Stochastic parametrization: an alternative to inflation in Ensemble Kalman filters

Benjamin Dufée, Etienne Mémin, Dan Crisan

Stochastic parametrization: an alternative to inflation in Ensemble Kalman filters

Benjamin Dufée[1]*, Etienne Mémin[1], and Dan Crisan[2]
[1]Inria/Irmar, Fluminance, Campus universitaire de Beaulieu, Rennes Cedex 35042, France
[2]Department of Mathematics, Imperial College, London, SW7 2AZ, UK

February 9, 2022

Abstract

We investigate the application of a stochastic dynamical model in ensemble Kalman filter methods. Ensemble Kalman filters are very popular in data assimilation because of their ability to handle the filtering of high-dimensional systems with reasonably small ensembles (especially when they are accompanied with so called localization techniques). The stochastic framework presented here relies on Location Uncertainty (LU) principles which model the effects of the model errors on the large-scale flow components. The experiments carried out on the Surface Quasi Geostrophic (SQG) model with the localized square root filter demonstrate two significant improvements compared to the deterministic framework. Firstly, as the uncertainty is a priori built into the model through the stochastic parametrization, there is no need for ad-hoc variance inflation or perturbation of the initial condition. Secondly, it yields better MSE results than the deterministic ones.

1 Introduction

Data assimilation uses observational data to correct the predictions of a numerical model, to get closer to the "true" state of the system. It has been applied for a long time in weather forecasting and other geophysical dynamical systems. Numerous approches from different methodological settings have been proposed to that end. They rely on optimal control methodology, functional approximation theory, stochastic processes and filtering, or geostatistics to cite only a few of them (see [15] for an overview of the techniques and recent trends in data assimilation). Among the different techniques proposed, a very popular data assimilation method is the ensemble Kalman filter (EnKF), initially devised by [23], and since then explored in many different variations (see the books

*Corresponding author, email benjamin.dufee@inria.fr

[5, 24, 46] and references therein for a precise description of all these different variations).

Ensemble methods, formulated either in variational or Bayesian forms are based on constructing an ensemble of possible realizations of a given dynamical system, with no need to explicitly compute its tangent linear model as it is required in optimal control methods [37]. This ability is very convenient in high-dimensional and complex models, as it enables in some way to free the data assimilation technique from the very likely evolution of the code with time, in terms of update, additional modules or features, requiring a new computation of the tangent linear model at each update. On the contrary, adjoint based optimization is fully attached to a given version of the numerical code of the model. For a routine use in meteorological centers, the separation between data assimilation and the numerical codes of the dynamics constitutes an undeniable advantage. As such these methods can cope, through an extension of the linear superposition principle, with non-linear models, even if in itself, the Kalman equations are theoretically set in linear and Gaussian hypothesis. Among the different versions of the proposed ensemble Kalman filter, the so called square-root filters (SRF) consist in formulating the square root of the forecast and analysis covariance matrices in terms of the ensemble members' anomaly matrix, obtained from the residual of each member with the empirical mean [2, 9, 43, 54, 55]. The non uniqueness of the square-root covariance has lead to several variations of the method. The ensemble transform Kalman filter (ETKF) proposed in [9] is a popular and efficient technique based on this principle. It aims at describing the posterior ensemble members as linear combinations of the prior ones in an optimal way. Indeed, square root schemes written in terms of the anomaly matrix are particular cases of ETKF.

Applying these methods to very high-dimensional geophysical systems (see [30, 31, 33, 32] in routine weather forecast applications) requires modest size ensembles (from tens to 100-200 at best). This is a true limitation of such ensemble Kalman filters as they might lead to strong sampling errors of the empirical covariance matrices involved in the filtering. These covariances have indeed a rank of at most $p-1$, with p denoting the number of ensemble members. In order to tackle this limitation and to extend their rank, so called localization approaches have been introduced [3, 4, 27, 52]. These techniques filter out spurious correlations between two far away points and are implemented either through a Schur product of the empirical covariance matrices with a finite support positive definite kernel or by localizing explicitly the area of influence of a set of observation on a given site of the spatial domain. Both methods, although formally very different, have been shown to lead in practice to equivalent results [52]. The latter techniques are computationally efficient as they lead to local Kalman filters, with the risk of producing a global solution that is not a solution of the partial differential equations driving the system. The Schur product conditioning of the empirical covariance is more secured from a formal point of view but cannot be used in conjunction with squared-root filtering techniques [52].

Ensemble Kalman filters are, by nature, stochastic filtering techniques. They

rely in one way or another on randomness as a third ingredient. The randomness considered is, in the vast majority of cases, related to the initial condition of the deterministic dynamics, with possibly another additive random variable representative of the error carried by the model itself. Such perturbations correspond to the errors attached to the physical or numerical approximation performed in the constitution steps of the system's dynamics. One of the main difficulty faced by the random dynamics built from perturbations of the initial condition is that they usually exhibit an insufficient spread of the ensemble members. This flaw leads to serious problems for data assimilation as all the ensemble members may quickly drift far away from the current observation. In high dimensional spaces, it cannot correct sufficiently the ensemble toward the current observation and therefore can yield to filter divergence. The introduction of additive random variables or inflating the ensemble variance by some factor is called variance inflation and is commonly used to tackle deficient variance representation. The problem associated with such a correction is that variance is gradually added to the system across time without being explicitly dissipated. This injection of energy leads eventually to a blowup of the dynamics simulation if it is not controlled. As a result, only very low (relatively to a given dynamics) factors of inflation can be considered in practice. Another problem, more specifically related to additive inflation, is that it corresponds to a model error variable that is blindly added to the system through an increase of variance. No particular spatial structure for the error is indeed injected. The last potential pitfall associated with such a technique is that the injection of an additive random forcing may completely change the attractor or the probability distribution of the underlying system's dynamics even for a noise of small amplitude [17], leading consequently to a quite different system from the one of interest. Still this is different from the multiplicative inflation mentioned before, which is the most commonly used in practice.

In this work, we explore an alternative to the randomization of the dynamics described above. Compared to the techniques mentioned previously, the error term in the dynamics is not added as an additional forcing but is, instead, part of the model. In a similar way to the Reynolds decomposition principle used in large eddies simulation setting, the stochastic formalism on which we rely, referred to as modeling under location uncertainty (LU) [40] in the following, consists in decomposing the dynamics in terms of a smooth in time component and a highly oscillating random component that is assumed uncorrelated at the characteristic time scale of the smooth component. A large-scale dynamics which takes into account the random components is then inferred from this decomposition and usual conservation laws shall hold for the stochastic model. This setting offers a stochastic parametrization, representing the impact of the small-scale flow on the global dynamics by a random small-scale velocity. This way to represent the small scales independently, and not only subordinate to the large scales, can be put in the broad context of stochastic physics. Still we only aim at addressing the subgrid scale effects, which only constitute a subset of all the "hidden physics" considered in techniques such as the stochastically perturbed parametrized tendency (SPPT) (see [14, 41, 42, 53] for general insight

on this). Uncertainty attached to hidden physics include among others deep convection in ocean, ocean-atmosphere interaction, cloud convection scheme, boundary layer turbulence etc. For a broad view of the different stochastic frameworks proposed in climate sciences, the reader may refer to several recent reviews on the subject [8, 25, 26, 41]

The LU framework entails deep changes in the classical formalism. Among them, ensemble members are now semi-martingales and the material derivative is replaced by its stochastic equivalent when it comes to deal with transport equations. LU has the advantage to be flow-driven and physically relevant as it truly encodes transport through the stochastic flow and relies on physical conservation laws. This stochastic parametrization has been shown to produce a higher spread than perturbation of the initial condition while keeping a good trade-off with the error representation [6, 7, 48, 51]. A similar random decomposition is also at the center of the variational stochastic framework of [29]. As our setting, this latter approach constitutes a general formal framework to derive stochastic representation of fluid flow dynamics. Both frameworks rely on a stochastic transport principle. In [29], it is dedicated to Hamiltonian dynamical systems and defined from a circulation preserving constrained variational formulation, while the one in [40] is based on Newtonian principles and built upon classical physical conservation laws. This latter scheme has been used as a fundamental tool to derive stochastic representations of large-scale geophysical dynamics [6, 7, 13, 49, 50, 51] or to define large eddy simulation models of turbulent flows [16, 34]

In this study, we would like to answer the following question: Does a stochastic parameterization enable us to replace in a more secure and efficient way the classical (multiplicative) inflation techniques? A general answer is obviously difficult to provide. However the experiments carried out here on the Surface Quasi-Geostrophic (SQG) model indicate that this effort seems worth being done. The SQG model is simple as it is a 2D model, yet it captures highly non linear realistic features of ocean surface dynamics [10, 18, 19, 20, 28, 36]. This model combines a buoyancy transport equation with a kinematic relation between buoyancy and velocity. For this example, we present numerical evidence that LU brings many improvements to classical ensemble simulation strategy [48, 50]. We show that this parametrization is also useful in a data assimilation context, again, by comparing to variance inflation. We present numerical evidence that the Mean Square Error performances are systematically better than all the deterministic inflation parameters tested here. We will also highlight the flaws of both inflation and localization parameter tuning when comparing them to a flow-driven stochastic model and point out that the latter seems to stabilize the localization drawbacks we observe in deterministic settings.

