Ensemble Filter on dynamically driven RKHS: Application to a Multilayer Quasi-Geostrophic ocean model

Maël Jaouen, Benjamin Dufée, Etienne Mémin, Gilles Tissot

Ensemble Filter on dynamically driven RKHS: Application to a Multilayer Quasi-Geostrophic ocean model

Maël Jaouen[1], Benjamin Dufée[2,3], Etienne Mémin[1], Gilles Tissot[1]

[1]Univ Rennes, CNRS, Inria, IRMAR - UMR CNRS 6625, F-35000 Rennes, France
[2]Univ Bretagne Sud, IUT Vannes, Department of Data Science, 8 rue Michel de Montaigne, 56000, Vannes, France
[3]Laboratory of Mathematics in Bretagne Atlantique, Rue de Saint-Maudé, 56100 Lorient, France

Key Points:

- An ensemble filter defined on a time evolving RKHS family driven by the system dynamics is presented.
- The corresponding filter enables extending classical Bayesian ensemble filters to efficiently solve a 4D Var assimilation problem.
- Assessed on a multilayer quasi-geostrophic simulation for the North Atlantic Ocean basin in an OSSE framework with SSH observations.

Corresponding author: Maël Jaouen, `mael.jaouen@inria.fr`

Abstract

Estimating the ocean's dynamic state, particularly at the mesoscale, is challenging due to the non-linear nature of ocean processes. Traditional data assimilation methods like optimal interpolation (OI) and four-dimensional variational assimilation (4D-Var) improve ocean state estimates but face trade-offs: OI is computationally efficient but struggles with small-scale features, while 4D-Var is more accurate but computationally demanding. Ensemble methods offer a compromise, with ensemble Kalman filters assuming Gaussianity and particle filters limited by high-dimensional systems.

This article introduces a novel ensemble filter using a Reproducing Kernel Hilbert Spaces (RKHS) framework, embedding observables in time-evolving RKHSs to relax Gaussianity assumptions and provide computational efficiency comparable to 3D-Var while addressing 4D-Var problems. A key innovation is a tiled assimilation approach that enhances the ensemble's physical properties over time. Applied to a quasi-geostrophic model of the North Atlantic with synthetic sea surface height (SSH) data, this RKHS-based filter accurately assimilates observations with non-Gaussian noise, improving both the accuracy and efficiency of ocean state estimations in complex systems.

Plain Language Summary

Estimating the ocean's dynamic state, such as eddies and currents, is challenging due to the complexity of ocean processes. This paper introduces a new data assimilation method, grounded in a rigorous theoretical framework, that uses an ensemble of simulations to improve these estimates. It builds new trajectories by combining ensemble members in a time-invariant way, enabling efficient assimilation of observations across space and time.

By updating the ensemble at the start of each observation period and simulating forward, the method enhances physical consistency. Tests with a North Atlantic model and synthetic satellite data demonstrate its ability to accurately estimate the ocean's state, even with noisy, non-Gaussian observations, while improving both accuracy and efficiency compared to traditional methods.

1 Introduction

Data assimilation methods in ocean modeling aim to integrate sparse observations, both in time and space, of the ocean surface into numerical circulation models to produce accurate estimates of the ocean's state. In this context, a major challenge is the assimilation of sea surface height (SSH) observations, collected over time from orbiting satellite swaths, to accurately represent ocean mesoscale dynamics. The mesoscale, characterized by features such as eddies, fronts, and jets, is crucial for understanding and modeling global ocean circulation. To address this complex problem, several methods proposed by the data assimilation community (Carrassi et al., 2018; Evensen, 2009; Reich & Cotter, 2015) have been adapted and are now routinely used for various applications.

Optimal interpolation (OI) is a geostatistical method used to reconstruct a dense field of relevant oceanic variables from spatially sparse satellite observations by considering predetermined statistical properties (Le Traon et al., 2003). This method integrates observations collected at various times, assuming spatiotemporal correlation function between the observations and the estimated field. It is in particular employed to construct the Data Unification and Altimeter Combination System (DUACS) products (Taburet et al., 2019). OI's reliance on statistical interpolation to merge observations from varying times often results in an over-smooth degraded representation of small-scale structures. This limitation is particularly evident in the reconstruction of mesoscale features, where finer details are crucial for accurate modeling. Ubelmann et al. (2015) observed that optimal interpolation often misrepresents the transport over time of small-scale eddies. To address this issue, they

introduced dynamic interpolation, which offers a more accurate reconstruction of small-scale structure transport between two observations. Assuming an invertible dynamical system, the method involves running the model forward and backward between two observation times, and then reconstructing the sea surface height as the mean of the results from both simulations. Although this method better represents transport in the reconstructions, it still requires two space interpolations to obtain the complete fields needed to initialize the forward and backward simulations. Consequently, the effective resolution remains relatively low.

Variational methods aim to minimize a cost function that penalizes the distance between the observations and the estimated state. Among these, four-dimensional variational assimilation (4D-Var), assimilates spatially and temporally distributed observations. This method uses optimizations tools, such as adjoint methods (Le Dimet & Talagrand, 1986), which efficiently compute the gradient of the cost function with respect to its parameters by running the dynamics forward and then the tangent linear adjoint system backward. Despite the effectiveness of these methods, the associated computational costs can be prohibitively high, hardly parallelizable, and constructing a numerically exact tangent linear operator can be a challenging task. To address this, neural networks such as 4DVarNet (Fablet et al., 2021) can be use to efficiently solve it after a learning phase. Recently, 4DVarNet has been applied to the reconstruction of the sea surface currents from sea surface height (SSH) and sea surface temperature (SST) observations (Fablet et al., 2024), demonstrating reconstructions with a higher effective resolution than DUACS. Machine learning methods offer powerful tools for efficient data assimilation (Cheng et al., 2023). These approaches can be integrated with classical assimilation methods to improve future state predictions (Brajard et al., 2020; Farchi et al., 2021). It can also be used to interpolate partial observations to retrieve the ocean mesoscale dynamic state (Ouala et al., 2018; Manucharyan et al., 2021). In these methods, neural networks are employed to model the system's dynamics, thereby bypassing the need to solve the dynamical system equations directly. However, this class of approach opens the question how to assess, without the help of a dynamical model, the physical quality of the generated predictions. In particular, neural networks introduce generalization challenges, which may lead to poor performance when the model encounters new or previously unseen physical phenomena.

Another strategy, called *nudging*, introduces a correction term into the dynamics equations to fit better with the observations (Hoke & Anthes, 1976). Back-and-forth nudging (BFN) (Auroux & Blum, 2008), involves iteratively solving the forced dynamical equation forward and backward in time until the algorithm converges toward a trajectory. This method circumvents the need for minimization tools used in 4D-Var methods, but it requires uncontrolled modifications of the dynamical system, and hence of the physical prior we have on the system's dynamics. Le Guillou, Metref, et al. (2021) apply BFN to data from the Surface Water Ocean Topography (SWOT) mission to assimilate a quasi-geostrophic model, demonstrating the potential of SWOT data to enhance BFN performance. It can be remarked, that BFN incurs computational costs due to the necessity of running the model multiple times until convergence. This makes it impractical as it is for application to complex models. To address this limitation, Le Guillou, Lahaye, et al. (2021) proposed a hybrid approach that combines BFN with 4D-Var. In this approach, BFN is used to assimilate balanced motions in a quasi-geostrophic model, while 4D-Var is employed to assimilate internal tides in a shallow water model.

To manage computational complexity in high-dimensional ocean models, one effective approach is the use of ensemble Kalman filter (EnKF) methods, which operate locally in the phase space (Evensen, 2009; Carrassi et al., 2018), in the sense that the solution is assumed to be well characterized by the span of an ensemble of realizations of the dynamical model relatively close from each other. The use of Kalman filters in nonlinear systems is justified by their ability to provide the best linear unbiased estimator within a Gaussian framework. EnKFs assimilate observations sequentially as they become available, assuming

both the state estimation error and observation error to be Gaussian. They approximate the state estimate and its uncertainty using an ensemble of realizations, where the analyzed ensemble members are constructed as linear combinations of the forecast ensemble. However, this superposition principle lacks theoretical justification for highly nonlinear systems and it has been shown that EnKF do not converge toward the true filtering distribution (Le Gland et al., 2011). For a general review on the asymptotic results of the EnKF in a Gaussian context, interested readers may refer to (A. N. Bishop & Del Moral, 2023). Despite this limitation, EnKFs have been successfully applied in various oceanographic studies, demonstrating their efficiency in estimating the ocean state by assimilating SST or SSH data (Leeuwenburgh et al., 2005; Wan et al., 2010). A significant limitation of classical EnKFs is their sequential nature, which leads to correct the model at the final observation time of a given time window (and not at the window's initial time). To address this, smoothing methods based on forward-backward assimilation of trajectories from data collected over the whole time window have been developed. Among these, ensemble extension of the 4DVar techniques (4DEnVar) have been proposed (Lorenc, 2003; Liu et al., 2008; Desroziers et al., 2014). Such techniques replace the exact numerical tangent linear operator by an ensemble approximation, leading to much cheaper and practical methods (Yang et al., 2015). Other methods use time-constant weights to reconstruct trajectories, and formulate a cost function of this weight vector (Hunt et al., 2004; Bocquet & Sakov, 2014). Connections between these methods and 4DEnVar can be established. However, the assumption of constant weights over time lacks clear theoretical justification in a classical setting.

Alternatively, particle filters (Doucet et al., 2001; Gordon et al., 1993; Van Leeuwen et al., 2019) relax the Gaussianity and linearity assumptions of Kalman filters by representing the state distribution with an ensemble of particles, each associated with a weight. Unlike EnKFs, particle filters do not alter the ensemble's physical properties by creating members as linear combinations of existing members. However, particle filters are subject to the curse of dimensionality, meaning that in high-dimensional state spaces, a huge number of particles is required to accurately represent the state distribution. This makes particle filters computationally expensive and challenging to apply to high-dimensional systems.

Reproducing Kernel Hilbert Spaces (RKHS) have been employed in data assimilation methods to address system non-linearity and to relax the inherent assumptions of Kalman filters. Mauran et al. (2023) propose a reformulation of the Ensemble Transform Kalman Filter (ETKF) (C. H. Bishop et al., 2001; Hunt et al., 2007), operating within a RKHS framework that uses feature maps (i.e. the mapping from the phase space onto the RKHS) defined from the ensemble state space and observations. Gottwald and Reich (2021) introduced an ensemble Kalman filter designed to simultaneously retrieve the state variable and learn the dynamical propagator, using random feature map bases. Both approaches aim to determine weight vectors whose size scales with the state space, making them computationally expensive for high-dimensional systems. This issue of increasing computational cost with system dimensionality is also observed in other kernel-based assimilation methods (Luo, 2019; Zhang et al., 2022).

RKHS-based assimilation methods share similarities with the decomposition method introduced by Ubelmann et al. (2021), where the state variable of the dynamical system is decomposed into a reduced space, and optimal interpolation is performed directly on the reduced state vector to define the analyzed state. In RKHS approaches, the reduced state vector corresponds to the weight vector in the RKHS, and the ensemble filter operates directly in this space.

The method proposed in this study relies on the theoretical framework developed in (Dufée et al., 2024), which differs from usual RKHS based estimation. In this setting the dynamical system is embedded within a family of time-varying reproducing kernel Hilbert spaces (RKHS). This framework enables establishing an evolution law for the RKHS bases across time. The RKHS are transported along time by the dynamical system, defining an isometry between the kernel evaluations over time. This approach enables the use of embar-

rassingly simple data assimilation schemes, reconstructing the trajectory as a time-invariant linear combination of the ensemble members. These schemes rely on a fully justified superposition principle allowing to compare the trajectories at initial time, leading consequently to effectively solve a 4D-Var problem with a time series of observations at the computational cost of a three-dimensional variational (3D-Var) methods, in which time evolution is not considered. Compared to the other RKHS schemes proposed in the literature our method works on a small dimensional vector of function evaluations corresponding to the ensemble dimension. The small dimension of these evaluation vectors allows for direct matrix inversion in the weight space making our approach applicable to high-dimensional systems. Similar to particle filters, each ensemble member in the RKHS are assigned with (a signed) weight, representing the state distribution.

This method has been applied to a multilayer quasi-geostrophic system in the North Atlantic basin, using synthetic observations generated from real-world satellite tracks, following an observing system simulation experiment (OSSE) methodology. We show that, within this framework, the proposed method leads to improved reconstruction resolution.

The paper is structured as follows: section 2 provides a brief overview of the RKHSs, the Koopman operator and how with the combination of these two tools a RKHS family transported by the dynamics is defined. Section 3 focuses on assimilation methods. Section 4 introduces the experimental framework. Finally, section 5 describes the results obtained.

2 Theoretical framework

This section introduces the RKHS theory, the Koopman operator, and the framework established by Dufée et al. (2024), which defines a family of RKHSs, built from the transport by the dynamical system of an initial kernel at the time instant t_0.

2.1 RKHS

A Reproducing Kernel Hilbert Space (RKHS), $\mathcal{H}$, is a Hilbert space of functions $f : E \to \mathbb{C}$, where E is a non-empty set, equipped with a inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. A map $k : E, E \to \mathbb{C}$ is defined as being a reproducing kernel of $\mathcal{H}$ if it has the two following properties:

- membership property: $\forall x \in E, k(\cdot, x) \in \mathcal{H}$,
- reproducing property: $\forall x \in E, \forall f \in \mathcal{H},\ \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$.

The reproducing property implies that the kernel satisfies $k(x, y) = \overline{k(y, x)}$, where $\bar{\bullet}$ denotes the complex conjugate. The RKHS is fully and uniquely determined by its kernel k (Aronszajn, 1950). The functions $k(\cdot, x)$ for $x \in E$ are usually referred to as feature maps. It can be shown that their span is dense in $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$. The integral operator associated with the kernel k, is defined as $\mathcal{L}_k : L^2(E, \nu) \to L^2(E, \nu)$, such that for any $f \in L^2(E, \nu)$,

$$(\mathcal{L}_k f)(x) = \int_E k(x, y) f(y) \nu(dy) \ , \tag{1}$$

where L^2 denotes the space of square-integrable functions, and ν is a measure on E. The inner product in $L^2(E, \nu)$ for f and g in $L^2(E, \nu)$ is given by

$$\langle f, g \rangle_{L^2(E,\nu)} = \int_E f(y) \overline{g(y)} \nu(dy) \ . \tag{2}$$

Consequently, the inner product between a function $f \in L^2(E, \nu)$ and a kernel function $k(\cdot, x)$ becomes:

$$\langle f, k(\cdot, x) \rangle_{L^2(E,\nu)} = \int_E f(y) \overline{k(y, x)} \nu(dy) = \int_E k(x, y) f(y) \nu(dy) = (\mathcal{L}_k f)(x) \ . \tag{3}$$

The range of $\mathcal{L}_k$ is assumed to be dense in $L^2(E,\nu)$, implying that its eigenvalues are strictly positive. By Mercer's theorem as k is a bounded, continuous, symmetric, positive semi-definite kernel, $\mathcal{L}_k$ eigenfunctions $\{\varphi_i\}_{i\in\mathbb{N}}$ and eigenvalues $\{\mu_i\}_{i\in\mathbb{N}}$ define the RKHS $\mathcal{H}$ as

$$\mathcal{H} = \left\{ f \in L^2(E,\nu), f = \sum_{i=1}^{\infty} a_i \varphi_i \ s.t. \ \sum_i \frac{|a_i|^2}{\mu_i} < \infty \right\} . \tag{4}$$

This last expression highlights the fact that $\mathcal{H}$ is a space of functions with fast decaying coefficients within $L^2(E,\nu)$.

