

Data Assimilation by Alignment for Lagrangian Simulation

Marius Duvillard¹ and Olivier Le Maître²

¹CEA, DES, IRESNE, DEC, SESC, LMCP, Cadarache, F-13108
Saint-Paul-Lez-Durance, France

²CNRS, Inria, Centre de Mathématiques Appliquées, Ecole Polytechnique,
IPP, Route de Saclay, 91128, Palaiseau Cedex, France

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 an adapted EnKF update. 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.

Keywords: Lagrangian Simulation, Particle-based Method, Data Assimilation, Position Correction, Ensemble Methods, Vortex Methods, Field Alignment.

Contents

1	Introduction	2
2	Background	6
2.1	Model setting	6
2.2	Classical ensemble data assimilation method	6
2.3	Ensemble Filtering with Displacement Error Adjustment	8

3 Methods	9
3.1 Position transformation	9
3.2 Alignment step	10
3.3 Ensemble-based velocity field decomposition	10
3.4 Decomposition with the Singular Value Decomposition	11
3.5 Gradient computation	13
3.6 Selection of the penalization coefficient	13
4 Applications	14
4.1 Problem description	14
4.2 Assimilation of vortex centers	14
4.3 Amplitude filters	17
4.4 Alignment step	18
4.5 SVD decomposition	22
4.6 Assimilation with position-strength correction	24
4.7 Filters performances and parametric studies	25
5 Conclusion	29
A Parameters	30
B Gradient computation	30

1 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 [9]. 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 in [22], 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 [53]. 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 [27, 28, 35]. 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 [49]. 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 [11, 59], complex geometries [33, 10, 5], fragmentation [44, 36], free-surface flows [38], multi-phase flows [52], granular flows [34, 20], and various applications in geotechnical engineering [3, 51], biomechanics [25], manufacturing [1, 48], and computer graphics [30].

Applying ensemble methods to Lagrangian frameworks presents significant challenges when handling an ensemble of realizations discretized on different particle configurations. Although some approaches attempt to incorporate the positions and strengths of particles within the state definition [12, 17], 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 [21] 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 a regular discretization of particle positions, and the Part-EnKF, which focuses solely on updating particle strengths on each given member particle discretization.

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 the context of the Ensemble Kalman Filter (EnKF), which explicitly combines ensemble members [28]. 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 [21], which corrects intensities without altering particle positions, projects an analyzed solution onto a particle support that may not be suitable as illustrated in Figure 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.

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. 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 [42, 37, 45, 13]. Nevertheless, current assimilation methods still fall short in adequately addressing these errors. These methods predominantly focus on correcting

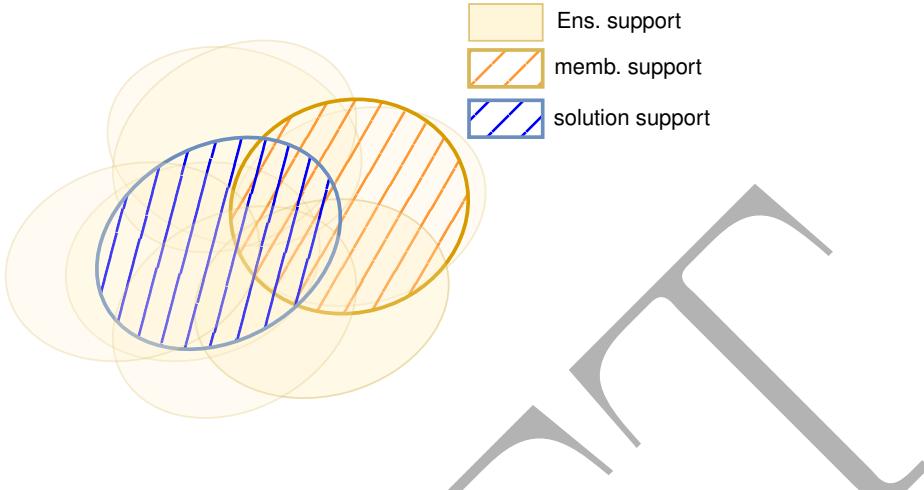


Figure 1: Although the union of the particle supports of all ensemble members overlaps with the true solution, individual ensemble members may have disjoint supports that do not fully cover the solution.

field intensities, with errors assessed using an L_2 norm. However, this approach tends to overestimate errors arising from misalignment, as illustrated in Figure 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 [2]. The double penalty effect significantly contributes to representation error, as highlighted in [29].

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 [55, 6]. This method involves substituting the quadratic norm with the Wasserstein distance. For instance, the following thesis [24] investigated the use of Wasserstein distance for data assimilation in images. From there, [8] adapted the 3DVar method to optimal transport by applying this approach to state and observation distributions with potentially differing masses. Although optimal transport methods are increasingly computationally efficient [16, 41, 50], they remain in their early stages of data assimilation.

Other approaches incorporate an alignment variable determined through displacement assimilation methods. For instance, [40] solve a rearrangement problem in order to apply a coordinate transformation. In the case the structure features are directly described, [14] propose directly assimilating the position of a tornado using the Ensemble Kalman Filter (EnKF). When the features are implicit, [46] 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[57, 58]. Finally, [47] developed a two-step sequential method to align and correct intensities successively. Their goal is to preserve the morphological properties of

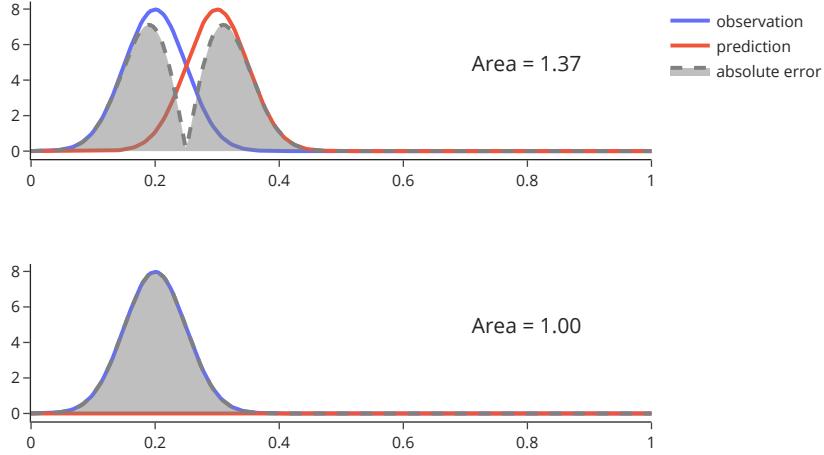


Figure 2: Illustration of the double penalty effect: A shift in the solution results in an error measurement that exceeds the error associated with the null function.

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 [46, 47], 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 in Section 2.1 to solve the Navier-Stokes equations for an inviscid, incompressible flow. In Section 2.2, we provide background on the intensity correction algorithm for Lagrangian simulations, followed by the field alignment paradigm introduced by Ravela [46]. In Section 3, 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 [21]. In Section 4, 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.

2 Background

2.1 Model setting

In this article, we employ the vortex method to solve the Navier-Stokes equation for inviscid incompressible flow. The Vortex Method [15] 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

$$\begin{cases} \frac{\partial \omega}{\partial t} + (\mathbf{v} \cdot \nabla) \omega = 0, \\ \Delta \mathbf{v} = -\nabla \times \omega, \end{cases} \quad (1)$$

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 $\Delta \mathbf{v} = -\nabla \times \omega$ and determined \mathbf{v} . The equation of evolution is defined by using the Lagrangian version of Equation (1) such as

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

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

2.2 Classical ensemble data assimilation method

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

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

$$p(\omega | \mathbf{y}) \propto p(\mathbf{y} | \omega) p(\omega), \quad (2)$$

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

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

where $\|\cdot\|_{\mathbf{A}}^2$ denote the weight norm defined by $\|\mathbf{v}\|_{\mathbf{A}}^2 = \mathbf{v}^T \mathbf{A}^{-1} \mathbf{v}$ for any positive-definite \mathbf{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 [32] and [39]. In the weakly non-linear case, the posterior distribution is approximately sampled by minimizing the following ensemble of cost functions

$$\mathcal{L}_{\omega,i}(\boldsymbol{\omega}) = \frac{1}{2} \|\mathbf{d} - \mathcal{H}(\boldsymbol{\omega})\|_{\mathbf{R}}^2 + \frac{1}{2} \|\boldsymbol{\omega} - \boldsymbol{\omega}_i^f\|_{\mathbf{P}^f}^2, \quad \text{for } i = 1, \dots, N, \quad (3)$$

which differ from the cost function \mathcal{L}_ω by substituting \mathbf{y} with $\mathbf{d}_i \sim \mathcal{N}(\mathbf{y}, \mathbf{R})$, representing the perturbed measurements, and the mean $\boldsymbol{\omega}^f$ with random samples $\boldsymbol{\omega}_i^f \sim \mathcal{N}(\boldsymbol{\omega}^f, \mathbf{P}^f)$. The solutions $\boldsymbol{\omega}_i^a$ 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 (3). This limitation has been addressed in the work [21] by utilizing the Ensemble Kalman Filter (EnKF) update. The EnKF, a widely used filtering method introduced by Evensen [22], provides a solution to (3) in the form of the Kalman update [31]. 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 [23]. Through this approach, the updated states $\boldsymbol{\omega}_i^a$ are directly determined as a linear combination of the ensemble members, such that for $\mathbf{x} \in \Omega$

$$\boldsymbol{\omega}_i^a(\mathbf{x}) = \boldsymbol{\omega}_i^f(\mathbf{x}) + \sum_j \mathbf{F}_{ji} \boldsymbol{\omega}_j^f(\mathbf{x}) \quad i = 1, \dots, N,$$

where \mathbf{F} is defined as

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

where \mathbf{Y} is the matrix where each column contains the centered prediction of each realization. Thus, the update relies solely on the observation $\boldsymbol{\omega}$, the predictions $\{\mathbf{y}_i = \mathcal{H}(\boldsymbol{\omega}_i^f)\}_{i=1}^N$, and noise samples $\{\boldsymbol{\varepsilon}_i\}_{i=1}^N$ such as $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. In [21], two formulations of the EnKF filter are introduced, depending on the choice of discretization for $\boldsymbol{\omega}_i^a$ 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.

