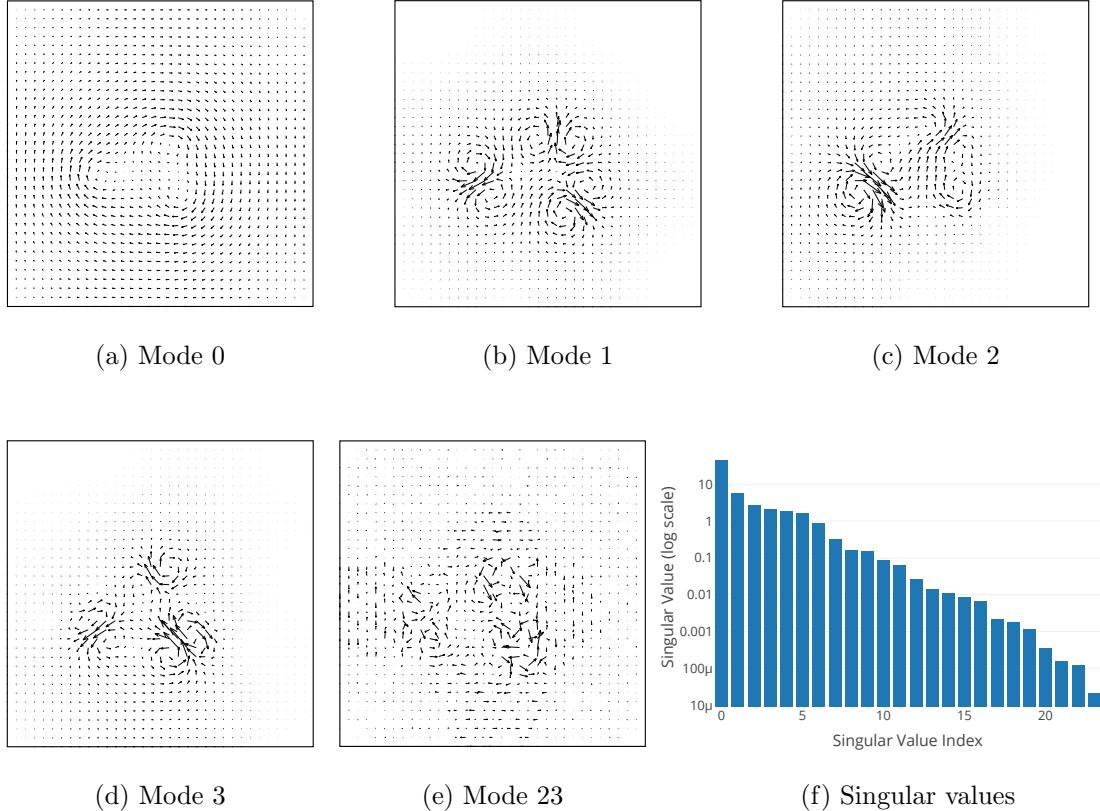
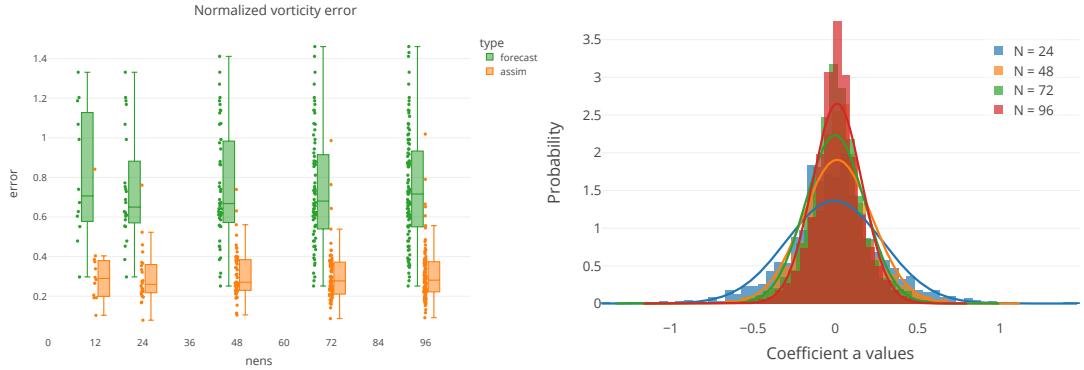


Figure 5.10: Singular Value Decomposition (SVD) of alignment fields. The first five subfigures illustrate the principal modes of alignment, showing the predominant correction patterns, while the last subfigure depicts the corresponding singular values. The modes reveal various types of corrections applied to the vortex distributions, including phase corrections and stretching.

observed that as the ensemble size increases, the coefficients become smaller. Analysis of the coefficient distribution shows that as the ensemble grows, the coefficients diminish such as the distribution remains normal, and the standard deviation σ_a depends on N such that $\sigma(N) \propto 1/\sqrt{N}$. This implies that many coefficients are not necessary and tend to take smaller values. This phenomenon primarily arises because, even in the forecast, the velocity fields exhibit strong similarities. Applying SVD to the ensemble of forecast velocity fields reveals a similar mode structure. This is particularly true in subsequent assimilation steps. As variance decreases, only a few modes contribute significantly to correcting the positions, which results in a sharper distribution.

These findings indicate that, in this specific context, the number of terms in the decomposition can be reduced without compromising the quality of the alignment process. This result is promising as it suggests that the current method could be enhanced by considering a smaller number of modes, $N_{\text{mode}} \ll N$. This would reduce the computational complexity from $O(N^2)$ à $O(NN_{\text{mode}})$ for solving the nonlinear optimization problem. Employing principal directions to define a reduced basis while focusing on



(a) Forward and assimilation error for different ensemble sizes. Increasing ensemble size does not improve the alignment error.
 (b) Distribution of coefficients a by ensemble size

Figure 5.11: Analysis of ensemble size impact on alignment and coefficient distribution.
 (a) Error after alignment versus ensemble size and (b) Distribution of coefficients a for varying ensemble sizes.

modes with high singular values appears to be a promising strategy for optimizing alignment efficiency.

Finally, by implementing the position correction step, the particle support is better aligned, facilitating the EnKF correction on an admissible particle distribution. This improvement is particularly evident in the corrected particle positions of previous member 2, where the Part-EnKF correction is now effectively projected, as illustrated in Figure 5.12. The alignment step thus represents a viable solution to address the limitations discussed in Section 5.6.3.

5.6.5 Assimilation with position-strength correction

After demonstrating the capability to correct the vortex position by adjusting the particle discretization at the initial assimilation step, we now integrate this adjustment throughout the assimilation process. We first compare the results obtained by performing only the position correction (Align-Filter) with those achieved when combining position correction with the amplitude filter (Remesh-Align-EnKF or Part-Align-EnKF).

The Align-Filter alone, as depicted in the first row of Figure 5.14, can sequentially improve the alignment over time and ultimately achieve the correct shape. However, it still exhibits significant uncertainty at the boundaries of each vortex, as shown in Figure 5.14d. This residual error is further illustrated by the position error box plot over time in Figure 5.13a. Although the position error e_c decreases with the assimilation, it does not entirely vanish. Besides occasional distortions in vortex shape and loss of

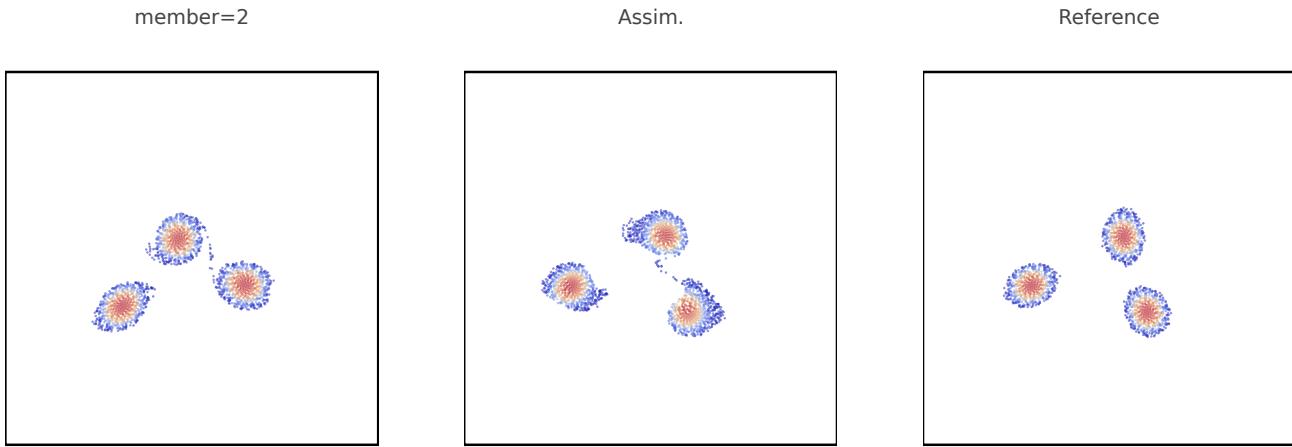


Figure 5.12: Update of member 2 using the Part-Align-EnKF during the initial assimilation. The particle configuration has been adjusted to effectively accommodate the correction.

symmetry due to alignment, a primary reason for the residual error is the inability to reduce amplitude errors associated with the intensity variations. Specifically, if the blobs differ in total circulation and core size, a transformation is required to adjust the volumes accordingly. Therefore, combining the alignment step with amplitude correction filters is crucial, as it addresses the uncertainty induced by amplitude variations and improves overall accuracy.

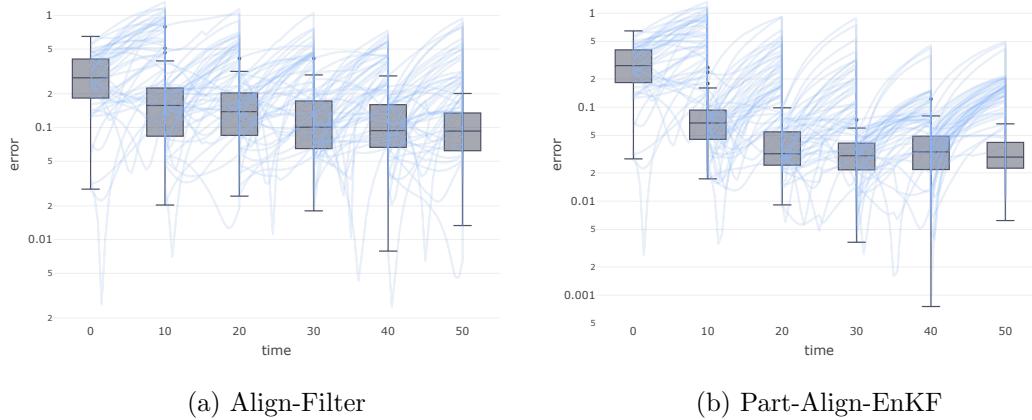


Figure 5.13: Position error e_c for different filtering methods over time. (a) Error for the Align-Filter. (b) Error for the Part-Align-EnKF. The plots show how the position error evolves with each assimilation step for both filtering approaches.

We apply the Part-EnKF-filter in this context. As observed in Figure 5.13b the filter converges effectively and now produces results comparable to what we get in Figure 5.6a.