2.2 Dynamical system

Let consider a dynamical system defined by a continuous flow Φ_t, operating on functions $\boldsymbol{X} \in L^2\left(\mathbb{R}^+ \times \Omega_x, \mathbb{R}^d\right)$ included in an invariant phase space manifold, with Ω_x representing the spatial support. The state of the system at time $t \in \mathbb{R}^+$, denoted by $\boldsymbol{X}(t) = \boldsymbol{X}_t$ is given by $\boldsymbol{X}_t = \Phi_t(\boldsymbol{X}_0)$, where $\boldsymbol{X}_0$ is the initial condition. In differential form the flow is given by the system:

$$\begin{cases} \dfrac{\partial \boldsymbol{X}_t}{\partial t} = \mathcal{M}(\boldsymbol{X}_t) \\ \boldsymbol{X}(0) = \boldsymbol{X}_0 \end{cases} , \tag{5}$$

where $\boldsymbol{X}_t$ belongs to Ω a compact (invariant) subset of the space of square-integrable functions $L^2\left(\Omega_x, \mathbb{R}^d\right)$. The nonlinear operator $\mathcal{M}$ is assumed to be differentiable with continuous derivatives. An observable of this system is defined as a square-integrable function that takes as input a function from Ω and outputs a real valued vector. In the following, we assume that the dynamical system is invertible, and that there exists a measure ν preserved by this dynamical system.

Remark 1 *Let us point out that in our data assimilation setting, the invertibility condition will be required only on the assimilation time window. Existence of invariant measures for continuous system is ubiquitous.*

The Koopman operator U_t is an infinite-dimensional linear operator mapping $L^2(\Omega,\nu)$ into itself. It propagates the observables over time, such that for every observable f and every $\boldsymbol{X}$ in Ω, we have $U_t f(\boldsymbol{X}) = f(\Phi_t(\boldsymbol{X}))$.

2.3 Dynamical system embedded in a RKHS

With all the previous ingredients, we are now able to present the framework proposed by Dufée et al. (2024), on which our assimilation schemes are built.

Let Ω_0 denote the ensemble of initial conditions, and define $\Omega_t := \Phi_t(\Omega_0)$ as the ensemble of states (or snapshots) at time $t \in \mathbb{R}^+$. Since the system is invertible, for any snapshot $\boldsymbol{X}_t \in \Omega_t$, there exists a unique initial condition $\boldsymbol{X}_0 \in \Omega_0$ associated with it. Moreover in the present framework, we consider that for each state $\boldsymbol{X}_t$, the current time t is known, in such a way it can be uniquely associated with its initial condition $\boldsymbol{X}_0$. Given an initial kernel $k_0 : \Omega_0 \times \Omega_0$ it is then possible to define a positive Hermitian kernel $k_t : \Omega_t \times \Omega_t$ such that,

$$\forall \boldsymbol{X}_t, \boldsymbol{Y}_t \in \Omega_t, \ k_t(\boldsymbol{X}_t, \boldsymbol{Y}_t) = k_0\left(\Phi_t^{-1}(\boldsymbol{X}_t), \Phi_t^{-1}(\boldsymbol{Y}_t)\right) . \tag{6}$$

Let $(\mathcal{H}_0, \langle\cdot,\cdot\rangle_{\mathcal{H}_0})$ and $(\mathcal{H}_t, \langle\cdot,\cdot\rangle_{\mathcal{H}_t})$ represent the RKHSs associated with k_0 and k_t, respectively. A RKHS family $\mathcal{W} = (\mathcal{H}_t)_{t\geq 0}$ can be constructed by considering the RKHSs along time.

The operator $\mathcal{U}_t : \mathcal{H}_0 \to \mathcal{H}_t$ is defined as:

$$\forall \boldsymbol{X}_0 \in \Omega_0, \ \mathcal{U}_t k_0(\cdot, \boldsymbol{X}_0) = k_t(\cdot, \Phi_t(\boldsymbol{X}_0)) . \tag{7}$$

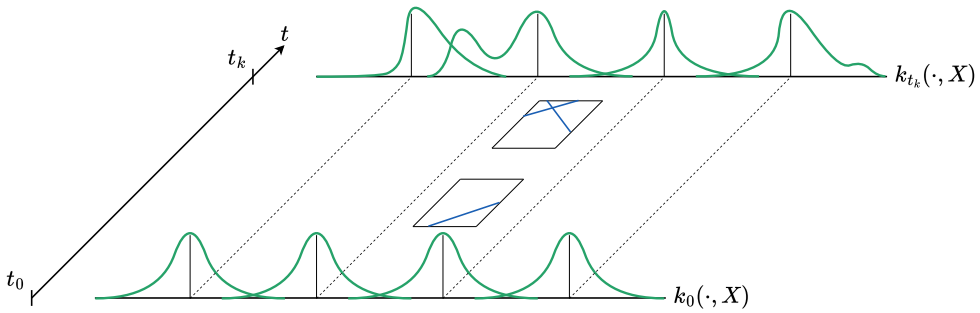

Figure 1: Feature maps transported over time from t_0 to t_k by the dynamical system, with the observations collected between t_0 and t_k represented at the center of the figure in the form of satellite swath.

This operator propagates the features maps over time. It can be seen as a useful tool to define the kernel k_t at any time instant. Operator $\mathcal{U}_t$ is closely related to the adjoint of the Koopman operator U_t, know as the Perron Frobenius operator. It follows from equation (6) that $\mathcal{U}_t$ defines an isometry between $\mathcal{H}_0$ and $\mathcal{H}_t$.

The theory presented here operates on infinite ensembles of functions, given that both Ω and $\mathcal{H}$ are infinite dimensional spaces. However, in practice, we are constrained to consider only a finite number of trajectories, denoted as $\{\boldsymbol{X}^{(i)}, i = 1, \ldots, p\}$. Figure 1 illustrates the transported feature maps associated with the ensemble members, which will be employed to solve the data assimilation problem in section 3.2.

In ensemble data assimilation methods, a superposition principle is most of the time assumed, which in theory is not applicable for non-linear dynamical systems. However, this principle can be rigorously justified within the RKHS family described previously (Dufée et al., 2024). Considering an observable $g_x : \Omega_t \to \mathbb{R}^d$ such that for $t \geq 0$, $\forall \boldsymbol{X}_t \in \Omega_t, \forall x \in \Omega_x$, $g_x(\boldsymbol{X}_t) = \boldsymbol{X}_t(x)$. Assuming g_x is smooth enough to belong to $\mathcal{H}_t$, the kernel reproducing property allows the following decomposition:

$$g_x(\boldsymbol{X}_t) = \langle g_x(\cdot), k_t(\cdot, \boldsymbol{X}_t) \rangle_{\mathcal{H}_t} . \tag{8}$$

Since $k_t(\cdot, \boldsymbol{X}_t)$ is included in $\mathcal{H}_t$ it admits a decomposition such that

$$k_t(\cdot, \boldsymbol{X}_t) = \sum_{i=1}^{p} \boldsymbol{\beta}_i k_t\left(\cdot, \boldsymbol{X}_t^{(i)}\right) , \tag{9}$$

where $\boldsymbol{\beta}$ is a vector of length p. At the continuous level, the sum involved in (9) is infinite, however with the finite number of ensemble members considered, it reduces to a finite sum approximation. By the inner product bilinearity and the kernel reproducing property it follows that:

$$g_x(\boldsymbol{X}_t) = \sum_{i=1}^{p} \boldsymbol{\beta}_i g_x\left(\boldsymbol{X}_t^{(i)}\right) . \tag{10}$$

As this holds for every $x \in \Omega_x$, it can be concluded that:

$$\boldsymbol{X}_t = \sum_{i=1}^{p} \boldsymbol{\beta}_i \boldsymbol{X}_t^{(i)} , \tag{11}$$

at condition that all the observables constituting the state vector – or at least the variables we aim at estimating – belong to $\mathcal{H}_t$. This superposition principle allows the construction of new trajectories as linear combinations of previously computed ones, with the linear

combination being constant over time due to equation (6). As further described in the following sections this property leads to the setting of very appealing data assimilation methods.

Remark 2 *From a machine learning perspective, the RKHS family can be viewed as latent spaces defined from the dynamics, where the weights in the latent spaces are equal to those in the physical space, and characterize a given trajectory.*

In the next section we provide a brief recap on ensemble data assimilation techniques.

3 Assimilation methods

3.1 Classical methods

Although the previous setting is expressed in full generality within functional spaces, for the sake of simplicity, we will describe the assimilation techniques in the following sections using finite-dimensional spaces.

Assimilation methods aim to provide the best estimate of the state of a dynamical system from observations. The observations $\boldsymbol{Y} \in \mathbb{R}^m$ are assumed to be related to the state of the system $\boldsymbol{X} \in \mathbb{R}^n$, as follows,

$$\boldsymbol{Y} = H\left(\boldsymbol{X}\right) + \epsilon \,, \tag{12}$$

where $H : \mathbb{R}^m \mapsto \mathbb{R}^n$, with usually $n \gg m$, denotes the observation operator, $\epsilon \in \mathbb{R}^m$ represent an additive centred Gaussian observation noise with covariance matrix $\boldsymbol{R} \in \mathbb{R}^{m \times m}$. Classical assimilation methods aim to solve the ill-posed problem $H\left(\boldsymbol{X}\right) \simeq \boldsymbol{Y}$, seeking to find the state that best fits the observations. For sake of simplicity, H will be assumed to be linear in the following, and thus represented through matrix $\mathbb{H} \in \mathbb{R}^{m \times n}$ such that $H\left(\boldsymbol{X}\right) = \mathbb{H}\boldsymbol{X}$. To have a well-posed problem and therefore a unique solution, one searches for the closest solution to the forecasted state $\boldsymbol{X}^f$. Consequently, assimilation methods generally seek to minimize a cost function as follows

$$J\left(\boldsymbol{X}\right) = \frac{1}{2}\left\|\mathbb{H}\boldsymbol{X} - \boldsymbol{Y}\right\|^2_{\boldsymbol{R}^{-1}} + \frac{1}{2}\left\|\boldsymbol{X} - \boldsymbol{X}^f\right\|^2_{\boldsymbol{P}^{-1}} \,, \tag{13}$$

where $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is the state variable covariance matrix, and as usual $\| \bullet \|_{\boldsymbol{A}}$ stands for the norm $(\bullet^T \boldsymbol{A} \bullet)$, with upperscript, T, denoting the transpose operator.

Cancellation of the cost function gradient gives the following expression of the optimal analyzed state

$$\boldsymbol{X}^a = -\left(\nabla^2 J\left(\boldsymbol{X}\right)\right)^{-1}\left(\nabla J\left(0\right)\right) = \left(\mathbb{H}^T\boldsymbol{R}^{-1}\mathbb{H} + \boldsymbol{P}^{-1}\right)^{-1}\left(\mathbb{H}^T\boldsymbol{R}^{-1}\boldsymbol{Y} + \boldsymbol{P}^{-1}\boldsymbol{X}^f\right) \,. \tag{14}$$

Using the Sherman-Morrison-Woodbury formula, the previous equation can be rewritten,

$$\boldsymbol{X}^a = \boldsymbol{X}^f - \boldsymbol{G}\left(\mathbb{H}\boldsymbol{X}^f - \boldsymbol{Y}\right) \,, \tag{15}$$

with the Kalman gain expressed as

$$\boldsymbol{G} = \boldsymbol{P}\mathbb{H}^T\left(\boldsymbol{R} + \mathbb{H}\boldsymbol{P}\mathbb{H}^T\right)^{-1} \,. \tag{16}$$

3.1.1 Ensemble Square root Filter (ESRF)

Ensemble methods consider p realizations (or members) in the state space, whose mean and covariance matrix represent the estimated state and covariance. Note that in high-dimensional problems, such as those encountered in ocean dynamics, we have $p \ll m \ll n$, with typical values of $p \sim 50$, $m \sim 10^3$, $n \sim 10^6$.

The ensemble matrix $\boldsymbol{X}$ is defined as a $n \times p$ matrix. Each columns i of $\boldsymbol{X}$ contains a member of the ensemble noted $\boldsymbol{X}^{(i)}$, with $1 \leq i \leq p$. We briefly present in the following the ensemble square root filter (Whitaker & Hamill, 2002), which will serve as a reference in this study. Its name comes from the fact that the state covariance matrix is expressed through its square root, the anomaly matrix $\boldsymbol{A} = [\left(\boldsymbol{X}^{(1)} - \overline{\boldsymbol{X}}\right), \left(\boldsymbol{X}^{(2)} - \overline{\boldsymbol{X}}\right), \ldots, \left(\boldsymbol{X}^{(p)} - \overline{\boldsymbol{X}}\right)]$, where $\overline{\boldsymbol{X}}$ denotes the ensemble mean, which represents the estimated state. The empirical covariance matrix is then expressed as:

$$\boldsymbol{P} = \frac{1}{p-1}\boldsymbol{A}\boldsymbol{A}^T \ . \tag{17}$$

With this decomposition, $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ is stored instead of $\boldsymbol{P} \in \mathbb{R}^{n \times n}$.

The analyzed covariance matrix is expressed as the inverse of the Hessian of cost function J defined in equation (13). Using the Sherman-Morrison-Woodbury formula the covariance matrix update reads

$$\boldsymbol{P}^a = (\boldsymbol{I} - \boldsymbol{G}\mathbb{H})\,\boldsymbol{P}^f \ , \tag{18}$$

where $\boldsymbol{G}$ is the gain matrix expressed in equation (16). Exponents $\bullet^f$ and $\bullet^a$ stand for *forecast* and *analyzed*, respectively. Substituting the gain into the covariance matrix formula leads to,

$$\begin{aligned} &\frac{1}{p-1}\boldsymbol{A}^a\left(\boldsymbol{A}^a\right)^T \\ &\quad = \boldsymbol{A}^f\left(\boldsymbol{I} - \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)^T\left(\boldsymbol{R} + \mathbb{H}\frac{1}{p-1}\boldsymbol{A}^f\left(\boldsymbol{A}^f\right)^T\mathbb{H}^T\right)^{-1}\left(\mathbb{H}\boldsymbol{A}^f\right)\right)\frac{1}{p-1}\left(\boldsymbol{A}^f\right)^T \ . \end{aligned} \tag{19}$$

Applying the Sherman–Morrison–Woodbury formula the previous equation can be written as

$$\boldsymbol{A}^a\left(\boldsymbol{A}^a\right)^T = \boldsymbol{A}^f\left(\boldsymbol{I} + \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)^T\boldsymbol{R}^{-1}\left(\mathbb{H}\boldsymbol{A}^f\right)\right)^{-1}\left(\boldsymbol{A}^f\right)^T \ . \tag{20}$$

Using the square-root matrix $\boldsymbol{S} \in \mathbb{R}^{p \times p}$ defined as

$$\boldsymbol{S} = \left(\boldsymbol{I} + \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)^T\boldsymbol{R}^{-1}\left(\mathbb{H}\boldsymbol{A}^f\right)\right)^{-1/2} \ . \tag{21}$$

The analyzed anomaly is expressed from the forecasted anomaly, as:

$$\boldsymbol{A}^a = \boldsymbol{A}^f\boldsymbol{S} \ . \tag{22}$$

The analyzed ensemble mean is defined using equation (15),

$$\overline{\boldsymbol{X}^a} = \overline{\boldsymbol{X}^f} - \boldsymbol{G}\left(\overline{\mathbb{H}\boldsymbol{X}^f} - \boldsymbol{Y}\right) \ , \tag{23}$$

and the analyzed ensemble members are obtained by adding to the analyzed mean (23) the analyzed anomaly (22),

$$\begin{aligned} \boldsymbol{X}^a &= \overline{\boldsymbol{X}^a} + \boldsymbol{A}^a & (24)\\ &= \overline{\boldsymbol{X}^a} + \boldsymbol{A}^f\boldsymbol{S} \ . & (25) \end{aligned}$$

It is important to note that (25) defines the update of the ensemble in the state space. In ocean models, as indicated previously the state is typically very large. To address this issue it is preferable to work in the weight space.

3.1.2 ESRF in the weight space

We now express directly the ESRF in the weight space. To that end, let $\boldsymbol{\beta}^f$ and $\boldsymbol{\beta}^a$ denote the $p \times p$ weight matrices representing the forecast and analyzed ensembles, such that $\boldsymbol{X}^f = \boldsymbol{X}^f \boldsymbol{\beta}^f$ and $\boldsymbol{X}^a = \boldsymbol{X}^f \boldsymbol{\beta}^a$. The i^{th} column of the weight matrix contains the coefficients to construct the i^{th} member of the ensemble from the forecasted ensemble members. It appears clearly from its definition, that $\boldsymbol{\beta}^f$ is the identity matrix. This formulation is similar to decompositions found in the literature, such as in Lorenc (2003), Hunt et al. (2004), and Liu et al. (2008).