2.3 Ensemble Filtering with Displacement Error Adjustment

We first address the alignment error by decomposing the state error as $\eta = \eta^d + \eta^a$, where η^d represents the displacement or position error and η^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 Equation (2), the posterior distribution is given by

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

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

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

$$p(\Psi) \propto \exp(-L(\Psi)),$$

where L is a regularization function. For instance, in [46], the transformation is penalized for smoothness, while in [47], 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

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

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

A two-step approach is generally used [46, 47, 57] to determine ω' and Ψ' sequentially:

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

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

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

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

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

As in equation 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 2.2, but is applied within the aligned coordinate space. This means that the amplitude step may be performed by using classical filters on the state $\omega_{\Psi'}$.

3 Methods

Our contribution focuses on defining an alignment process to correct both particle positions and strengths in Lagrangian simulations. We build on the two-stage formulation previously introduced, which allows for sequential correction of position and strength. Our approach leverages the methods discussed in Section 2.2 for the strength 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.

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

$$\begin{cases} \mathbf{x}'(\tau = 0) = \mathbf{x}, \\ \frac{d\mathbf{x}'}{d\tau}(\tau) = \mathbf{u}(\mathbf{x}'(\tau); \mathbf{a}), \quad \Psi(\mathbf{x}) = \mathbf{x}'(\tau = 1). \end{cases} \quad (4)$$

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.

3.2 Alignment step

The cost function defined in (2.3) can be reformulated using the previously defined transformation. For each member $\{\omega_i^f\}_{i=1}^N$ 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, $\mathcal{H}(\omega_i^f | \mathbf{u})$ is the observation operator conditioned by \mathbf{u} . Specifically, $\mathcal{H}(\omega^f | \mathbf{u}) = \mathcal{H}(\omega_i^f \circ \Psi^{-1}(\cdot; \mathbf{u}))$ represents the observation of ω^f after applying the position correction. The regularization based on Ψ 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 strength correction strategy described in Section 2.2. 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.

3.3 Ensemble-based velocity field 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 $\mathcal{V} = \text{Span}(\{\mathbf{v}_j\}_{j=1}^N)$, 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

$$\mathbf{u} = \sum_{j=1}^N a_j \frac{\mathbf{v}_j}{\sqrt{N}}. \quad (5)$$

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}_i 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, 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 $\mathbf{B} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i \otimes \mathbf{v}_i$. This reflects how much each direction in the space is amplified or attenuated by the transformation given by \mathbf{B} . Figure 3 shows that a non-regularized solution is usually unrealistic.

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

$$\mathcal{L}_i(\mathbf{a}) = \|h(\mathbf{a}) - \mathbf{d}_i\|_{\mathbf{R}}^2 + \frac{\lambda}{2} \|\mathbf{a}\|_2^2, \quad (6)$$

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

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

3.4 Decomposition with the Singular Value Decomposition

In the previous paragraph, we established a coherent space for aligning the field, which is both kinematically admissible and locally defined around the particles. As the data assimilation progresses and the members become aligned, the need for a large number of members in the decomposition of \mathbf{u} decreases. To improve computational efficiency, we aim to reduce the dimensionality of the problem.

Therefore, we chose to introduce a method based on Singular Value Decomposition (SVD), which orthogonalizes the initial decomposition. This decomposition allows us to identify

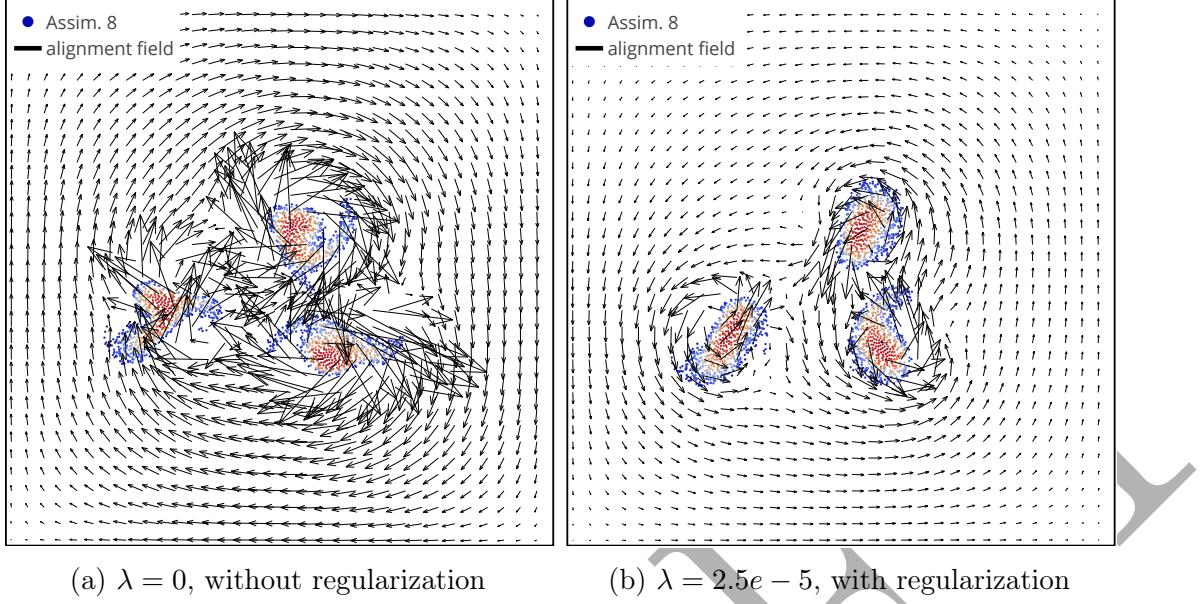


Figure 3: Alignment fields and particle distributions with and without regularization in the simulation. Without regularization, the field is noisier, leading to distortions in particle distribution.

the most significant modes, i.e., those that contribute the most to the information captured in the ensemble velocity space.

We start by discretizing the fields and representing them as vectors. Let $\mathbf{A} = \frac{1}{\sqrt{N}} \mathbf{X}$, where the i -th column of $\mathbf{X} \in \mathbb{R}^{m \times N}$ contains the values of the velocity field \mathbf{v}_j over the discretization. The matrix \mathbf{A} is factorized, such as

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T,$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V}^T \in \mathbb{R}^{N \times N}$ are two unitary matrices and $\Sigma \in \mathbb{R}^{m \times N}$ a rectangular diagonal matrix whose entries $\Sigma_{ii} = \sigma_i$ are the singular values sorted in descending order. The columns of \mathbf{U} represent the modes in the discretized space, and \mathbf{V}^T is the transpose matrix whose rows represent the modes in the reduced space.

To reduce dimensionality, we only keep the first highest r singular values with their corresponding modes approximating the matrix \mathbf{A} as

$$\mathbf{A} \approx \mathbf{U}_r \Sigma_r \mathbf{V}_r^T$$

where Σ_r is the upper-left $r \times r$ block of Σ , \mathbf{U}_r and \mathbf{V}_r contain respectively the first columns and rows of \mathbf{U} and \mathbf{V} .

In this truncated form, $\mathbf{U}_r \Sigma_r$ represents the principal modes in the discretized space, while the columns of \mathbf{V}_r represent the principal modes in the ensemble space. Therefore, the modes can be directly expressed as

$$\mathbf{m}_j = \sum_{j=1}^N V_{ij} \mathbf{v}_i \quad j = 1, \dots, r$$

The decomposition of the solution \mathbf{u} is then given by

$$\mathbf{u} = \sum_{j=1}^r a_j \mathbf{m}_j.$$

To determine the number of dominant modes r , we use a criterion based on the cumulative ratio of the singular values squares, which represents the energy captured by the retained modes. Specifically, we retain the highest r modes such that

$$\frac{\sum_{i=1}^r \sigma_i^2}{\sum_{i=1}^N \sigma_i^2} < 1 - \tau,$$

where τ is a positive threshold indicating the allowing loss of energy.

3.5 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

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

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 B. This method efficiently computes the gradient by leveraging information from the perturbations, thereby reducing computational costs. Another alternative is to use auto-differentiation methods, in the case where the simulation is based on appropriate programming tools.

3.6 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 Equation (4), and the norm of the member velocity fields \mathbf{v}_i . The normalization of the decomposition by \sqrt{N} is used to scale with the ensemble size. Therefore, this hyperparameter must be carefully calibrated to avoid overfitting. Several methods can be used for this purpose, including L-curve analysis [26], cross-validation [19], or the discrepancy principle [7]. In our study, we qualitatively selected the coefficient that facilitated correction while minimizing significant distortion to the problem under investigation.