In fact, the result is better the the Remesh-EnKF due to the alignment step. Compared to the Align-Filter, incorporating amplitude correction enhances the control over vortex spreading during the forecast phase. Consequently, the boundaries of the vortices are more precisely delineated, and the solution provides a more accurate representation of their geometry, as illustrated in the second row of Figure 5.14. This demonstrates that the assimilation procedure successfully addresses both alignment errors, as achieved by the Align-Filter, and uncertainties related to vortex strength.

In this manner, we define two filters for position-strength correction. In addition to enabling the use of the Part-EnKF filter and preventing divergence, we aim to assess the performance improvements achieved by applying the alignment step prior to the strength correction filter, compared to using the strength correction filter alone.

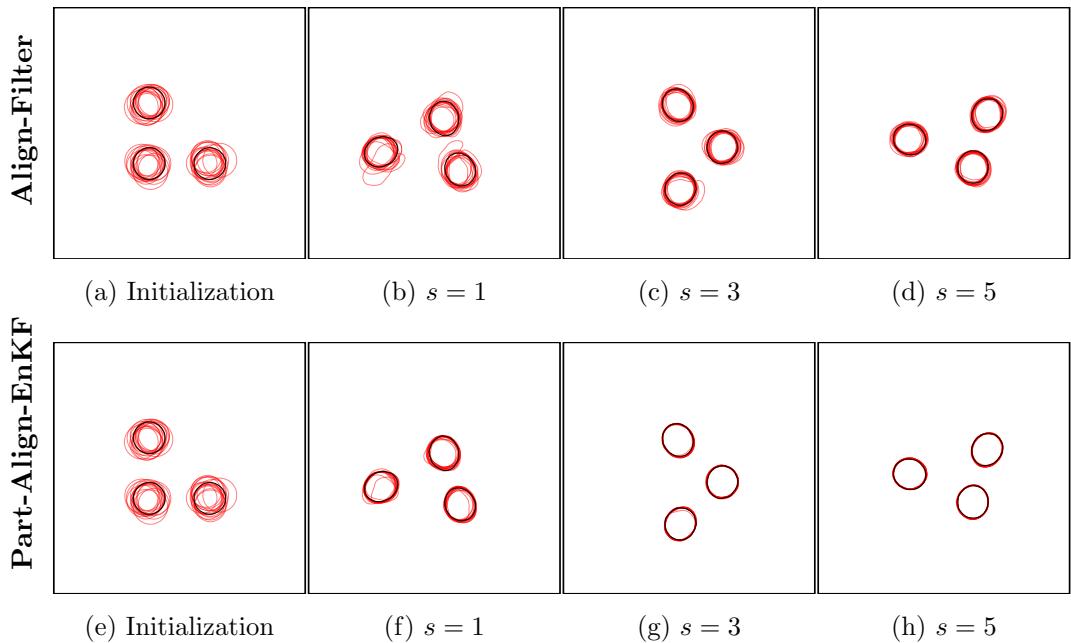
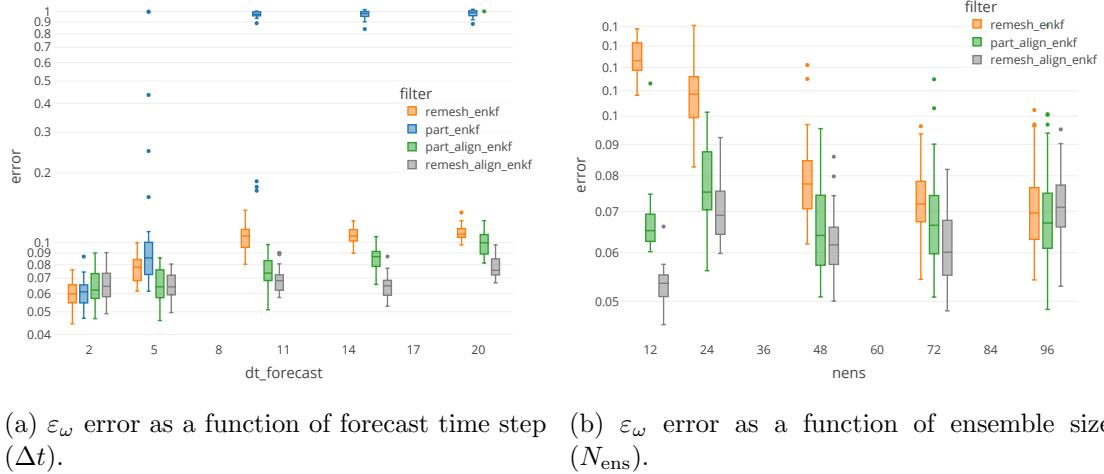


Figure 5.14: Comparison of vortex contour plots for different assimilation methods and steps. The first row corresponds to the Align-Filter, and the second row to the Part-Align-EnKF filter.

5.6.6 Filters performances and parametric studies

Observation frequency The first effect we aim to evaluate is the influence of the forecast time step, Δt , on the performance of the various filters. As expected, an increase in Δt leads to a greater ensemble spread. Figure 5.15a demonstrates that as Δt increases, the final error in the vorticity field, e_ω , also tends to increase. This is particularly noticeable for the Part-EnKF filter, which initially performs comparably to other filters but gradually diverges as the observation frequency decreases, especially for the previously used value of $\Delta t = 10$. Due to its inability to converge under these conditions, the Part-EnKF filter will not be presented further in the study. The other

filters are significantly less sensitive to changes in Δt , particularly the Part-Align-EnKF filter. The alignment step in this filter ensures that its performance remains comparable to the two other Remesh filters, which employ a remeshing strategy.



(a) ε_ω error as a function of forecast time step (b) ε_ω error as a function of ensemble size (N_{ens}).

Figure 5.15: Normalized vorticity error (ε_ω) error for varying forecast time steps and ensemble sizes. Alignment-based filters consistently demonstrate superior performance, achieving convergence with fewer ensemble members compared to the Remesh-EnKF filter.

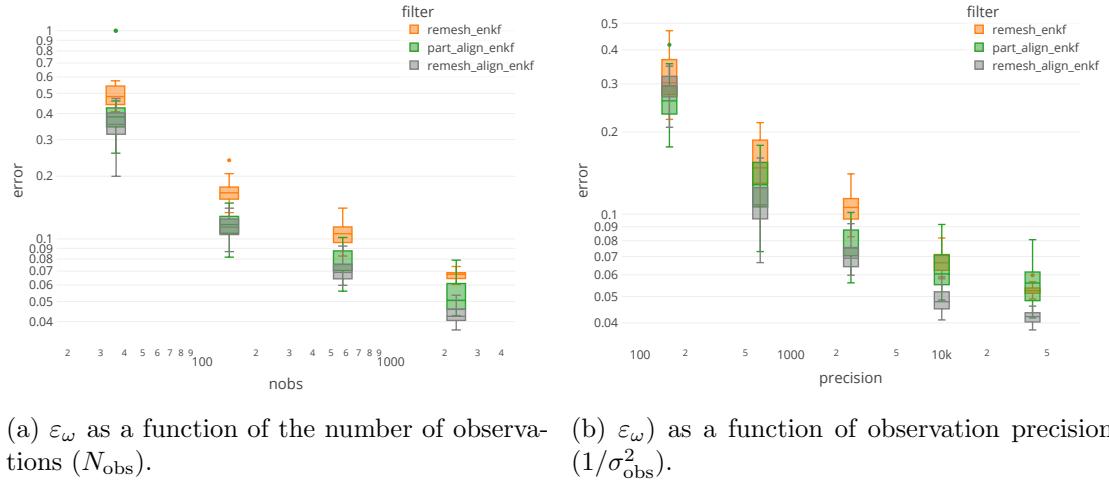
Ensemble size Beyond the sensitivity to observation frequency, we also investigate the convergence of the filter with respect to ensemble size. It is particularly important to determine the optimal balance between computational cost and the ability to represent system uncertainties and correlations accurately. Figure 5.15b shows that fewer ensemble members are required for convergence when using filters with an alignment step. Specifically, the Remesh-EnKF filter requires at least 48 members to achieve stable error levels around 0.07, while the alignment-based filters can effectively converge with as few as 12 members. This improvement is due to the richer correction space available with the alignment-based filters compared to the Remesh-EnKF for the same ensemble size. Consequently, by using a smaller ensemble size and incorporating an alignment step, significant computational savings can be achieved. Although the alignment step involves solving a nonlinear optimization problem, there is potential for further computational efficiency improvements, mainly if an SVD basis is applied, as discussed in the final remark of Section 5.6.4.

Finally, we assess the influence of the number of observations N_{obs} and the observation precision, defined as $\text{Prec.} = 1/\sigma_{\text{obs}}^2$. As shown in Figure 5.16b, all alignment-based filters exhibit comparable convergence rates. Notably, in all cases, these filters outperform the Remesh-EnKF filter, demonstrating superior robustness to variations in observation quantity and precision.

These plots highlight the improvement introduced by incorporating alignment corrections, as seen in the comparison between the Remesh-EnKF and Remesh-Align-EnKF

filters. Specifically, we observe a consistent relative improvement of XXX between these two filters. As for the Part-EnKF, its performance lies in between, offering the advantage of using a purely Lagrangian filter while still outperforming the Remesh-EnKF in this context. We observe that the interquartile range is generally larger for the Part-Align-EnKF. This is likely due to the fact that each ensemble member uses its own discretization, which introduces variability even when representing the same continuous solution. In contrast, for the Remesh filters, the state is ultimately represented on the same regular particle discretization, leading to less variability from discretization. In the case of Remesh filters, the state is finally described on the same particle discretization, leading to less discretization variability.

While for low values of σ_{obs} , the Remesh-EnKF filter performs better than the Part-Align-EnKF, it is important to note that in the case of high-precision observations, this filter requires an inflation process with a multiplicative factor of 1.05 to prevent ensemble collapse. This was not necessary for the two alignment-based filters. In fact, we observe that the alignment step proves beneficial, especially in cases where the low-rank approximation of the EnKF filter would otherwise prevent convergence.



(a) ε_ω as a function of the number of observations (N_{obs}). (b) ε_ω) as a function of observation precision ($1/\sigma_{\text{obs}}^2$).

Figure 5.16: Normalized vorticity error (ε_ω) as a function of different parameters: number of observations and precision, both plotted on a logarithmic scale. Alignment-based filters consistently show improved performance compared to the Remesh-EnKF filter.

5.7 Conclusion

In this work, we developed a new data assimilation technique aimed at simultaneously correcting both particle positions and intensities in Lagrangian simulations. Traditional filters, such as Part-EnKF, which focus solely on correcting intensity, face limitations when the particle support does not align with the analysis solution. Recognizing that these limitations stem primarily from alignment errors during the forecast, we introduced a method that directly addresses the alignment of particle positions through a sequential

two-step correction. By leveraging ensemble velocity fields, we established a kinematically admissible transformation for particle positions, while correcting intensities using existing Lagrangian simulation filters such as Remesh-EnKF and Part-EnKF.

Our results show that the proposed method effectively corrects the entire Lagrangian discretization, making Part-EnKF applicable even when particle supports are initially disjoint. Moreover, our approach outperforms traditional intensity-based filters by compensating for position errors, delivering comparable results with fewer members. Although this method was applied to the Vortex Method, it is adaptable to other Lagrangian data assimilation frameworks where velocity fields are used to update particle positions, extending its applicability to cases involving disjoint subdomains of the global space.