Injecting $\boldsymbol{P}^f = \boldsymbol{A}^f \left(\boldsymbol{A}^f\right)^T$ in the gain definition leads to

$$\boldsymbol{G} = \frac{1}{p-1} \boldsymbol{A}^f \left(\boldsymbol{A}^f\right)^T \mathbb{H}^T \left(\boldsymbol{R} + \mathbb{H} \frac{1}{p-1} \boldsymbol{A}^f \left(\boldsymbol{A}^f\right)^T \mathbb{H}^T \right)^{-1} . \tag{26}$$

Writing $\boldsymbol{A} = \boldsymbol{X}\left(\boldsymbol{I} - \mathbb{1}/p\right)$, where $\mathbb{1}$ denotes the $p \times p$ matrix filled with one, the previous equation reads then

$$\boldsymbol{G} = \frac{1}{p-1} \boldsymbol{X}^f \left(\boldsymbol{I} - \mathbb{1}/p\right) \left(\boldsymbol{I} - \mathbb{1}/p\right)^T \left(\boldsymbol{X}^f\right)^T \mathbb{H}^T \left(\boldsymbol{R} + \mathbb{H} \frac{1}{p-1} \boldsymbol{A}^f \left(\boldsymbol{A}^f\right)^T \mathbb{H}^T \right)^{-1} . \tag{27}$$

A gain matrix in the weight space, denoted $\boldsymbol{G}_w \in \mathbb{R}^{p \times m}$, can thus be defined such that $\boldsymbol{G} = \boldsymbol{X}^f \boldsymbol{G}_w$, with,

$$\boldsymbol{G}_w = \frac{1}{p-1} \left(\mathbb{H}\boldsymbol{A}\right)^T \left(\boldsymbol{R} + \frac{1}{p-1} \mathbb{H}\boldsymbol{A}^f \left(\mathbb{H}\boldsymbol{A}^f\right)^T \right)^{-1} . \tag{28}$$

Considering the mean weight vectors $\overline{\boldsymbol{\beta}^f}$ and $\overline{\boldsymbol{\beta}^a}$, such that $\overline{\boldsymbol{X}^f} = \boldsymbol{X}^f \overline{\boldsymbol{\beta}^f}$ and $\overline{\boldsymbol{X}^a} = \boldsymbol{X}^f \overline{\boldsymbol{\beta}^a}$, which correspond to the column averages of the associated weight matrices. As $\boldsymbol{\beta}^f$ is the identity, $\overline{\boldsymbol{\beta}^f}$ is a p length vector filled with $\frac{1}{p}$. The equation (23) can thus be rewritten in the weight space:

$$\overline{\boldsymbol{\beta}^a} = \overline{\boldsymbol{\beta}^f} - \boldsymbol{G}_w \left(\mathbb{H}\boldsymbol{X}^f \overline{\boldsymbol{\beta}^f} - \boldsymbol{Y} \right) . \tag{29}$$

As $\boldsymbol{X}^a = \boldsymbol{X}^f \boldsymbol{\beta}^a$ and $\boldsymbol{A}^f = \boldsymbol{X}^f \left(\boldsymbol{\beta}^f\right)'$, where $\bullet'$ denotes the right matrix multiplication of the operand by the matrix $\boldsymbol{I} - \mathbb{1}/p$, equation (25) can be rewritten

$$\boldsymbol{X}^f \boldsymbol{\beta}^a = \boldsymbol{X}^f \overline{\boldsymbol{\beta}^a} + \boldsymbol{X}^f \left(\boldsymbol{\beta}^f\right)' \boldsymbol{S}, \tag{30}$$

which yields the weight update

$$\boldsymbol{\beta}^a = \overline{\boldsymbol{\beta}^a} + \left(\boldsymbol{\beta}^f\right)' \boldsymbol{S} . \tag{31}$$

3.2 Ensemble filter on RKHS family

The formalism proposed by Dufée et al. (2024), as detailed in section 2, has the advantage to be directly formulated in terms of piece-wise in time, linear combination of trajectories. It naturally enables the estimation of an optimal trajectory from time-series of observations. More precisely, this ensemble filter (EnF) defined on the RKHS family, operates on the weights, within the RKHS family, assuming the observation operator is well defined on the RKHS family through the smooth evaluation operator assumption (8). In the following, we will assume without loss of generality that the kernel is real-valued. The filter, referred to as RKHS-EnF in the following, is built from a 4DVar-like cost function, accounting for time-series of observations, and incorporating a regularization term based on the norm of the kernel function in the initial RKHS $\mathcal{H}_0$:

$$J\left(\boldsymbol{X}\right) = \frac{1}{2} \sum_{j=0}^{J} \left\| \mathbb{H}_{t_j} \boldsymbol{X}_{t_j} - \boldsymbol{Y}_{t_j} \right\|^2_{\boldsymbol{R}_{t_j}^{-1}} + \frac{1}{2} \left\| k_0 \left(\cdot, \boldsymbol{X}_0 - \overline{\boldsymbol{X}_0^f} \right) \right\|^2_{\mathcal{H}_0} , \tag{32}$$

where $\mathbb{H}_{t_j}$, $\boldsymbol{X}_{t_j}$, $\boldsymbol{Y}_{t_j}$ stand respectively for the observation matrix, the ensemble matrix and the observation vector at time t_j. Given $\boldsymbol{\beta}$ the weight vector such that $\boldsymbol{X}_{t_j} = \sum_{i=1}^{p} \boldsymbol{\beta}_i \boldsymbol{X}_{t_j}^{(i)}$, and using equations (9), the second term of the right-hand side can be expressed as

$$\left\| k_0\left(\cdot, \boldsymbol{X}_0 - \overline{\boldsymbol{X}_0}\right)\right\|_{\mathcal{H}_0}^2 = \left\| \sum_{i=1}^{p} \left(\boldsymbol{\beta}_i - \overline{\boldsymbol{\beta}_i^f}\right) k_0\left(\cdot, \boldsymbol{X}^{(i)}\right)\right\|_{\mathcal{H}_0}^2 . \tag{33}$$

By the kernel reproducing property, we obtain

$$\left\| k_0\left(\cdot, \boldsymbol{X}_0 - \overline{\boldsymbol{X}_0}\right)\right\|_{\mathcal{H}_0}^2 = \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}^f}\right)^T \boldsymbol{K} \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}^f}\right) , \tag{34}$$

where $\boldsymbol{K}$ is the kernel matrix that contains the evaluation of the initial kernel between every member's pair of the ensemble, $\boldsymbol{K}_{i,j} = k_0\left(\boldsymbol{X}_0^{(i)}, \boldsymbol{X}_0^{(j)}\right)$, for $1 \leq i, j \leq p$. The second member of the cost function penalizes hence the weight anomaly in the RKHS.

The previous cost function can thus be rewritten as a function of the weight vector in the RKHS, thereby formulating the assimilation problem directly in the weight space:

$$J_{\text{RKHS}}\left(\boldsymbol{\beta}\right) = \frac{1}{2} \sum_{j=0}^{J} \left\| \sum_{i=1}^{p} \boldsymbol{\beta}_i \mathbb{H}_{t_j} \boldsymbol{X}_{t_j}^{(i)} - \boldsymbol{Y}_{t_j} \right\|_{\boldsymbol{R}_{t_j}^{-1}} + \frac{1}{2} \left\| \boldsymbol{\beta} - \overline{\boldsymbol{\beta}^f} \right\|_{\boldsymbol{K}}^2 . \tag{35}$$

Let us define $\boldsymbol{V}$ such that $\mathbb{H}_t \boldsymbol{X}_t = \boldsymbol{V}_t \boldsymbol{k}$ and $\boldsymbol{V} = \left(\boldsymbol{V}_{t_j}\right)_{j \in [\![1,J]\!]}$ where $\boldsymbol{k} = \left(k_0\left(\boldsymbol{X}_0, \boldsymbol{X}^{(i)}\right)\right)_{i \in [\![1,p]\!]}$. By the reproducing property, we have $\boldsymbol{k} = \boldsymbol{K}\boldsymbol{\beta}$. Denoting $\boldsymbol{Y} = \left(\boldsymbol{Y}_{t_j}\right)_{j \in [\![1,J]\!]}$ and $\widetilde{\boldsymbol{R}} = \boldsymbol{I}_J \otimes \left(\boldsymbol{R}_{t_j}\right)_{j \in [\![0,J]\!]}$, where $\otimes$ represents the Kronecker product, the cost function can be finally compactly expressed as:

$$J_{\text{RKHS}}\left(\boldsymbol{\beta}\right) = \frac{1}{2} \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{\beta} - \boldsymbol{Y}\right)^T \widetilde{\boldsymbol{R}}^{-1} \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{\beta} - \boldsymbol{Y}\right) + \frac{1}{2} \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}^f}\right)^T \boldsymbol{K} \left(\boldsymbol{\beta} - \overline{\boldsymbol{\beta}^f}\right) . \tag{36}$$

Remark 3 *In equation* (35)*, a linear observation operator is considered through the matrix* $\mathbb{H}_{t_j}$*. However, the expression* (36) *remains valid for non-linear observation operators, at the condition that the observables associated with these observations belong to* $\mathcal{H}_t$*. This is the reason of introducing explicitly the notation* $\mathbb{H}_t \boldsymbol{X}_t = \boldsymbol{V}_t \boldsymbol{K} \boldsymbol{\beta}$*. This linear combination representation is a direct consequence of the reproducing property. The fact that these coefficients are constant in time ensues from the framework proposed by Dufée et al. (2024) presented in section 2.3.*

3.2.1 Updating the kernel matrix

The inverse of the kernel matrix $\boldsymbol{K} \in \mathbb{R}^{p \times p}$ can be interpreted as the weight covariance matrix, $\boldsymbol{P}_w^f = (\boldsymbol{K})^{-1}$. We notice that due to the RKHS property expressed in equation (4), $\boldsymbol{K}$ is invertible. To simplify the notation, let us denote $\boldsymbol{B}^f = \boldsymbol{K}^{-1/2}$ as the square root matrix of $\boldsymbol{P}_w^f$, so that $\boldsymbol{P}_w^f = \boldsymbol{B}^f \left(\boldsymbol{B}^f\right)^T$. Similarly, we denote the analyzed covariance square root matrix as $\boldsymbol{B}^a = \left(\boldsymbol{P}_w^a\right)^{1/2}$. In practice, the square root of the covariance matrix is computed using an eigenvalue decomposition. Analogous to the ESRF, we can then define the square root matrix $\boldsymbol{S}_{\text{RKHS}}$,

$$\boldsymbol{B}^a = \boldsymbol{B}^f \boldsymbol{S}_{\text{RKHS}} , \tag{37}$$

the EnF gain,

$$\boldsymbol{G}_{\text{RKHS}} = \boldsymbol{P}_w^f \left(\boldsymbol{V}\boldsymbol{K}\right)^T \left(\widetilde{\boldsymbol{R}} + \left(\boldsymbol{V}\boldsymbol{K}\right) \boldsymbol{P}_w^f \left(\boldsymbol{V}\boldsymbol{K}\right)^T\right)^{-1} , \tag{38}$$

and the covariance matrix, which is updated according to:

$$\boldsymbol{P}_w^a = \left(\boldsymbol{I} - \boldsymbol{G}_{\text{RKHS}} \left(\boldsymbol{V}\boldsymbol{K}\right)\right) \boldsymbol{P}_w^f . \tag{39}$$

Substituting the gain into this equation yields:

$$\boldsymbol{B}^a \left(\boldsymbol{B}^a\right)^T = \boldsymbol{B}^f \left(\boldsymbol{I} - \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right)^T \left(\widetilde{\boldsymbol{R}} + \left(\boldsymbol{V}\boldsymbol{K}\right) \boldsymbol{B}^f \left(\boldsymbol{B}^f\right)^T \left(\boldsymbol{V}\boldsymbol{K}\right)^T \right)^{-1} \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right) \right) \left(\boldsymbol{B}^f\right)^T \ . \tag{40}$$

Upon applying the Sherman–Morrison–Woodbury formula, the above equation can be rewritten as:

$$\boldsymbol{B}^a \left(\boldsymbol{B}^a\right)^T = \boldsymbol{B}^f \left(\boldsymbol{I} + \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right)^T \widetilde{\boldsymbol{R}}^{-1} \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right) \right)^{-1} \left(\boldsymbol{B}^f\right)^T \ . \tag{41}$$

As $\boldsymbol{B}^a \left(\boldsymbol{B}^a\right)^T = \boldsymbol{B}^f \boldsymbol{S}_{\text{RKHS}} \left(\boldsymbol{S}_{\text{RKHS}}\right)^T \left(\boldsymbol{B}^f\right)^T$, the RKHS EnF square root matrix is finally identified as:

$$\boldsymbol{S}_{\text{RKHS}} = \left(\boldsymbol{I} + \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right)^T \widetilde{\boldsymbol{R}}^{-1} \left(\boldsymbol{V}\boldsymbol{K}\boldsymbol{B}^f\right) \right)^{-1/2} \ . \tag{42}$$

The matrix $\boldsymbol{S}_{\text{RKHS}}$ is then used to update the covariance square root matrix, from which the analyzed covariance matrix can be derived. Both the gain $\boldsymbol{G}_{\text{RKHS}}$ and transition matrix $\boldsymbol{S}_{\text{RKHS}}$ are defined for assimilating an observation time series, with these matrices being weighted by the inverse of the kernel matrix.

3.2.2 Updating the weights

It is important to note that unlike the ESRF, the RKHS-based approach does not directly provide a formula for updating the ensemble weight anomalies. The matrix $\boldsymbol{S}_{\text{RKHS}}$ does not update the anomalies in the same direct manner. However, since the anomaly can be treated as a centered random vector with a non-singular covariance matrix, it is possible to apply a whitening transformation to this vector. This is achieved by multiplying it by the square root of its covariance matrix inverse, $\boldsymbol{K}^{1/2}$, resulting in a decorrelated vector (Koivunen & Kostinski, 1999; Friedman, 1987). To transform this vector into one that aligns with the analyzed covariance matrix $\boldsymbol{P}_w^a$, a subsequent coloring transformation is applied. This transformation, being the inverse of the whitening process, involves multiplying the vector by the square root of the analyzed covariance matrix $\left(\boldsymbol{P}_w^a\right)^{1/2}$. Consequently, the updated weight anomaly can be expressed as follows:

$$\begin{aligned} \left(\boldsymbol{\beta}^a\right)' &= \left(\boldsymbol{\beta}^f\right)' \left(\boldsymbol{P}_w^f\right)^{-1/2} \left(\boldsymbol{P}_w^a\right)^{1/2} \\ &= \left(\boldsymbol{\beta}^f\right)' \boldsymbol{K}^{1/2} \left(\boldsymbol{B}^f \boldsymbol{S}_{\text{RKHS}}\right) \ . \end{aligned} \tag{43}$$

Given that $\boldsymbol{B}^f = \boldsymbol{K}^{-1/2}$, it follows that the previous equation can be simplified to

$$\left(\boldsymbol{\beta}^a\right)' = \left(\boldsymbol{\beta}^f\right)' \boldsymbol{S}_{\text{RKHS}} \ . \tag{44}$$

The ensemble mean can be then updated in a manner consistent with classical ensemble Kalman filters:

$$\overline{\boldsymbol{\beta}^a} = \overline{\boldsymbol{\beta}^f} - \boldsymbol{G}_{\text{RKHS}} \left(\boldsymbol{V}\boldsymbol{K}\overline{\boldsymbol{\beta}^f} - \boldsymbol{Y} \right) \ , \tag{45}$$

and the ensemble weights are updated as:

$$\boldsymbol{\beta}^a = \overline{\boldsymbol{\beta}^a} + \left(\boldsymbol{\beta}^f\right)' \boldsymbol{S}_{\text{RKHS}} \ , \tag{46}$$

similarly as in (31).