4 Applications

4.1 Problem description

We assess the performance of the filters by using the three-vortex problem [4, 56]. 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 [43, 18].

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 [54]. A Bessel vortex is defined by a continuous vorticity field on a circle of radius R , such as

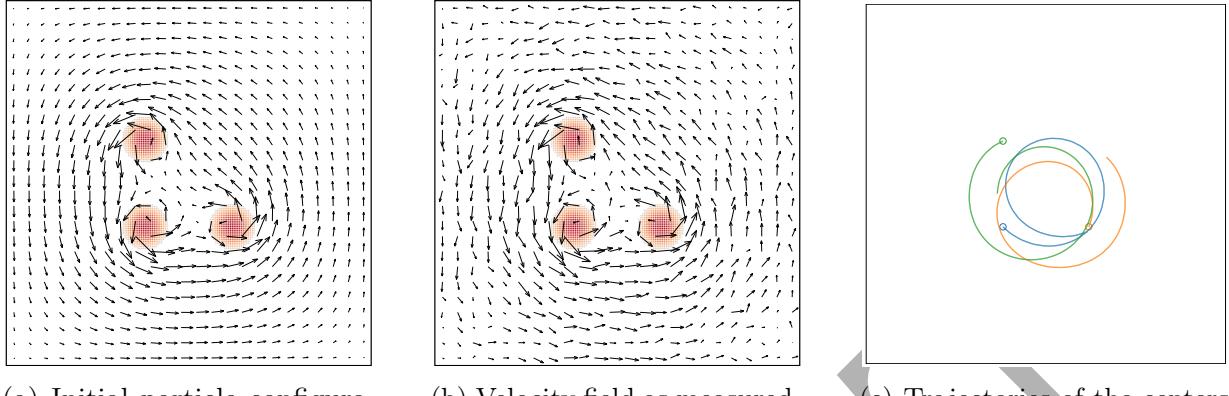
$$\omega(r) = \begin{cases} \Gamma J_0\left(\frac{kr}{R}\right), & r < R, \\ 0 & \text{else,} \end{cases}$$

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 4a and 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 2.

4.2 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 2. Even if the positions of the vortex centers, the core size R , as well as their strengths Γ , are close to the reference configuration, the Figure 5 shows that at the end of the forecast, the vortices are extensively dispersed. They are no longer closely



(a) Initial particle configuration along with the induced velocity field.

(b) Velocity field as measured.

(c) Trajectories of the centers of the three vortices over time, $t \in [0, 50]$

Figure 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.

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.

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 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 strength corrections, as summarized in Table 1.

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

Table 1: Summary of different filter combinations.

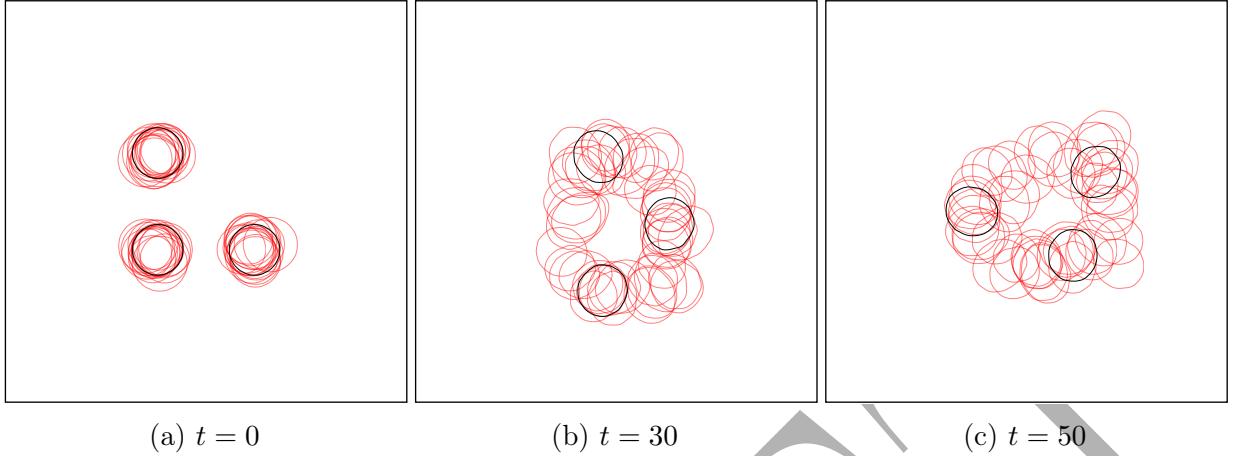


Figure 5: Evolution of the dispersion of the ensemble without assimilation. The figures illustrate the dispersion at different time points: $t = 0$, $t = 30$, and $t = 50$.

Error definitions

We assess the alignment errors of the vortex centers for each vortex $v = 1, 2, 3$ of member $i = 1, \dots, N$. Each particle p within the particle set \mathcal{P}^i is assigned to a vortex v during initialization, forming the subset \mathcal{P}_v^i . It induces a partition of the particle set such that $\mathcal{P}^i = \mathcal{P}_1^i \cup \mathcal{P}_2^i \cup \mathcal{P}_3^i$. The vortex number is treated as a particle variable and is consequently added to the state vector during assimilation as well as the remeshing process. The alignment error for ensemble member i is then defined as the normalized quantity

$$e_i^c = \frac{1}{3R} \sum_{v=1}^3 \left\| \mathbf{x}_{v,gt}^c - \mathbf{x}_{v,i}^c \right\|_2^2, \quad \text{with } \mathbf{x}_{v,i}^c = \frac{M_v^1}{M_v^0} = \frac{\sum_{p \in \mathcal{P}_v^i} \Gamma_p \mathbf{x}_p}{\sum_{p \in \mathcal{P}_v^i} \Gamma_p}$$

where R is the core size of the ground truth vortices and $\mathbf{x}_{c,v}^i$ denoting the normalized first moment of vorticity for member i (i.e., the barycenter of the vortex). We evaluate the errors within the ensemble to assess the variability, and use box plots to illustrate the distribution of errors across the ensemble, providing insights into the uncertainties of vortex position estimation. This enables us to determine whether the ensemble successfully recovers the mean vortex positions.

To globally assess the performance of the different methods, we evaluate the error on the vorticity field. The normalized root mean square error (RMSE) over the entire domain is defined as

$$e_\omega = \frac{1}{\|\omega_{gt}\|_{L_2}} \left(\frac{1}{N} \sum_{i=1}^N \|\omega_i - \omega_{gt}\|_{L_2}^2 \right)^{1/2}$$

where ω_i is the i -th ensemble member and the norm $\|\omega\|_{L_2}^2 = \int_{\Omega} \|\omega(\mathbf{x})\|^2 d\mathbf{x}$ is computed over the entire domain. Due to variability introduced by the initial ensemble sampling and noise realization, the mean and variance of e_ω are estimated by performing $M = 10$

experiments, each with different ensemble member samples and noise realizations, while maintaining the same twin experiment for comparison. The use of RMSE allows for a combined evaluation of both the mean estimate from the ensemble Kalman filter (EnKF) and the associated uncertainty across the distribution.

4.3 Amplitude filters

We begin by performing data assimilation using two previously developed EnKF-based filters: Part-EnKF and Remesh-EnKF [21]. 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 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.

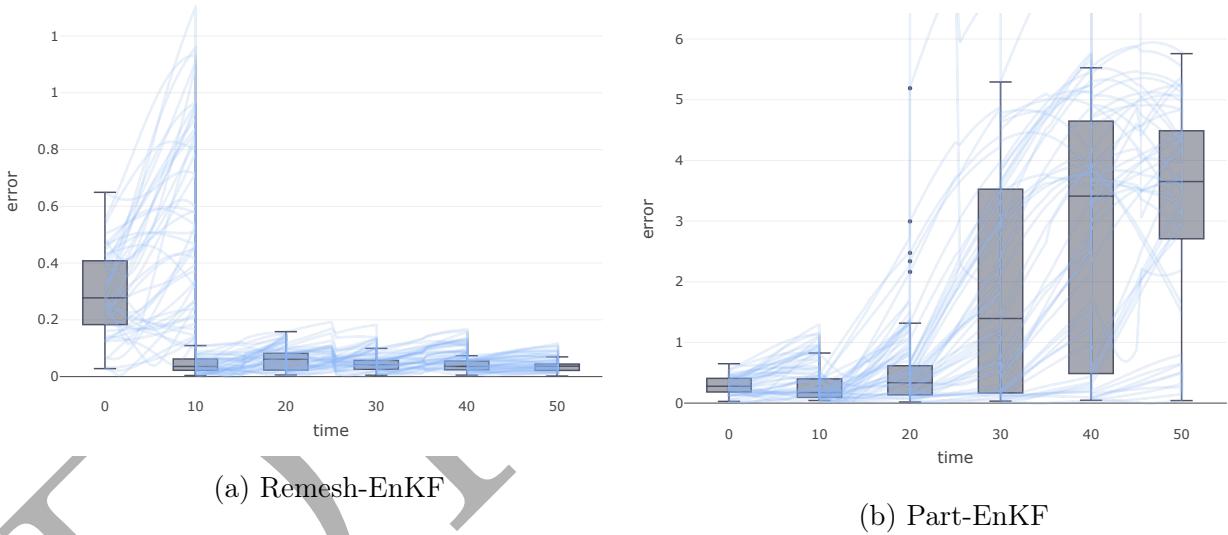


Figure 6: Position error e_c over time for the strength correction filters. The Remesh-EnKF filter demonstrates convergence for the given configuration, whereas the Part-EnKF filter shows divergence as time progresses.

