State and Stochastic Parameters Estimation with Combined Ensemble Kalman and Particle Filters

Jules Guillot, Pierre Ailliot, Emmanuel Frénod, Juan Ruiz, Pierre Tandeo

State and Stochastic Parameters Estimation with Combined Ensemble Kalman and Particle Filters

Jules Guillot,[a,⋆] Pierre Ailliot,[b] Emmanuel Frénod,[a,*] Juan Ruiz,[c] and Pierre Tandeo[d]

[a] *Univ Bretagne - Sud, CNRS UMR 6205, LMBA, F-56000 Vannes, France*

[b] *Univ Brest, CNRS UMR 6205, Laboratoire de Mathematiques de Bretagne Atlantique*

[c] *CNRS-IRD-CONICET-UBA, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI), Buenos Aires, Argentina*

[d] *IMT Atlantique, Lab-STICC, UMR CNRS, 6285, France*

[⋆] *eOdyn, Brest, France*

[*] *See-d, Vannes, France*

Corresponding author: Jules Guillot, jules.guillot56@orange.fr

ABSTRACT: Quantifying uncertainties is a key aspect of data assimilation systems since it has a large impact on the quality of the forecasts and analyses. Sequential data assimilation algorithms, such as the Ensemble Kalman Filter (EnKF), describe the model and observation errors as additive Gaussian noises and use both inflation and localization to avoid filter degeneracy and compensate for misspecifications. This introduces different stochastic parameters which need to be carefully estimated in order to get a reliable estimate of the latent state of the system. A classical approach to estimate unknown parameters in data assimilation consists in using state-augmentation, where the unknown parameters are included in the latent space and are updated at each iteration of the EnKF. However, it is well-known that this approach is not efficient to estimate stochastic parameters because of the complex (non-Gaussian and non-linear) relationship between the observations and the stochastic parameters which can not be handled by the EnKF. A natural alternative for non-Gaussian and non-linear state-space models is to use a particle filter (PF), but this algorithm fails to estimate high-dimensional systems due to the curse of dimensionality. The strengths of these two methods are gathered in the proposed algorithm, where the PF first generates the particles that estimate the stochastic parameters, then using the mean particle the EnKF generates the members that estimate the geophysical variables. This generic method is first detailed for the estimation of parameters related to the model or observation error and then for the joint estimation of inflation and localization parameters. Numerical experiments are performed using the Lorenz-96 model to compare our approach with state-of-the-art methods. The results show the ability of the new method to retrieve the geophysical state and to estimate online time-dependent stochastic parameters. The algorithm can be easily built from an existing EnKF with low additional cost and without further running the dynamical model.

1. Introduction

Data assimilation consists in combining a dynamical model with observation data to retrieve the latent true state of a system. The dynamical model is generally based on equations with physical assumptions (see Carrassi et al. 2018), though data-driven approaches have gained a lot of research attention in recent years (see Lguensat et al. 2017). In both approaches the dynamical model is generally misspecified, generating an error called the model error. It may represent for example unknown or unresolved physical phenomena as in Guillot et al. (2022), unknown physical parameters (Smith et al. 2013), or calibration errors in the data driven approach. Also, the observations are often imperfect, because of measurement and representation errors (see Janjić et al. 2018), leading to observation errors. Quantifying the model and observation errors for data assimilation remains a difficult task as explained in Tandeo et al. (2020). These two sources of uncertainties are usually represented as additive Gaussian white noises whose covariance matrices $\boldsymbol{Q}$ (for the model error) and $\boldsymbol{R}$ (for the observation error) respectively depend on vectors of stochastic parameters $\boldsymbol{\theta_Q}$ and $\boldsymbol{\theta_R}$ which need to be calibrated. This estimation problem has been addressed by many authors and a recent review of the existing methods can be found in Tandeo et al. (2020).

Instead of directly calibrating $\boldsymbol{\theta_Q}$ and $\boldsymbol{\theta_R}$, we can focus on the covariance matrix of the forecasted members. Covariance inflation (see Anderson and Anderson 1999) consists in multiplying this covariance matrix by an inflation parameter to reproduce the effect of the model error on it. This strategy also allows to reduce the impact of the sampling error (Anderson 2012), which is due to the use of a too small ensemble size (generally for computational reasons), even in situations where the matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are known. The inflation parameter is usually estimated online using the innovation statistics of Desroziers et al. (2005), as in Miyoshi et al. (2013) where the joint estimation of the inflation parameter and $\boldsymbol{\theta_R}$ is discussed. The hierarchical Bayesian approaches of Anderson (2009) and El Gharamti (2018) allow to estimate an inflation parameter which varies in space and time but with a not negligible computational cost. Covariance inflation is generally combined with covariance localization (see Houtekamer and Mitchell 2001) to further mitigate the impact of the sampling error and avoid filter degeneracy. The covariance localization eliminates the long-range spurious correlations in the covariance matrix of the forecasted members which is generally rank-deficient. For that, a mask depending on a localization parameter is applied to this

covariance matrix. Usually a grid search is used to estimate the localization parameter but it is computationally expensive.