The outline of the paper is as follows. We first briefly present the Location Uncertainty principles and its mathematical implications. [1] Section 3 explains the numerical setup and details the results mentioned just above. The ensemble

[1] For interested reader, thorough descriptions with detailed derivation of the stochastic representation of geophysical flow dynamics can be found in [6, 13, 50].

methods used in our experiments are recaped in appendix. The following is a summary of the numerical experiments carried out in this article :

- Numerical validation that the deterministic inflation parameter is difficult to tune , we show divergence of the filter in finite time (Figures 3 and 4).
- Performance comparison between LU and deterministic settings both in terms of MSE (Figures 5 and 6) and spread quality (Figure 7).

A conclusion with some perspectives closes the paper.

2 Transport under Location Uncertainty (LU)

Location Uncertainty models the impact of the small scales on the global flow dynamics. It is a stochastic framework that relies on the following decomposition of the Lagrangian velocity of a fluid particle positioned at x_t in a spatial domain $\Omega \subset \mathbb{R}^d$:

$$\mathrm{d}x_t = v(x_t, t)\mathrm{d}t + \sigma(x_t, t)\mathrm{d}B_t, \tag{1}$$

where v is a smooth-in-time component (referred to as the large-scale component in the following) and $\sigma \mathrm{d}B_t$ is a highly oscillating random component, built from a (cylindrical) Wiener process B_t (ie a well-defined Brownian motion taking values in a functional space) [44]. This latter component is uncorrelated in time. Let us point out that the above relation is rigorously understood in its integral form and corresponds thus only to a practical shortcut notation. The correlation operator σ is defined through a bounded matrix kernel $\hat{\sigma}$, for any function $f \in (L^2(\Omega))^d$, by

$$\sigma(x, t)f = \int_\Omega \hat{\sigma}(x, y, t)f(y)\mathrm{d}y. \tag{2}$$

From this correlation operator, the covariance matrix kernel $\hat{q}$ reads

$$\hat{q}(x, y, t) = \int_\Omega \hat{\sigma}(x, x', t)\hat{\sigma}(y, x', t)^T \mathrm{d}x', \tag{3}$$

and the associated covariance operator Q is given by

$$Q(x, t)f = \int_\Omega \hat{q}(x, y, t)f(y)\mathrm{d}y. \tag{4}$$

The random velocity is Gaussian and distributed as

$$\sigma \mathrm{d}B_t \sim \mathcal{N}(0, Q\mathrm{d}t). \tag{5}$$

Moreover, at each time t, the covariance operator $Q(., t)$ is self-adjoint, non-negative definite and compact. Thus it admits an orthonormal eigenfunction

basis $\{\phi_n(\cdot,t)\}_{n\in\mathbb{N}}$ with non-negative eigenvalues $(\lambda_n(t))_{n\in\mathbb{N}}$. This entails an alternative convenient spectral definition of the noise as

$$\sigma(x,t)\mathrm{d}B_t = \sum_{n\in\mathbb{N}} \sqrt{\lambda_n(t)}\phi_n(x,t)d\beta_t^n, \tag{6}$$

where the β^n are i.i.d standard Brownian motions. From (6), the noise variance tensor a is then defined by

$$a(x,t) = \sum_{n\in\mathbb{N}} \lambda_n(t)\phi_n(x,t)\phi_n(x,t)^T. \tag{7}$$

It can be noticed the variance tensor has the physical dimension of a viscosity (ie m^2/s). The properties and structure of the noise will of course depend on the procedure used to generate the orthonormal basis functions. The one used in our experiments will be presented later. In the deterministic case, a transported tracer Θ has zero material derivative :$D_t\Theta = \partial_t\Theta + v\cdot\nabla\Theta = 0$. In the LU framework, the material derivative is replaced by the stochastic transport operator

$$\mathbb{D}_t\Theta = \mathrm{d}_t\Theta + (v^*\mathrm{d}t + \sigma\mathrm{d}B_t)\cdot\nabla\Theta - \frac{1}{2}\nabla\cdot(a\nabla\Theta)\mathrm{d}t, \tag{8}$$

where

- $\mathrm{d}_t\Theta = \Theta(x,t+\delta t) - \Theta(x,t)$ is the forward time increment of the tracer.
- The effective advection velocity is defined by

$$v^* = v - \frac{1}{2}\nabla\cdot a. \tag{9}$$

- The term $\sigma\mathrm{d}B_t\cdot\nabla\Theta$ is a non-Gaussian multiplicative noise corresponding to the tracer's transport by the small-scale flow.
- The last term is a diffusion term, as the variance tensor a is definite positive.

Note that the expression of the transport operator is given here for a divergence-free noise. In the case of a compressible random field, the modified advection incorporates an additional term related to the noise divergence [49]. One key property of LU (for an incompressible random term) is that under the same ideal boundary conditions as in the deterministic case, it conserves the energy of the transported random tracer Θ:

$$\mathrm{d}\int_\Omega \Theta^2(x)\mathrm{d}x = 0, \tag{10}$$

and, very importantly, this energy conservation property holds pathwise (i.e for any realization of the Brownian noise) [6, 49]. A few general remarks on the stochastic transport operator can be done at this point. Compared to the

usual material derivative, it encompasses additional new terms. The viscosity is now described by the variance tensor a. It is no longer a scalar variable but a (positive definite) diffusion matrix. It is thus richer than the classical eddy viscosity models [12] derived from an analogy with the molecular friction (the so called Boussinesq's assumption). The multiplicative noise corresponds to a backscattering of energy that is exactly compensated by the loss of energy brought by the stochastic diffusion (meaning the second order differential term associated to the stochastic transport operator in equation (8), which is indeed diffusive, in the sense that it dissipates energy). This balance is the reason why we have the energy conservation property and can be seen as a redistribution of the tracer (in the same way as a deterministic advection equation). Finally, the modified advection corresponds to the statistical effect induced by the small-scale inhomogeneity on the large scale component structuration [6]. With this term, the particles are statistically encouraged to migrate from regions with high variance (ie of high kinetic turbulent energy) to regions with low variance. This modification of advection, the backscattering carried by the advection noise and the balanced diffusion matrix are the three marked ingredients of the modeling under location uncertainty. It is important to point out that this balance only holds globally. Locally, these three terms may play their own role without any balance. This scheme has been used for the modeling of large scale flow dynamics for numerous flow configurations. From a data assimilation point of view, it has been used in an optimal control setting for 3D flow reconstruction and for joint parameter estimation in a 1D shallow water model [16, 56], where it was shown to provide an interesting trade-off with the so called weak constraint variational assimilation.

The LU scheme obviously depends on the noise parametrization chosen. For instance, for a homogeneous noise associated to Fourier basis functions, the variance tensor is homogeneous and constant (even diagonal for a divergence free flow). Hence, there is no modified advection. For a stationary noise, the variance field is constant in time and thus not related to the evolving large-scale components. The ability to build a flow-dependent noise enables us to improve probabilistic forecasting skills [13]. For the SQG dynamics several noise parametrizations have been compared and assessed through different statistical proper scores [48]. One of the main findings of this work is that a time-varying inhomogeneous parametrization, termed as SVD (Singular Value Decomposition) noise, provides the best quantification results.

SVD noise : In this paragraph, we describe the generation of the noise used in the experiments. It relies on the creation of pseudo-observations at each point of the simulation grid, and then on the diagonalization of the associated empirical matrix to extract a proper basis to support the noise. Here the domain Ω is a two-dimensional regular grid of size $n_x \times n_y$. The pseudo observations are constructed from the running velocity fluctuations around a sample mean, more precisely around a velocity field composed of local spatial means computed at each grid point. At each time t and each grid point $x_{i,j}$, a spatial window $W_{i,j}$ of size $n_w \times n_w$ (with n_w odd), much smaller than the whole simulation grid, is built around the point, together with the model boundary conditions

(periodicity, replication,...) if the current point is on the border. Then a pseudo-observation is given by a draw of the velocity in the following set :

$$I(x_{i,j},t) = \{v(x_{k,l},t)\,|k,l \in W_{i,j}\}\ . \tag{11}$$

Proceeding to n_o draws within the window, and iterating over all the grid points, a global pseudo-observation matrix V is built :

$$V = \begin{pmatrix} v_1^1 & \cdots & v_1^{n_o} \\ \vdots & \dots & \vdots \\ v_{n_x n_y}^1 & \cdots & v_{n_x n_y}^{n_o} \end{pmatrix}, \tag{12}$$

whose size is $(2 \times n_x \times n_y) \times n_o$ (the 2 comes from the two components of the velocity v).
Then the mean over the n_o pseudo-observations $\langle V \rangle$ is retrieved

$$V' = V - \langle V \rangle \tag{13}$$

and Singular Value Decomposition is applied to the fluctuation matrix V' to diagonalize the corresponding second order empirical moment. This way, the matrix Φ of the left eigenvectors on which we can decompose the noise as in (6) is obtained.
Let us denote ℓ the simulation grid scale and $L = n_w \ell$ the spatial scale of the sliding window used to compute the noise. The procedure described above provides a noise $\sigma_L \mathrm{d}B_t$ at scale L, which is artificial and it must be downscaled to the true simulation scale ℓ. [34] proposed a rescaling of the variance tensors based on 2D turbulent cascade assumption :

$$a_\ell = \left(\frac{\ell}{L}\right)^{\frac{2}{3}} a_L,$$

which relies on an estimation of the velocity fluctuations at the simulation scale ℓ. Then the effective noise for the simulation grid is

$$\sigma \mathrm{d}B_t = n_w^{-\frac{1}{3}} \sigma_L \mathrm{d}B_t. \tag{14}$$

[48] also show that the Uncertainty Quantification (UQ) results for the SVD noise forecast reliability are good for all the metrics presented, in particular much better than the Perturbation on Initial Conditions (PIC), often used in ensemble data assimilation techniques. Some other procedures exist to generate the noise (6), relying for example on off-line learning on high-resolution datasets, or energy-budget based method (again see [48] for further details). This one has the advantage to be flow-driven and quite simple. As the basis depends both on time and on the ensemble member at hand, some adaptations should probably be devised for realistic models. One can envisage, for instance, a combination with wavelet basis or flow-based criterion to decide across time when the noise basis should be updated.