A similar approach can be considered in the ESRF context to express the ensemble update. Given that $\boldsymbol{A}^a = \boldsymbol{A}^f \boldsymbol{S}$, the transition matrix reads $\boldsymbol{S} = \left(\boldsymbol{A}^f\right)^{-1} \boldsymbol{A}^a$. Within the ESRF framework, where the anomaly is assumed to be the square root of the covariance matrix, the update of the anomaly can also be interpreted as a whitening transformation followed by a coloring transformation. In the ESRF setting, the change of variable $\left(\boldsymbol{P}_w^f\right)^{-1/2}\boldsymbol{\beta}$ allows to work with variables decorrelated from each other – through the diagonalisation of the covariance matrix. In the RKHS-based case, $\left(\boldsymbol{P}_w^f\right)^{-1/2} = \boldsymbol{K}^{1/2}$, and the equivalent change of variable procedure constitutes an embedding in the RKHS.

3.3 Link between ESRF and EnF on a RKHS family

The ESRF can be viewed as a specific instance of RKHS-EnF, given an appropriately selected initial kernel. To establish the connection between these two techniques, we derive in this section the ESRF equations from the RKHS EnF equations by selecting a suitable initial kernel.

Let us consider a RKHS EnF which assimilates observations directly at the time they are available, as a consequence $\boldsymbol{V} = \boldsymbol{V}_t$ and $\widetilde{\boldsymbol{R}} = \boldsymbol{R}_t$ for a fixed time instant t. As $\boldsymbol{k} = \boldsymbol{K}\boldsymbol{\beta}$, by the definition of $\boldsymbol{V}$, we have

$$\boldsymbol{V}_t\boldsymbol{K}\boldsymbol{\beta} = \mathbb{H}\boldsymbol{X}_t \,, \tag{47}$$

and it follows that $\boldsymbol{V}_t\boldsymbol{K}\boldsymbol{\beta}^f = \mathbb{H}\boldsymbol{X}_t^f$. However, as seen previously $\boldsymbol{\beta}^f$ is equal to the identity matrix. This represents the forecast members' weights expressed in their own basis, therefore $\boldsymbol{V}_t\boldsymbol{K} = \mathbb{H}\boldsymbol{X}_t^f$. For simplicity, the time index t is omitted in the subsequent discussion since all observations are assumed to be taken simultaneously. Assuming the kernel is chosen such that $(\boldsymbol{K})^{-1} = \frac{1}{p-1}(\boldsymbol{I} - \mathbb{1}/p)\,\boldsymbol{I}\,(\boldsymbol{I} - \mathbb{1}/p)$, the square root of $\boldsymbol{K}$ is given by $\boldsymbol{B}^f = \frac{1}{\sqrt{p-1}}(\boldsymbol{I} - \mathbb{1}/p)$, and the RKHS-EnF gain reads

$$\begin{aligned}\boldsymbol{G}_{\text{RKHS}} = &\frac{1}{p-1}(\boldsymbol{I} - \mathbb{1}/p)\,\boldsymbol{I}\,(\boldsymbol{I} - \mathbb{1}/p)\left(\mathbb{H}\boldsymbol{X}^f\right)^T \\ &\left(\widetilde{\boldsymbol{R}} + \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{X}^f\right)(\boldsymbol{I} - \mathbb{1}/p)\,\boldsymbol{I}\,(\boldsymbol{I} - \mathbb{1}/p)\left(\mathbb{H}\boldsymbol{X}^f\right)^T\right)^{-1} .\end{aligned} \tag{48}$$

It is worth noting that the kernel K defined here has a Mercer spectrum $\mu_i = 1$ for $0 \leq i \leq p$, and can be interpreted as the zero-correlation limit of a Gaussian, tending towards a Dirac measure, which is not a RKHS reproducing kernel (as it is unbounded). As $\boldsymbol{X}^f(\boldsymbol{I} - \mathbb{1}/p) = \boldsymbol{A}^f$ and $(\boldsymbol{I} - \mathbb{1}/p)^2 = (\boldsymbol{I} - \mathbb{1}/p)$, we have

$$\boldsymbol{G}_{\text{RKHS}} = \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)^T\left(\widetilde{\boldsymbol{R}} + \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)\left(\mathbb{H}\boldsymbol{A}^f\right)^T\right)^{-1} = \boldsymbol{G}_w \,. \tag{49}$$

Considering the matrix $\boldsymbol{S}_{\text{RKHS}}$ under the same assumption, we get:

$$\boldsymbol{S}_{\text{RKHS}} = \left(\boldsymbol{I} + \left(\mathbb{H}\boldsymbol{X}^f\frac{1}{\sqrt{p-1}}(\boldsymbol{I} - \mathbb{1}/p)\right)^T\widetilde{\boldsymbol{R}}^{-1}\left(\mathbb{H}\boldsymbol{X}^f\frac{1}{\sqrt{p-1}}(\boldsymbol{I} - \mathbb{1}/p)\right)\right)^{-1/2}, \tag{50}$$

which simplifies to:

$$\boldsymbol{S}_{\text{RKHS}} = \left(\boldsymbol{I} + \frac{1}{p-1}\left(\mathbb{H}\boldsymbol{A}^f\right)^T\widetilde{\boldsymbol{R}}^{-1}\left(\mathbb{H}\boldsymbol{A}^f\right)\right)^{-1/2} = \boldsymbol{S} \,. \tag{51}$$

Consequently, performing the assimilation step each time an observation is available, using a well-chosen kernel, provides an RKHS EnF gain and a square root matrix that are equivalent to those of a classical square root filter. Furthermore, since under these conditions $\boldsymbol{V}\boldsymbol{K} = \mathbb{H}\boldsymbol{X}$, the two formulas used to update the ensemble are equivalent and the two filters are equivalent.

3.4 Ensemble update strategy

The embedding of the dynamics into the RKHS family ensures that the ensemble weights remain invariant over time. This invariance provides the reconstruction of the analyzed ensemble members at any time of the assimilation window, as a linear combination of the forecast ensemble. The weights are calculated with respect to the initial kernel, and consequently, in the method proposed herein, it is natural to construct the analyzed members at the beginning of the assimilation window. Due to the non-linear dynamics of the system and the finite number of ensemble members, the analyzed members reconstructed at the beginning of the assimilation window and subsequently propagated forward in time may

differ at the end of the assimilation window from the analyzed members. It is noteworthy that, according to the theoretical framework, if an infinite number of ensemble members were available, for an invertible dynamics the reconstruction at the beginning of the assimilation window, when transported by the dynamical system, would be identical to the reconstruction at the end of the assimilation window. However, obviously due to the finite size of the ensemble members and due to noninversible numerical dissipation process, this is only approximately true. The ability of a direct reconstruction of the initial condition at the beginning of the assimilation window is used to forecast updated trajectories of the ensemble members. This approach referred to as the tiled RKHS ensemble filter is described in Algorithm 1.

Remark 4 *The tiled RKHS EnF is consistent with the theoretical framework, as all estimations are re-expressed at the initial time t_0 through the isometry between the evolving RKHSs. This procedure has also the advantage of introducing a spin-up phase during the assimilation window, where spurious numerical artefacts induced by the finite size linear-combination reconstructions are filtered-out by the dynamical system, and are ultimately replaced by physically relevant features.*

Remark 5 *The resulting tiled RKHS ensemble filter shares similarities with Bayesian smoothers and 4D-Var methods, with the key distinction that it avoids multiple forward-backward integrations of the dynamical system, as we infer directly the initial condition of the system (at the beginning of the assimilation window). In the proposed method, the model is integrated forward only once to propagate the analyzed ensemble to the next assimilation window. An intermediate method, which can be referred to as sequential RKHS EnF, can also be devised by reconstructing the ensemble at the end of the assimilation window, akin to classical ensemble filters. Although this approach does not fully leverage the theoretical framework's potential, it maintains the same computational cost as a classical sequential filter.*

Algorithm 1 Tiled RKHS EnF

Require: Ensemble X_0, initial time t_0, end time t_{end}, time step dt, assimilation interval T_a.
  $X_t^f \leftarrow X_0$
  $t \leftarrow t_0$
  $t_a \leftarrow T_a$ the assimilation time horizon.
  Save X_t for future reconstruction.
  Construct the kernel matrix K.
  Initialize VK and Y as empty.
  while $t < t_{end}$ **do**
    $t \leftarrow t + dt$
    Propagates the ensemble members in time to update X_t^f.
    Append to VK the observations of the ensemble members.
    Append to Y the reference observations.
    if It is time to assimilate i.e. t is a multiple of t_a **then**
      $\beta^a \leftarrow \text{assimilate}(VK, Y, K)$ ▷ Equations (45) and (46).
      Empty VK and Y.
      Apply the weights β^a to $X_{t-t_a}^f$ saved at $t - t_a$, to obtain $X_{t-t_a}^a$.
      Propagates $X_{t-t_a}^a$ on the time interval $[t - t_a, t]$, to obtain X_t^a.
      $X_t^f \leftarrow X_t^a$
      Save X_t^f for future reconstruction.
      Update the kernel matrix K.
    end if
  end while

4 Experimental framework

In the following we first describe the experimental setup we built to assess the data assimilation we propose.

4.1 Dynamical model

The data assimilation methods previously presented are compared using a North Atlantic Ocean model, covering the region between 9°N and 48°N in latitude and -98°E and -3.84°E in longitude with realistic coasts. This area represents a domain of 4329 km by 8658 km. The ocean bottom is assumed to be flat, and all boundaries are considered closed, even those at the equator and in the north, which are normally open boundaries. Due to the flat bottom, the Gulf Stream develops close to the North American coast. To achieve a well-positioned Gulf Stream, areas with a bathymetry shallower than -100 meters are treated as land. A realistic wind forcing is considered. A multi-layer quasi-geostrophic model implemented as the projection of a rotating shallow-water (Thiry et al., 2024) is considered. The rotating shallow water is simulated with the model proposed by Roullet and Gaillard (2022), while the QG projection uses the elliptic solver of Thiry et al. (2023). Three layers of heights $\boldsymbol{H} = (400, 1100, 2600)$ are considered. The evolution of three prognostic variables are simulated: $\boldsymbol{u}$, $\boldsymbol{v}$ the two components of the horizontal velocity vector, and $\boldsymbol{h}$ the height anomaly in each layers. These variables are propagated over time according to the following equations:

$$\begin{cases} \partial_t \boldsymbol{u} = (\boldsymbol{\omega} + \boldsymbol{f})\,\boldsymbol{v} - \partial_x (\boldsymbol{p} + \boldsymbol{k}) \\ \partial_t \boldsymbol{v} = -(\boldsymbol{\omega} + \boldsymbol{f})\,\boldsymbol{u} - \partial_y (\boldsymbol{p} + \boldsymbol{k}) \\ \partial_t \boldsymbol{h} = -\boldsymbol{H}\,(\partial_x \boldsymbol{u} + \partial_y \boldsymbol{v}) - \partial_x (\boldsymbol{u}\boldsymbol{h}) - \partial_y (\boldsymbol{v}\boldsymbol{h}) \end{cases}, \tag{52}$$

with $\boldsymbol{\omega} = \partial_x \boldsymbol{v} - \partial_y \boldsymbol{u}$ the relative vorticity, $\boldsymbol{p}$ the pressure vector, $\boldsymbol{k} = \left(\boldsymbol{u}^2 + \boldsymbol{v}^2\right)/2$ the kinetic energy and $\boldsymbol{f}_i = f_0 + \beta y$ for $0 \leq i \leq 3$ the Coriolis parameter under a beta plane assumption, with $f_0 = 9.375 \times 10^{-5}\,\mathrm{s}^{-1}$ the mean Coriolis parameter, and $\beta = 1.754 \times 10^{-11}\,\mathrm{m}^{-1}\mathrm{s}^{-1}$ the Coriolis parameter gradient. The pressure values in layer $1 \leq i \leq 3$ are define such that $p_i = \rho_1 \sum_{j=1}^{i} \left(g'_j \sum_{k=j}^{3} h_k\right)$ with $g_1 = g = 9.81\,\mathrm{m}\cdot\mathrm{s}^{-2}$ and $g'_i = g\,(\rho_i - \rho_{i-1})/\rho_1$ for $i = 2, 3$. ρ is the density in the different layers.

The QG equations are then obtained by projecting this shallow water system using a projector involving an elliptic operator. The full details of this procedure as well as the numerical codes can be found in Thiry et al. (2024). The simulation is performed on a $512 \times 1024 \times 3$ grid ($n_x = 512$, $n_y = 1024$, $n_l = 3$), resulting in a mesh size of approximately 8.46×8.46 kilometers square. The time step is set to 2000 seconds.

4.2 Ensemble generation

The initial ensemble $\boldsymbol{X}_0$ is generated using a noise based on the spectral decomposition of random local fluctuations (Bauer et al., 2020; Brecht et al., 2021). This technique is applied to perturb the velocity field. Given a velocity field $\boldsymbol{u}$ defined on a $n = n_l \times n_x \times n_y$ grid, a local spatial window $W_{i,j}$ of size $n_w \times n_w$ (n_w being odd) in the horizontal direction is considered around each grid point $x_{i,j}$, with $1 < i < n_x$ and $1 < j < n_y$. Within each window, n_o samples are randomly picked, where $n_o > 1$. The sampled values within the window $W_{i,j}$ are denoted as $\boldsymbol{w}_k^{(i,j)}$, where $1 \leq k \leq n_o$, and these values represent n_l-dimensional vectors associated with the number of vertical layers. These values are centered by subtracting their mean as follows:

$$\left(\boldsymbol{w}_k^{(i,j)}\right)' = \boldsymbol{w}_k^{(i,j)} - \overline{\boldsymbol{w}}^{(i,j)} \quad \text{with} \quad \overline{\boldsymbol{w}}^{(i,j)} = \frac{1}{n_o} \sum_{\ell=1}^{n_o} \boldsymbol{w}_\ell^{(i,j)}. \tag{53}$$

Hence a matrix $\boldsymbol{Z} \in \mathbb{R}^{n \times n_o}(\mathbb{R})$ is constructed by organizing the centered values $\left(\boldsymbol{w}_k^{(i,j)}\right)'$ for each grid point in the $n_l \times n_x \times n_y$ grid. A singular value decomposition (SVD) is performed

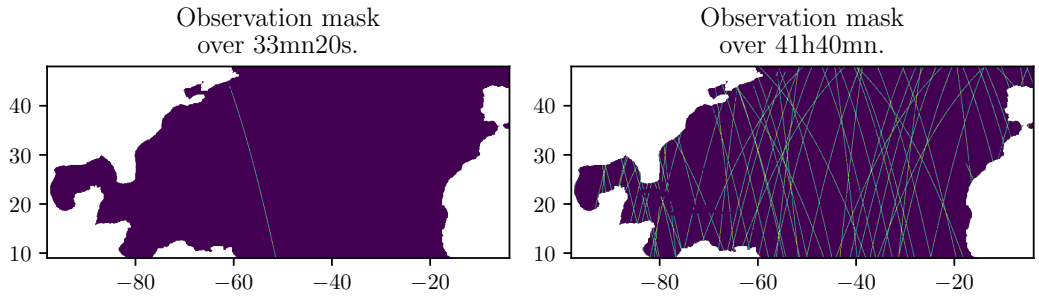

Figure 2: Observation masks for the periods of 33 minutes and 20 seconds (left) and 41 hours and 40 minutes (right), both from January 1, 2022.

$\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, allowing to define perturbation modes $\boldsymbol{U}_i$ from the columns of $\boldsymbol{U}$ and the diagonal matrix of singular values $\boldsymbol{\Sigma}_{jj} = \lambda_j$. The perturbation, $\{\boldsymbol{X}_0^{'(i)}, i = 1\ldots, p\}$ are generated from the p eigenpairs of the empirical covariance matrix:

$$\boldsymbol{X}^{'(i)} = \sum_{j=1}^{n_o} \lambda_j \boldsymbol{U}_j \, \eta_j^{(i)}, \tag{54}$$

where $\eta \in \mathbb{R}^{n_o}$ is a multivariate random vector composed of n_o standard Gaussian variables and $\eta^{(i)}$ denotes the i th realization. The initial ensemble members are obtained by adding these random fields to the initial velocity component. In the experiments, 76 members have been generated using this method, with $n_w = 5$ and $n_o = 21$. One member is retained as the reference, while the remaining 75 members constitute the training set. The reference member is not included in the training ensemble.