Despite the promising results, some limitations remain. The position alignment step introduces a non-linear optimization process, which could limit the scalability of the method with larger ensemble sizes. Additionally, the velocity space spanned by the ensemble at the end of the forecast may not provide sufficient flexibility to resolve the correction field in more complex scenarios.

Future improvements should focus on enriching the velocity search space, possibly by incorporating velocity fields throughout the entire forecast, or replacing fixed velocity fields with iteratively determined sequences, allowing for more precise position updates. To further enhance computational efficiency, SVD decomposition could be implemented during optimization, particularly when the ensemble velocity field spans a reduced search space.

5.A Parameters

Variables	Vortex 1	Vortex 2	Vortex 3
Vortex center (x^i, y^i)	$(\pi/2 - \delta, \pi/2 - \delta)$	$(\pi/2 + \delta, \pi/2 - \delta)$	$(\pi/2 - \delta, \pi/2 + \delta)$
Core size (R)		0.2	
Amplitude (Γ)		4	

Table 5.2: Initial three-vortex true configuration with $\delta = 0.375$

Variables	Distributions
Vortex Center	$\mathcal{N}(x^i, 0.05^2), \mathcal{N}(y^i, 0.05^2) \quad (i = 1, 2, 3)$
Core Size (R)	$\mathcal{N}(0.2, 0.01^2)$
Amplitude (Γ)	$\mathcal{N}(4.0, 0.08^2)$

Table 5.3: Initial distributions for the vortex parameters. We sample the vortex center positions, core size, and also amplitude.

5.B Gradient computation

The calculation of the derivative by finite difference requires performing N additional forward evaluations, which can be costly. An alternative is to determine a main trajectory and to consider N perturbations around this trajectory.

To calculate $\nabla_a h$, we define a perturbation of the alignment velocity field $\mathbf{u}(\mathbf{a})$ such as $\mathbf{a}' = \mathbf{a} + \varepsilon \mathbf{e}_i$. It induces a perturbation around the trajectory that we note \mathbf{x}_a such as $\mathbf{x}_{a'}(t) = \mathbf{x}_a(t) + \varepsilon \mathbf{x}_1(t) + o(\varepsilon)$ which give the following evolution model

$$\frac{d}{dt}(\mathbf{x}_a(t) + \varepsilon \mathbf{x}_1(t), \dots) = \sum_{j=1}^N a_j \mathbf{u}_j(\mathbf{x}_a + \varepsilon \mathbf{x}_1 + \varepsilon^2 \mathbf{x}_2, \dots) + \varepsilon a_i \mathbf{u}_i(\mathbf{x}_a + \varepsilon \mathbf{x}_1 + \dots). \quad (5.6)$$

We develop the perturbed velocity fields by a first order Taylor expansion such as

$$u_j(\mathbf{x}_a + \varepsilon \mathbf{x}_1 + \dots) = u_j(\mathbf{x}_a) + \varepsilon \mathbf{x}_1 \nabla_x u_j(\mathbf{x}_a) + o(\varepsilon).$$

By substituting in (5.6), we finally get couple problem where the particles are

$$\begin{cases} \frac{d\mathbf{x}_a}{dt} &= \sum_j a_j u_j(\mathbf{x}_a), \\ \mathbf{x}_a(t=0) &= \mathbf{x}_a^0, \\ i = 1, \dots, M : \\ \frac{d\hat{\mathbf{x}}_i}{dt} &= \hat{\mathbf{x}}_i \sum_j a_j \nabla_x \mathbf{u}_j(\mathbf{x}_a) + a_i \mathbf{u}_i(\mathbf{x}_a), \\ \hat{\mathbf{x}}_i(t=0) &= 0. \end{cases}$$

Where we only need to know the gradient and velocity field of each component \mathbf{u}_j for the trajectory \mathbf{x}_a , and track N perturbations \mathbf{x}^i . So that, $\frac{\partial \mathbf{x}}{\partial \mathbf{v}}$ is the matrix where the columns are $(\mathbf{x}_1, \dots, \mathbf{x}^N)$ such that the gradient of observation is

$$\frac{\partial \mathbf{h}}{\partial \mathbf{v}} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{v}}$$

where $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ could be estimate by finite difference such as

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}_i} \approx \frac{\mathcal{H}(\mathbf{x}_a + \varepsilon \mathbf{x}_i) - \mathcal{H}(\mathbf{x}_a)}{\varepsilon}.$$

Conclusion générale

6.1 Synthèse des résultats

Dans cette thèse, nous avons développé des méthodes d’assimilation de données visant à améliorer l’estimation de l’état dans des simulations reposant sur une discrétisation particulaire lagrangienne. La principale limitation des méthodes ensemblistes actuellement utilisées, telles que le filtre de Kalman d’ensemble (EnKF), réside dans le fait que chaque membre de l’ensemble dispose de sa propre configuration particulière. Cela rend impossible l’application directe de la correction basée sur une combinaison des membres sans entraîner une augmentation exponentielle du nombre de particules.

Ainsi, dans le chapitre 4, nous avons proposé deux adaptations du filtre EnKF par correction d’intensité. Le filtre Remesh-EnKF repose sur une étape de remaillage pour générer une nouvelle distribution particulaire commune à tous les membres de l’ensemble, sur laquelle la mise à jour peut être appliquée. D’autre part, le filtre Part-EnKF consiste à approximer la solution analysée continue sur la discrétisation de chaque membre. Dans cette seconde approche, la formulation reste purement particulaire et consiste en une mise à jour des amplitudes de chaque membre. Nous avons comparé ces différentes formulations avec un filtre défini sur une discrétisation eulérienne dans le cadre d’un cas d’advection unidimensionnel. Le filtre Remesh-EnKF présente de bonnes performances, à hauteur de l’ordre de précision du schéma de remaillage. Le filtre Part-EnKF, quant à lui, est sensible au recouvrement des distributions particulières au sein de l’ensemble. Ces filtres ont été testés avec succès dans le cadre de la méthode vortex pour la simulation d’écoulements incompressibles.

Dans le chapitre 5 partant de l’hypothèse que le recouvrement pouvait provenir d’une erreur d’alignement, nous avons proposé une méthode capable de corriger également les positions des particules dans la discrétisation. Une approche variationnelle a été développée pour déterminer une transformation physiquement admissible. Dans le cadre de la méthode vortex, cette transformation a été définie par l’intégration d’un champ de vitesse à divergence nulle. Dans notre développement, l’espace de recherche prend en compte les champs de vitesse générés par l’ensemble afin de définir un problème d’optimisation de rang faible. Cette méthode a été combinée aux deux filtres précédents, formant ainsi deux nouveaux filtres : Part-Align-EnKF et Remesh-Align-EnKF. L’état défini par la discrétisation particulaire est ainsi complètement mis à jour. Nous avons montré que l’étape d’alignement permettait effectivement un meilleur recouvrement des

distributions, rendant ainsi possible la correction d'amplitude avec le Part-EnKF. De plus, les performances sont améliorées en offrant un espace de correction plus riche, même avec une taille d'ensemble réduite.

6.2 Orientations futures

Les perspectives futures concernent à la fois l'amélioration des méthodes de filtrage précédemment développées et leur adaptation à diverses simulations. En particulier, nous souhaitons explorer des adaptations potentielles pour les méthodes MPM, SPH ou DEM.

Amélioration des méthodes d'alignement

Pour perfectionner les méthodes d'alignement, plusieurs axes d'amélioration peuvent être envisagés. Actuellement, nous nous appuyons uniquement sur le champ de vitesse actuel pour définir la base réduite. Cependant, l'intégration des champs de vitesse au cours de la phase de prévision (forecast) pourrait enrichir cette base. De plus, dans le cas de scénarios d'alignement complexes, il serait avantageux d'autoriser une variation du champ de vitesse pendant l'intégration. Une variante incrémentale de la procédure d'alignement pourrait être développée, où la transformation serait exprimée comme une composition de transformations successives $\Psi(x) = \Psi_N \circ \dots \circ \Psi_1(x)$, associées à l'intégration de champs de vitesse \mathbf{v}_i sur des pas de temps successifs. Chaque transformation Ψ_i correspondrait à l'intégration du champ \mathbf{v}_i sur un pas de temps $\frac{1}{N}$. Cette approche permettrait de mieux s'adapter à des situations d'alignement plus dynamiques et complexes.

Application aux méthodes de simulation d'écoulement granulaire continues

L'application des filtres par correction d'amplitude aux simulations d'écoulement granulaire continu, telles que celles basées sur les méthodes SPH et MPM, ouvre plusieurs perspectives intéressantes. Pour SPH, les développements appliqués dans le cas scalaire peuvent être étendus au cas vectoriel, en les appliquant à la densité ρ , à la vitesse \mathbf{v} , au champ de contrainte σ , ainsi qu'aux autres propriétés nécessaires pour mettre à jour la loi de comportement. En ce qui concerne la méthode MPM, celle-ci repose sur un schéma intrinsèquement basé sur un transfert particules à grille, similaire à celui utilisé dans le filtre Remesh-EnKK 4.5.1. Le filtre Remesh-EnKF semble donc particulièrement adapté à ce type de simulation, avec l'extension nécessaire pour traiter également les champs vectoriels. Dans les deux cas, il est essentiel de mettre à jour des propriétés qui, traditionnellement, restent fixes afin de garantir le respect des principes de conservation.

Application aux méthodes discrètes de simulation d'écoulement granulaire

Dans le contexte des méthodes de simulation d'écoulement granulaire discret, telles que la méthode des éléments discrets (DEM), des ajustements spécifiques sont nécessaires. Contrairement aux méthodes à champ continu, la discréttisation en DEM dépend des interactions entre particules, et il est crucial de prendre en compte les contraintes géométriques de ces interactions. En effet, la mise à jour des simulations DEM repose sur le calcul des forces interparticulaires, qui dépendent de la distance entre les particules. Cette contrainte est particulièrement sensible aux chevauchements géométriques et ne tolère pas les recouvrements. Ainsi, l'application des méthodes de correction d'intensité n'est pas directement envisageable, mais une correction des positions pourrait être envisagée, à condition de respecter la contrainte de distance interparticulaire.

Dans ce cadre, à l'instar de la méthode vortex, les champs de vitesse générés par les ensembles peuvent être utilisés pour aligner les particules. Ces champs de vitesse seraient déterminés sur une grille régulière à l'aide de fonctions de transfert particules-grille, analogues à celles utilisées dans la méthode MPM.

Cependant, en raison des contraintes associées aux interactions particules-particules, les forces d'interaction interparticulaires doivent être prises en compte lors de l'intégration du champ de vitesse d'alignement \mathbf{v}_{corr} . Pour ce faire, un schéma de prédition-correction pourrait être mis en place [Allen and Tildesley, 1987b]. Un schéma classique pourrait se dérouler ainsi :

- *Prédiction des positions en utilisant la vitesse prescrite:* Pour chaque particule p , on prédit d'abord sa nouvelle position $\mathbf{x}_p^* = \mathbf{x}_p + \mathbf{v}_p \Delta t$ et sa vitesse \mathbf{v}_p^* en utilisant la vitesse prescrite \mathbf{v}_{corr} ;
- *Calcul des forces d'interaction:* On calcule les forces d'interaction \mathbf{F}_p basées sur les positions prédictes ;
- *Correction avec les interactions* La vitesse est corrigées tel que $\mathbf{v}_p = \mathbf{v}_p^* + \mathbf{F}_p \Delta t$. La position prédictive est corrigée à partir de la différence entre les vitesses corrigées et celles prédictes $\mathbf{x}_p = \mathbf{x}_p^* + \Delta t(\mathbf{v}_p - \mathbf{v}_p^*)$.