3 Experiments on both deterministic and stochastic models

This section details the numerical experiments carried out in this study. The main aim is to study the benefits brought by a stochastic dynamics in an up-to-date version of ensemble Kalman filter with localization and inflation (details in appendix). In particular we wish to observe whether or not the stochastic dynamics brings by itself an efficient and practical inflation procedure for ensemble Kalman filtering.

In this work we apply the LU procedure to the 2D Surface Quasi-Geostrophic (SQG) model. This model constitutes an idealized dynamics for surface oceanic currents. Yet, it involves many realistic features of real world phenomena such as front, strong multiscale eddies, driven by a 3D like turbulent energy cascade, see [20, 36] for details.

The SQG model relies on a (deterministic or stochastic) transport equation of the buoyancy field b

$$D_t b = 0, \tag{15}$$

coupled with a kinematic equation

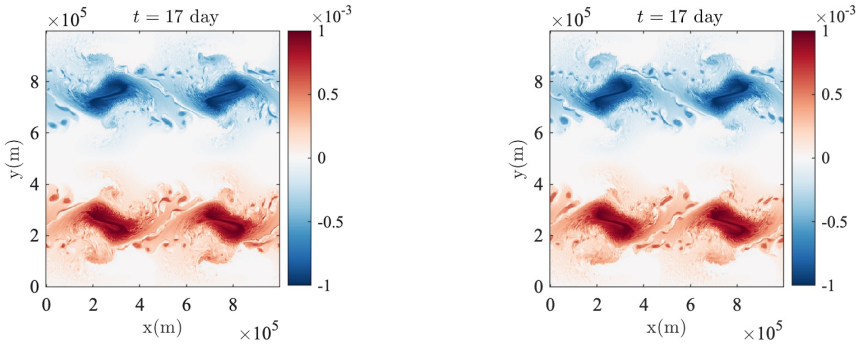
$$b = \frac{N_{strat}}{f_0}(-\Delta)^{\frac{1}{2}}\psi \tag{16}$$

and the incompressibility constraints

$$v = \nabla^{\perp}\psi; \quad \nabla \cdot \sigma dB_t = 0, \tag{17}$$

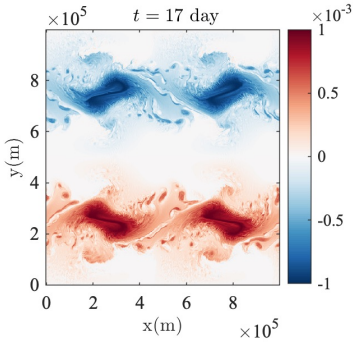
linking the buoyancy field to the velocity field v, where ψ is the stream function, N_{strat} is the stratification, f_0 is the Coriolis frequency and $\nabla f = (\partial_x f, \partial_y f)^T$, $\nabla^{\perp} f = (-\partial_y f, \partial_x f)$ respectively stand for the classical and orthogonal gradients, while Δ denotes the Laplace operator. The stochastic dynamics simply consist in replacing the material derivative D_t by the stochastic transport operator $\mathbb{D}_t$ given by (8). We show in Figure 1 several realizations of the stochastic dynamics.

It can be noticed that the large-scale component of the different particles remain quite close after 17 days. They mainly differ by their small-scale features and vortices (subfigures (a,b,c,d)). We also added the states of these realizations at day 72 as an example to point out that the system is chaotic and may lead to significant large scale differences (subfigures (e,f,g,h)), at least on the time range studied in this work. This system is unforced and involves in its numerical implementation a small hyperviscosity term. It is hence decaying in the long run. However, as shown in Figure 1, the system remains turbulent at the end of the time period on which we will focus in this study. Mesoscale eddies as well as submesoscale eddies, filaments and fronts can be observed for the different realizations at day 72 of Figure 1. The stochastic and deterministic simulation is run on a simulation grid, G_s, of size 64×64, meaning that each realization x_n is a 64×64 matrix, or equivalently a vector of size 4096 (because here only the

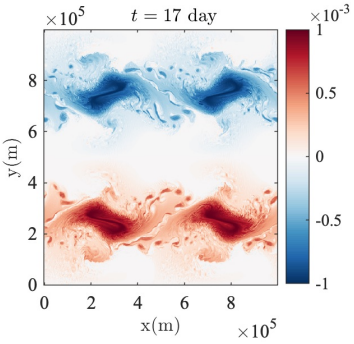


(a) Realization 1

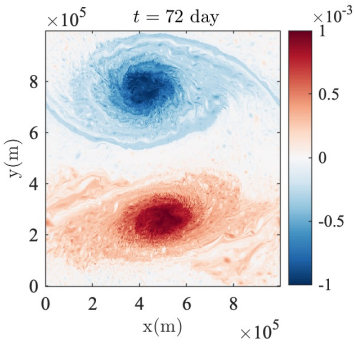
(b) Realization 2



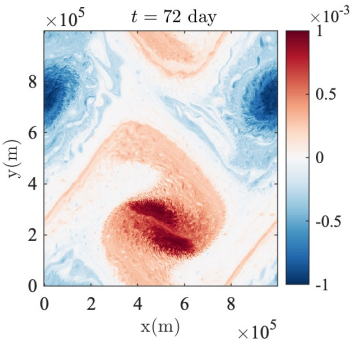
(c) Realization 3



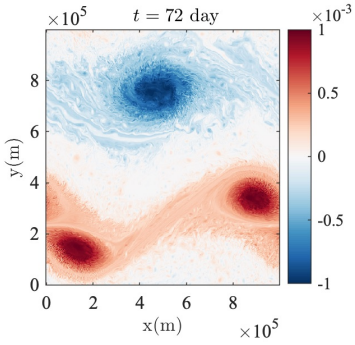
(d) Realization 4



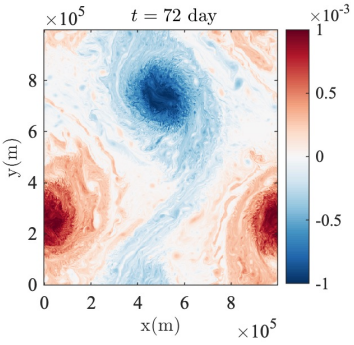
(e) Realization 1



(f) Realization 2



(g) Realization 3



(h) Realization 4

Figure 1: Four different realizations for the stochastic dynamics at day 17 and day 72.

buoyancy field is observed and simulated, the velocity being obtained through the inversion of a fractional Laplacian). The actual physical size of the domain being 1000km $\times$ 1000km, two neighbor grid points are distant of about 15km. An observation y^{obs} on a coarser observation grid, G_o, of size 16×16, is assumed to be available every day (i.e. every 600 time steps of the dynamics). It is is generated as follows:

- A trajectory $(z_t)_t$ is run from the deterministic model (PDE) at a very fine resolution grid G_f, of size 512×512
- Then a convolution-decimation procedure [39] is applied, which is the composition of a Gaussian filter G_σ and a decimation operator D subsampling one pixel out of two. The Gaussian filtering writes

$$G_\sigma(z_t)(x) = (g_\sigma * z_t)(x),$$

  where g_σ is a two-dimensional Gaussian function. For any observable f defined on a grid $(x_{11}, ..., x_{2n,2n})$, we define Dg on a decimated subgrid $(x'_{11}, ..., x'_{n,n})$ by

$$Df(x'_{ij}) = f(x_{2i,2j}).$$

  This convolution-decimation operator $D \circ G_\sigma$ is then performed three times in order to fit to the targeted simulation grid.

- Finally, a projection operator P is applied from G_s to G_o, the latter being a subset of the first, and the observation is defined, for all t, by

$$y_t^{obs} = P \circ (D \circ G_\sigma)^3(z_t) + \eta_t,$$

  where

$$\eta_t \sim \mathcal{N}(\mathbf{0}, R) \quad \text{and} \quad R = diag(r_1^2, ..., r_M^2) \tag{18}$$

  is the diagonal observation covariance matrix. It will often be considered that $R = r^2 I_M$, where M is the number of points on the observation grid. As a consequence, the operator P plays the role of the observation operator H in the Kalman equations.