4.3 Observations simulation

An Observing System Simulation Experiment (OSSE) was conducted using sea surface height (SSH) as the observed variable. The SSH corresponds to the sum of the layer height anomalies $\boldsymbol{h}$ along the vertical. Synthetic observations were generated from the reference simulation by applying spatial masks derived from realistic data provided by the Copernicus along-track sea surface height database (DOI https://doi.org/10.48670/moi-00147). The initial observation date was set to January 1, 2022. Figure 2 illustrates the observation masks for two periods: one over 33 minutes and 20 seconds, corresponding to a single simulation time step, and another over 41 hours and 40 minutes. It can be shown that over 41h40min, the altimeters swaths constitute a reasonable coarse grid observational coverage of the basin. Usual OI methods, provide a dense field of observation, by interpolating these observation spatially assuming they are all acquired at the same time. In the technique we propose, the observations are treated as a time-series and are not interpolated.

Four distinct experiments were carried out. In the first two, observations were perturbed with additive independent white Gaussian noise. The noise standard deviations were set to 1% and 10% of the root mean square (RMS) of the reference member's SSH at the initial time, corresponding to standard deviations of $\sigma_o = 0.18\,\text{cm}$ and $\sigma_o = 1.8\,\text{cm}$, respectively. Considering spatially uncorrelated white Gaussian noise is a strong idealization of real-world data noise. In operational setups, the noise is correlated, as raw sea surface height measurements are corrected to remove barotropic tidal effects and the predictible internal tide waves, compensate for satellite jitters, and account for atmospheric turbulence.

Due to this, two additional experiments with spatially correlated noise were conducted. The noise was generated by smoothing a field of white Gaussian noise using a Gaussian

filter with a standard deviation of 250 km and a support length of 300 km. This approach produced noise patterns with spatial correlations and, notably, the resulting noise was non-Gaussian due to the relatively limited support area of the filter. The initial variance of the white Gaussian noise was calibrated so that the mean covariance square root of the resulting spatially correlated noise was approximately of 1% and 10% of the SSH RMS (0.18 cm and 1.8 cm respectively). Although this noise is more realistic than white Gaussian noise, it is still obviously idealized. In all experiments, the observation error covariance matrix $\widetilde{\boldsymbol{R}}$ used in the filters was assumed to be diagonal, with all observations having the same error variance σ_o^2. As a result, the filter was not informed of the correlations between observations, posing a challenge for the data assimilation method to remain robust despite this misinformation.

4.4 Filters Setting

4.4.1 Assimilation frequency

In the context of the RKHS EnF, the assimilation frequency can be adjusted to optimize the reconstruction quality. In the one hand, if the assimilation window is too short, the observation time-series might be too short to exploit all the potential benefit of the RKHS EnF. In the limit case where observations are sequentially assimilated at each time step and the kernel is defined as described in section 3.3, the RKHS EnF and ESRF filters are equivalent. However, an excessively long assimilation window may violate the invertibility assumption of the dynamical system, leading to poorer outcomes for the ensemble linear combination of trajectories over that period. Conversely, for the tiled version of the filter, a longer window can enhance the development of small-scale feature variability and assist in eliminating non-physical structures within the dynamical system.

In the following, for the untiled sequential version of the RKHS EnF, the assimilation window length is set to 8 hours and 20 minutes when the observations are perturbed by white Gaussian noise, and to 2 hours, 46 minutes, and 40 seconds when perturbed by spatially correlated noise. These correspond to analysis updates every 15 and 5 time steps of the dynamics, respectively. The presence of spatially correlated noise degrades the filter's performance, necessitating more frequent corrections to the ensemble. For the tiled version, the assimilation window is set to 41 hours and 40 minutes, which corresponds to 75 time steps of the dynamics, regardless of the type of noise affecting the observations. This suggests a robustness in the tiled EnF's configuration, making it less sensitive to the noise characteristics. The ESRF is updated every time step (*i.e.* every 33 minutes and 20 seconds).

Remark 6 *It is important to note that a tiled version of the ESRF is theoretically not directly feasible. The possibility of using this strategy with the RKHS EnF arises from the stationary weights associated with the linear combination of trajectories and the time series of observations. Obtaining the assimilated state at the beginning of the assimilation windows requires either a forward-backward Bayesian smoother or the application of 4D VAR techniques.*

4.4.2 Kernel selection

In the experiments, a centered Gaussian kernel , defined by the following expression, was employed:

$$\forall \boldsymbol{X}, \boldsymbol{Y} \in \Omega_0,\ k_g\left(\boldsymbol{X}, \boldsymbol{Y}\right) = \exp\left(-\frac{\|\boldsymbol{u_X} - \boldsymbol{u_Y}\|^2_{\Omega_x} + \|\boldsymbol{v_X} - \boldsymbol{v_Y}\|^2_{\Omega_x}}{l_G^2}\right), \tag{55}$$

where $(\boldsymbol{u_X}, \boldsymbol{v_X})$ and $(\boldsymbol{u_Y}, \boldsymbol{v_Y})$ denote the velocity vectors associated with states $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, and $\|\bullet\|_{\Omega_x}$ denoting the L^2 norm over the domain Ω_x. The kernel matrix associated with this kernel is denoted $\boldsymbol{K}_G$. The length scale parameter l_G is chosen to maintain a constant ratio between the smallest and largest Mercer eigenvalues of $\boldsymbol{K}_G$ which

is set to 10^{-2}. The Gaussian kernel matrix is then normalized by multiplying it by $p-1$ and then centered by pre- and post-multiplying it by the projection operator $(\boldsymbol{I} - \mathbb{1}/p)$:

$$\boldsymbol{K} = \alpha(p-1)\,(\boldsymbol{I} - \mathbb{1}/p)\,\boldsymbol{K}_G\,(\boldsymbol{I} - \mathbb{1}/p)\ , \tag{56}$$

where α is a constant. This centering procedure is simply equivalent to a change of variable toward the anomaly of the coefficients $\boldsymbol{\beta} \mapsto \boldsymbol{\beta}'$. Consequently, this is equivalent to work with a full rank kernel matrix (*i.e.* the kernel and the associated RKHS is well defined), which operates on the weight anomalies. Specifically, α is set to 0.75 for uncorrelated Gaussian noise, 2 for correlated (non-Gaussian) low noise, and 5 for correlated (non-Gaussian) high noise. As a rule of thumb, for correlated noise, α is set to a value greater than 1. This adjustment compensates for the correlations not accounted for in the $\underset{\sim}{\boldsymbol{R}}$ matrix. The higher the correlation between observations, the more α should be increased.

Otherwise, since the rank of $(\boldsymbol{I} - \mathbb{1}/p)$ is $p-1$, the kernel matrix $\boldsymbol{K}$ is not invertible. Thus, its Moore-Penrose pseudo-inverse must be used. Given that $(\boldsymbol{I} - \mathbb{1}/p)$ is a projection matrix, it equals its own pseudo-inverse, thus the Moore-Penrose inverse of $\boldsymbol{K}$ is given by:

$$\boldsymbol{K}^{\dagger} = \frac{1}{\alpha \cdot (p-1)}\,(\boldsymbol{I} - \mathbb{1}/p)\,\boldsymbol{K}_G^{-1}\,(\boldsymbol{I} - \mathbb{1}/p)\,. \tag{57}$$

The square root of the kernel matrix pseudo-inverse is typically computed through SVD decomposition. The kernel pseudo-inverse and its square root are used to compute the gain (38) and the covariance square root (42). The centering of the kernel ensures consistency with classical square root filtering methods. The kernel used in section 3.3 to demonstrate the connection between the ESRF and the RKHS EnF is a specific case of the kernel presented here, with l_G tending towards 0.

4.5 Diagnostics

To assess the reconstruction quality of each methods, we consider the reconstructed relative vorticity fields. Relative vorticity is a less smooth variable compared to velocity or water height, making it a more effective discriminator between the different reconstruction methods. Three metrics are considered to evaluate the effectiveness of the methods. In this subsection the vorticity of the ith ensemble member to be evaluated is denoted as $\boldsymbol{\omega}^{(i)}$ and the reference is denoted as $\boldsymbol{\omega}_{\text{ref}}$.
The root mean square error (RMSE) is a widely used metric for quantifying the discrepancy between the ensemble members and the reference state. It is defined as:

$$e_{\text{RMSE}} = \sqrt{\frac{1}{p}\sum_{i=1}^{p} \left\|\boldsymbol{\omega}^{(i)} - \boldsymbol{\omega}_{\text{ref}}\right\|_{\Omega_x}^2}\ . \tag{58}$$

It penalizes the deviation of each ensemble member from the reference, providing a measure of the overall accuracy of the ensemble. To improve interpretability, the RMSE can be expressed as a percentage by normalizing it with respect to the norm of the reference state, as follows:

$$e_{\text{NRMSE}} = \frac{e_{\text{RMSE}}}{\|\boldsymbol{\omega}_{\text{ref}}\|_{\Omega_x}}. \tag{59}$$

This normalized RMSE offers a more intuitive understanding of the ensemble's mean performance. It is important to note that this measure reflects the mean accuracy of the individual ensemble members, rather than the accuracy of the ensemble mean, which is captured by the bias. The bias is defined as the distance between the reference and the ensemble mean estimation:

$$\boldsymbol{\omega}_{\text{bias}} = \overline{\boldsymbol{\omega}} - \boldsymbol{\omega}_{\text{ref}}\ . \tag{60}$$

By computing the norm of the bias vorticity field over the domain Ω_x, a scalar value can be obtained to represent the bias error across the entire spatial domain:

$$e_{\text{bias}} = \|\boldsymbol{\omega}_{\text{bias}}\|_{\Omega_x}\ . \tag{61}$$

This indicator represent the accuracy of the ensemble, the lower it is the more accurate is the estimation. The standard deviation is expressed as

$$\sigma_e = \sqrt{\frac{1}{p}\sum_{i=1}^{p} \left\|\boldsymbol{\omega}^{(i)} - \overline{\boldsymbol{\omega}}\right\|_{\Omega_x}^2} \,. \tag{62}$$

The standard deviation represents the dispersion of the ensemble. A small standard deviation indicates that the ensemble members are close to each others. The three metrics presented here are interconnected, considering the following inequality for $1 \leq i \leq p$,

$$\left\|\boldsymbol{\omega}^{(i)} - \boldsymbol{\omega}_{\text{ref}}\right\|_{\Omega_x}^2 \leq \left\|\boldsymbol{\omega}^{(i)} - \overline{\boldsymbol{\omega}}\right\|_{\Omega_x}^2 + \left\|\overline{\boldsymbol{\omega}} - \boldsymbol{\omega}_{\text{ref}}\right\|_{\Omega_x}^2 , \tag{63}$$

which gives with an empirical mean

$$e_{\text{RMSE}}^2 \leq \sigma_e^2 + e_{\text{bias}}^2 \,. \tag{64}$$

A low standard deviation may indicate a good result if the bias is also low, meaning that all ensemble members are close to the reference, resulting in a low RMSE. However, if the ensemble spread is too small, there is a risk that the reference may fall outside the ensemble's range, making it difficult for the method to adjust the ensemble members to track the reference in the future. This can lead to divergence over time. Therefore, a larger standard deviation within the ensemble can contribute to more robust predictions.

Striking a balance between standard deviation and bias, known as the bias-variance trade-off, is crucial. A standard deviation that is too low can lead to divergence, while one that is too high may indicate that individual ensemble members poorly represent the system. A wide spread can also impact the bias, as overly dispersed ensemble members may produce an over-smoothed mean, failing to capture small-scale structures accurately.

5 Results

In this section, we compare the performance of the data assimilation methods presented in this study over 100 days. Let us mention that the state space dimension is large, while the number of ensemble members is limited, making this scenario particularly challenging for ensemble methods. It is well known that in such cases, localization techniques are recommended to prevent filter divergence. However, we will not consider them here, as the focus is on the efficiency and robustness provided by the RKHS embedding, as well as its capacity to handle a 4D VAR problem at low cost using the tiled simulation strategy. While extending the proposed methods with localization techniques is possible, it lies beyond the scope of this paper and presents a clear avenue for future research.

As an initial comparison, Figure 3 shows the NRMSE for both the classical ESRF and the RKHS EnF. The NRMSE is averaged over five different reference datasets, each with its own initial ensemble, generated as described in Section 4.2. The shaded areas around the mean NRMSE curves represent the standard deviation across these experiments. A wide shaded area indicates high sensitivity of the results to both the initial reference (from which the observations are extracted) and the initial ensemble, while a narrow band highlights low variability in performance, suggesting greater robustness.

In the experiments with 1% Gaussian observation noise, the classical ESRF shows a constant increase in error and significant variability, indicating a lack of robustness. The RKHS EnF performs similarly, though with slightly lower variability. In contrast, the tiled RKHS EnF demonstrates remarkable long-term stability, maintaining consistent performance along time.

For higher, decorrelated noise (10% of the reference height field RMS), all methods experienced an increase in error. However, the tiled version of the RKHS EnF outperformed

the ESRF, achieving lower NRMSE with smaller standard deviation across different experiments, demonstrating greater resilience to high noise levels. The ESRF performs better with uncorrelated noise compared to the RKHS EnF, but similarly to the lower noise level, it exhibits slightly higher robustness in terms of a tighter standard deviation.

For 1% spatially correlated non-Gaussian noise, the tiled RKHS EnF significantly outperforms both ESRF and RKHS EnF, maintaining consistently low error and much smaller standard deviation. This highlights its strong robustness against misrepresentation of the observation covariance, which was assumed diagonal and hence representative of independent, uncorrelated spatial noise. ESRF and RKHS EnF perform similarly in terms of NRMSE average and standard deviation for this noise type.

With higher spatially correlated noise (10%), all methods see a more substantial increase in NRMSE. However, both the RKHS EnF and its tiled version significantly outperform the ESRF, showing greater robustness (with a tighter standard deviation) and maintaining a constant NRMSE at the noise level up to day 50. In contrast, the ESRF shows increasing error from around day 15. This experiment reveals a clear separation in the filters' performance. Theoretically, ESRF is not well-suited to nonlinear, non-Gaussian systems, whereas the RKHS EnF appears better equipped to handle non-Gaussianity and correlated noise.

As noted earlier, the performance of these filters could likely be improved with the use of localization techniques. For the RKHS EnF, localization could be achieved by localizing the observables and with the consequence of introducing localized stationary weights (i.e stationnary spatial vectors of weights) instead of global weights.

These experiments demonstrate that the tiled version of the RKHS EnF offers clear improvements. This approach consistently produces better results, showing less sensitivity to initial conditions (*i.e.*, the initial ensemble) by showcasing a tighter standard deviation for all noise levels. Additionally, the RKHS EnF is notably less sensitive to errors in the noise covariance modelling. The tiled RKHS EnF benefits from the theoretical properties of embedding the dynamical system in the dynamically evolving RKHS family, allowing it to formulate a 4D-Var assimilation as a single-iteration filtering process, similar to sequential Bayesian filters.

Figure 4 presents the analyzed ensemble mean of the upper-layer relative vorticity after 100 days, using observations perturbed by the 10% non-Gaussian, spatially correlated noise. This figure shows that all assimilation methods successfully reconstruct a vorticity map close to the reference vorticity field. At naked eye the estimations are almost indistinguishable from the reference. The enstrophy spectra from the upper layer, calculated within the rectangular region bounded by -72.25°E to -42.28°E longitude and 14.64°N to 39.47°N latitude, following the method described by Durran et al. (2017) are shown in Figure 5. This figure reveals that the enstrophy spectra for the three methods are nearly identical to the reference one, indicating accurate reproduction of the reference field energy across all scales. Enstrophy spectra for the other layers lead to the same conclusion (not shown). Importantly, the effective resolution of the reference data has not been compromised. This suggests that employing ensemble methods, without resorting to optimal interpolations or gridded data prior to assimilation, maintains the resolution integrity.