Une autre approche consisterait à déterminer directement un champ de forces de correction F_{corr} au lieu d'utiliser un champ de vitesse pour ajuster les positions des particules. De cette manière, la correction serait intégrée directement dans le calcul de l'accélération lors de l'intégration temporelle.

APPENDIX A

Conservation des moments particulaires du schéma de remaillage

Le m -ième moment d'une distribution de particules est défini comme la quantité $\sum_p z_p^\alpha \mathbf{U}_p$.

Tout d'abord, nous voyons que la partition de l'unité est nécessaire

$$\sum_{I \in \Lambda} W\left(\frac{z - z_I}{\ell_I}\right) = 1, \quad z \in \Omega \quad (\text{A.1})$$

en raison de l'arrangement final des particules \mathcal{P}' sur une grille de taille d_p , cela conduit à la propriété suivante

$$\sum_{p' \in \mathcal{P}'} W\left(\frac{z - z_{p'}}{\ell_I}\right) = \frac{V_I}{V_p}, \quad z \in \Omega. \quad (\text{A.2})$$

L'attention doit être concentrée sur la frontière. L'extension du domaine avec des particules ou des nœuds "fantômes" permet de vérifier les propriétés à l'intérieur de Ω .

Cette propriété est la condition nécessaire pour la conservation du premier moment. Principalement pour l'affectation

$$\sum_{I \in \Lambda} \mathbf{u}_I V_I = \sum_{p \in \Lambda} \mathbf{U}_p W\left(\frac{z_I - z_p}{\ell_I}\right) = \sum_{p \in \mathcal{P}} \mathbf{U}_p \sum_{I \in \Lambda} W\left(\frac{z_I - z_p}{\ell_I}\right) = \sum_{p \in \mathcal{P}} \mathbf{U}_p.$$

en utilisant la propriété (A.1). Deuxièmement, pour le processus d'interpolation

$$\begin{aligned}
 \sum_{p' \in \mathcal{P}'} \mathbf{U}_{p'} &= \sum_{p' \in \mathcal{P}'} \mathbf{u}_g(z_{p'}) V_{p'} = \sum_{p' \in \mathcal{P}'} V_{p'} \sum_{I \in \Lambda} \mathbf{u}_I W\left(\frac{z_{p'} - z_I}{\ell_I}\right) \\
 &= \sum_{I \in \Lambda} \mathbf{u}_I V_{p'} \sum_{p' \in \mathcal{P}'} W\left(\frac{z_{p'} - z_I}{\ell_I}\right) \\
 &= \sum_{I \in \Lambda} \frac{V_I}{V_p'} V_{p'} \mathbf{u}_I = \sum_{I \in \Lambda} \mathbf{u}_I V_I = \sum_{p \in \mathcal{P}} \mathbf{U}_p
 \end{aligned}$$

en utilisant l'équation (A.2).

On peut montrer de plus que si pour $1 \leq |\alpha| \leq m - 1$, W satisfait,

$$\sum_{I \in \Lambda} (\mathbf{z} - \mathbf{z}_I)^\alpha W\left(\frac{\mathbf{z} - \mathbf{z}_I}{\ell_I}\right) = 0, \quad (\text{A.3})$$

La procédure de regrillage sera ordonnée à m . De manière équivalente, l'égalité précédente conduit, pour $0 \leq |\alpha| \leq m - 1$, à

$$\sum_{I \in \Lambda} \mathbf{z}_I^\alpha W\left(\frac{\mathbf{z}_p - \mathbf{z}_I}{\ell_I}\right) = \mathbf{z}^\alpha,$$

obtenue en développant $(\mathbf{z} - \mathbf{z}_q)^\alpha$ et en utilisant une récurrence sur les ordres précédents. Cela signifie que l'interpolation est exacte pour les polynômes de degrés inférieurs ou égaux à $m - 1$ ou que le moment d'ordre $m - 1$ est conservé.

Bibliography

- Abe and Asakura (2012). “2.15 - Uranium Oxide and MOX Production”. In: *Comprehensive Nuclear Materials*. Ed. by Konings. Oxford: Elsevier, pp. 393–422. ISBN: 978-0-08-056033-5. DOI: <https://doi.org/10.1016/B978-0-08-056033-5.00036-7> (cit. on p. 12).
- Adepu et al. (May 2021). “Wall to particle bed contact conduction heat transfer in a rotary drum using DEM”. en. In: *Computational Particle Mechanics* 8.3, pp. 589–599. ISSN: 2196-4386. DOI: [10.1007/s40571-020-00356-z](https://doi.org/10.1007/s40571-020-00356-z) (cit. on p. 17).
- Agence internationale de l'énergie atomique (AIEA) (Sept. 2023). *Rapport d'ensemble sur la technologie nucléaire 2023*. Rapport GC(67)/INF/4. Autriche: Agence internationale de l'énergie atomique (AIEA) (cit. on p. 11).
- AIAA Digital Engineering Integration Committee (2020). *Digital twin: Definition and value*. Position Paper. American Institute of Aeronautics, Astronautics (AIAA), and Aerospace Industries Association (AIA) (cit. on p. 18).
- Alder and Wainwright (1959). “Studies in molecular dynamics. I. General method”. In: *Journal of Chemical Physics* 31.2, p. 459 (cit. on p. 39).
- Allaire and Craig (July 2007). *Numerical Analysis and Optimization: An Introduction to Mathematical Modelling and Numerical Simulation*. English. 1st edition. Mahwah, N.J: Oxford University Press. ISBN: 978-0-8058-3985-2 (cit. on p. 59).
- Allen and Tildesley (1987a). *Computer Simulation of Liquids*. Oxford: Oxford University Press (cit. on p. 39).
- (1987b). *Computer Simulation of Liquids*. Oxford: Oxford University Press (cit. on p. 135).
- Almond and Valderrama (2004). “Performance Enhancement Tools For Grinding Mills”. In: *The Southern African Institute of Mining and Metallurgy* (cit. on p. 17).
- Ambati et al. (2012). “Application of material point methods for cutting process simulations”. In: *Computational Materials Science* 57, pp. 102–110 (cit. on pp. 38, 107).
- Amodei and Stein (2009). “Deterministic and fuzzy verification methods for a hierarchy of numerical models”. In: *Meteorological Applications* 16, pp. 191–203. DOI: [10.1002/met.101](https://doi.org/10.1002/met.101) (cit. on pp. 65, 108).
- Andersen and Andersen (2010). “Modelling of landslides with the material-point method”. In: *Computational Geosciences* 14.1, pp. 137–147 (cit. on pp. 38, 107).
- Anderson and Anderson (1999). “A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts”. In: *Monthly Weather Review* 127.12, pp. 2741–2758. DOI: [10.1175/1520-0493\(1999\)127<2741:AMCIOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2) (cit. on p. 36).
- Aref (Mar. 1979). “Motion of three vortices”. en. In: *The Physics of Fluids* 22.3, pp. 393–400. ISSN: 0031-9171. DOI: [10.1063/1.862605](https://doi.org/10.1063/1.862605) (cit. on p. 117).
- Arseni et al. (Sept. 2020). “Granular flow in rotating drums through simulations adopting a continuum constitutive equation”. In: *Physics of Fluids* 32.9. P. 093305. ISSN: 1070-6631. DOI: [10.1063/5.0018694](https://doi.org/10.1063/5.0018694) (cit. on p. 16).

- Asch, Bocquet, and Nodet (2016). *Data assimilation: methods, algorithms, and applications*. en. Fundamentals of algorithms 11. Philadelphia: Society for Industrial and Applied Mathematics. ISBN: 978-1-61197-454-6 (cit. on p. 24).
- Austin and Bagga (1981). “An analysis of fine dry grinding in ball mills”. In: *Powder Technology* 28.1, pp. 83–90 (cit. on p. 13).
- Banerjee (2004). “Material point method simulations of fragmenting cylinders”. In: *17th ASCE Engineering Mechanics Conference*. University of Delaware, Newark, DE (cit. on p. 38).
- Barba (May 2004). “Vortex Method for Computing High-Reynolds Number Flows: Increased Accuracy with a Fully Mesh-Less Formulation”. en. PhD thesis. California Institute of Technology. DOI: [10.7907/TSR5-DE67](https://doi.org/10.7907/TSR5-DE67) (cit. on pp. 48, 73).
- Bardenhagen, Brydon, and Guilkey (2005). “Insight into the physics of foam densification via numerical simulation”. In: *Journal of the Mechanics and Physics of Solids* 53.3, pp. 597–617 (cit. on pp. 38, 107).
- Bardenhagen and Kober (June 2004). “The Generalized Interpolation Material Point Method”. In: *CMES - Computer Modeling in Engineering and Sciences* 5 (cit. on pp. 42, 51).
- Beale (1988). “On the Accuracy of Vortex Methods at Large Times”. en. In: *Computational Fluid Dynamics and Reacting Gas Flows*. Ed. by Engquist, Majda, and Luskin. The IMA Volumes in Mathematics and Its Applications. New York, NY: Springer, pp. 19–32. ISBN: 978-1-4612-3882-9. DOI: [10.1007/978-1-4612-3882-9_2](https://doi.org/10.1007/978-1-4612-3882-9_2) (cit. on pp. 48, 73, 77).
- Ben-Israel and Greville (2003). *Generalized Inverses: Theory and Applications*. 2nd. Springer-Verlag (cit. on p. 35).
- Benamou and Brenier (Jan. 2000). “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. en. In: *Numerische Mathematik* 84.3, pp. 375–393. ISSN: 0945-3245. DOI: [10.1007/s002110050002](https://doi.org/10.1007/s002110050002) (cit. on pp. 65, 109).
- Birdsall and Fuss (1969). “Clouds-in-clouds, clouds-in-cells physics for many-body plasma simulation”. In: *Journal of Computational Physics* 3.4, pp. 494–511. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(69\)90058-8](https://doi.org/10.1016/0021-9991(69)90058-8) (cit. on pp. 54, 91).
- Bishop, Etherton, and Majumdar (2001). “Adaptive sampling with the ensemble transform Kalman filter Part I: Theoretical aspects”. English. In: *Monthly Weather Review* 129.3, pp. 420–436. ISSN: 0027-0644. DOI: [10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2) (cit. on pp. 31, 75).
- Blanchard and Mathé (Nov. 2012). “Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration”. en. In: *Inverse Problems* 28.11, p. 115011. ISSN: 0266-5611, 1361-6420. DOI: [10.1088/0266-5611/28/11/115011](https://doi.org/10.1088/0266-5611/28/11/115011) (cit. on p. 117).
- Boateng (1998). “Boundary layer modeling of granular flow in the transverse plane of a partially filled rotating cylinder”. In: *International Journal of Multiphase Flow* 24.3, pp. 499–521 (cit. on p. 15).
- Bocquet et al. (2024). “Bridging classical data assimilation and optimal transport: the 3D-Var case”. In: *Nonlinear Processes in Geophysics* 31.3, pp. 335–357. DOI: [10.5194/npg-31-335-2024](https://doi.org/10.5194/npg-31-335-2024) (cit. on pp. 65, 100, 109).