Numerical setup : The simulations have been performed with a pseudo-spectral code in space [48]. The time-scheme is a fourth-order Runge-Kutta scheme for the deterministic PDE, and an Euler-Maruyama scheme for the SPDEs [35]. Regardless of the resolution and stochasticity, we use a standard hyperviscosity model to dissipate the energy at the resolution cut-off. The resulting implemented dynamics is :

$$\mathbb{D}_t b = \alpha \Delta^2 b \mathrm{d}t, \tag{19}$$

with a hyperviscosity coefficient $\alpha = (5 \times 10^{29} m^8.s^{-1}) M_x^{-8}$, where M_x is the grid resolution (here 512 for the fine-resolution PDE used to generate the observations, and 64 for the ensemble members). The boundary conditions are double-periodic. As mentioned before, the equations are mostly handled in the

Fourier space, where the following SQG relation between velocity and buoyancy Fourier transforms can be used :

$$\hat{v} = ik^{\perp}\frac{\hat{b}}{N_{strat}||k||}, \tag{20}$$

with k is the horizontal wave-vector, $k^{\perp}$ the orthogonal horizontal wave-vector and N_{strat} is the stratification. The test case considered in this study is the following : an ensemble of $N = 100$ particles is started from the very same initial condition at day 0, which consists in two cold vortices to the north and two warm vortices to the south (figure 2) :

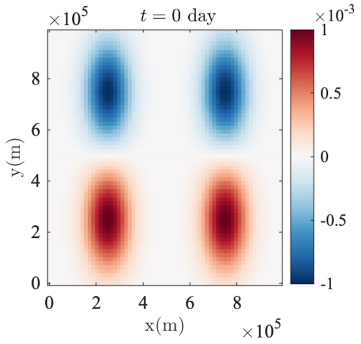


Figure 2: Initial condition of the buoyancy field for all particles.

The initial field is mathematically defined on each grid point $r = \begin{pmatrix} x \\ y \end{pmatrix}$ by the following formula :

$$b_0(r) = F\left(r - r_1\right) - F(r - r_2) - F(r - r_3) - F(r - r_4), \tag{21}$$

where the vortices initial cores are

$$r_1 = \begin{pmatrix} 250\text{km} \\ 250\text{km} \end{pmatrix} \quad ; \quad r_2 = \begin{pmatrix} 750\text{km} \\ 250\text{km} \end{pmatrix} \quad ; \quad r_3 = \begin{pmatrix} 250\text{km} \\ 750\text{km} \end{pmatrix} \quad ; \quad r_4 = \begin{pmatrix} 750\text{km} \\ 750\text{km} \end{pmatrix}$$

and the function F is defined by

$$F(r) = B_0 \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right) \tag{22}$$

with $B_0 = 10^{-3} m.s^{-1}$, $\sigma_x = 67$ km and $\sigma_y = 133$ km. We also set the Coriolis frequency to $f_0 = 1.028 \times 10^{-4} s^{-1}$ and the stratification to $N_{strat} = 3f_0$.
In this experiment, we study the differences of efficiency of the localized Ensemble Square Root Filter (cf appendix) with inflated deterministic forecast and non inflated stochastic simulations. In both cases, the initial ensemble is generated

as follows. Starting from the initial condition, the stochastic dynamics is simulated using the SVD noise for 3 days (meaning 1800 time steps for the (S)PDE), without data assimilation. This way, a random ensemble is generated, and the performances of the SVD noise indicate that this ensemble is well-spread around the truth, which will be confirmed in Figure 7. An observation is provided each day (i.e. every 600 time steps of the SPDE (or PDE)), with an observation error covariance set to $r_i = 10^{-5}$ for $i = 1, ..., M$ in equation (18), which corresponds to a weak (but not negligible) noise on the observation. For the remaining of the simulation (100 days), this ensemble is used for two experiments :

- **Experiment 1 (deterministic dynamics with inflation):** The stochasticity is shut down after day 3, the forecast is then generated by the deterministic PDE, and prior inflation is used to artificially increase the ensemble spread. Namely, before the assimilation step, given a forecast ensemble x^f, the inflated ensemble $x^{f,inf}$ is defined by

$$x_j^{f,inf} = \overline{x}^f + \alpha(x_j^f - \overline{x}^f), \tag{23}$$

  where $\alpha > 1$ is the inflation parameter. Then the localized ESRF is applied to $x^{f,inf}$ and the same procedure is done each day at each new observation.

- **Experiment 2 (stochastic dynamics without inflation):** The same SPDE drives the particles for the whole simulation, and the localized ESRF is applied with an observation each day.

The localization radius was set to l_{obs} here, where $l_{obs} \simeq 60km$ denotes the distance between two neighboring observational sites, as it provided the best results for both cases. The inflation procedure is very sensitive to the parameter α. It must be finely tuned to have the best results. If it is too small, then the spread is not large enough. If it is too large, then it could give rise to a divergence of the filter (cf figure 3). For our SQG configuration, it turns out that the range of validity of α is between 1 and 1.08 approximately. Starting from $\alpha = 1.09$, the filter starts diverging in the long term. In the context of our model, this small range for the inflation parameter and the tuning procedure is in itself a drawback of the inflation method. Still this tuning can be very different depending on the model at hand, so we do not claim that this range is small in general. Moreover, when α is badly chosen, it brings in the additional problem of strong divergence of particles of localized ESRF, as shown in Figure 4. In this example ($r_{loc} = 3l_{obs}$ and $\alpha = 1.15$), starting from day 50 or so, strong gradients between neighboring grid points can be observed, and progressively lead to non physical predictions. Obviously the same kind of behaviour also occurs in our setting with a smaller value of r_{loc}. This was theoretically expected as, in the localization procedure, the posterior ensemble members are combinations by block of linear combinations of the prior ensemble members (cf equation (35) in appendix). This formula relies on the strong assumption that such a block recombination remains a solution of the underlying PDE (or SPDE), which suggests that the

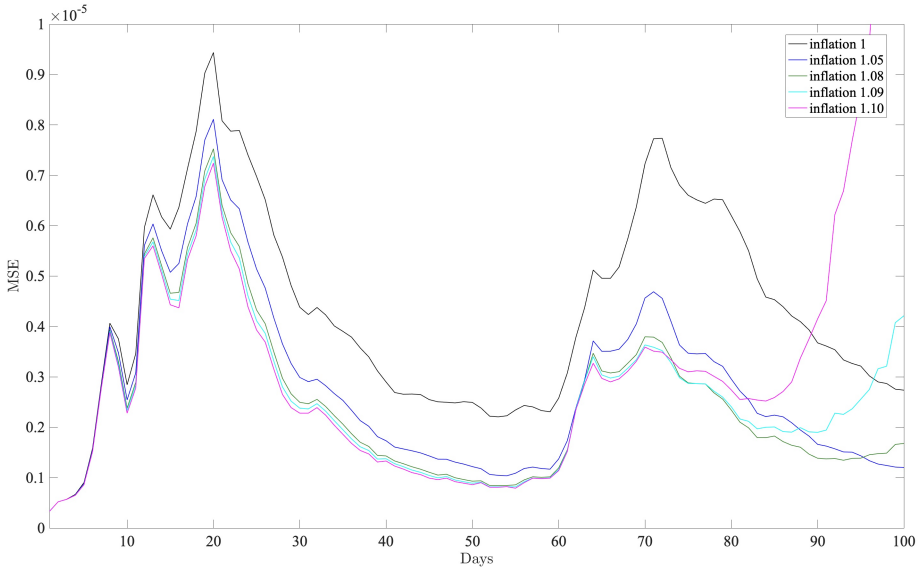


Figure 3: Mean Square Error curves for different values of the inflation parameter α in the deterministic case (case 1): magenta ($\alpha = 1.10$), cyan ($\alpha = 1.09$), green ($\alpha = 1.08$), blue ($\alpha = 1.05$) and black (no inflation: $\alpha = 1.0$). Inflation is overall very beneficial compared to the black line (no inflation), but too big values of α lead to a long-term divergence of the filter (magenta and cyan).

global equation is equivalent to a combination of local ones. Let us note that for instance for fluid dynamics equations, the presence of a pressure term, solution of an elliptic problem, is intrinsically global and theoretically prevents such a local modeling. A careless application of localization may lead to the appearance of some discontinuities or gradients due to very different decisions taken by the filter at neighboring points. This may entail, after subsequent iterations of the underlying dynamics, major errors, numerical divergence of some ensemble members, leading eventually to unphysical realizations.