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 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 strength, 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.

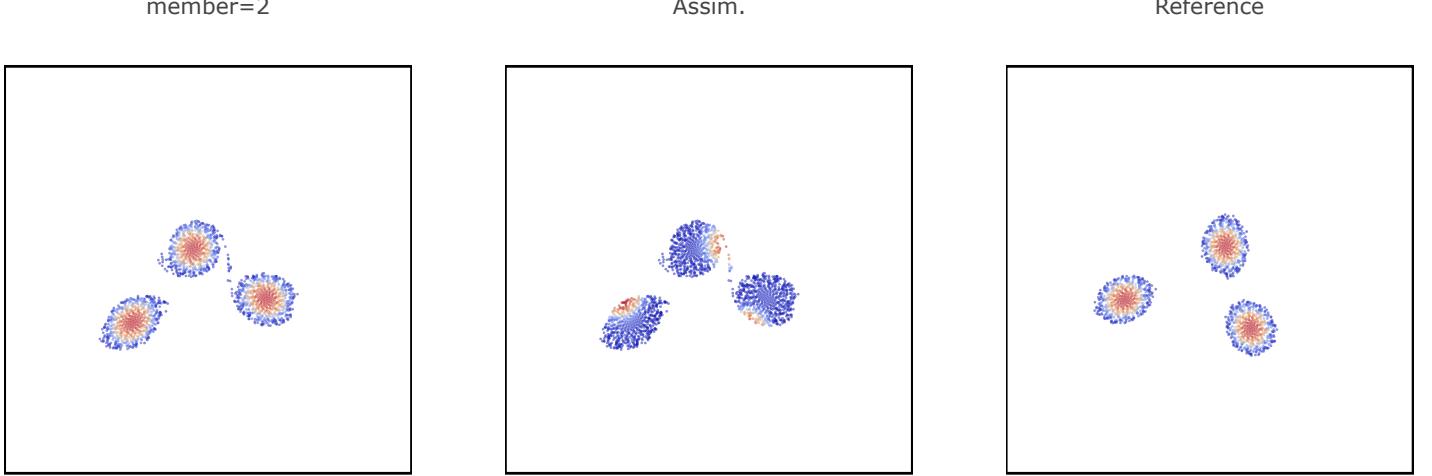


Figure 7: Update of member 2 using Part-EnKF. The particle configuration does not align with the correction, resulting in partial recovery of the continuous assimilation solution.

4.4 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 (6) over an ensemble of 24 members. In this way, we determined for each member i the alignment velocity field \mathbf{u}_i first define like in Equation (5).

In Figure 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 strength. Although the alignment step may introduce changes in the vortex shape, the subsequent strength correction step can project necessary adjustments to refine the vortex geometry and enhance the alignment. The alignment step is illustrated explicitly in Figure 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

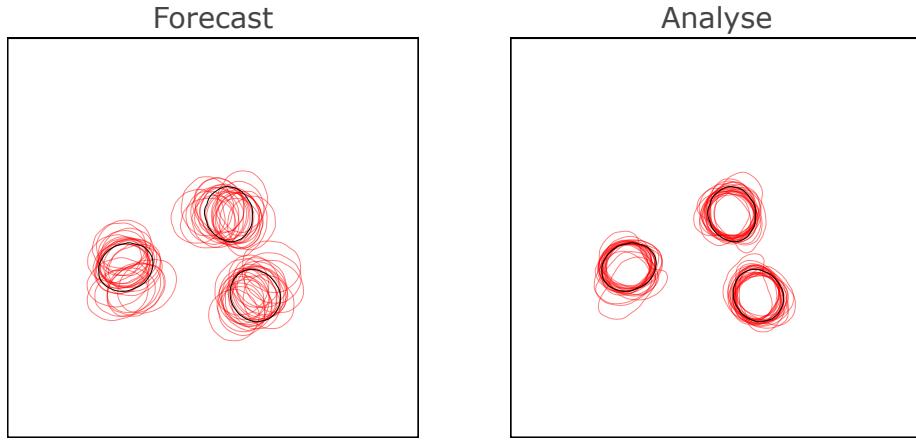


Figure 8: Contour plots of the vortex ensemble before and after the alignment process at the first assimilation step. The alignment step significantly improves the match between the dispersed ensemble and the reference configuration.

the 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 on Figure 9d. 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. 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 Figure 10a. We get that the first singular mode 10a corresponds to a phase correction and looks close to the mean velocity field. In contrast, the mode depicted in Figure 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.

The last modes are residual ones. In fact, it is observed that very few 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 10f. For instance, the mode 23 10e 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

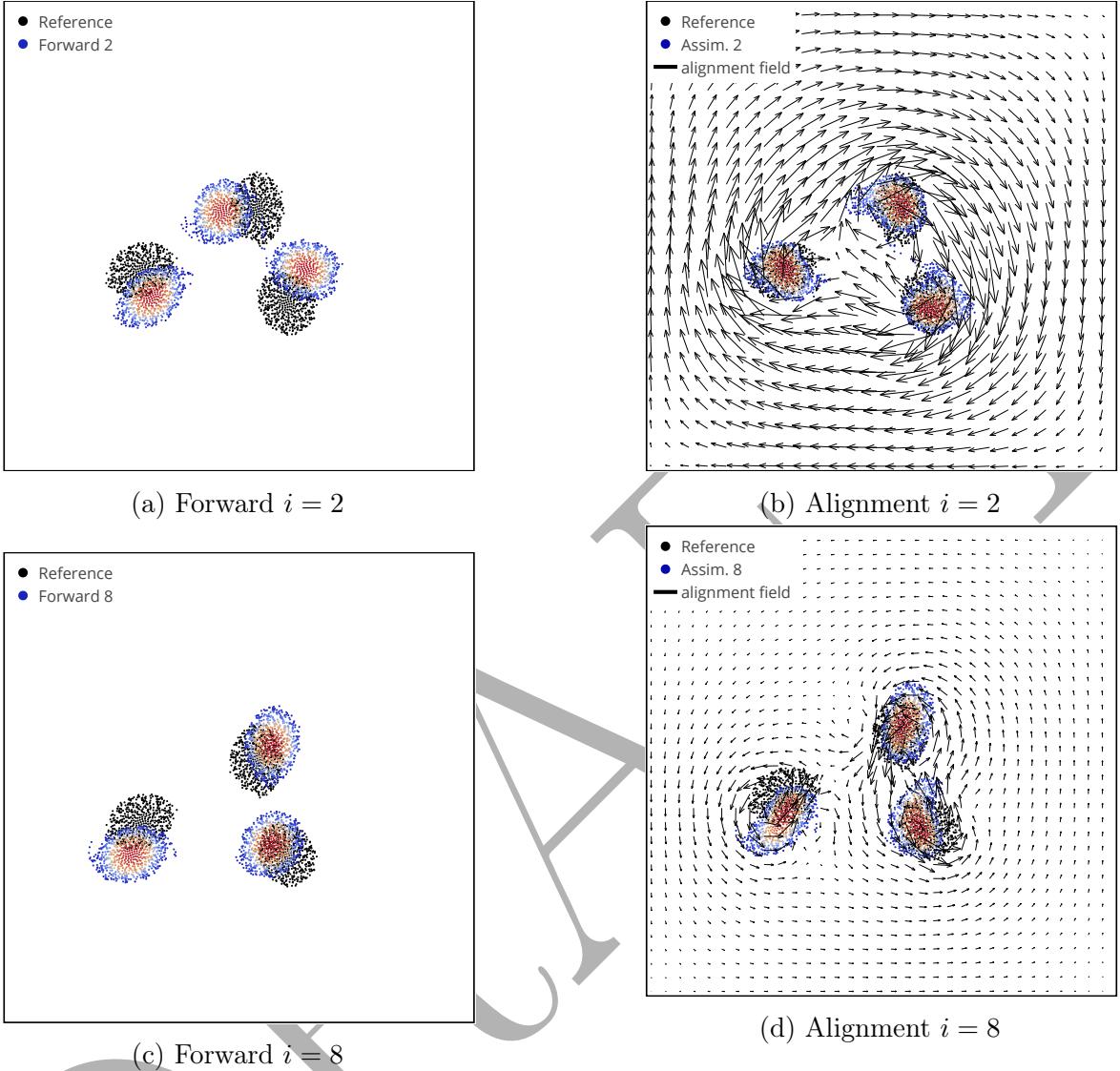


Figure 9: Particle alignment for two different members ($i = 2$ and $i = 8$). The left column shows the particle configurations before alignment (forward particles), while the right column displays the configurations after alignment, including the velocity fields \mathbf{u}_i used for the alignment. 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.

do not induce position 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 11a. This observation is corroborated by examining the distribution of coefficients relative to ensemble size in Figure 11b. It is observed that as the ensemble size increases,