- Bocquet (2014). “Introduction to the principles and methods of data assimilation in the geosciences”. en. In: p. 89 (cit. on pp. 21, 71, 106).
- Bonan et al. (Sept. 2017). “Data assimilation for moving mesh methods with an application to ice sheet modelling”. In: *Nonlinear Processes in Geophysics* 24, pp. 515–534. DOI: [10.5194/npg-24-515-2017](https://doi.org/10.5194/npg-24-515-2017) (cit. on pp. 64, 72).
- Bouloré et al. (2015). “Advanced characterization of MIMAS MOX fuel microstructure to quantify the HBS formation”. In: *Nuclear Engineering and Design* 281, pp. 79–87. ISSN: 0029-5493. DOI: <https://doi.org/10.1016/j.nucengdes.2014.11.020> (cit. on p. 12).
- Box (1979). “Some Problems of Statistics and Everyday Life”. In: *Journal of the American Statistical Association* 74.365, pp. 1–4. DOI: [10.1080/01621459.1979.10481600](https://doi.org/10.1080/01621459.1979.10481600) (cit. on p. 19).
- Brackbill, Kothe, and Ruppel (Jan. 1988). “Flip: A low-dissipation, particle-in-cell method for fluid flow”. In: *Computer Physics Communications* 48.1, pp. 25–38. ISSN: 0010-4655. DOI: [10.1016/0010-4655\(88\)90020-3](https://doi.org/10.1016/0010-4655(88)90020-3) (cit. on pp. 50, 73).
- Brandaو et al. (2020). “Experimental study and dem analysis of granular segregation in a rotating drum”. In: *Powder Technology* 364, pp. 1–12 (cit. on p. 13).
- Brezzi et al. (Oct. 2016). “A new data assimilation procedure to develop a debris flow run-out model”. en. In: *Landslides* 13.5, pp. 1083–1096. ISSN: 1612-5118. DOI: [10.1007/s10346-015-0625-y](https://doi.org/10.1007/s10346-015-0625-y) (cit. on p. 58).
- Brydon et al. (2005). “Simulation of the densification of real open-celled foam microstructures”. In: *Journal of the Mechanics and Physics of Solids* 53.12, pp. 2638–2660 (cit. on pp. 38, 107).
- Campbell (2006). “Granular material flows - An overview”. In: *Powder Technology* 162.3, pp. 208–229 (cit. on p. 14).
- Capece, Bilgili, and Davé (2014). “Insight into first-order breakage kinetics using a particle-scale breakage rate constant”. In: *Chemical Engineering Science* 117, pp. 318–330 (cit. on p. 13).
- Chandra et al. (Aug. 2021). “Nonconforming Dirichlet boundary conditions in implicit material point method by means of penalty augmentation”. en. In: *Acta Geotechnica* 16.8, pp. 2315–2335. ISSN: 1861-1125, 1861-1133. DOI: [10.1007/s11440-020-01123-3](https://doi.org/10.1007/s11440-020-01123-3) (cit. on p. 16).
- Chen, Pan, and Wu (1997). “Large deformation analysis of rubber based on a reproducing kernel particle method”. In: *Computational Mechanics* 19.3, pp. 211–227 (cit. on pp. 38, 107).
- Chen, Deng, and Stechmann (2022). “Superfloe Parameterization with Physics Constraints for Uncertainty Quantification of Sea Ice Floes”. In: *SIAM/ASA Journal on Uncertainty Quantification* 10.4, pp. 1384–1409. DOI: [10.1137/21M1428777](https://doi.org/10.1137/21M1428777) (cit. on pp. 21, 58, 61, 62, 107).
- Chen and Snyder (2007a). “Assimilating Vortex Position with an Ensemble Kalman Filter”. In: *Monthly Weather Review* 135.5, pp. 1828–1845. DOI: [10.1175/MWR3351.1](https://doi.org/10.1175/MWR3351.1) (cit. on p. 108).
- (2007b). “Assimilating Vortex Position with an Ensemble Kalman Filter”. In: *Monthly Weather Review* 135.5, pp. 1828–1845. DOI: [10.1175/MWR3351.1](https://doi.org/10.1175/MWR3351.1) (cit. on p. 109).
- Cheng et al. (June 2019). “An iterative Bayesian filtering framework for fast and automated calibration of DEM models”. en. In: *Computer Methods in Applied Mechanics*

- and Engineering* 350, pp. 268–294. ISSN: 0045-7825. DOI: [10.1016/j.cma.2019.01.027](https://doi.org/10.1016/j.cma.2019.01.027) (cit. on p. 58).
- Chorin and Bernard (Nov. 1973). “Discretization of a vortex sheet, with an example of roll-up”. In: *Journal of Computational Physics* 13.3, pp. 423–429. ISSN: 0021-9991. DOI: [10.1016/0021-9991\(73\)90045-4](https://doi.org/10.1016/0021-9991(73)90045-4) (cit. on p. 54).
- Chou and Lee (Jan. 2009). “Cross-sectional and axial flow characteristics of dry granular material in rotating drums”. In: *Granular Matter* 11.1, pp. 13–32. ISSN: 1434-7636. DOI: [10.1007/s10035-008-0118-y](https://doi.org/10.1007/s10035-008-0118-y) (cit. on p. 15).
- Christiansen (1973). “Numerical simulation of hydrodynamics by the method of point vortices”. In: *Journal of Computational Physics* 13.3, pp. 363–379. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(73\)90042-9](https://doi.org/10.1016/0021-9991(73)90042-9) (cit. on pp. 54, 91).
- Cortet et al. (Oct. 2009). “Relevance of visco-plastic theory in a multi-directional inhomogeneous granular flow”. In: *Europhysics Letters* 88.1, p. 14001. DOI: [10.1209/0295-5075/88/14001](https://doi.org/10.1209/0295-5075/88/14001) (cit. on p. 15).
- Cottet and Mas-Gallic (Dec. 1990). “A particle method to solve the Navier-Stokes system”. In: *Numerische Mathematik* 57.1, pp. 805–827. ISSN: 0029-599X. DOI: [10.1007/BF01386445](https://doi.org/10.1007/BF01386445) (cit. on pp. 86, 91).
- Cottet, Ould Salihi, and El Hamroui (1999). “Multi-purpose regridding in vortex methods”. en. In: *ESAIM: Proceedings* 7. Ed. by Giovannini et al., pp. 94–103. ISSN: 1270-900X. DOI: [10.1051/proc:1999009](https://doi.org/10.1051/proc:1999009) (cit. on pp. 72, 73, 78).
- Cottet and Koumoutsakos (2000a). *Vortex Methods: Theory and Practice*. Cambridge University Press (cit. on pp. 40, 71, 76, 78, 91, 110).
- (2000b). *Vortex Methods: Theory and Practice*. Cambridge University Press (cit. on p. 52).
- Cundall and Strack (Mar. 1979). “A discrete numerical model for granular assemblies”. In: *Géotechnique* 29.1. Pp. 47–65. ISSN: 0016-8505. DOI: [10.1680/geot.1979.29.1.47](https://doi.org/10.1680/geot.1979.29.1.47) (cit. on pp. 16, 39).
- Cuturi and Doucet (June 2014). “Fast Computation of Wasserstein Barycenters”. In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Xing and Jebara. Vol. 32. Proceedings of Machine Learning Research. Bejing, China: PMLR, pp. 685–693 (cit. on pp. 66, 109).
- Darakananda et al. (Nov. 2018). “Data-assimilated low-order vortex modeling of separated flows”. In: *Physical Review Fluids*. DOI: [10.1103/PhysRevFluids.3.124701](https://doi.org/10.1103/PhysRevFluids.3.124701) (cit. on pp. 21, 58, 61, 62, 72, 107).
- Datta and Rajamani (2002). “A direct approach of modeling batch grinding in ball mills using population balance principles and impact energy distribution”. In: *International Journal of Mineral Processing* 64.4, pp. 181–200 (cit. on p. 13).
- Davey et al. (2012). “Response of primary grinding mill performance to changes in operating conditions using an on-line surface vibration monitor”. In: *Proceeding of international mineral processing congress (IMPC)*. New Delhi India, pp. 1064–1074 (cit. on p. 17).
- Degond and Mas-Gallic (Oct. 1989a). “The weighted particle method for convection-diffusion equations. I - The case of an isotropic viscosity. II - The anisotropic case”. In: *Mathematics of Computation* 53, pp. 485–507 (cit. on p. 54).

- (1989b). “The Weighted Particle Method for Convection-Diffusion Equations. Part 1: The Case of an Isotropic Viscosity”. In: *Mathematics of Computation* 53.188, pp. 485–507. ISSN: 00255718, 10886842 (cit. on p. 86).
- Diacu and Holmes (1996). *Celestial Encounters: The Origin of Chaos*. Princeton University Press. ISBN: 0-691-00545-1 (cit. on p. 117).
- Ding et al. (2001). “Solids motion in rolling mode rotating drums operated at low to medium rotational speeds”. In: *Chemical Engineering Science* 56.5, pp. 1769–1780 (cit. on p. 15).
- Doucet, Godsill, and Andrieu (2000). “On sequential Monte Carlo sampling methods for Bayesian filtering”. In: *Statistics and Computing* 10.3, pp. 197–208. ISSN: 1573-1375. DOI: [10.1023/A:1008935410038](https://doi.org/10.1023/A:1008935410038) (cit. on p. 29).
- Drucker and Prager (1952). “Soil mechanics and plastic analysis for limit design”. In: *Quarterly of Applied Mathematics* 10.2, pp. 157–165 (cit. on p. 14).
- Duarte and Oden (1996). “An hp adaptive method using clouds”. In: *Computer methods in applied mechanics and engineering* 139.1-4, pp. 237–262 (cit. on p. 45).
- Dubrule (1983). “Cross validation of kriging in a unique neighborhood”. In: *Journal of the International Association for Mathematical Geology* 15.6, pp. 687–699. ISSN: 1573-8868. DOI: [10.1007/BF01033232](https://doi.org/10.1007/BF01033232) (cit. on p. 117).
- Dunatunga and Kamrin (2015). “Continuum modeling and simulation of granular flows through their many phases”. In: *Journal of Fluids Mechanics* (cit. on pp. 38, 107).
- Duvillard, Giraldi, and Maître (2024). “Ensemble Data Assimilation for Particle-based Methods”. en. In: (cit. on pp. 107, 108, 110, 112, 120).
- EDF R&D (May 2022). *ACV du kWh nucléaire EDF - version 2022*. Tech. rep. Avenue des Renardières - Écuelles, 77818 Moret sur Loing CEDEX, France: EDF R&D, Technologies et Recherche pour l’Efficacité Énergétique, Modélisation, Environnement & Bâtiments (cit. on p. 11).
- Evensen (1994). “Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics”. en. In: *Journal of Geophysical Research: Oceans* 99.C5, pp. 10143–10162. ISSN: 2156-2202. DOI: [10.1029/94JC00572](https://doi.org/10.1029/94JC00572) (cit. on pp. 29, 72, 75, 106, 112).
- Evensen, Vossepoel, and Leeuwen (2022a). *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. en. Springer Textbooks in Earth Sciences, Geography and Environment. Cham: Springer International Publishing. ISBN: 978-3-030-96708-6. DOI: [10.1007/978-3-030-96709-3](https://doi.org/10.1007/978-3-030-96709-3) (cit. on p. 24).
- (2022b). “Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem”. en. In: *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Textbooks in Earth Sciences, Geography and Environment. Cham: Springer International Publishing. Chap. 8. ISBN: 978-3-030-96708-6. DOI: [10.1007/978-3-030-96709-3](https://doi.org/10.1007/978-3-030-96709-3) (cit. on p. 112).
- Eymard, Gallouët, and Herbin (2000). “Finite Volume Methods”. In: *Techniques of Scientific Computing (Part 3)*. Ed. by Lions and Ciarlet. Vol. 7. Handbook of Numerical Analysis. Elsevier, pp. 713–1020. DOI: [10.1016/S1570-8659\(00\)07005-8](https://doi.org/10.1016/S1570-8659(00)07005-8) (cit. on p. 59).
- Feyeux (Dec. 2016). “Transport optimal pour l’assimilation de données images”. Theses. Université Grenoble Alpes (cit. on pp. 65, 109).