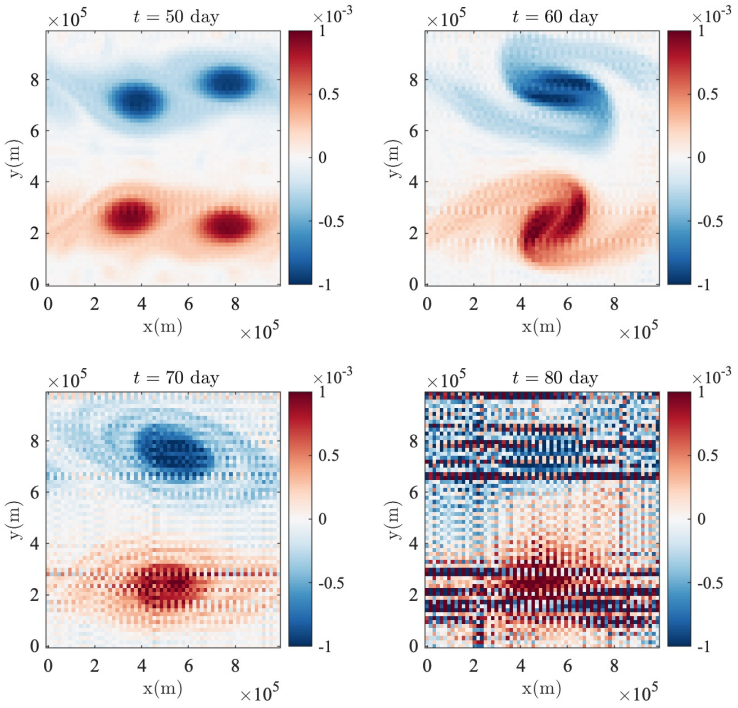


Figure 4: Examples of diverging realizations resulting from LESRF with $r_{loc} = 3l_{obs} \simeq 180km$ and $\alpha = 1.15$.

The long-term instability of inflation is not a good sign, especially considering we deal here with a coarse-scale diffusive simulation. For finer resolutions with less diffusion and much more pronounced non-linear features, divergence is likely to occur sooner for the same inflation value. This instability is theoretically expected, as the artificial variance injection entailed by inflation is never counterbalanced in any way in the model. At a coarse resolution, even for low inflation parameters, the typical slow growth of error at the end of the simulation seems to indicate that long-time divergence as in figures 5 and 4 should be observed extending the simulation time (Figure 6 equivalently shows that increasing the inflation parameter makes the divergence occur earlier). The

divergence caused by variance inflation can be attenuated by temporal adaptive schemes with introduction of diagnostic criteria [38, 45]. Another alternative consists in changing the hypothesis on the prior distribution accounting for the sampling errors in the ensemble, which was shown in [11] to make multiplicative inflation optional on Lorenz models.

An increase of variance without controlling the global energy by a balanced dissipation raises the question of the mathematical well-posedness of the numerical scheme but also on the modified physics undergone by this forcing. These two questions are far from cosmetic. The first one is related to the generality of the numerical scheme (ie it must be valid for any noise and at any resolution, etc.). The second point is about the error terms, they should not change dramatically the targeted underlying physical system [17]. These two points can hardly be met by an artificial increase of variance or by non physical multiplicative noise (see [17] for an example on the Lorenz63 model).

On the opposite, the LU setting brings a natural balance between the energy brought by the noise and the amount dissipated by the stochastic diffusion. In addition, as shown in Figure 5, the simulation is stable while bringing an equivalent spread as a relatively strongly inflated deterministic simulation value (with respect to the SQG dynamics studied here). It leads also to better MSE results than the deterministic setting for the different values of inflation experimented here.

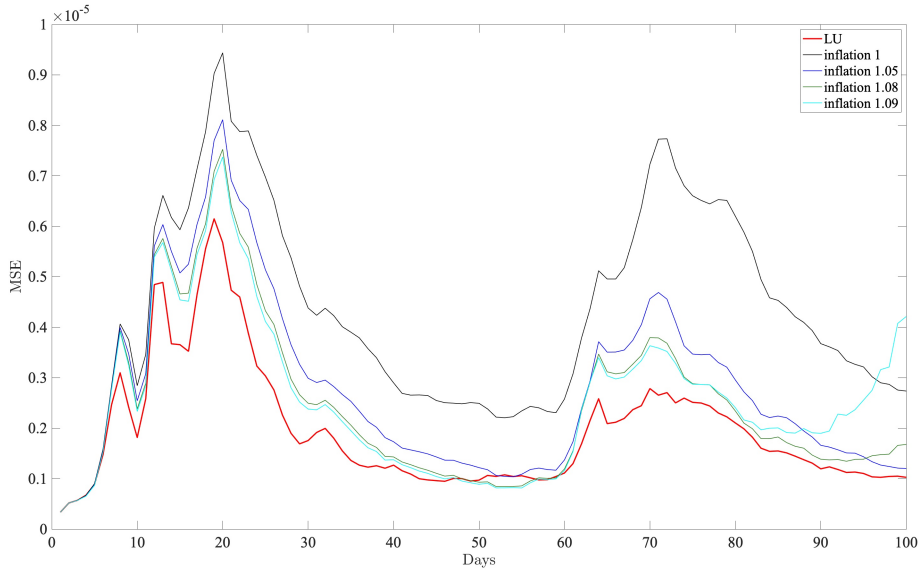


Figure 5: The LU framework (in red) performs better than the deterministic cases for all the reasonable inflation parameters tested and plotted in cyan ($\alpha = 1.09$), green ($\alpha = 1.08$), blue ($\alpha = 1.05$) and black (no inflation: $\alpha = 1.0$) color.

Figure 6 shows the same comparison with a greater inflation parameter set to $\alpha = 1.20$. The time window is reduced to 50 days because, as expected, increasing the inflation parameter leads to an earlier divergence, here starting from days 40-50. Still it has comparable results with LU in the first 30-35 days. This means that LU has short-term MSE performances comparable with a very strong deterministic inflation parameter, and avoids long-term divergence as well.

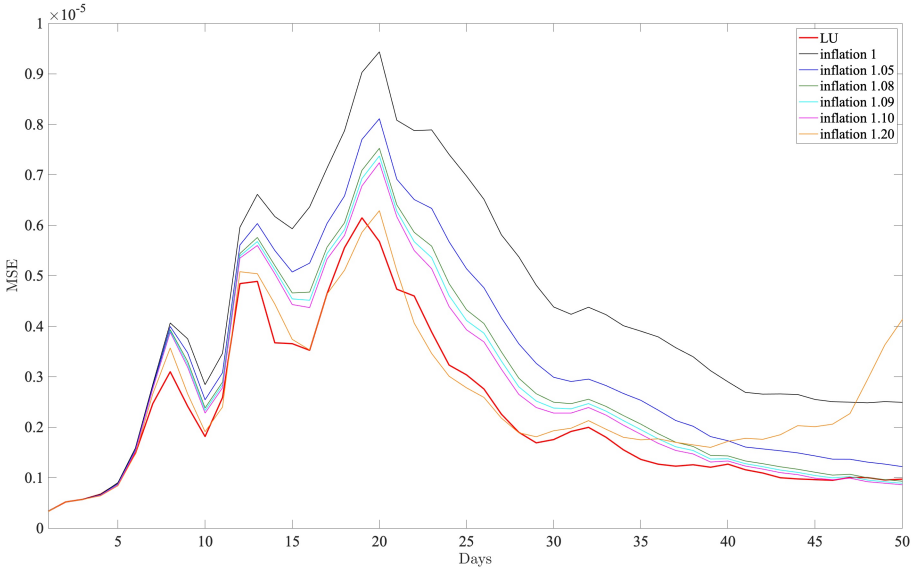
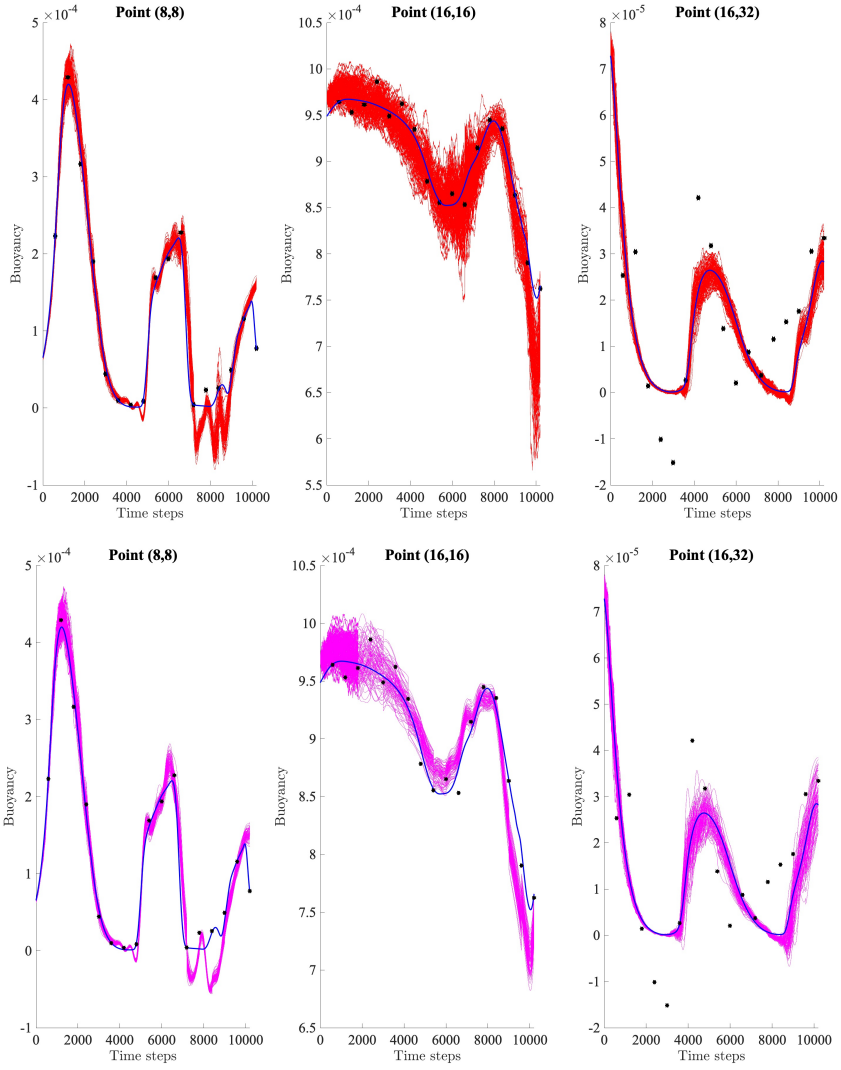


Figure 6: Same figure as 5 with additional inflation parameters $\alpha = 1.10$ (in magenta) and $\alpha = 1.20$ (in orange). The time window is reduced to 50 days as the orange curve diverges much sooner than the others.