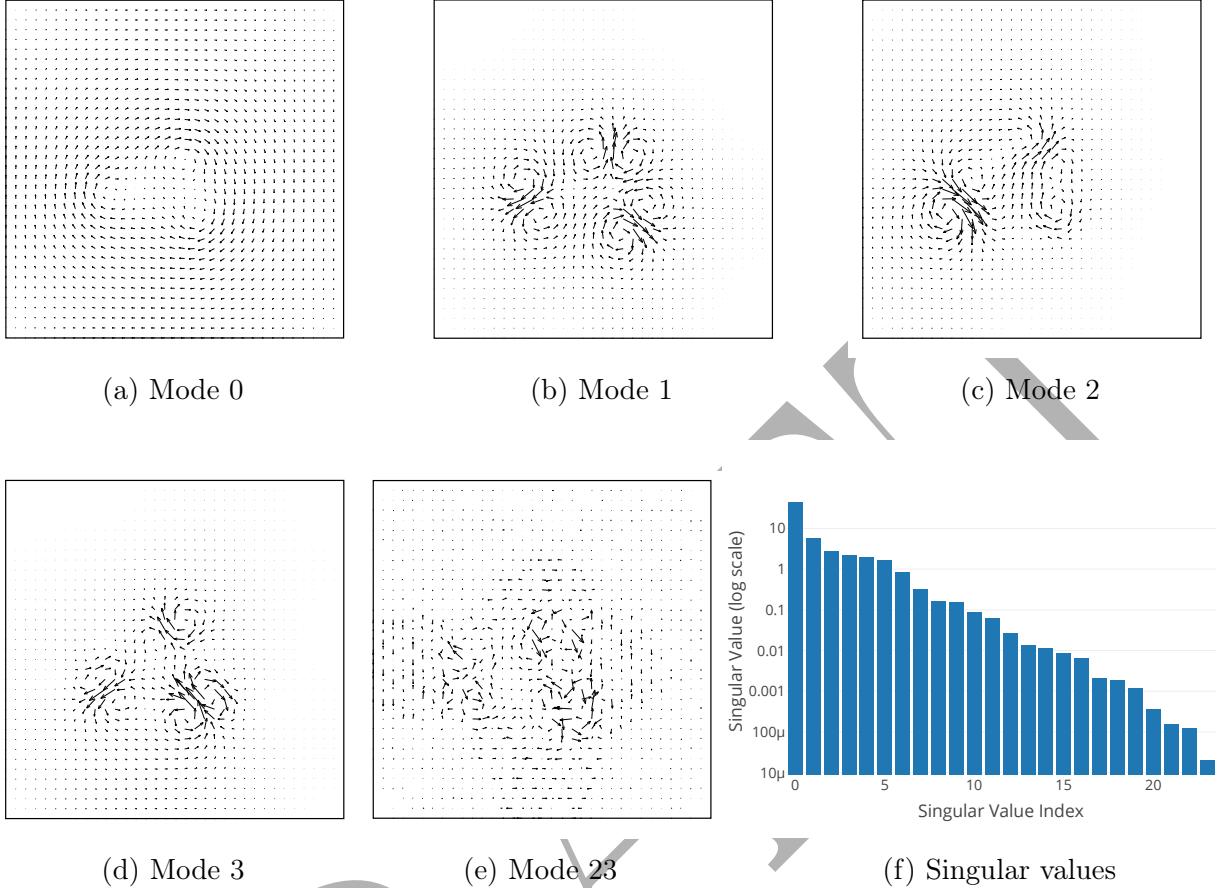
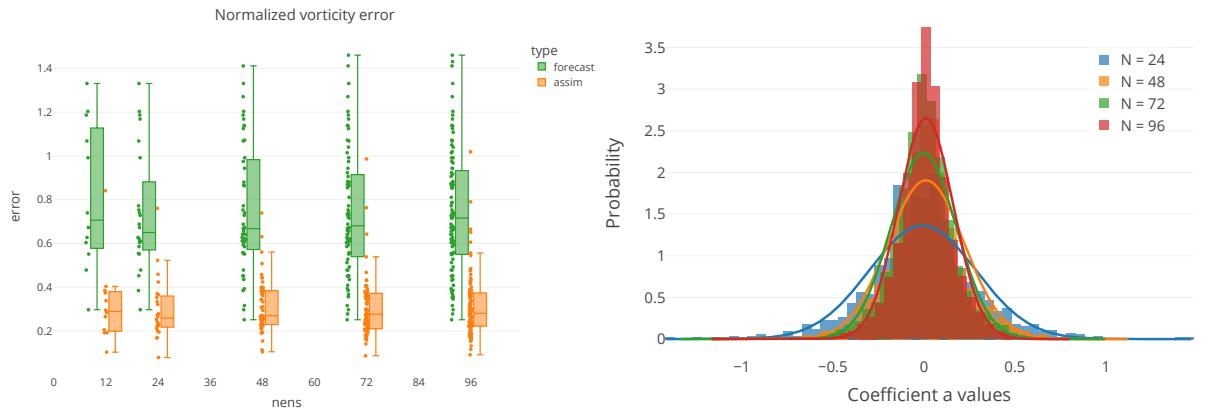


Figure 10: Singular Value Decomposition (SVD) of alignment fields. The first five Figures illustrate the principal modes of alignment, showing the predominant correction patterns, while the last Figure depicts the corresponding singular values. The modes reveal various types of corrections applied to the vortex distributions, including phase corrections and stretching.

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.

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 12. The alignment step thus represents a viable solution to address the limitations discussed in Section 4.3.



(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 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.

member=2

Assim.

Reference

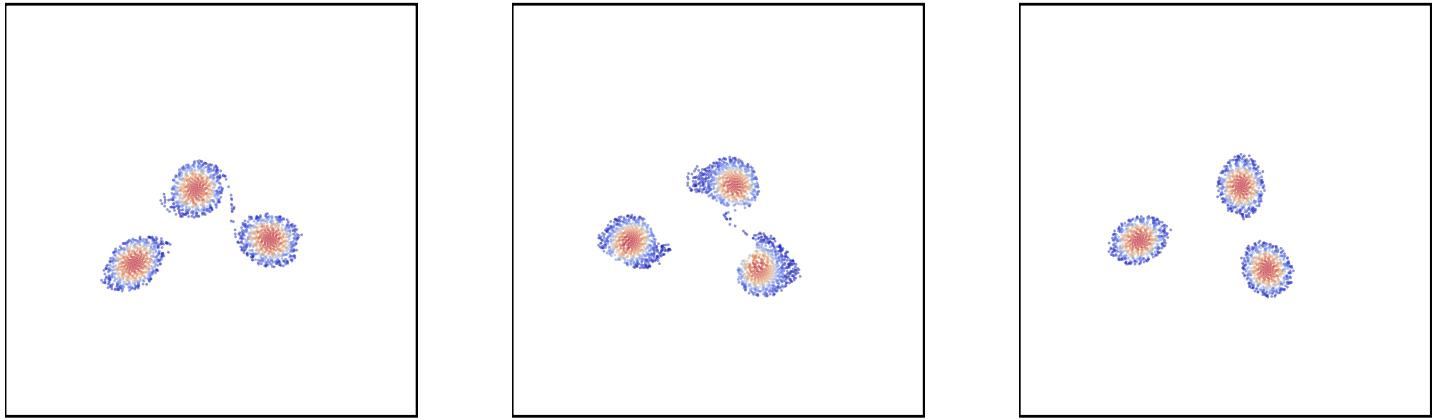


Figure 12: Update of member 2 using the Part-Align-EnKF during the initial assimilation. The particle configuration has been adjusted to effectively accommodate the correction.

4.5 SVD decomposition

The previous result motivates us to use less modes, as discussed in Section 3.4. To determine an appropriate threshold, we compare the RMSE after the first alignment, varying the

threshold values, as shown in Figure 13 for an ensemble of 72 members. Simultaneously, we track the number of selected modes r . By decreasing the threshold value, the RMSE reaches a minimum rapidly. Further increases in the number of modes produce only marginal changes in the RMSE. This is directly linked to the exponential decreasing rate of the singular values observed in the Figure 10f. As a result, we select a threshold of 0.0025, which minimizes the error with just 12 modes, yielding a compression ratio of 15% when compared to the total number of ensemble members.