- Fornberg and Flyer (2015). “Solving PDEs with radial basis functions”. In: *Acta Numerica* 24, pp. 215–258. DOI: [10.1017/S0962492914000130](https://doi.org/10.1017/S0962492914000130) (cit. on pp. 48, 60, 88).
- Forsterre and Pouliquen (Jan. 2008). “Flows of Dense Granular Media”. en. In: *Annual Review of Fluid Mechanics* 40.1, pp. 1–24. ISSN: 0066-4189, 1545-4479. DOI: [10.1146/annurev.fluid.40.111406.102142](https://doi.org/10.1146/annurev.fluid.40.111406.102142) (cit. on p. 14).
- Fu et al. (Nov. 2017). “A polynomial particle-in-cell method”. en. In: *ACM Transactions on Graphics* 36.6, pp. 1–12. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/3130800.3130878](https://doi.org/10.1145/3130800.3130878) (cit. on p. 52).
- Gaume et al. (2018). “Dynamic anticrack propagation in snow”. In: *Nature Communications* 9.1, p. 3047 (cit. on p. 38).
- GDR MiDi (Aug. 2004). “On dense granular flows”. en. In: *The European Physical Journal E* 14.4, pp. 341–365. ISSN: 1292-8941, 1292-895X. DOI: [10.1140/epje/i2003-10153-0](https://doi.org/10.1140/epje/i2003-10153-0) (cit. on p. 14).
- Geers, Kouznetsova, and Brekelmans (2010). “Multi-scale computational homogenization: Trends and challenges”. In: *Journal of Computational and Applied Mathematics* 234.7, pp. 2175–2182 (cit. on p. 38).
- Gingold and Monaghan (Dec. 1977). “Smoothed particle hydrodynamics: theory and application to non-spherical stars”. In: *Monthly Notices of the Royal Astronomical Society* 181.3, pp. 375–389. ISSN: 0035-8711. DOI: [10.1093/mnras/181.3.375](https://doi.org/10.1093/mnras/181.3.375). eprint: <https://academic.oup.com/mnras/article-pdf/181/3/375/3104055/mnras181-0375.pdf> (cit. on pp. 40, 49, 76).
- Giraud (Dec. 2020). “Analyse du comportement rhéologique des poudres à partir des propriétés des grains, application à l'étude d'un procédé de broyage/mélange pour la préparation du combustible nucléaire MOX”. fr. PhD thesis, p. 269 (cit. on p. 13).
- Guilkey, Hoying, and Weiss (2006). “Computational modeling of multicellular constructs with the material point method”. In: *Journal of Biomechanics* 39.11, pp. 2074–2086 (cit. on pp. 38, 107).
- Hamill and Snyder (2000). “A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme”. In: *Monthly Weather Review* 128.8, pp. 2905–2919. DOI: [10.1175/1520-0493\(2000\)128<2905:AHEKFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2) (cit. on p. 34).
- Hamill, Whitaker, and Snyder (2001). “Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter”. In: *Monthly Weather Review* 129.11, pp. 2776–2790. DOI: [10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2) (cit. on p. 36).
- Han and Meng (2001). “Error analysis of the reproducing kernel particle method”. In: *Computer Methods in Applied Mechanics and Engineering* 190.46, pp. 6157–6181. ISSN: 0045-7825. DOI: [https://doi.org/10.1016/S0045-7825\(01\)00214-6](https://doi.org/10.1016/S0045-7825(01)00214-6) (cit. on p. 45).
- Hansen (2000). “The L-curve and its use in the numerical treatment of inverse problems”. English. In: *Invited Computational Inverse Problems in Electrophysiology*. WIT Press (cit. on p. 117).
- Haugen and Evensen (June 2002). “Assimilation of SLA and SST data into an OGCM for the Indian Ocean”. In: *Ocean Dynamics* 52.3, pp. 133–151. ISSN: 1616-7228. DOI: [10.1007/s10236-002-0014-7](https://doi.org/10.1007/s10236-002-0014-7) (cit. on p. 36).
- Hoerner (1960). “Die numerische Integration des N-Körper-Problems für Sternhaufen”. In: *Zeitschrift für Astrophysik* 50 (cit. on p. 39).

- Hoteit and Köhl (2006). “Efficiency of reduced-order, time-dependent adjoint data assimilation approaches”. In: *Journal of Oceanography* 62, pp. 539–550 (cit. on pp. 34, 107).
- Houtekamer and Mitchell (2001). “A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation”. In: *Monthly Weather Review* 129.1, pp. 123–137. DOI: [10.1175/1520-0493\(2001\)129<123:ASEKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<123:ASEKFF>2.0.CO;2) (cit. on p. 36).
- Hu, Fang, et al. (Aug. 2018). “A moving least squares material point method with displacement discontinuity and two-way rigid body coupling”. en. In: *ACM Transactions on Graphics* 37.4, pp. 1–14. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/3197517.3201293](https://doi.org/10.1145/3197517.3201293) (cit. on p. 52).
- Hu, Liu, et al. (2019). “ChainQueen: A Real-Time Differentiable Physical Simulator for Soft Robotics”. In: *2019 International Conference on Robotics and Automation (ICRA)*. Montreal, QC, Canada: IEEE Press, pp. 6265–6271. DOI: [10.1109/ICRA.2019.8794333](https://doi.org/10.1109/ICRA.2019.8794333) (cit. on p. 38).
- Hultman and Pharasyn (1999). “Hierarchical, dissipative formation of elliptical galaxies: is thermal instability the key mechanism?. Hydrodynamical simulations including supernova feedback, multi-phase gas and metal enrichment in CDM: structure and dynamics of elliptical galaxies”. In: *Astronomy and Astrophysics*, v. 347, p. 769-798 (1999) 347, pp. 769–798 (cit. on pp. 38, 49).
- Hunt, Kostelich, and Szunyogh (2007). “Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter”. In: *Physica D* 230, pp. 112–126 (cit. on pp. 31, 34, 107, 108).
- Iglesias, Law, and Stuart (Mar. 2013). “Ensemble Kalman methods for inverse problems”. en. In: *Inverse Problems* 29.4. P. 045001. ISSN: 0266-5611. DOI: [10.1088/0266-5611/29/4/045001](https://doi.org/10.1088/0266-5611/29/4/045001) (cit. on p. 27).
- Janjić et al. (2018). “On the representation error in data assimilation”. In: *Quarterly Journal of the Royal Meteorological Society* 144, pp. 1257–1278. DOI: [10.1002/qj.3130](https://doi.org/10.1002/qj.3130) (cit. on pp. 65, 109).
- Jarray, Magnanimo, and Luding (Jan. 2019). “Wet granular flow control through liquid induced cohesion”. In: *Powder Technology*. Continuous Manufacturing/Processing 341, pp. 126–139. ISSN: 0032-5910. DOI: [10.1016/j.powtec.2018.02.045](https://doi.org/10.1016/j.powtec.2018.02.045) (cit. on p. 17).
- Jiang, Schroeder, Selle, et al. (July 2015). “The affine particle-in-cell method”. en. In: *ACM Transactions on Graphics* 34.4, pp. 1–10. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/2766996](https://doi.org/10.1145/2766996) (cit. on p. 52).
- Jiang, Schroeder, Teran, et al. (July 2016). “The material point method for simulating continuum materials”. en. In: *ACM SIGGRAPH 2016 Courses*. Anaheim California: ACM, pp. 1–52. ISBN: 978-1-4503-4289-6. DOI: [10.1145/2897826.2927348](https://doi.org/10.1145/2897826.2927348) (cit. on pp. 38, 107).
- Jop, Forterre, and Pouliquen (June 2006). “A constitutive law for dense granular flows”. en. In: *Nature* 441.7094. Pp. 727–730. ISSN: 1476-4687. DOI: [10.1038/nature04801](https://doi.org/10.1038/nature04801) (cit. on p. 14).
- Juvinal and Marshek (1991). *Fundamentals of Machine Component Design*. 2nd. New York: John Wiley & Sons, p. 217. ISBN: 0-471-62281-8 (cit. on p. 14).

- Kalman (Mar. 1960). “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1, pp. 35–45. ISSN: 0021-9223. DOI: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552) (cit. on pp. 21, 28, 72, 75, 112).
- Kapteyn, Pretorius, and Willcox (May 2021). “A probabilistic graphical model foundation for enabling predictive digital twins at scale”. en. In: *Nature Computational Science* 1.5, pp. 337–347. ISSN: 2662-8457. DOI: [10.1038/s43588-021-00069-0](https://doi.org/10.1038/s43588-021-00069-0) (cit. on p. 19).
- Kitanidis (1995). “Quasi-Linear Geostatistical Theory for Inversing”. In: *Water Resources Research* 31.10, pp. 2411–2419. DOI: <https://doi.org/10.1029/95WR01945>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95WR01945> (cit. on pp. 34, 111).
- Klaas and Shephard (2000). “Automatic generation of octree-based three-dimensional discretizations for partition of unity methods”. In: *Computational Mechanics* 25.2, pp. 296–304 (cit. on pp. 38, 107).
- Kum, Hoover, and Posch (Nov. 1995). “Viscous conducting flows with smooth-particle applied mechanics”. In: *Phys. Rev. E* 52 (5), pp. 4899–4908. DOI: [10.1103/PhysRevE.52.4899](https://doi.org/10.1103/PhysRevE.52.4899) (cit. on p. 49).
- Kumar et al. (2017). “Modelling Transient Dynamics of Granular Slopes: MPM and DEM”. In: *Procedia Engineering* 175. Pp. 94–101. ISSN: 1877-7058. DOI: <https://doi.org/10.1016/j.proeng.2017.01.032> (cit. on pp. 16, 38, 107).
- Law, Stuart, and Zygalakis (2015). *Data Assimilation: A Mathematical Introduction*. Springer. ISBN: 978-3-319-20325-6. DOI: [10.1007/978-3-319-20325-6](https://doi.org/10.1007/978-3-319-20325-6) (cit. on p. 24).
- Le Provost et al. (Jan. 2021). *A low-rank nonlinear ensemble filter for vortex models of aerodynamic flows*. DOI: [10.2514/6.2021-1937](https://doi.org/10.2514/6.2021-1937) (cit. on p. 72).
- Le Provost and Eldredge (May 2021). “Ensemble Kalman filter for vortex models of disturbed aerodynamic flows”. In: *Physical Review Fluids* 6. DOI: [10.1103/PhysRevFluids.6.050506](https://doi.org/10.1103/PhysRevFluids.6.050506) (cit. on pp. 58, 61).
- Leeuwen (2020). “A consistent interpretation of the stochastic version of the Ensemble Kalman Filter”. en. In: *Quarterly Journal of the Royal Meteorological Society* 146.731. Pp. 2815–2825. ISSN: 1477-870X. DOI: [10.1002/qj.3819](https://doi.org/10.1002/qj.3819) (cit. on p. 31).
- Lei and Bickel (2011). “A Moment Matching Ensemble Filter for Nonlinear Non-Gaussian Data Assimilation”. In: *Monthly Weather Review* 139.12, pp. 3964–3973. DOI: [10.1175/2011MWR3553.1](https://doi.org/10.1175/2011MWR3553.1) (cit. on p. 29).
- Leonard (1980). “Vortex methods for flow simulation”. In: *Journal of Computational Physics* 37.3, pp. 289–335. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(80\)90040-6](https://doi.org/10.1016/0021-9991(80)90040-6) (cit. on p. 54).
- Levy et al. (2010). “Physically-based data assimilation”. In: *Geoscientific Model Development* 3.2, pp. 669–677. DOI: [10.5194/gmd-3-669-2010](https://doi.org/10.5194/gmd-3-669-2010) (cit. on p. 58).
- Liu, Xiao, and Wang (2008). “An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test”. In: *Monthly Weather Review* 136.9, pp. 3363–3373 (cit. on pp. 34, 107).
- Lucy (1977). “A numerical approach to the testing of the fission hypothesis.” In: *The Astronomical Journal* 82, pp. 1013–1024 (cit. on pp. 40, 49, 76).