Additionally, we examine the spread of the ensemble members around the truth and observation points. We chose three characteristic grid points, corresponding to the center, north and southwest of the southwest warm vortex of the initial condition (Figure 2). In the following figure (Figure 7), we compare the behavior of the spread of LU ensemble members (in red) and deterministically inflated ones with parameter $\alpha = 1.10$ (in magenta) and $\alpha = 1.20$ (in orange) around the truth (blue dots) and observation points (black dots). Note that the two chosen inflation parameters are quite strong and lead both to divergence of the filter, with the second one exhibiting a divergence much sooner than the first one.

Figure 7 shows this comparison for the first 17 days of simulation. Although both spreads of subfigures (a) and (b) seem very similar for the figures of left and right columns, we observe on the center column figure that deterministic inflation does not provide enough spread to contain the truth, while LU does (see for instance time window 8,000-10,000). It can also be noticed that the


Point (8,8)
Point (16,16)
Point (16,32)
Buoyancy
Time steps
×10^-4
×10^-5

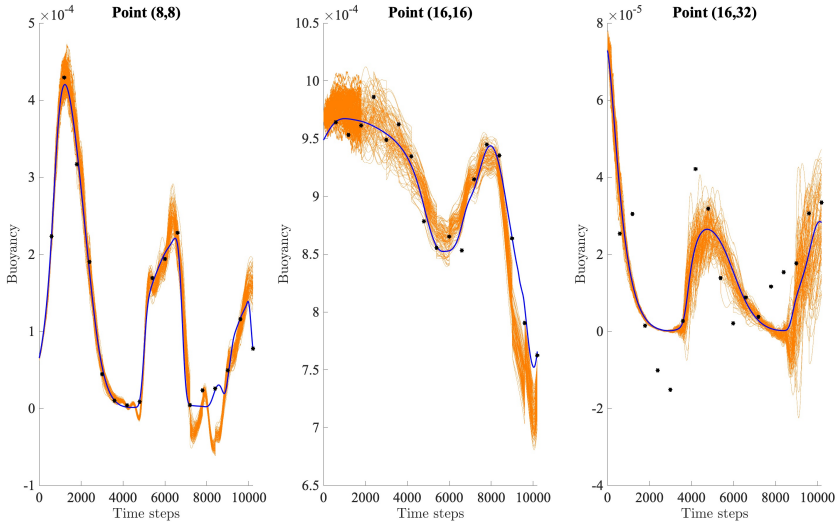


Figure 7: Comparison of spread between LU and deterministic inflation on the first 17 days of the dynamics for three points located at the center, top and bottom left of the bottom left warm vortex of the initial condition. The upper row shows in red the buoyancy values at these points for the stochastic ensemble. The two lower ones show the buoyancy values at the same points for the deterministically inflated ensembles for $\alpha = 1.10$ (in magenta) and $\alpha = 1.20$ (in orange). The black dots are the observations; the blue line stand for the truth.

truth at time step 8400 is completely skipped by magenta and orange spreads, while LU manages to reach it.
We see that increasing the inflation parameter does not counteract the flaws of the smaller inflation parameter $\alpha = 1.10$ (magenta spread, Figure 7), meaning that increasing α does not entail a richer ensemble. To that extent, we can also notice that when the stochasticity is shut down at day 3, the deterministic sets of trajectories immediately become less dense than the stochastic one. This means that the neighborhood of the truth/observation is visited much more often (in time) by stochastic trajectories. This improved "density" of the stochastic ensemble in the state space is an interesting feature, not exploited yet but it could be relevant in a particle filter framework.

Deterministic inflation sometimes offers more spread than LU. We observe this situation for example at the time window 9,000-10,000 on the right column figure. But this spread seems bigger than necessary, as LU has a smaller spread, but already well-centered around the truth. This supports the idea that deterministic inflation is blind, in the sense that it is decorrelated from the dynamics and the current state of the ensemble. LU does not only provide a spread of the ensemble equivalent to a strong inflation parameter, it brings an ensemble

of better quality. This ensemble is constructed from the large-scale fluctuation and, in a way, fits the physics encoded by the original dynamics.

Another indicator of this is the following spread-error consistency in Figure 8. We compared the ensemble bias absolute value with the error estimated from the ensemble spread. We did this for LU and for the deterministic setting with inflation parameters $\alpha = 1.10$ and $\alpha = 1.20$ at day 17, when the three models are very close in terms of global MSE (see Figure 6). The results clearly point out that the LU framework provides much more spread and a better estimation of the error. The spread induced by inflation has much less physical structure and relevance.

Let us outline that the noise experimented here is free from any parameter.

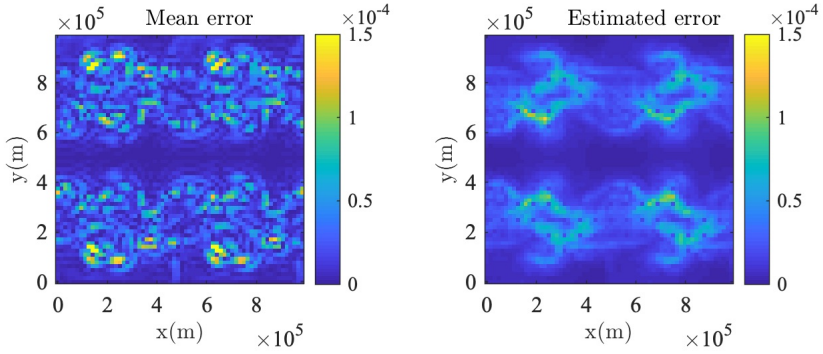
4 Conclusion

We have shown in this study that the introduction of a stochastic parameterization of the flow dynamics within the location uncertainty setting enables to improve localized ESRF results when compared to an ensemble simulation of the original deterministic dynamics. Our main findings are as follows :

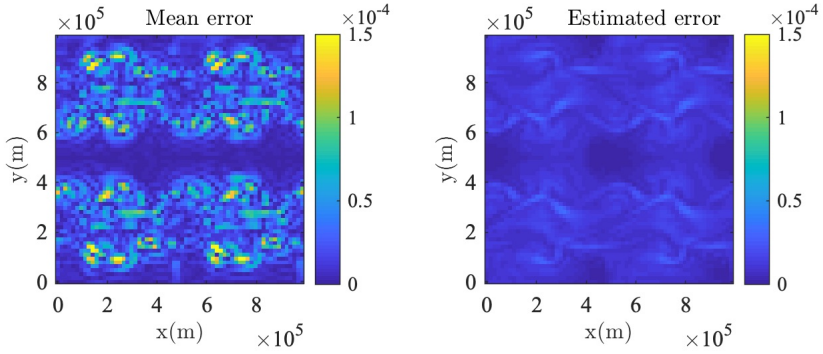
- The LU framework provides good spread, good MSE performances, and does not require variance inflation contrary to the deterministic ensemble simulations. This constitutes a first clear advantage of such a stochastic parametrization over an artificial increase of the ensemble spread. The parameter of such inflation is known to be difficult to fix and requires a fine tuning to get good performances.

- In addition, even for low values of the inflation parameter, we have shown that such a steady increase of variance leads to filter divergence at finite time. For finer resolutions with less diffusion and much more non-linear features, this time is likely to be much shorter for the same inflation value. Augmentation of variance through inflation corresponds to an error model with no particular spatial patterns. On the contrary, the stochastic dynamics consider a more subtle transport noise directly set from the large-scale on-going simulation. The fact that the injection of variance is naturally counterbalanced by a diffusion term prevents long-term divergence. It creates also a more physical and useful spread, whereas the spread entailed by deterministic inflation is sometimes too small and sometimes bigger than necessary.