However, a clearer understanding of the differences between the methods is provided by the bias fields, which represent the discrepancies between the reference and analyzed fields. Figures 6, figure 7, and 8 illustrate the relative vorticity bias fields for the three assimilation methods across the upper, middle, and lower layers, respectively. Notably, the middle and lower layers are better reconstructed by all methods compared to the upper layer, likely due to the reduced variability and the fewer small-scale structures in these layers. To improve clarity, the color scale for the bias fields in these two layers has been refined. An analysis of the bias fields reveals similar trends across all three layers. The classical ESRF

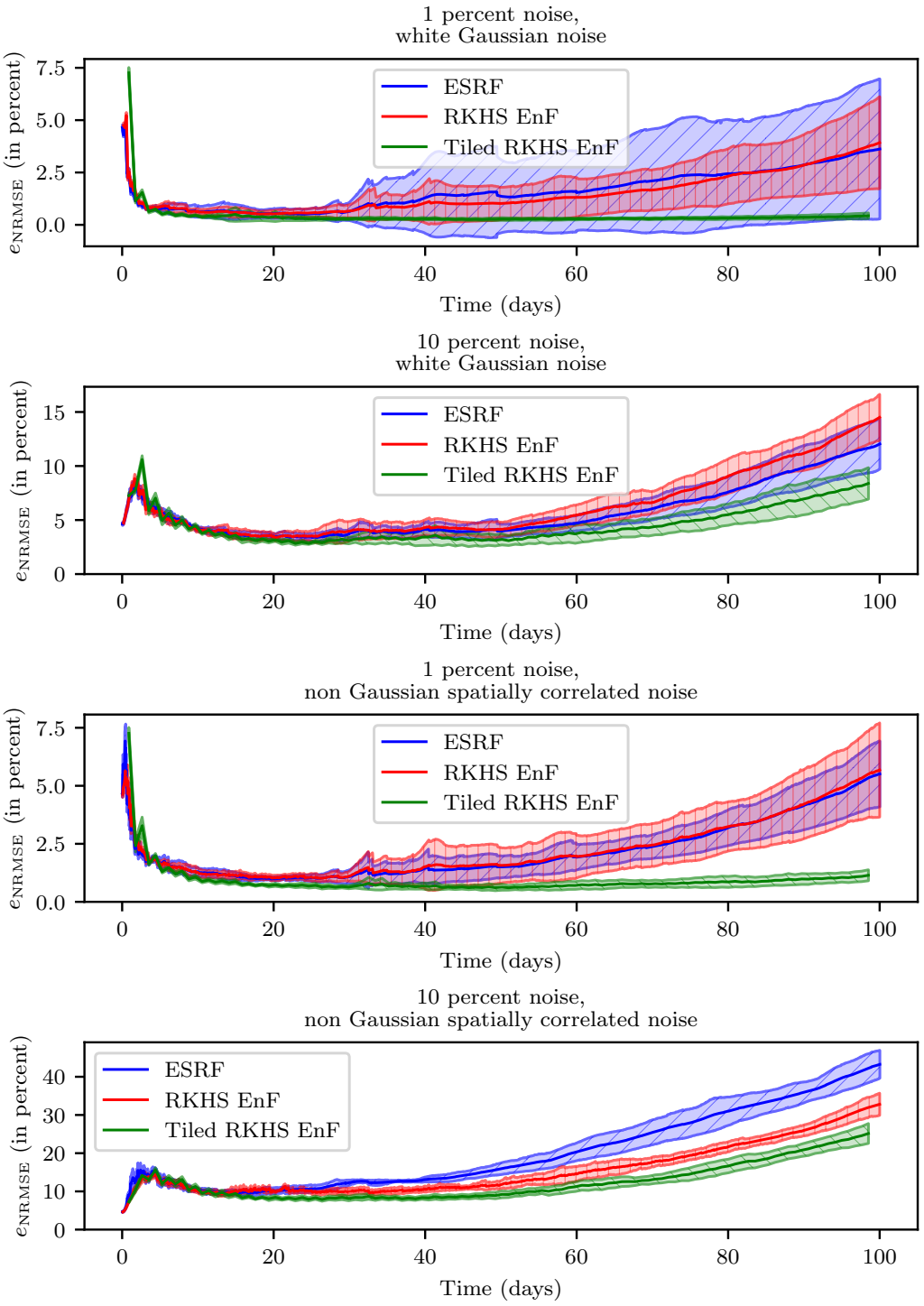

Figure 3: Comparison of the Normalized Root Mean Square Error (NRMSE) for the ESRF (blue), RKHS EnF (red), and Tiled RKHS EnF (green). The first two rows display results for experiments with additive white Gaussian noise, while the last two rows present results for experiments with non-Gaussian, spatially correlated noise. Rows one and three correspond to experiments with 1% standard deviation noise, and rows two and four to experiments with 10% standard deviation noise.

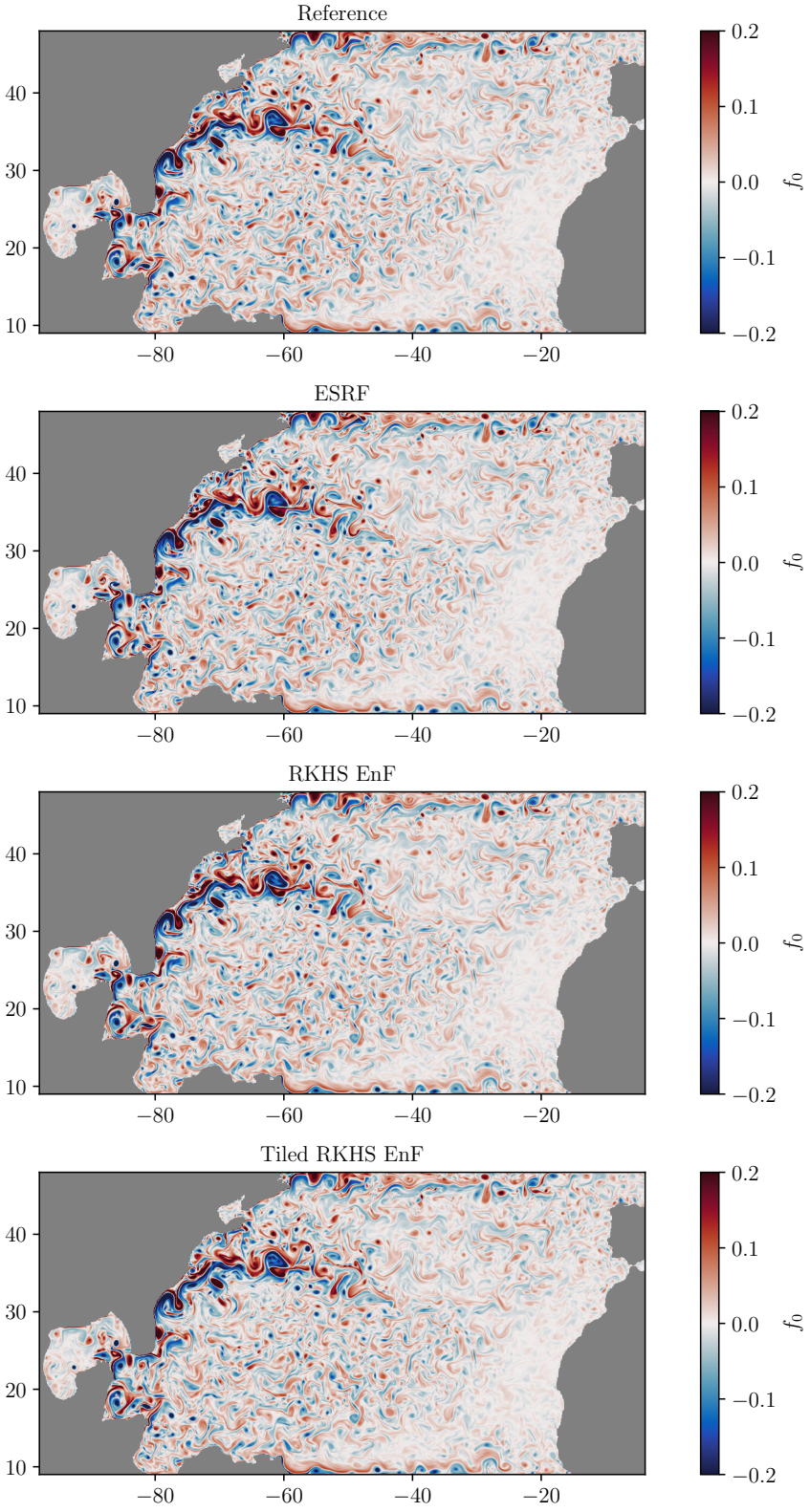

Figure 4: Reference upper layer relative vorticity field (top row) after 100 days, analyzed fields for the ESRF (second row), RKHS EnF (third row), and tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

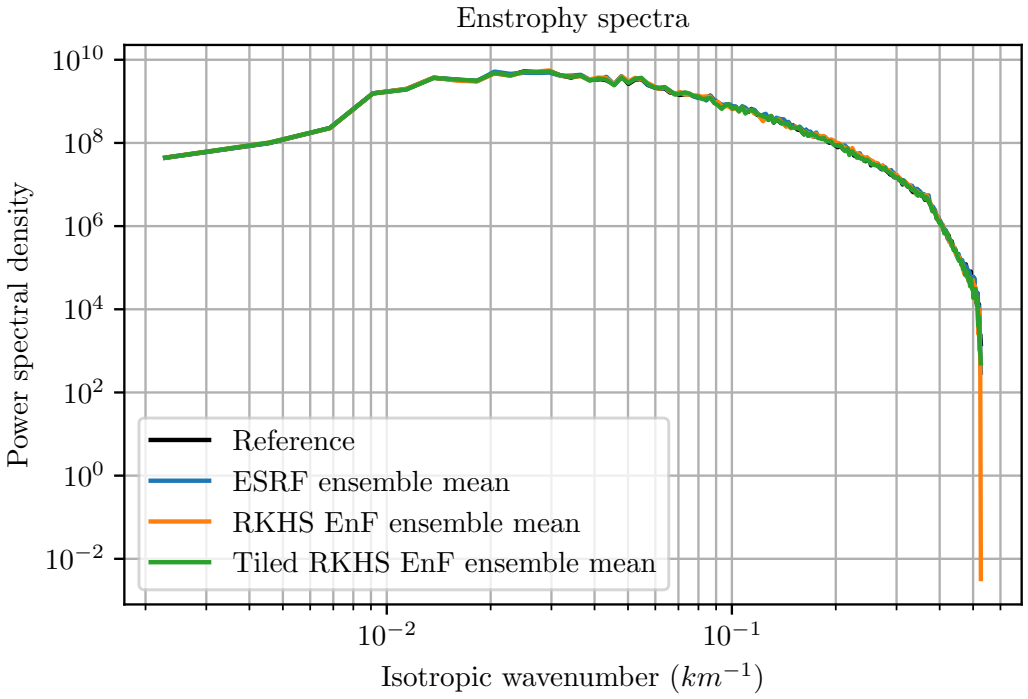

Figure 5: Upper layer enstrophy spectra after 100 days within the region defined by -72.25°E to -42.28°E longitude and 14.64°N to 39.47°N latitude, presented as a function of wavenumber. The reference spectra (black) is compared to the results from the ESRF (blue), RKHS EnF (orange), and tiled RKHS EnF (green), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

exhibits significant bias, particularly along the Gulf Stream, indicating poorer performance in capturing the dynamics of this region. In contrast, the RKHS EnF demonstrates improved results, with lower bias compared to the ESRF. The tiled RKHS EnF outperforms both methods, producing substantially lower biases. This suggests that the tiled method assimilates the data more effectively, leading to more accurate reconstructions. The study of the bias fields aligns with the conclusions drawn from the RMSE analysis done in the previous paragraph. This consistency is expected, given the inherent connection between these diagnostic variables.

To further quantify the errors, Figure 9 presents the enstrophy bias spectra for the upper layer. The spectra indicate that most errors occur at wavenumbers greater than 10^{-2}, corresponding to structures smaller than 600 km. The bias error energy of the RKHS EnF is slightly lower than that of the ESRF, while the tiled RKHS EnF exhibits significantly reduced error energy. Notably, the improvement is substantial across all scales, from large-scale features to the eddy-diffusion cut-off, with the greatest enhancement around 300 km. This highlights the superior performance of the tiled method in capturing both large-scale and small-scale features in the vorticity field.

To finally complement these analysis results, we present in Figure 10 the RMSE obtained for the upper layer. As observed, the ESRF exhibits the highest RMSE, while the RKHS EnF shows a lower RMSE, and the Tiled RKHS EnF further improves upon it. The other layers, which follow the same trend of lower RMSE, are not shown. The spread (standard deviation) of the ensemble for the upper layer is also displayed in Figure 11. We observe that the spread is attenuated for the ESRF, suggesting a collapse of the ensemble. Both the RKHS EnF and Tiled RKHS EnF display some variation in spread. While the spread is weak for the Tiled RKHS EnF, the bias is also minimal, indicating a less problematic situation compared to the ESRF.

6 Conclusion and perspectives

This study introduces a novel approach for assimilating observations of a non-linear dynamical system by using a family $\mathcal{W}$ of reproducing kernel Hilbert spaces (RKHS). Each RKHS in this family is associated with kernels k_t, which are defined by transporting an initial kernel k_0 along the system's dynamics up to time t. This framework establishes an isometry between the evolving RKHS, fully justifying the superposition principle, allowing new system trajectories to be constructed as linear combinations of known ensemble members trajectories. Importantly, the isometry ensures these linear combinations remain invariant over time.

An ensemble filter is developed within this RKHS family. The theoretical properties of this framework naturally enable the assimilation of time series of observations, effectively accumulating data at different time instants before analyzing the entire trajectory within the assimilation window. This leads to more robust outcomes. By leveraging the framework's properties, the analyzed ensemble can be provided at the start of the assimilation window and propagated forward. This tiled version of the filter has demonstrated superior performance in reconstructing physical structures and improving accuracy in estimating the reference state.

The methods presented in this study have been applied to a quasi-geostrophic model of the North Atlantic Ocean and compared to a classical square root filter (ESRF), which can be viewed as a limit of the sequential version of the RKHS-based ensemble filter.