- Ma, Zhang, and Qiu (2009). “Comparison study of MPM and SPH in modeling hypervelocity impact problems”. In: *International Journal of Impact Engineering* 36.2, pp. 272–282 (cit. on pp. 38, 107).
- Mankosa, Adel, and Yoon (1986). “Effect of media size in stirred ball mill grinding of coal”. In: *Powder Technology* 49.1, pp. 75–82 (cit. on p. 13).
- Max (1950). “Inverting modified matrices”. In: *Memorandum Rept. 42, Statistical Research Group*. Princeton Univ., p. 4 (cit. on p. 31).
- Mellmann (2001). “The transverse motion of solids in rotating cylinders—forms of motion and transition behavior”. In: *Powder Technology* 118.3, pp. 251–270. ISSN: 0032-5910. DOI: [https://doi.org/10.1016/S0032-5910\(00\)00402-2](https://doi.org/10.1016/S0032-5910(00)00402-2) (cit. on p. 15).
- Michel (2011). “Displacing Potential Vorticity Structures by the Assimilation of Pseudo-Observations”. In: *Monthly Weather Review* 139.2, pp. 549–565. DOI: [10.1175/2010MWR3395.1](https://doi.org/10.1175/2010MWR3395.1) (cit. on p. 108).
- Mimeau and Mortazavi (Feb. 2021). “A Review of Vortex Methods and Their Applications: From Creation to Recent Advances”. en. In: *Fluids* 6.2, p. 68. ISSN: 2311-5521. DOI: [10.3390/fluids6020068](https://doi.org/10.3390/fluids6020068) (cit. on pp. 71, 73).
- Mishra and Rajamani (1992). “The discrete element method for the simulation of ball mills”. In: *Applied Mathematical Modelling* 16, pp. 598–604 (cit. on p. 16).
- Monaghan (1994). “Simulating Free Surface Flows with SPH”. In: *Journal of Computational Physics* 110, pp. 399–406. DOI: [10.1006/JCPH.1994.1034](https://doi.org/10.1006/JCPH.1994.1034) (cit. on pp. 38, 107).
- Monaghan (Sept. 1985). “Extrapolating B splines for interpolation”. en. In: *Journal of Computational Physics* 60.2, pp. 253–262. ISSN: 0021-9991. DOI: [10.1016/0021-9991\(85\)90006-3](https://doi.org/10.1016/0021-9991(85)90006-3) (cit. on pp. 46, 79).
- Monaghan and Gingold (1983). “Shock simulation by the particle method SPH”. In: *Journal of Computational Physics* 52.2, pp. 374–389. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(83\)90036-0](https://doi.org/10.1016/0021-9991(83)90036-0) (cit. on p. 50).
- Nairn, Bardenhagen, and Smith (2018). “Generalized contact and improved frictional heating in the material point method”. In: *Computational Particle Mechanics* 5.3, pp. 285–296 (cit. on p. 38).
- Ng, Ding, and Ghadiri (2008). “Assessment of the kinetic-frictional model for dense granular flow”. In: *Particuology* 6.1, pp. 50–58 (cit. on p. 15).
- Nguyen, Stroeven, and Sluys (2012). “Multiscale continuous and discontinuous modelling of heterogeneous materials: A review on recent developments”. In: *Journal of Multiscale Modelling* 3.4, pp. 1–42 (cit. on p. 38).
- Nicholas and Gray (2001). “Granular flow in partially filled slowly rotating drums”. In: *Journal of Fluid Mechanics* 441, pp. 1–29 (cit. on p. 15).
- Oliver, He, and Reynolds (1996). “Conditioning permeability fields to pressure data”. In: *ECMOR V-5th European conference on the mathematics of oil recovery*. European Association of Geoscientists & Engineers, cp-101 (cit. on pp. 34, 111).
- ONU - Assemblée générale (June 1968). *Résolution 2373 (XXII) - Traité sur la non-prolifération des armes nucléaires* (cit. on p. 11).
- Orlandi (Aug. 1990). “Vortex dipole rebound from a wall”. In: *Physics of Fluids A: Fluid Dynamics* 2.8, pp. 1429–1436. ISSN: 0899-8213. DOI: [10.1063/1.857591](https://doi.org/10.1063/1.857591) (cit. on p. 91).

- Orozco (2019). "Numerical modeling and rheology of crushable granular flows: application to ball mills". PhD thesis. Université de Montpellier (cit. on p. 13).
- Orozco et al. (Feb. 2020). "Discrete-element simulations of comminution in rotating drums: Effects of grinding media". In: *Powder Technology* 362, pp. 157–167. DOI: [10.1016/j.powtec.2019.12.014](https://doi.org/10.1016/j.powtec.2019.12.014) (cit. on p. 16).
- Oudinet et al. (2015). "Characterization of Plutonium Distribution in MIMAS MOX by Image Analysis". In: *Nuclear Engineering and Design* (cit. on p. 12).
- Owusu et al. (Sept. 2021). "Predicting mill feed grind characteristics through acoustic measurements". en. In: *Minerals Engineering* 171, p. 107099. ISSN: 0892-6875. DOI: [10.1016/j.mineng.2021.107099](https://doi.org/10.1016/j.mineng.2021.107099) (cit. on p. 17).
- Pedrayes et al. (Oct. 2017). "Frequency domain characterization of torque in tumbling ball mills using DEM modelling: Application to filling level monitoring". In: *Powder Technology* 323. DOI: [10.1016/j.powtec.2017.10.026](https://doi.org/10.1016/j.powtec.2017.10.026) (cit. on p. 17).
- Percival (2008). "Displacement assimilation for ocean models". PhD thesis. Citeseer (cit. on pp. 66, 109).
- Peyré and Cuturi (2019). "Computational Optimal Transport". In: *Foundations and Trends in Machine Learning* 11 (5-6), pp. 355–602 (cit. on pp. 66, 109).
- Plu (Oct. 2013). "A variational formulation for translation and assimilation of coherent structures". English. In: *Nonlinear Processes in Geophysics* 20.5. Pp. 793–801. ISSN: 1023-5809. DOI: [10.5194/npg-20-793-2013](https://doi.org/10.5194/npg-20-793-2013) (cit. on p. 108).
- Poincaré (1890). "Sur le problème des trois corps et les équations de la dynamique". In: *Acta Mathematica* 13, pp. 1–270 (cit. on p. 117).
- Prager (1928). "Die Druckverteilung an Körpern in ebener Potentialströmung". In: *Phys. Z.* 29, pp. 865–869 (cit. on p. 52).
- Prospective en Énergie de l'Académie des sciences (Oct. 2022). *Les réacteurs nucléaires modulaires de faible puissance (SMR) : État des lieux et perspectives*. Avis et Rapport. Académie des sciences (cit. on p. 11).
- Radjai and Dubois (2011). *Discrete-element modeling of granular materials*. Wiley-Iste, 425 p. (Cit. on p. 39).
- Rao and Rahman (2004). "An enriched meshless method for non-linear fracture mechanics". In: *International Journal for Numerical Methods in Engineering* 59. DOI: [10.1002/NME.868](https://doi.org/10.1002/NME.868) (cit. on pp. 38, 107).
- Ravela (Mar. 2014). "Spatial inference for coherent geophysical fluids by appearance and geometry". en. In: *IEEE Winter Conference on Applications of Computer Vision*. Steamboat Springs, CO, USA: IEEE, pp. 925–932. ISBN: 978-1-4799-4985-4. DOI: [10.1109/WACV.2014.6836005](https://doi.org/10.1109/WACV.2014.6836005) (cit. on p. 108).
- Ravela, Emanuel, and McLaughlin (June 2007). "Data assimilation by field alignment". en. In: *Physica D: Nonlinear Phenomena* 230.1-2, pp. 127–145. ISSN: 01672789. DOI: [10.1016/j.physd.2006.09.035](https://doi.org/10.1016/j.physd.2006.09.035) (cit. on pp. 66, 67, 100, 109, 110, 113).
- Réseau de Transport d'Electricité (RTE) (2023). *Bilan électrique 2023*. Tech. rep. Paris, France: Réseau de Transport d'Electricité (RTE) (cit. on p. 11).
- Rognon et al. (2015). "Long-range wall perturbations in dense granular flows". In: *Journal of Fluid Mechanics* 764, pp. 171–192. DOI: [10.1017/jfm.2014.707](https://doi.org/10.1017/jfm.2014.707) (cit. on p. 15).
- Rosenhead (1931). "The formation of vortices from a surface of discontinuity". In: *Proc. R. Soc. Lond. A* 134, pp. 170–192. DOI: [10.1098/rspa.1931.0189](https://doi.org/10.1098/rspa.1931.0189) (cit. on p. 52).