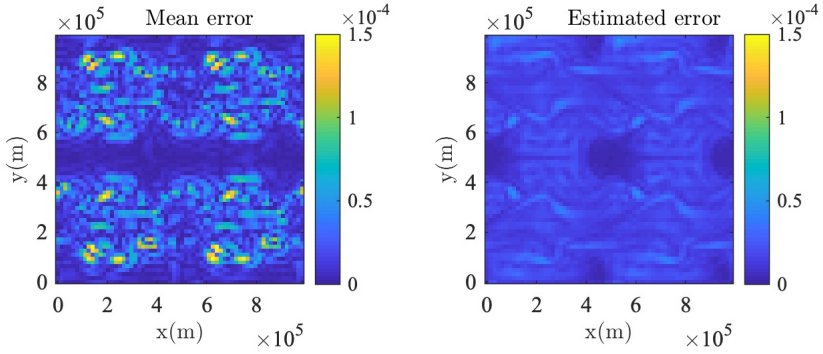
- We also point out the potential drawbacks of localization techniques when accompanied with localization in the observation space, due to the block recombination of the solution. An augmentation of spread in a blind and constant way leads the local filter to take many different decisions, yielding a strong local augmentation of the gradients. When the dynamics involves an elliptic equation, as it is the case in the surface quasi-geostrophic model, this leads to dramatic divergence in finite time.



(a) LU



(b) Deterministic inflation $\alpha = 1.10$



(c) Deterministic inflation $\alpha = 1.20$

Figure 8: Comparison between the ensemble bias absolute value $e(x) = |\overline{b}(x) - b^{truth}(x)|$ (left maps) compared to the estimate error (1.96× the standard deviation of the ensemble) evaluated at each grid point (right maps) at day 17. The upper row shows this comparison for LU, the other two show the same comparison with the deterministic setting respectively for inflation parameters of $\alpha = 1.10$ (central row) and $\alpha = 1.20$ (bottom row).

- The localized Ensemble Square Root Filter with a stochastic dynamics seems to be able to handle sparse observations in time – in the experimental context used here only one observation field per day was available – meaning every 600 time steps of the stochastic partial differential equation. This ability could be very helpful in the context of realistic forecasting of sea surface state. This study provides encouraging results and constitutes a first step toward this goal. Time interval of 3 to 10 days together with higher resolution observation will have to be investigated.

The stochastic SQG model experimented here constitutes an interesting model with highly non-linear features and a global elliptic equation. The resolution grid considered here is relatively small (ie 64×64) but yet of dimension larger than usual benchmarks elaborated from the different Lorenz models. We are confident the techniques could be extended to higher dimension.

As for the intrinsic nonlinear features of the dynamics, we intend to explore other data assimilation techniques like particle filtering for the stochastic SQG forecast. The particle filtering framework is promising as it constitutes a flexible data assimilation technique in terms of prior assumptions on both the linearity (or not) of the model and the Gaussianity of the errors. Even if it is known to struggle with high-dimensional systems, we expect the additional sophisticated methods devised in [21, 22] to counteract the curse of dimensionality.

Acknowledgments

The authors acknowledge the support of the ERC project 856408-STUOD

A Appendix: Ensemble methods

In this section we recall briefly the ensemble methods on which we focus in this study. We start by introducing the notations used here. Ensembles denoted x will contain N members. Each member lives in the state space $\mathbb{R}^d$. The empirical mean will be denoted by $\overline{x}$ and the (unbiased) ensemble covariance matrix is defined by

$$P = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \overline{x})(x_n - \overline{x})^T. \tag{24}$$

We are given an observation y^{obs} living in $\mathbb{R}^M$ with R its observation covariance matrix.
The observation operator H maps the state space $\mathbb{R}^d$ to the observation space $\mathbb{R}^M$.
Let us do a recap on different ensemble methods, including the classical EnKF and the localized square root filter used in the experiments.

Classical EnKF (with perturbed observations) Starting from a forecast ensemble denoted x^f, in the classical EnKF the ensemble is updated as follows :

- The so called Kalman gain matrix is defined by

$$K = P^f H^T (H P^f H^T + R)^{-1}. \tag{25}$$

- The members of the ensemble are updated by :

$$x_n^a = x_n^f - K(H x_n^f - y^{obs} - \xi_n), \tag{26}$$

  with $\xi_n \sim \mathcal{N}(0, R)$, which corresponds to random draws of the observation error.

Ensemble Square Root Filter (ESRF) The square root scheme, developed in [55], provides a deterministic way to generate the posterior ensemble, without sampling of the ξ_ns. The method is the following:

- The innovation term $H\overline{x}^f - y^{obs}$ is multiplied by K in order to update the mean of the ensemble :

$$\overline{x}^a = \overline{x}^f - K(H\overline{x}^f - y^{obs}). \tag{27}$$

- The ensemble covariance matrix is updated as well :

$$P^a = P^f - KHP^f. \tag{28}$$

- Each member of the ensemble is given by

$$x_n^a = \overline{x}^a + (P^a)^{\frac{1}{2}} (P^f)^{-\frac{1}{2}} (x_n^f - \overline{x}^f). \tag{29}$$

Since the ensemble covariance matrices are, in practice, of very high dimension, they cannot be neither stored nor manipulated (remember that for a state of dimension 10^{11}, which is usual in weather forecasting, then this gives a covariance matrix of size 10^{22}, that is to say of 10^7 peta (storage unit of the state variable)). An alternative is to rewrite the filtering equations only in terms of the ensemble anomaly matrix

$$A^f = [(x_1^f - \overline{x}^f), (x_2^f - \overline{x}^f), \ldots, (x_N^f - \overline{x}^f)],$$

which is a $d \times N$ matrix, so a more convenient one to store. In the following, we provide the expression of the different covariance matrices involved in this setting. An exhaustive presentation of the full computational details may be found in [47].
The ensemble covariance matrix is linked to the anomaly matrix by

$$P^f = \frac{1}{N-1} A^f (A^f)^T,$$

and the same relation holds for the posterior (or analysis) matrices. Injecting this relation in equation (28) of the update of P^a, one can find, after some algebraic manipulations, that

$$A^a(A^a)^T = A^f\left(I - \frac{1}{N-1}(HA^f)^T(HP^fH+R)^{-1}HA^f\right)(A^f)^T.$$

So, if S is such that

$$SS^T = I - \frac{1}{N-1}(HA^f)^T(HP^fH+R)^{-1}HA^f,$$

then

$$A^a = A^fS.$$

Using the Sherman-Morrison-Woodbury formula [1], we deduce that

$$S = \left(I + \frac{1}{N-1}(HA^f)^TR^{-1}HA^f\right)^{-\frac{1}{2}}. \tag{30}$$

Let us remark that this matrix has size $N \times N$, and Singular Value Decomposition can be used in order to compute the inverse square root. By definition of A^a, the posterior forecast is given by

$$x^a = \overline{x}^a + A^a = \overline{x}^a + A^fS. \tag{31}$$

It can be shown that the mean $\overline{x}^a$ can be written as a linear combination of the members of the forecast ensemble, namely

$$\overline{x}^a = \sum_{i=1}^{N}\omega_i x_i^f,$$

where ω_i is the i-th coordinate of the column vector

$$\omega = \frac{1}{N}\mathbf{1} - \frac{1}{N-1}S^2(A^f)^TH^TR^{-1}(H\overline{x}^f - y^{obs}).$$

Finally, combining this with equation (31), the members of the posterior ensemble can be computed, for $n = 1, \ldots, N$,

$$x_n^a = \sum_{i=1}^{N}\omega_i x_i^f + \sum_{i=1}^{N}(x_i^f - \overline{x}^f)S_{in}. \tag{32}$$

This filter is part of what is generally called Ensemble Transform Kalman Filter (ETKF), because members of the posterior ensemble are linear combinations of the forecast ones. Here

$$x_n^a = \sum_{i=1}^{N} d_{i,n}x_i^f,$$

with

$$d_{i,n} = \omega_i + S_{i,n} - \frac{1}{N}\sum_{j=1}^{N} S_{j,n}.$$

Let us mention that such superposition principle (stability of solutions by linear combinations) is in theory only valid for linear dynamics. In the context of nonlinear dynamics, this setting has to be understood as constraining the solution to live in a particular Hilbert space of Gaussian processes spanned by the ensemble members and associated to a norm defined by the inverse ensemble covariance matrix.

Localized version of the ESRF There exist two main ways to localize the filter. The R-localization (or local analysis) works on the observation error covariance matrix R. Another way to achieve localization is to work on the ensemble covariance matrix P (B-localization or covariance localization). As equations (30-32) are only based on the use of the anomaly matrices instead of the whole covariance matrices P, R-localization is the only one compatible with our equations. [52] give details on both localization techniques and give numerical insights showing that both methods are likely to produce similar results.
The basic idea of R-localization is to update the ensemble members, not in one step and using all the grid points, but coordinate by coordinate, using for each one a local neighborhood of "relevance". By doing so, the dimension of the state space is dramatically reduced for each local update.
More precisely, for each grid point $k = 1, \ldots, d$, let us denote p_k the projection on the k-th coordinate (or grid point). For each k, the aim is to compute the local coefficients $d^k_{i,n}$ such that

$$p_k(x^a_n) = \sum_{i=1}^{N} d^k_{i,n} p_k(x^f_i). \tag{33}$$

In order to do this, R-localization is applied, meaning that the observation error covariance matrix will be modified for each coordinate, and then we will apply the previous formulas (30-32) for this new matrix.
For each $k = 1, \ldots, d$, a diagonal localization matrix C_k of size $p \times p$ is defined, where p is the number of points on which you have observations (possibly less than d if the observation grid is coarser), and the diagonal coefficients are given, for $l = 1, \ldots, p$, by

$$(C_k)_{ll} = \rho\left(\frac{d(l,k)}{r_{loc}}\right),$$

where :

- $d(l,k)$ denotes the distance between the grid point k and the observation site l.
- r_{loc} is the localization radius.