Improvements to the RKHS EnF could be made by exploring methods to optimize the choice of kernels. One such approach, as proposed by Akian et al. (2022), involves an algorithm that adjusts the kernel eigenvalues to minimize accuracy loss when data are removed from a training set. Investigating alternative kernels beyond the Gaussian could also be beneficial. For example, as noted by Stein (1999), the Gaussian kernel may be too

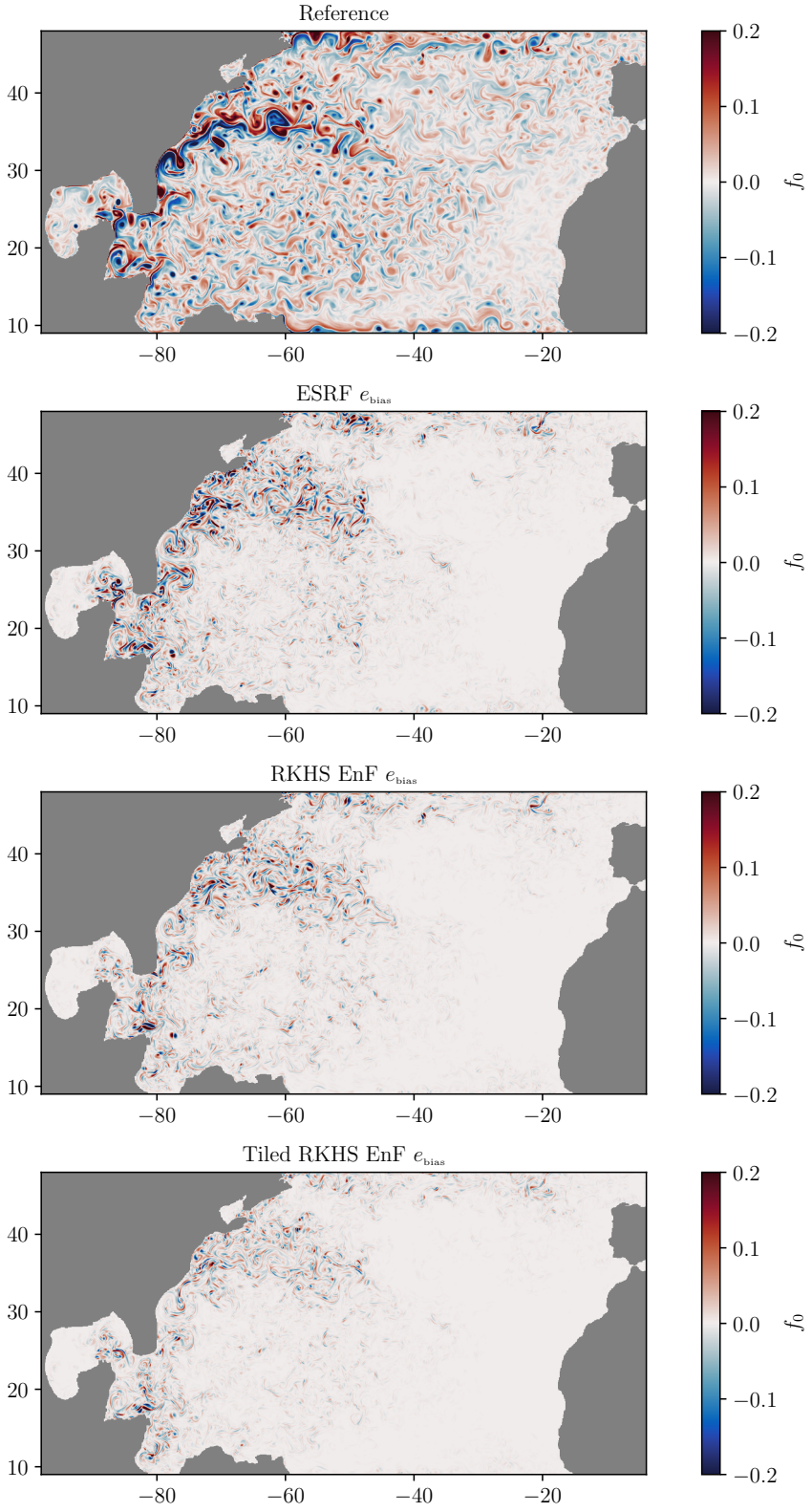

Figure 6: Reference upper layer relative vorticity field (top row) after 100 days, bias field corresponding to it for the ESRF (second row), RKHS EnF (third row), tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

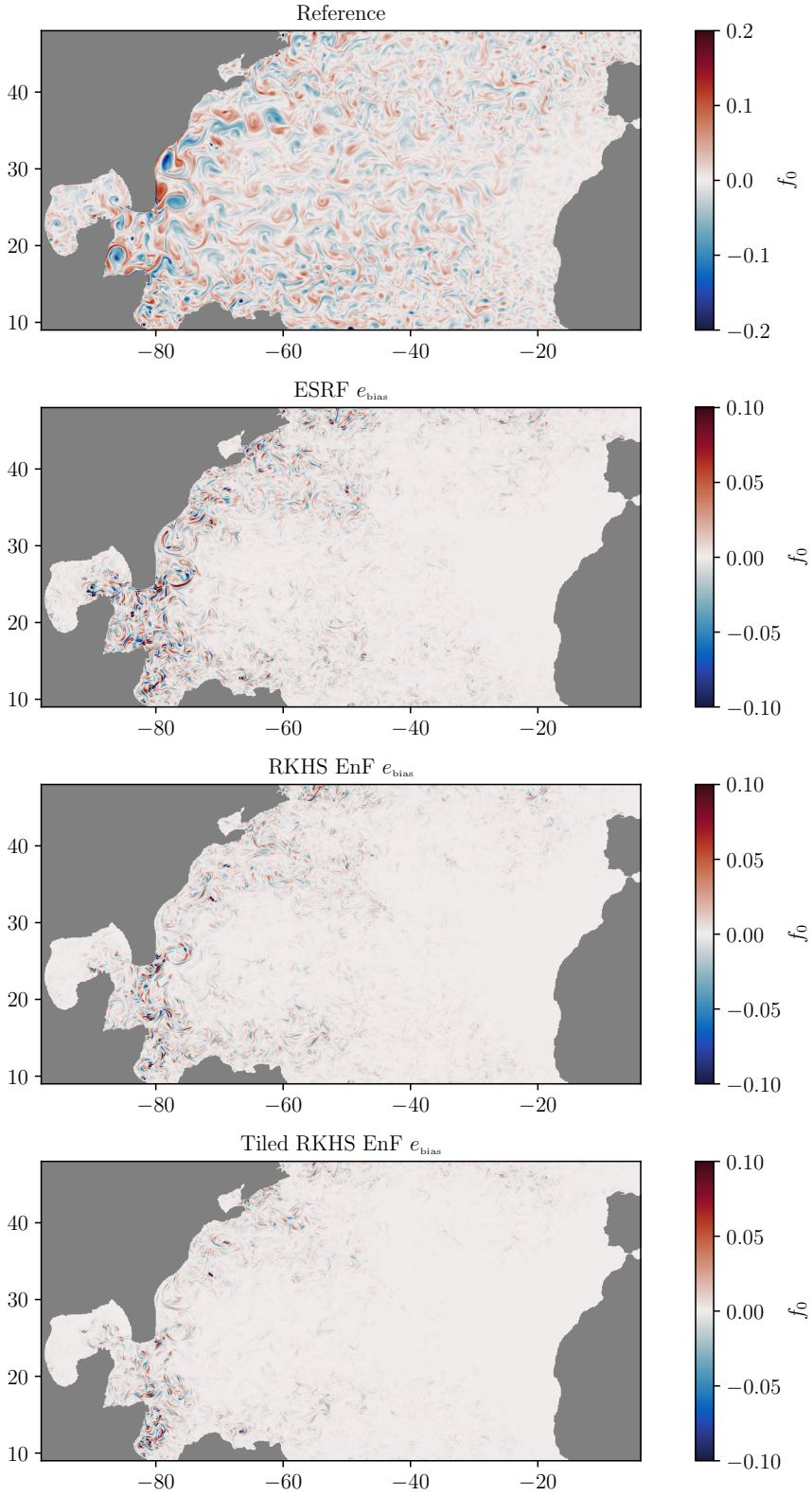

Figure 7: Reference middle layer realtive vorticity field (top row) after 100 days, bias field corresponding to it for the ESRF (second row), RKHS EnF (third row), tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

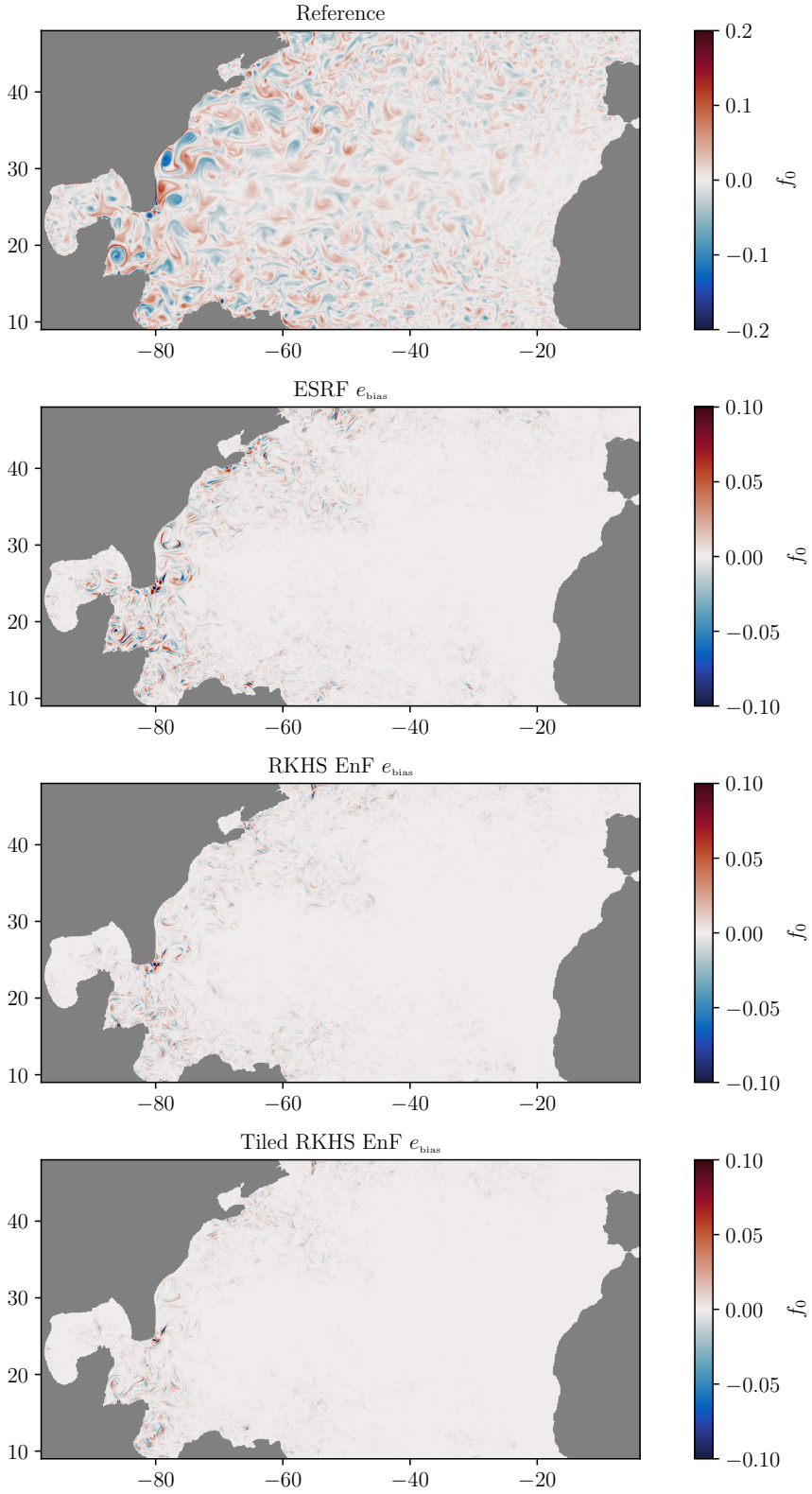

Figure 8: Reference lower layer realtive vorticity field (top row) after 100 days, bias field corresponding to it for the ESRF (second row), RKHS EnF (third row), tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

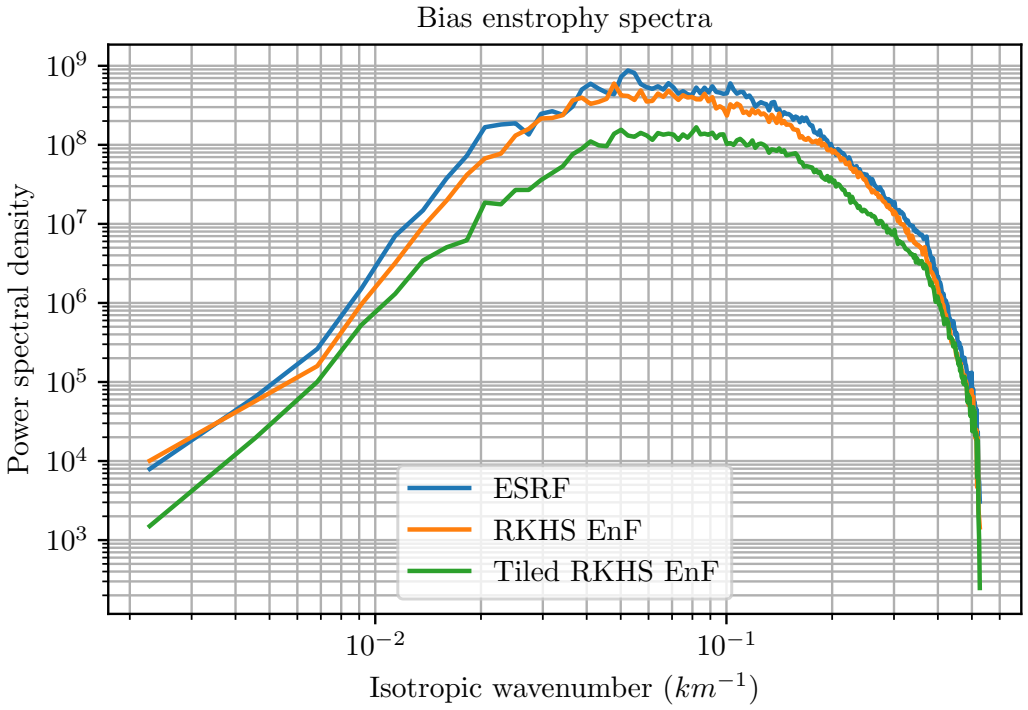

Figure 9: Upper layer enstrophy bias spectra after 100 days within the region bounded by -72.25°E to -42.28°E longitude and 14.64°N to 39.47°N latitude, plotted as a function of wavenumber. Results are shown for the ESRF (blue), EnF on an RKHS family (orange), and tiled EnF on an RKHS family (green), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

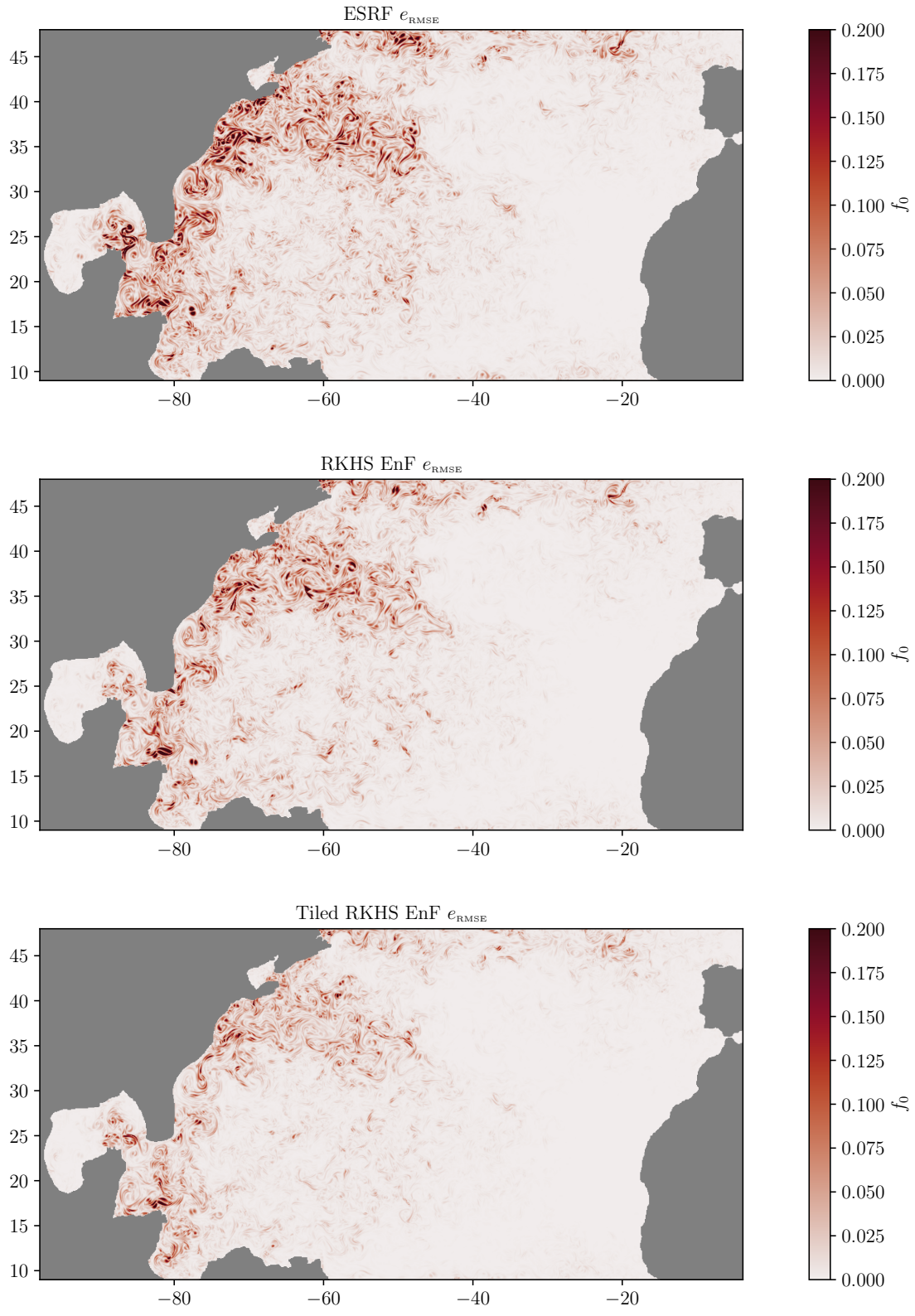

Figure 10: Upper layer relative vorticity RMSE field after 100 days for the ESRF (top row), RKHS EnF (second row), tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