- Rosenthal et al. (Feb. 2017). “Displacement data assimilation”. In: *Journal of Computational Physics* 330, pp. 594–614. ISSN: 0021-9991. DOI: [10.1016/j.jcp.2016.10.025](https://doi.org/10.1016/j.jcp.2016.10.025) (cit. on pp. 66, 67, 100, 109, 110, 113).
- Russell, Souto-Iglesias, and Zohdi (2018). “Numerical simulation of Laser Fusion Additive Manufacturing processes using the SPH method”. In: *Computer Methods in Applied Mechanics and Engineering*. DOI: [10.1016/J.CMA.2018.06.033](https://doi.org/10.1016/J.CMA.2018.06.033) (cit. on pp. 38, 107).
- Russo and Strain (1994). “Fast triangulated vortex methods for the 2D Euler equations”. In: *Journal of Computational Physics* 111, pp. 291–323 (cit. on p. 73).
- S. Li and W. K. Liu (2004). *Meshfree Particle Methods*. en. Berlin, Heidelberg: Springer. ISBN: 978-3-540-22256-9. DOI: [10.1007/978-3-540-71471-2](https://doi.org/10.1007/978-3-540-71471-2) (cit. on pp. 43, 107).
- Sakov and Bertino (Mar. 2011). “Relation between two common localisation methods for the EnKF”. In: *Computational Geosciences* 15.2, pp. 225–237. ISSN: 1573-1499. DOI: [10.1007/s10596-010-9202-6](https://doi.org/10.1007/s10596-010-9202-6) (cit. on p. 36).
- Santos et al. (2013). “Experimental and CFD study of the hydrodynamic behavior in a rotating drum”. In: *Powder Technology* 250, pp. 52–62 (cit. on p. 16).
- Sasaki (1958). “An ObJective Analysis Based on the Variational Method”. en. In: *Journal of the Meteorological Society of Japan. Ser. II* 36.3, pp. 77–88. ISSN: 0026-1165, 2186-9057. DOI: [10.2151/jmsj1923.36.3_77](https://doi.org/10.2151/jmsj1923.36.3_77) (cit. on pp. 21, 71).
- Schlömer et al. (2014). “Technology-specific Cost and Performance Parameters”. In: *Climate change 2014: mitigation of climate change: Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by Intergovernmental Panel on Climate Change and Edenhofer. New York, NY: Cambridge University Press, numéro de page début–numéro de page fin. ISBN: 978-1-107-05821-7 (cit. on p. 11).
- Shadloo, Oger, and Le Touzé (2016). “Smoothed particle hydrodynamics method for fluid flows, towards industrial applications: Motivations, current state, and challenges”. In: *Computers & Fluids* 136, pp. 11–34 (cit. on p. 38).
- Shadloo, Oger, and Le Touzé (2016). “Smoothed particle hydrodynamics method for fluid flows, towards industrial applications: Motivations, current state, and challenges”. In: *Computers & Fluids* 136, pp. 11–34. ISSN: 0045-7930. DOI: <https://doi.org/10.1016/j.compfluid.2016.05.029> (cit. on p. 49).
- Simsekli et al. (2018). “Sliced-Wasserstein Flows: Nonparametric Generative Modeling via Optimal Transport and Diffusions”. In: *International Conference on Machine Learning* (cit. on pp. 66, 109).
- Siripatana et al. (Dec. 2019). “Combining ensemble Kalman filter and multiresolution analysis for efficient assimilation into adaptive mesh models”. en. In: *Computational Geosciences* 23.6, pp. 1259–1276. ISSN: 1420-0597, 1573-1499. DOI: [10.1007/s10596-019-09882-z](https://doi.org/10.1007/s10596-019-09882-z) (cit. on pp. 64, 72).
- Sołowski and Sloan (2015). “Evaluation of material point method for use in geotechnics”. In: *International Journal for Numerical and Analytical Methods in Geomechanics* 39, pp. 685–701. DOI: [10.1002/nag.2321](https://doi.org/10.1002/nag.2321) (cit. on pp. 38, 107).
- Sperotto, Pieraccini, and Mendez (June 2022). “A meshless method to compute pressure fields from image velocimetry”. In: *Measurement Science and Technology* 33.9, p. 094005. DOI: [10.1088/1361-6501/ac70a9](https://doi.org/10.1088/1361-6501/ac70a9) (cit. on pp. 48, 73).

- Stomakhin et al. (2014). “Augmented MPM for phase-change and varied materials”. In: *ACM Transactions on Graphics (TOG)* 33, pp. 1–11. DOI: [10.1145/2601097.2601176](https://doi.org/10.1145/2601097.2601176) (cit. on pp. 38, 107).
- Stomakhin et al. (2013). “A material point method for snow simulation”. In: *ACM Transactions on Graphics (TOG)* 32.4, pp. 1–10 (cit. on p. 52).
- Sulsky, Chenb, and Schreyer (1994). “A particle method for history-dependent materials”. en. In: *Computer methods in applied mechanics and engineering*, p. 18 (cit. on pp. 38, 40, 42, 50, 73, 76).
- Swope et al. (Jan. 1982). “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters”. In: *The Journal of Chemical Physics* 76.1, pp. 637–649. ISSN: 0021-9606. DOI: [10.1063/1.442716](https://doi.org/10.1063/1.442716) (cit. on p. 39).
- Talagrand (1997). “Assimilation of Observations, an Introduction”. In: *Journal of the Meteorological Society of Japan* 75.1B, pp. 191–209 (cit. on pp. 21, 32, 71, 107).
- Tano, Pålsson, and Sellgren (Sept. 2005). “On-line lifter deflection measurements showing flow resistance effects in grinding mills”. In: *Minerals Engineering* 18.11, pp. 1077–1085. ISSN: 0892-6875. DOI: [10.1016/j.mineng.2005.02.005](https://doi.org/10.1016/j.mineng.2005.02.005) (cit. on p. 17).
- Van Geffen, Meleshko, and Van Heijst (1996). “Motion of a two-dimensional monopolar vortex in a bounded rectangular domain”. In: *Physics of Fluids* 8.10, pp. 2393–2399 (cit. on p. 117).
- Villani et al. (2009). *Optimal transport: old and new*. Vol. 338. Springer (cit. on pp. 65, 109).
- Vu (Nov. 2023). “Quasi-static and dynamic granular flows : scaling behavior, microstructure, and particle shape effects”. en. PhD thesis. Université de Montpellier (cit. on p. 13).
- Wang et al. (Nov. 2021). “Assessing load in ball mill using instrumented grinding media”. In: *Minerals Engineering* 173, p. 107198. ISSN: 0892-6875. DOI: [10.1016/j.mineng.2021.107198](https://doi.org/10.1016/j.mineng.2021.107198) (cit. on p. 18).
- Wang et al. (2016). “An overview of smoothed particle hydrodynamics for simulating multiphase flow”. In: *Applied Mathematical Modelling* 40, pp. 9625–9655. DOI: [10.1016/j.apm.2016.06.030](https://doi.org/10.1016/j.apm.2016.06.030) (cit. on p. 38).
- Y. et al. (Nov. 2015). “Continuum Foam: A Material Point Method for Shear-Dependent Flows”. en. In: *ACM Transactions on Graphics* 34.5, pp. 1–20. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/2751541](https://doi.org/10.1145/2751541) (cit. on p. 72).
- Yang et al. (2014). “Improved decohesion modeling with the material point method for simulating crack evolution”. In: *International Journal of Fracture* 186.1-2, pp. 177–184 (cit. on p. 38).
- Yim, Kim, and Sohn (Dec. 2022). “Motion of three geostrophic Bessel vortices”. In: *Physica D: Nonlinear Phenomena* 441, p. 133509. ISSN: 0167-2789. DOI: [10.1016/j.physd.2022.133509](https://doi.org/10.1016/j.physd.2022.133509) (cit. on p. 117).
- Ying (Dec. 2019). “A Multiscale Alignment Method for Ensemble Filtering with Displacement Errors”. EN. In: *Monthly Weather Review* 147.12. Pp. 4553–4565. ISSN: 1520-0493, 0027-0644. DOI: [10.1175/MWR-D-19-0170.1](https://doi.org/10.1175/MWR-D-19-0170.1) (cit. on pp. 66, 109, 113).
- Ying, Anderson, and Bertino (June 2023). “Improving Vortex Position Accuracy with a New Multiscale Alignment Ensemble Filter”. en. In: *Monthly Weather Review* 151.6,

- pp. 1387–1405. ISSN: 0027-0644, 1520-0493. doi: [10.1175/MWR-D-22-0140.1](https://doi.org/10.1175/MWR-D-22-0140.1) (cit. on pp. 66, 109).
- Zhang, Ming, and Cao (2014). “Total Lagrangian particle method for the large-deformation analyses of solids and curved shells”. In: *Acta Mechanica* 225.1, pp. 253–275 (cit. on pp. 38, 107).
- Zhang et al. (2008). “Material point method applied to multiphase flows”. In: *J. Comput. Phys.* 227, pp. 3159–3173. doi: [10.1016/j.jcp.2007.11.021](https://doi.org/10.1016/j.jcp.2007.11.021) (cit. on p. 38).
- Zheng and Yu (2015). “Modelling the granular flow in a rotating drum by the Eulerian finite element method”. In: *Powder Technology* 286, pp. 361–370. ISSN: 0032-5910. doi: <https://doi.org/10.1016/j.powtec.2015.08.025> (cit. on p. 16).
- Zhu, Peng, and Wu (Aug. 2022). “Lagrangian meshfree particle method (SPH) based simulation for granular flow in a rotating drum with regularized mu(I) elastoplastic model”. en. In: *Powder Technology* 408, p. 117699. ISSN: 0032-5910. doi: [10.1016/j.powtec.2022.117699](https://doi.org/10.1016/j.powtec.2022.117699) (cit. on p. 16).
- Zuo, Gong, and Xie (Mar. 2020). “Numerical simulation of granular mixing in a rotary drum using a generalized interpolation material point method”. en. In: *Asia-Pacific Journal of Chemical Engineering* 15.2. ISSN: 1932-2135, 1932-2143. doi: [10.1002/apj.2426](https://doi.org/10.1002/apj.2426) (cit. on p. 16).



Titre: Méthodes d'assimilation de données pour des simulations lagrangiennes particulières

Mots clés: Assimilation de données, Méthodes particulières, Méthodes d'ensemble, Simulation par la méthode vortex

Résumé: Cette thèse porte sur le développement de méthodes d'assimilation de données pour les simulations lagrangiennes basées sur une discréétisation particulaire, avec des applications en simulation en mécanique des fluides ou des écoulements granulaires. Nous étudions des situations où un ensemble de simulations est utilisé pour représenter la distribution de l'état de la simulation. Dans ce contexte, différentes discréétisations de support doivent être prises en compte lors du développement des méthodes de mise à jour et de l'analyse.

Nous avons d'abord d'adapter le filtre de Kalman d'ensemble pour corriger les champs en modifiant uniquement leurs intensités. La distribu-

tion de la position des particules reste alors inchangée ou est régénérée sur une grille régulière.

Ensuite, nous présentons une approche variationnelle d'ensemble pour corriger les positions du support particulaire. Nous montrons que cette implémentation peut être combinée avec le premier filtre pour corriger séquentiellement les positions et les intensités. Nous évaluons ces différentes implémentations sur des applications en dynamique des fluides incompressibles à l'aide de la méthode vortex, et nous démontrons l'efficacité du filtre sur des problèmes d'advection où une erreur de position peut être présente.

Title: Data Assimilation Methods for Lagrangian Simulations

Keywords: Data Assimilation, Particle-based Methods, Ensemble Methods, Vortex Simulation

Abstract: This thesis focuses on the development of data assimilation methods for Lagrangian particle-based simulations, with applications in fluid mechanics simulations or granular flow simulations. We study situations where an ensemble of simulations is used to represent the distribution of the simulation state. In this context, different support discretizations must be considered during the development of update methods and analysis.

We first adapt the ensemble Kalman filter to correct fields by modifying only their intensities.

The distribution of particle positions remains unchanged or is regenerated on a regular grid.

Next, we present a variational ensemble approach to correct the positions of the particle support. We show that this implementation can be combined with the first filter to sequentially correct positions and intensities. We evaluate these different implementations on applications in incompressible fluid dynamics using the vortex method, and we demonstrate the filter's effectiveness on advection problems where a positional error may be present.