- ρ is the Gaspari-Cohn function, defined for any $z \geq 0$ by

$$\rho(z) = \begin{cases} -\frac{1}{4}z^5 + \frac{1}{2}z^4 + \frac{5}{8}z^3 - \frac{5}{3}z^2 + 1 & \text{if } z < 1 \\ \frac{1}{12}z^5 - \frac{1}{2}z^4 + \frac{5}{8}z^3 + \frac{5}{3}z^2 - 5z + 4 - \frac{2}{3z} & \text{if } 1 \leq z \leq 2 \\ 0 & \text{if } z \geq 2 \end{cases}.$$

  It basically behaves like a Gaussian, but has compact support $[0, 2]$, so that the coefficients are set to zero as soon as $d(l, k)$ is twice as big as the localization radius.

Then the inverse observation error covariance matrix is modified by the Schur product

$$R_k^{-1} = C_k \circ R^{-1}. \tag{34}$$

By doing so, the grid points outside this area are no longer taken into account, as the observation error is set to infinity at these points (due to the cancellation of spurious correlations). Finally, R is replaced by R_k in formula (30), which gives a localized version S_k of S, and so a localized version of equation (32). In order to sum everything up, in this localized version, the update formula is given by

$$x_n^a = \sum_{k=1}^{d} p_k(x_n^a) = \sum_{k=1}^{d} \sum_{i=1}^{N} d_{i,n}^k \; p_k(x_i^f). \tag{35}$$

References

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Dover Books on Electrical Engineering. Dover Publications, 2012.

[2] J.L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001.

[3] J.L. Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, 59A:210–224, 2007.

[4] J.L. Anderson. Localization and sampling error correction in ensemble kalman filter data assimilation. *Monthly Weather Review*, 140(7):2359 – 2371, 2012.

[5] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. Fundamentals of algorithms. SIAM, 2016.

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

[7] W. Bauer, P. Chandramouli, L. Li, and E. Mémin. Stochastic representation of mesoscale eddy effects in coarse-resolution barotropic models. *Ocean Modelling*, 151:101646, 2020b.

[8] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. de la Cámara, H. M. Christensen, M. Colangeli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, M. Sakradzija, J-S. von Storch, A. Weisheimer, M. Weniger, P. D. Williams, and J-I. Yano. Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3):565 – 588, 2017.

[9] C.H. Bishop, B.J. Etherton, and S.J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. part I: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.

[10] W. Blumen. Uniform potential vorticity flow: part I. theory of wave interactions and two-dimensional turbulence. *Journal of the Atmospheric Sciences*, 35(5):774–783, 1978.

[11] M. Bocquet. Ensemble kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750, 2011.

[12] J. Boussinesq. Mémoires présentés par divers savants à l'Académie des Sciences, 23 (1): 1–680, 1877.

[13] R. Brecht, L. Li, W. Bauer, and E. Mémin. Rotating shallow water flow under location uncertainty with a structure-preserving discretization. arXiv, physics.flu-dyn, 2021.

[14] R. Buizza, M.J. Miller, and T.N. Palmer. Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, (279), 1999.

[15] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5):e535, 2018.

[16] P. Chandramouli, E. Memin, and D. Heitz. 4d large scale variational data assimilation of a turbulent flow with a dynamics error model. *Journal of Computational Physics*, 412:109446, 2020.

[17] B. Chapron, P. Dérian, E. Mémin, and V. Resseguier. Large-scale flows under location uncertainty: a consistent stochastic framework. *Quarterly Journal of the Royal Meteorological Society*, 144(710):251–260, 2018.

[18] P. Constantin, M. Lai, R. Sharma, Y. Tseng, and J. Wu. New numerical results for the surface quasi-geostrophic equation. *Journal of Scientific Computing*, 50(1):1–28, 2012.

[19] P. Constantin, A. Majda, and E. Tabak. Formation of strong fronts in the 2-D quasigeostrophic thermal active scalar. *Nonlinearity*, 7(6):1495, 1994.

[20] P. Constantin, Q. Nie, and N. Schörghofer. Front formation in an active scalar equation. *Physical Review E*, 60(3):2858, 1999.

[21] C. Cotter, D. Crisan, D. Holm, W. Pan, and I. Shevchenko. Data assimilation for a quasi-geostrophic model with circulation-preserving stochastic transport noise. *Journal of Statistical Physics*, 179(5):1186–1221, 2020a.

[22] C. Cotter, D. Crisan, D.D. Holm, W. Pan, and I. Shevchenko. A particle filter for stochastic advection by lie transport (salt): A case study for the damped and forced incompressible 2d euler equation, 2020b.

[23] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[24] G. Evensen. *Data assimilation: The ensemble Kalman filter.* Springer-Verlag, New-york, 2006.

[25] C. L. E. Franzke, T. J. O'Kane, J. Berner, P. D. Williams, and V. Lucarini. Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6:63–78, 2014.

[26] G. A. Gottwald, D. T. Crommelin, and C. L. E. Franzke. Stochastic climate theory, 2016.

[27] T.M. Hamill, J.S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129(11):2776 – 2790, 2001.

[28] I. Held, R. Pierrehumbert, S. Garner, and K. Swanson. Surface quasi-geostrophic dynamics. *Journal of Fluid Mechanics*, 282:1–20, 1995.

[29] D.D. Holm. Variational principles for stochastic fluid dynamics *Proceedings of the Royal society A : Mathematical, Physical and Engineering Science*,471 (20140963).

[30] P. Houtekamer and H. Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review* , 126, 1998.

[31] P. Houtekamer and H. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. 129, 2001.

[32] P. Houtekamer, H. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen. Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. *Monthly Weather Review*, 133:604–620, 2005.

[33] P. Houtekamer and H.L. Mitchell. Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3269–3289, 2005.

[34] S. Kadri Harouna and E. Mémin. Stochastic representation of the Reynolds transport theorem: revisiting large-scale modeling. *Computers and Fluids*, 156:456–469, 2017.

[35] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1999.

[36] G. Lapeyre and P. Klein. Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory. *Journal of physical oceanography*, 36(2):165–176, 2006.

[37] F. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A:97–110, 1986.

[38] Y. Lee, A. J. Majda, and D Qi. Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence. *Monthly Weather Review*, 145(2):669–682, 2017.

[39] T. Lindeberg. *Scale-Space Theory in Computer Vision*. The Springer International Series in Engineering and Computer Science. Springer US, 1993.

[40] E. Mémin. Fluid flow dynamics under location uncertainty. *Geophysical & Astrophysical Fluid Dynamics*, 108(2):119–146, 2014.

[41] T. N. Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 2019.

[42] T. N. Palmer and P. Williams Stochastic physics and climate modelling. *Cambridge Univ. Press*, 2017.

[43] D.T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194 – 1207, 2001.

[44] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 1992.

[45] P. N. Raanes, M. Bocquet, and A. Carrassi. Adaptive covariance inflation in the ensemble kalman filter by gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, 145(718):53–75, 2019.

[46] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, 2015.

[47] M. Reinhardt. *Hybrid filters and multi-scale models*. doctoralthesis, Universität Potsdam, 2020.

[48] V. Resseguier, L. Li, G. Jouan, P. Derian, E. Mémin, and B. Chapron. New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics. *Archives of Computational Methods in Engineering*, pages 1886–1784, 2020a.

[49] V. Resseguier, E. Mémin, and B. Chapron. Geophysical flows under location uncertainty, Part I Random transport and general models. *Geophys. & Astro. Fluid Dyn.*, 111(3):149–176, 2017a.

[50] V. Resseguier, E. Mémin, and B. Chapron. Geophysical flows under location uncertainty, Part II Quasi-geostrophy and efficient ensemble spreading. *Geophys. & Astro. Fluid Dyn.*, 111(3):177–208, 2017b.

[51] V. Resseguier, E. Mémin, and B. Chapron. Geophysical flows under location uncertainty, Part III SQG and frontal dynamics under strong turbulence conditions. *Geophys. & Astro. Fluid Dyn.*, 111(3):209–227, 2017c.

[52] P. Sakov and L. Bertino. Relation between two common localisation methods for the enkf. *Computational Geosciences*, 15(2):225–237, 2011.

[53] G. J. Shutts and T. N. Palmer. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *Journal of Climate*, 20(2):187 – 202, 2007.

[54] M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485 – 1490, 2003.

[55] J.S. Whitaker and T.M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 2002.

[56] Y. Yang and E. Mémin. Estimation of physical parameters under location uncertainty using an ensemble^2-expectation-maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*, 145(719):418–433, 2019.