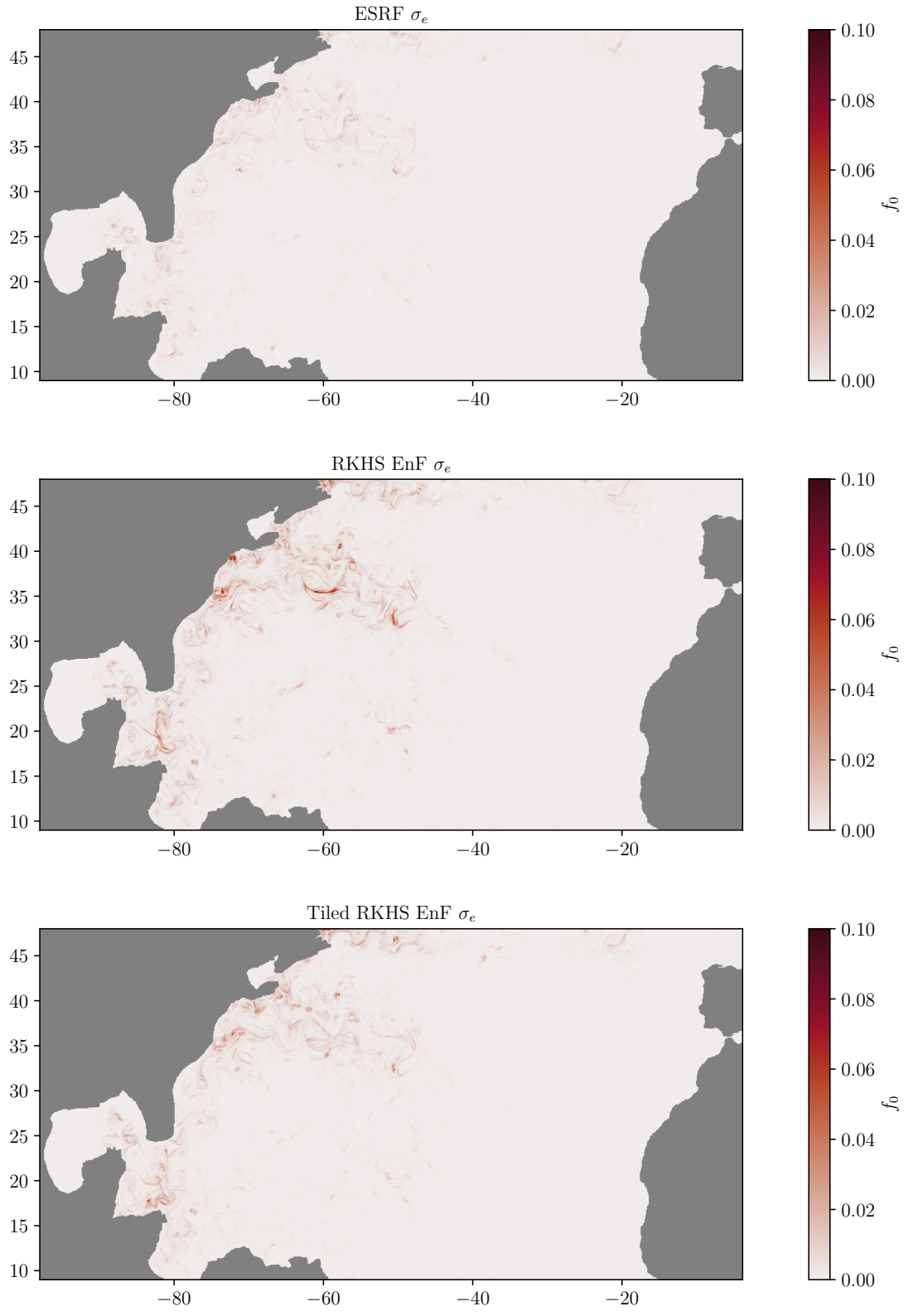

Figure 11: Upper layer relative vorticity standard deviation field after 100 days for the ESRF (top row), RKHS EnF (second row), tiled RKHS EnF (bottom row), all corresponding to the experiment with 10% standard deviation spatially correlated noise.

smooth, whereas the Matérn kernel offers greater flexibility in terms of smoothness while maintaining a reasonable number of parameters to configure.

One significant obstacle to improving state reconstructions in ensemble-based filtering methods is the inherent limited number of ensemble members that can be computationally afforded. Localization methods, as discussed by Sakov and Bertino (2010), help address this limitation. Localization is introduced either by localizing the observation in the assimilation process or by localizing the covariance matrix. This second approach seems more difficult to implement with the RKHS EnF, which operates directly in the weight space. The first approach, which result in local filter is more appealing and doable with the introduction of localized observable and considering spatial vector of stationary weights. In the same spirit Bocquet (2016) introduces a localization strategy that reconstructs trajectories through linear combinations, where the localization operator is propagated over time. This class of methods could potentially be naturally adapted within the time-evolving RKHS framework to introduce dynamical localization process.

In addition let us mention that, applying a tiled filter with local analysis could mitigate imbalances from localized approximate reconstructions.

Another perspective is that the framework allows direct access to the system's Lyapunov times (Dufée et al., 2024), which indicate when the predictability horizon has been exceeded. This information is valuable for determining the optimal moments for assimilation, facilitating the renewal of the ensemble and creating an adaptive assimilation step filter.

Additionally, the initial ensemble significantly impacts the method's performance, making it crucial to select a well-chosen starting set. Generative methods can be employed to establish an initial ensemble that best represents the reference simulation's initial conditions, thereby enhancing the filtering method's effectiveness.

In this study, the reference dynamics were modeled using the same model and resolution as the training set elements. However, modeling a reference that differs from the training set ensemble poses challenges. A different resolution can lead to variations in energy levels between the reference and training ensemble members; for instance, higher-resolution reference models generally exhibit more energy than their coarser counterparts. To address this, variability can be introduced into the training ensemble through methods such as inflation or by employing a stochastic model with calibrated noise (Dufée et al., 2022).

Moreover, to model the reference differently from the training set, the quasi-geostrophic projection can be omitted in the reference model run. This approach would enable the reference to capture fast waves, which would not be reconstructed by the training ensemble. Incorporating multi-modal data could further enhance the assimilation process. For example, combining sea surface height data with sea surface temperature and integrating data from the Surface Water Ocean Topography (SWOT) mission alongside traditional altimeter data could improve the assimilation of real data, as discussed by (Le Guillou, Lahaye, et al., 2021).

Open Research Section

The data and the code files necessary to reproduce the results presented in this paper are publicly available at: https://gitlab.inria.fr/mael.jaouen/RKHS_EnF/

Acknowledgments
We acknowledge the support of the ERC EU project 856408- STUOD, the French National program LEFE (Les Enveloppes Fluides et l'Environnement), as well as the support of Inria, which enabled the research presented in this paper through Inria exploratory actions.

The authors acknowledge the Pôle de Calcul et de Données Marines (PCDM) for providing DATARMOR computational resources. This study has been conducted using E.U. Copernicus Marine Service Information.

References

Akian, J.-L., Bonnet, L., Owhadi, H., & Savin, E. (2022). Learning "best" kernels from data in gaussian process regression. with application to aerodynamics. *Journal of Computational Physics*, *470*, 111595.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, *68*(3), 337–404.

Auroux, D., & Blum, J. (2008). A nudging-based data assimilation method: the back and forth nudging (bfn) algorithm. *Nonlinear Processes in Geophysics*, *15*(2), 305–319.

Bauer, W., Chandramouli, P., Chapron, B., Li, L., & Mémin, E. (2020). Deciphering the role of small-scale inhomogeneity on geophysical flow structuration: A stochastic approach. *Journal of Physical Oceanography*, *50*(4), 983 - 1003.

Bishop, A. N., & Del Moral, P. (2023). On the mathematical theory of ensemble (linear-gaussian) Kalman–bucy filtering. *Mathematics of Control, Signals, and Systems*, *35*(4), 835–903.

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. part I: Theoretical aspects. *Monthly Weather Review*, *129*(3), 420 - 436.

Bocquet, M. (2016). Localization and the iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 1075-1089.

Bocquet, M., & Sakov, P. (2014, July). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1521-1535.

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. *Journal of Computational Science*, *44*, 101171.

Brecht, R., Li, L., Bauer, W., & Mémin, E. (2021). Rotating shallow water flow under location uncertainty with a structure-preserving discretization. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002492.

Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Climate Change*, *9*.

Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., ... Arcucci, R. (2023). Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review. *IEEE/CAA Journal of Automatica Sinica*, *10*(6), 1361-1387.

Desroziers, G., Camino, J.-T., & Berre, L. (2014). 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, *140*(684), 2097-2110.

Doucet, A., Freitas, N. D., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer Verlag.

Dufée, B., Hug, B., Mémin, E., & Tissot, G. (2024). Ensemble forecasts in reproducing kernel Hilbert space family. *Physica D: Nonlinear Phenomena*, *459*, 134044.

Dufée, B., Mémin, E., & Crisan, D. (2022). Stochastic parametrization: An alternative to inflation in ensemble Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1075-1091.

Durran, D., Weyn, J. A., & Menchaca, M. Q. (2017). Practical considerations for computing dimensional spectra from gridded data. *Monthly Weather Review*, *145*(9), 3901 - 3910.

Evensen, G. (2009). *Data assimilation, the ensemble Kalman filter*. Springer Berlin, Heidelberg.

Fablet, R., Chapron, B., Drumetz, L., Mémin, E., Pannekoucke, O., & Rousseau, F. (2021). Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002572. (e2021MS002572 2021MS002572)

Fablet, R., Chapron, B., Le Sommer, J., & Sévellec, F. (2024). Inversion of sea surface currents from satellite-derived sst-ssh synergies with 4DVarNets. *Journal of Advances in Modeling Earth Systems*, *16*(6), e2023MS003609. (e2023MS003609 2023MS003609)

Farchi, A., Laloyaux, P., Bonavita, M., & Bocquet, M. (2021). Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3067-3084.

Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, *82*(397), 249–266.

Gordon, N., Salmond, D., & Smith, A. (1993, April). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, *140*, 107-113(6).

Gottwald, G. A., & Reich, S. (2021). Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *Physica D: Nonlinear Phenomena*, *423*, 132911.

Hoke, J. E., & Anthes, R. A. (1976). The initialization of numerical models by a dynamic-initialization technique. *Monthly Weather Review*, *104*(12), 1551 - 1556.

Hunt, B. R., Kalnay, E., Kostelich, E., Ott, E., Patil, D., Sauer, T., ... Zimin, A. (2004). Four-dimensional ensemble Kalman filtering. *Tellus A: Dynamic Meteorology and Oceanography*, *56*(4), 273–277.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1), 112-126. (Data Assimilation)

Koivunen, A. C., & Kostinski, A. B. (1999). The feasibility of data whitening to improve performance of weather radar. *Journal of Applied Meteorology*, *38*(6), 741 - 749.

Le Gland, F., Monbet, V., & Tran, V. (2011). Large sample asymptotics for the ensemble Kalman filter. In D. Crisan & B. Rozovskii (Eds.), *Handbook on nonlinear filtering.* Oxford University Press.

Le Dimet, F.-X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, *38A*(2), 97-110.

Leeuwenburgh, O., Evensen, G., & Bertino, L. (2005). The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical pacific. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3291-3300.

Le Guillou, F., Lahaye, N., Ubelmann, C., Metref, S., Cosme, E., Ponte, A., ... Vidard, A. (2021). Joint estimation of balanced motions and internal tides from future wide-swath altimetry. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002613. (e2021MS002613 2021MS002613)

Le Guillou, F., Metref, S., Cosme, E., Ubelmann, C., Ballarotta, M., Le Sommer, J., & Verron, J. (2021). Mapping altimetry in the forthcoming swot era by back-and-forth nudging a one-layer quasigeostrophic model. *Journal of Atmospheric and Oceanic Technology*, *38*(4), 697 - 710.

Le Traon, P. Y., Faugère, Y., Hernandez, F., Dorandeu, J., Mertz, F., & Ablain, M. (2003). Can we merge GEOSAT follow-on with TOPEX/Poseidon and ERS-2 for an improved description of the ocean circulation? *Journal of Atmospheric and Oceanic Technology*, *20*(6).

Liu, C., Xiao, Q., & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. part I: Technical formulation and preliminary test. *Monthly Weather Review*, *136*(9), 3363 - 3373.

Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, *129*(595), 3183-3203.

Luo, X. (2019, July). Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators. *PLOS ONE*, *14*(7), 1–40.

Manucharyan, G. E., Siegelman, L., & Klein, P. (2021). A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence. *Journal of Advances in Modeling Earth Systems*, *13*(1), e2019MS001965. (e2019MS001965 2019MS001965)

Mauran, S., Mouysset, S., Simon, E., & Bertino, L. (2023). A kernel extension of the ensemble transform Kalman filter. In J. Mikyška, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, & P. M. Sloot (Eds.), *Computational science – iccs 2023* (pp. 438–452). Cham: Springer Nature Switzerland.

Ouala, S., Fablet, R., Herzet, C., Chapron, B., Pascual, A., Collard, F., & Gaultier, L. (2018). Neural network based Kalman filters for the spatio-temporal interpolation of satellite-derived sea surface temperature. *Remote Sensing*, *10*(12).

Reich, S., & Cotter, C. (2015). *Probabilistic forecasting and bayesian data assimilation.* Cambridge University Press.

Roullet, G., & Gaillard, T. (2022). A fast monotone discretization of the rotating shallow water equations. *Journal of Advances in Modeling Earth Systems*, *14*(2), e2021MS002663. (e2021MS002663 2021MS002663)

Sakov, P., & Bertino, L. (2010). Relation between two common localisation methods for the enkf comput. *Computational Geosciences*, *15*, 225-237.

Stein, M. L. (1999). *Interpolation of spatial data.* Springer New York.

Taburet, G., Sanchez-Roman, A., Ballarotta, M., Pujol, M.-I., Legeais, J.-F., Fournier, F., ... Dibarboure, G. (2019). Duacs dt2018: 25 years of reprocessed sea level altimetry products. *Ocean Science*, *15*(5), 1207–1224.

Thiry, L., Li, L., Mémin, E., & Roullet, G. (2024). A unified formulation of quasi-geostrophic and shallow water equations via projection. *Journal of Advances in Modeling Earth Systems*, *16*(10), e2024MS004510.

Thiry, L., Li, L., Roullet, G., & Mémin, E. (2023). Mqgeometry-1.0: a multi-layer quasi-geostrophic solver on non-rectangular geometries. *EGUsphere*, *2023*, 1–25.

Ubelmann, C., Dibarboure, G., Gaultier, L., Ponte, A., Ardhuin, F., Ballarotta, M., & Faugère, Y. (2021). Reconstructing ocean surface current combining altimetry and future spaceborne doppler data. *Journal of Geophysical Research: Oceans*, *126*(3), e2020JC016560. (e2020JC016560 2020JC016560)

Ubelmann, C., Klein, P., & Fu, L.-L. (2015). Dynamic interpolation of sea surface height and potential applications for future high-resolution altimetry mapping. *Journal of Atmospheric and Oceanic Technology*, *32*(1).

Van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., & Reich, S. (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society*, *145*(723), 2335-2365.

Wan, L., Bertino, L., & Zhu, J. (2010). Assimilating altimetry data into a hycom model of the pacific: Ensemble optimal interpolation versus ensemble Kalman filter. *Journal of Atmospheric and Oceanic Technology*, *27*(4), 753 - 765.

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913 - 1924.

Yang, Y., Robinson, C., Heitz, D., & Mémin, E. (2015). Enhanced ensemble-based 4DVar scheme for data assimilation. *Computers & Fluids*, *115*, 201-210.

Zhang, Z., Archibald, R., & Bao, F. (2022). A pde-based adaptive kernel method for solving optimal filtering problems. *Journal of Machine Learning for Modeling and Computing*, *3*(3).