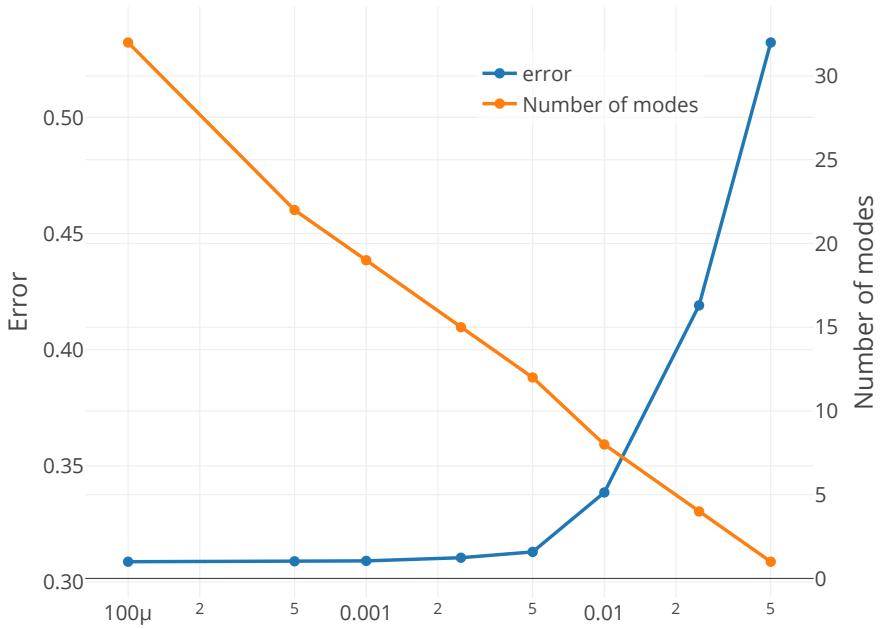
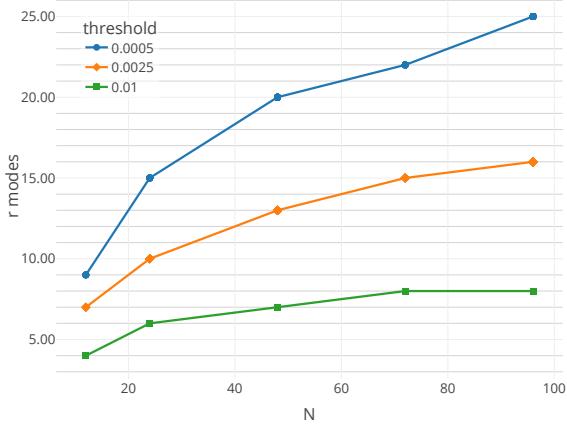


Figure 13: RMSE and selected number of modes for varying thresholds after the first alignment. The RMSE strongly decreases with a few modes.

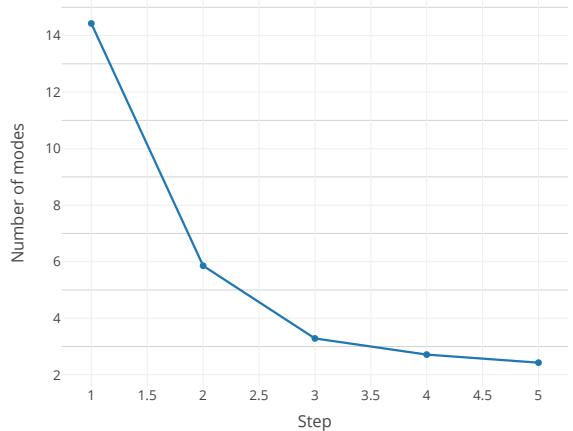
Additionally, as shown in Figure 14a, we varied the ensemble size to illustrate the number of modes required to capture a given energy ratio, as indicated by the threshold. On the one hand, we observe that the lower the threshold, the more modes are needed. However, more importantly, we notice that the number of modes for a given threshold eventually stops increasing. This is due to the properties of the SVD decomposition. This indicates that the rank of the velocity field space becomes independent of the ensemble size once a sufficiently large ensemble size is reached.

These results highlight the potential to reduce the number of terms in the decomposition without compromising the quality of the alignment process. This reduction leads to a significant decrease in computational complexity, from $O(N^2)$ to $O(NN_{\text{mode}})$ when solving the nonlinear optimization problem.

Furthermore, we assess the evolution of the number of modes r over time. Figure 14b reveals a consistent trend: as assimilation proceeds, the number of selected modes decreases for a



(a) Selected modes relative to ensemble size for different thresholds.



(b) The number of modes r over time for a given threshold. Fewer modes are needed as assimilation progresses.

given threshold. This progressive reduction in modes further decreases the computational complexity of the alignment step, making it comparable to traditional EnKF filtering.

4.6 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 strength filter (Remesh-Align-EnKF or Part-Align-EnKF).

The Align-Filter alone, as depicted in the first row of Figure 16, 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 16d. This residual error is further illustrated by the position error box plot over time in Figure 15a. Although the position error e_c decreases with the assimilation, it does not entirely vanish. Besides occasional distortions in vortex shape and loss of symmetry due to alignment, a primary reason for the residual error is the inability to reduce strength 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 strength correction filters is crucial, as it addresses the uncertainty induced by strength variations and improves overall accuracy.

We apply the Part-EnKF-filter in this context. As observed in Figure 15b the filter converges effectively and now produces results comparable to what we get in Figure 6a. In fact, the result is better than the Remesh-EnKF due to the alignment step. Compared to the Align-Filter, incorporating strength correction enhances the control over vortex spreading during the forecast phase. Consequently, the boundaries of the vortices are more precisely

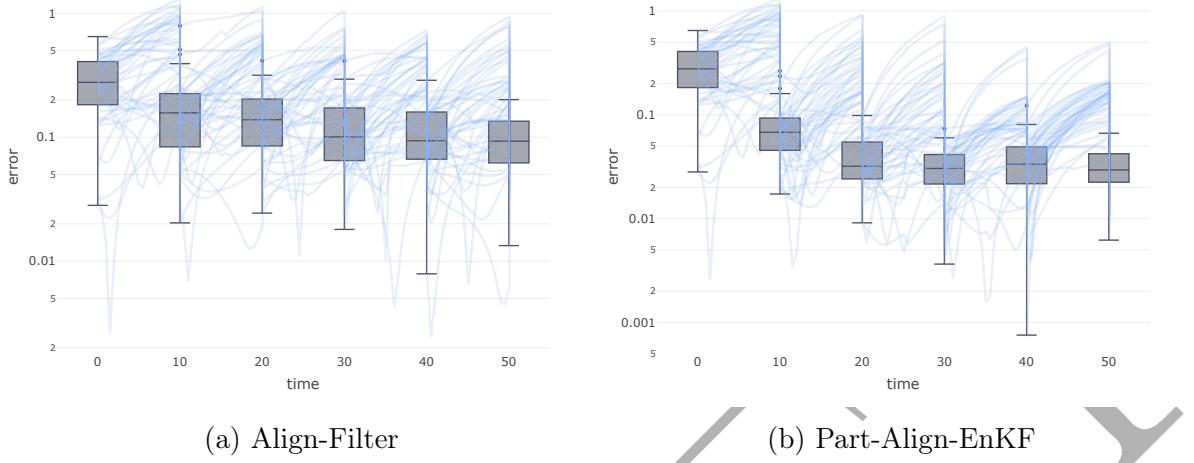


Figure 15: Position error e_c for different filtering methods over time. (a) Error for the Align-Filter. (b) Error for the Part-Align-EnKF. The plots show how the position error evolves with each assimilation step for both filtering approaches.

delineated, and the solution provides a more accurate representation of their geometry, as illustrated in the second row of Figure 16. 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.

4.7 Filters performances and parametric studies

Assimilation frequency f_{obs} The first effect we assess is the influence of the assimilation frequency, $f_{\text{obs}} = \frac{1}{\Delta t}$, on the performance of the various filters. As expected, a lower assimilation frequency leads to a greater ensemble spread. This is particularly noticeable for the Part-EnKF filter, which is similar to other filters but gradually diverges when the observation frequency falls below a certain threshold, such as $f_{\text{obs}} = 0.1$. Figure 17a shows that as f_{obs} increases, the final error in the vorticity field, e_ω , tends to decrease. Interestingly, the alignment-based filter demonstrates reduced sensitivity to observation frequency compared to the Remesh-EnKF filter, making it more robust in scenarios with infrequent observations.

Ensemble size N We also investigate the effect of ensemble size on filter convergence, aiming to balance computational cost with an accurate representation of system uncertainties.

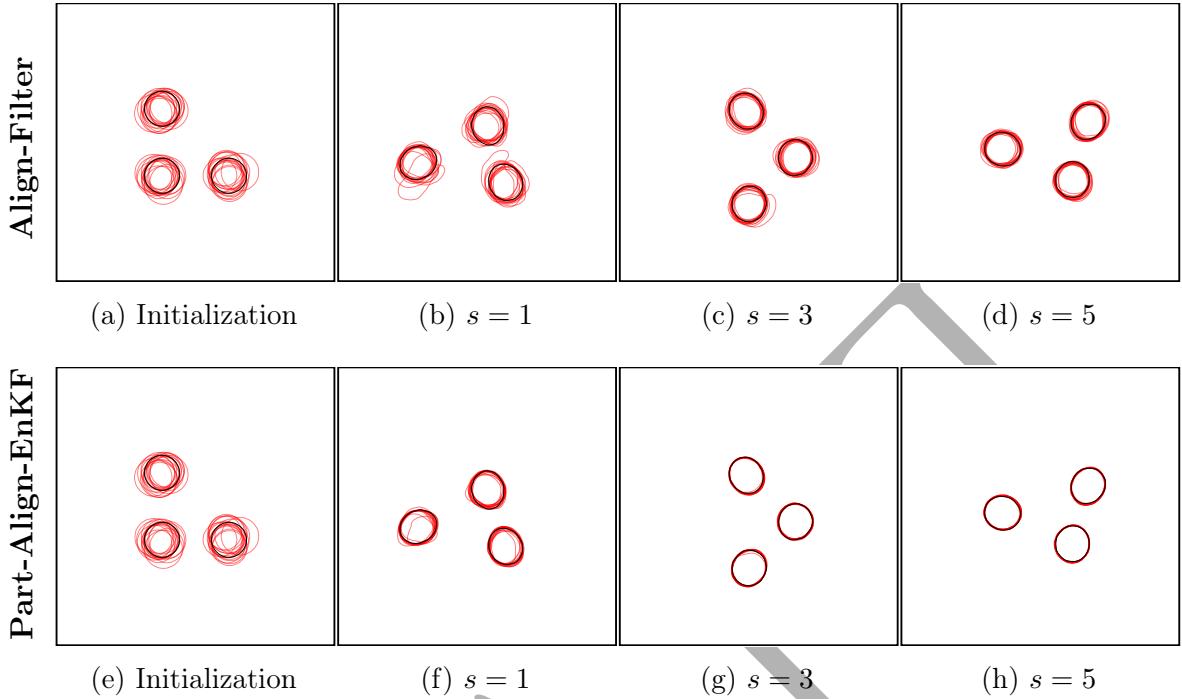
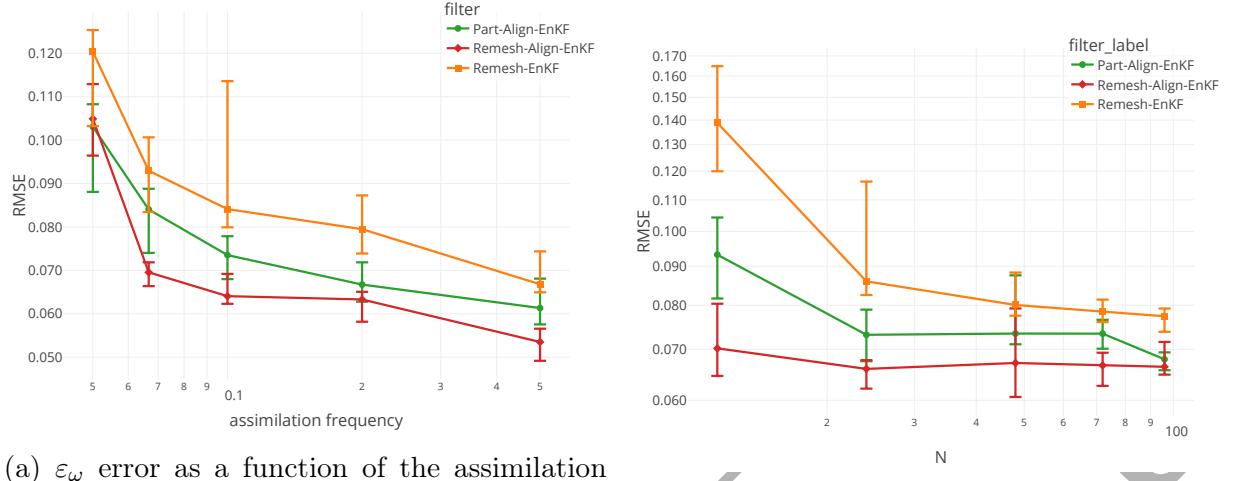


Figure 16: Comparison of vortex contour plots for different assimilation methods and steps. The first row corresponds to the Align-Filter, and the second row to the Part-Align-EnKF filter.

Figure 17b shows that fewer ensemble members are needed for convergence when using alignment-based filters. Specifically, while the Remesh-EnKF filter requires at least 48 members to achieve a stable RMSE of approximately 0.08, alignment-based filters converge effectively with as few as 12 members, reaching significantly lower RMSE values. This improvement stems from the alignment-based filters richer correction space, allowing them to achieve better accuracy with fewer ensemble members. Consequently, incorporating alignment steps allows for considerable computational savings, even though the alignment step involves solving a nonlinear optimization problem. The reduced dimensionality also mitigates the complexity through the use of SVD basis.

Number of observations N_{obs} and precision Prec. Finally, we assess the influence of the number of observations, N_{obs} , and the observation precision, defined as $\text{Prec.} = 1/\sigma_{\text{obs}}^2$. Figure 18b shows that while all filters exhibit similar convergence behavior, the alignment-based filters outperform the Remesh-EnKF filter, particularly under conditions of low precision and a smaller number of observations. These filters demonstrate significantly lower RMSE in these scenarios, highlighting their robustness and adaptability.

All these plots highlight the improvement introduced by incorporating alignment corrections, as seen in the comparison between the Remesh-EnKF and Remesh-Align-EnKF filters.

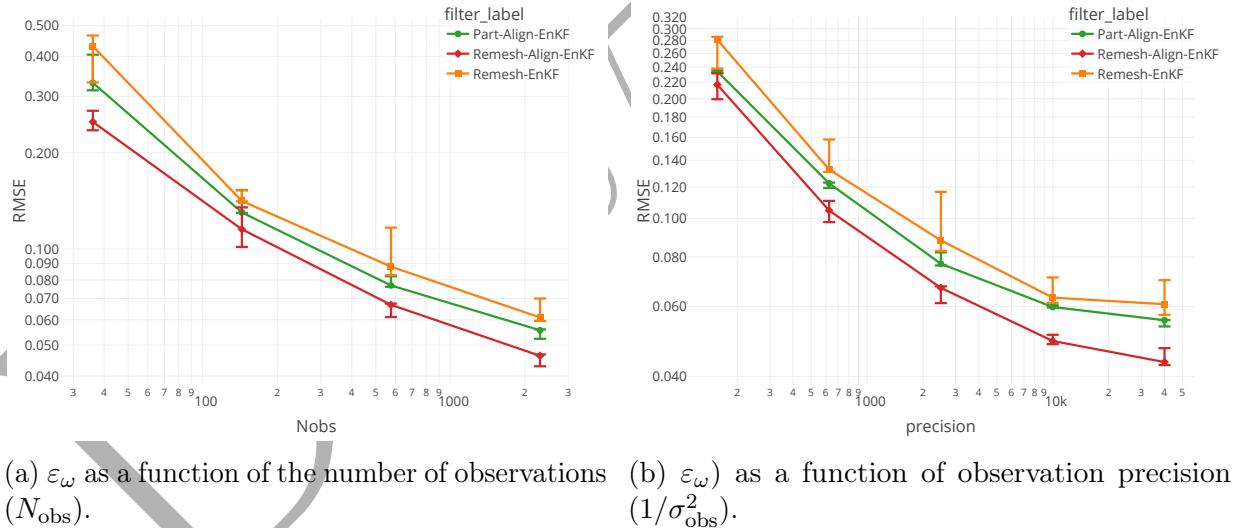


(a) ε_ω error as a function of the assimilation frequency (f_{obs}).

(b) ε_ω error as a function of ensemble size (N_{ens}).

Figure 17: RMSE (ε_ω) for varying assimilation frequencies and ensemble sizes. Alignment-based filters consistently demonstrate superior performance, achieving convergence with fewer ensemble members compared to the Remesh-EnKF filter.

The Part-EnKF performance lies in between. It offers the advantage of using a purely Lagrangian filter while still outperforming the Remesh-EnKF in this context.



(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 18: 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.

To conclude, in addition to allowing the convergence of Part-EnKF, the alignment step improves the accuracy of the filter state prediction. A direct comparison between the

Remesh-Align-EnKF and Remesh-EnKF filters reveals that the former consistently provides the best results across various studies. The Part-Align-EnKF filter generally performs in between, showing an improvement over traditional methods but not quite matching the Remesh-Align-EnKF. This difference is likely attributable to the fact that each ensemble member in Part-Align-EnKF employs its own discretization, introducing variability even when representing the same continuous solution. In contrast, Remesh filters use a unified particle discretization, resulting in less variability from discretization.



5 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.

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 2: Initial three-vortex true configuration with $\delta = 0.375$

Variables	Distributions
Vortex Center	$\mathcal{N}(x^i, 0.05^2), \mathcal{N}(y^i, 0.05^2)$ ($i = 1, 2, 3$)
Core Size (R)	$\mathcal{N}(0.2, 0.01^2)$
Amplitude (Γ)	$\mathcal{N}(4.0, 0.08^2)$

Table 3: Initial distributions for the vortex parameters. We sample the vortex center positions, core size, and also amplitude.

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

$$\frac{d}{dt}(\mathbf{x}_a(t) + \varepsilon \mathbf{x}_1(t), \dots) = \sum_{j=1}^N a_j \mathbf{u}_j(\mathbf{x}_a + \varepsilon \mathbf{x}_1 + \varepsilon^2 \mathbf{x}_2, \dots) + \varepsilon a_i \mathbf{u}_i(\mathbf{x}_a + \varepsilon \mathbf{x}_1 + \dots). \quad (7)$$

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 (7), we finally get couple problem where the particles are

$$\begin{cases} \frac{d\mathbf{x}_a}{dt} &= \sum_j a_j u_j(\mathbf{x}_a), \\ \mathbf{x}_a(t=0) &= \mathbf{x}_a^0, \\ i = 1, \dots, M : \\ \frac{d\hat{\mathbf{x}}_i}{dt} &= \hat{\mathbf{x}}_i \sum_j a_j \nabla_x u_j(\mathbf{x}_a) + a_i \mathbf{u}_i(\mathbf{x}_a), \\ \hat{\mathbf{x}}_i(t=0) &= 0. \end{cases}$$

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, $\frac{\partial \mathbf{h}}{\partial \mathbf{v}}$ 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

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}_i} \approx \frac{\mathcal{H}(\mathbf{x}_a + \varepsilon \mathbf{x}_i) - \mathcal{H}(\mathbf{x}_a)}{\varepsilon}.$$

References

- [1] R. Ambati, X. Pan, H. Yuan, and X. Zhang. Application of material point methods for cutting process simulations. *Computational Materials Science*, 57:102–110, 2012.
- [2] M. Amodei and J. Stein. Deterministic and fuzzy verification methods for a hierarchy of numerical models. *Meteorological Applications*, 16:191–203, 2009.
- [3] S. Andersen and L. Andersen. Modelling of landslides with the material-point method. *Computational Geosciences*, 14(1):137–147, 2010.
- [4] Hassan Aref. Motion of three vortices. *The Physics of Fluids*, 22(3):393–400, March 1979.
- [5] S. Bardenhagen, A. Brydon, and J. Guilkey. Insight into the physics of foam densification via numerical simulation. *Journal of the Mechanics and Physics of Solids*, 53(3):597–617, 2005.
- [6] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 1 2000.
- [7] G Blanchard and P Mathé. Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems*, 28(11):115011, November 2012.

- [8] M. Bocquet, P. J. Vanderbecken, A. Farchi, J. Dumont Le Brazidec, and Y. Roustan. Bridging classical data assimilation and optimal transport: the 3d-var case. *Nonlinear Processes in Geophysics*, 31(3):335–357, 2024.
- [9] Marc Bocquet. Introduction to the principles and methods of data assimilation in the geosciences. page 89, 2014.
- [10] A. Brydon, S. Bardenhagen, E. Miller, and G. Seidler. Simulation of the densification of real open-celled foam microstructures. *Journal of the Mechanics and Physics of Solids*, 53(12):2638–2660, 2005.
- [11] J-S Chen, C Pan, and C-T Wu. Large deformation analysis of rubber based on a reproducing kernel particle method. *Computational Mechanics*, 19(3):211–227, 1997.
- [12] Nan Chen, Quanling Deng, and Samuel N. Stechmann. Superfloe parameterization with physics constraints for uncertainty quantification of sea ice floes. *SIAM/ASA Journal on Uncertainty Quantification*, 10(4):1384–1409, 2022.
- [13] Yongsheng Chen and Chris Snyder. Assimilating vortex position with an ensemble kalman filter. *Monthly Weather Review*, 135(5):1828 – 1845, 2007.
- [14] Yongsheng Chen and Chris Snyder. Assimilating vortex position with an ensemble kalman filter. *Monthly Weather Review*, 135(5):1828 – 1845, 2007.
- [15] Georges-Henri Cottet and Petros D. Koumoutsakos. *Vortex Methods: Theory and Practice*. Cambridge University Press, 2000.
- [16] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 685–693, Bejing, China, 6 2014. PMLR.
- [17] Darwin Darakananda, Andre Fernando De Castro da Silva, Tim Colonius, and Jeff Eldredge. Data-assimilated low-order vortex modeling of separated flows. *Physical Review Fluids*, 11 2018.
- [18] Florin Diacu and Philip Holmes. *Celestial Encounters: The Origin of Chaos*. Princeton University Press, 1996.
- [19] Olivier Dubrule. Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15(6):687–699, 1983.
- [20] S. Dunatunga and K. Kamrin. Continuum modeling and simulation of granular flows through their many phases. *Journal of Fluids Mechanics*, 2015.
- [21] Marius Duvillard, Loïc Giraldi, and Olivier Le Maître. Ensemble Data Assimilation for Particle-based Methods. 2024. Draft version.

- [22] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- [23] Geir Evensen, Femke C. Vossepoel, and Peter Jan van Leeuwen. Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem. In *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*, Springer Textbooks in Earth Sciences, Geography and Environment, chapter 8. Springer International Publishing, Cham, 2022.
- [24] Nelson Feyeux. *Transport optimal pour l'assimilation de données images*. Theses, Université Grenoble Alpes, December 2016.
- [25] J. E. Guilkey, J. B. Hoying, and J. A. Weiss. Computational modeling of multicellular constructs with the material point method. *Journal of Biomechanics*, 39(11):2074–2086, 2006.
- [26] Per Christian Hansen. The l-curve and its use in the numerical treatment of inverse problems. In *InviteComputational Inverse Problems in Electrocadiology*. WIT Press, 2000. InviteComputational Inverse Problems in Electrocadiology ; Conference date: 01-01-2000.
- [27] I. Hoteit and A. Köhl. Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. *Journal of Oceanography*, 62:539–550, 2006.
- [28] B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D*, 230:112–126, 2007.
- [29] T. Janjić, N. Bormann, M. Bocquet, J. A. Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144:1257–1278, 2018.
- [30] Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The material point method for simulating continuum materials. In *ACM SIGGRAPH 2016 Courses*, pages 1–52, Anaheim California, July 2016. ACM.
- [31] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, March 1960.
- [32] Peter K. Kitanidis. Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419, 1995.
- [33] Ottmar Klaas and Mark S Shephard. Automatic generation of octree-based three-dimensional discretizations for partition of unity methods. *Computational Mechanics*, 25(2):296–304, 2000.

- [34] Krishna Kumar, Kenichi Soga, Jean-Yves Delenne, and Farhang Radjai. Modelling transient dynamics of granular slopes: Mpm and dem. *Procedia Engineering*, 175:94–101, 2017. Proceedings of the 1st International Conference on the Material Point Method (MPM 2017).
- [35] C. Liu, Q. Xiao, and B. Wang. An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Monthly Weather Review*, 136(9):3363–3373, 2008.
- [36] S. Ma, X. Zhang, and X. M. Qiu. Comparison study of mpm and sph in modeling hypervelocity impact problems. *International Journal of Impact Engineering*, 36(2):272–282, 2009.
- [37] Yann Michel. Displacing potential vorticity structures by the assimilation of pseudo-observations. *Monthly Weather Review*, 139(2):549 – 565, 2011.
- [38] J. Monaghan. Simulating free surface flows with sph. *Journal of Computational Physics*, 110:399–406, 1994.
- [39] Dean S Oliver, Nanqun He, and Albert C Reynolds. Conditioning permeability fields to pressure data. In *ECMOR V-5th European conference on the mathematics of oil recovery*, pages cp–101. European Association of Geoscientists & Engineers, 1996.
- [40] James R Percival. *Displacement assimilation for ocean models*. PhD thesis, Citeseer, 2008.
- [41] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11 (5-6):355–602, 2019.
- [42] M. Plu. A variational formulation for translation and assimilation of coherent structures. *Nonlinear Processes in Geophysics*, 20(5):793–801, October 2013. Publisher: Copernicus GmbH.
- [43] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890.
- [44] B. Rao and S. Rahman. An enriched meshless method for non-linear fracture mechanics. *International Journal for Numerical Methods in Engineering*, 59, 2004.
- [45] Sai Ravela. Spatial inference for coherent geophysical fluids by appearance and geometry. In *IEEE Winter Conference on Applications of Computer Vision*, pages 925–932, Steamboat Springs, CO, USA, March 2014. IEEE.
- [46] Sai Ravela, Kerry Emanuel, and Dennis McLaughlin. Data assimilation by field alignment. *Physica D: Nonlinear Phenomena*, 230(1-2):127–145, June 2007.

- [47] W. Steven Rosenthal, Shankar Venkataramani, Arthur J. Mariano, and Juan M. Restrepo. Displacement data assimilation. *Journal of Computational Physics*, 330:594–614, February 2017.
- [48] M. A. Russell, A. Souto-Iglesias, and T. Zohdi. Numerical simulation of laser fusion additive manufacturing processes using the sph method. *Computer Methods in Applied Mechanics and Engineering*, 2018.
- [49] S. Li and W. K. Liu. *Meshfree Particle Methods*. Springer, Berlin, Heidelberg, 2004.
- [50] Umut Simsekli, Antoine Liutkus, Szymon Majewski, and Alain Durmus. Sliced-wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *International Conference on Machine Learning*, 2018.
- [51] W. Sołowski and S. Sloan. Evaluation of material point method for use in geotechnics. *International Journal for Numerical and Analytical Methods in Geomechanics*, 39:685 – 701, 2015.
- [52] A. Stomakhin, Craig A. Schroeder, Chenfanfu Jiang, Lawrence Chai, J. Teran, and Andrew Selle. Augmented mpm for phase-change and varied materials. *ACM Transactions on Graphics (TOG)*, 33:1 – 11, 2014.
- [53] O. Talagrand. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75(1B):191–209, 1997.
- [54] J. H. G. M. Van Geffen, V. V. Meleshko, and G. J. F. Van Heijst. Motion of a two-dimensional monopolar vortex in a bounded rectangular domain. *Physics of Fluids*, 8(10):2393–2399, 1996.
- [55] Cedric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [56] Habin Yim, Sun-Chul Kim, and Sung-Ik Sohn. Motion of three geostrophic Bessel vortices. *Physica D: Nonlinear Phenomena*, 441:133509, December 2022.
- [57] Yue Ying. A Multiscale Alignment Method for Ensemble Filtering with Displacement Errors. *Monthly Weather Review*, 147(12):4553–4565, December 2019. Publisher: American Meteorological Society Section: Monthly Weather Review.
- [58] Yue Ying, Jeffrey L. Anderson, and Laurent Bertino. Improving Vortex Position Accuracy with a New Multiscale Alignment Ensemble Filter. *Monthly Weather Review*, 151(6):1387–1405, June 2023.
- [59] Aman Zhang, Furen Ming, and Xueyan Cao. Total lagrangian particle method for the large-deformation analyses of solids and curved shells. *Acta Mechanica*, 225(1):253–275, 2014.