The aim of this work is to develop a generic method for estimating a set of stochastic parameters related to the parametrization of $\boldsymbol{Q}_t$, $\boldsymbol{R}_t$, or covariance inflation and localization. The proposed methodology is online, meaning that the stochastic parameters are allowed to vary in time and are estimated adaptively. The need for considering time-dependent covariance matrices for the model and observation errors is discussed in Dee (1995). Our method is based on state augmentation which consists in augmenting the latent state with the unknown parameters and assuming that these latter follow a simple dynamical model (e.g., a random walk). Then a data assimilation algorithm is used to estimate simultaneously the geophysical variables and the parameters. This method is widely used in data assimilation since it is easy to implement and it was found to be efficient for the online estimation of dynamical parameters (see Anderson 2001). However it is not appropriate for the estimation of stochastic parameters when using the EnKF because this algorithm can not handle the complex (non-Gaussian and non-linear) relationship between the observations and the stochastic parameters (see DelSole and Yang 2010). In order to circumvent this issue, it has been proposed to combine the strengths of the EnKF, which is known to be efficient for high-dimensional geophysical systems and of the PF, which is suitable to handle non-linearities and non-Gaussianity in low-dimensional systems (Van Leeuwen 2009) but is prone to the curse of dimensionality (Snyder et al. 2008). Different combinations of these two algorithms have already been studied for the state estimation (Papadakis et al. 2010; Slivinski et al. 2015).

In this paper, a new combination of the PF with the EnKF, called the PF-EnKF algorithm, is detailed for estimating time-dependent stochastic parameters of the data assimilation system. The EnKF is used for the state estimation whereas the PF is used for the stochastic parameters estimation. At each time step, a set of particles is first generated to represent the distribution of the unknown stochastic parameters and weighted using the likelihood of the observations. Then, one step of the EnKF is run to update the ensemble members using the weighted mean of the particles as a value for the stochastic parameters. The proposed approach shares similarities with existing methods (Frei and Künsch 2012; Stroud et al. 2018; Ait-El-Fquih and Hoteit 2020) where the EnKF is used for the state estimation and the PF is used for the stochastic parameters estimation. The methods of Frei and Künsch (2012) allow to estimate a static (i.e. constant in time and

space) $\boldsymbol{\theta_R}$ and the method of Stroud et al. (2018) provides estimates of static $\boldsymbol{\theta_Q}$ and $\boldsymbol{\theta_R}$. Our methodology can handle the time dependency of these parameters, with a lower computational cost than Algorithm 1 of Frei and Künsch (2012) and of Stroud et al. (2018) because only a single Kalman gain matrix is computed in our algorithm. The method of Ait-El-Fquih and Hoteit (2020) allows to estimate a static inflation parameter with an inverse-gamma prior distribution, whereas our approach deals with a time-varying inflation parameter. The main contribution of this work is to propose a generic method to estimate time-dependent stochastic parameters, including inflation and localization, which can be easily appended on an existing EnKF-based data assimilation system with low additional cost.

The rest of the paper is organized as follows. The new method is developed in Section 2 for different case studies. Numerical results with the Lorenz-96 model are then discussed in Section 3. Finally, concluding remarks and perspectives are given in Section 4.

2. The new PF-EnKF method

The proposed methodology is based on the following state-space model

$$\begin{cases} \boldsymbol{x}_t = M(\boldsymbol{x}_{t-1}) + \boldsymbol{\eta}_t \quad \text{with } \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t), & \text{(1a)} \\ \boldsymbol{y}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \quad \text{with } \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t) & \text{(1b)} \end{cases}$$

where $t \in \{1, ..., T\}$ denotes the time, $\boldsymbol{x}_t \in \mathbb{R}^n$ the latent state, $\boldsymbol{y}_t \in \mathbb{R}^p$ the observations available at time t and $M(\cdot)$ the dynamical model. For the sake of simplicity, we consider a linear observation operator $\boldsymbol{H}$, but the new method also works when this operator is non-linear. $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$ denote respectively the model and observation errors with covariance matrices $\boldsymbol{Q}_t = \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t})$ and $\boldsymbol{R}_t = \boldsymbol{R}(\boldsymbol{\theta}_{\boldsymbol{R},t})$ depending on vectors of stochastic parameters $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ and $\boldsymbol{\theta}_{\boldsymbol{R},t}$.

a. PF-EnKF for estimating the model or observation error

In this section, we first assume that the vector of stochastic parameters $\boldsymbol{\theta}_{\boldsymbol{R},t}$ related to the observation error is known and we discuss the online estimation of $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ related to the model error. The estimation of $\boldsymbol{\theta}_{\boldsymbol{R},t}$ when $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ is known is similar and detailed in Appendix A.

$\boldsymbol{\theta}_{\boldsymbol{Q},t}$ is assumed to be a Markov process as it is usually done when using state-augmentation. In terms of conditional distributions, the state-space model can be written as

$$
\begin{aligned}
p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{\theta}_{\boldsymbol{Q},1:t-1}, \boldsymbol{x}_{1:t-1}) &\sim p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{\theta}_{\boldsymbol{Q},t-1}),\\
p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \boldsymbol{\theta}_{\boldsymbol{Q},1:t}) &\sim \phi(\boldsymbol{x}_t; M(\boldsymbol{x}_{t-1}), \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t})),\\
p(\boldsymbol{y}_t|\boldsymbol{x}_{1:t}, \boldsymbol{\theta}_{\boldsymbol{Q},1:t}) &\sim \phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t, \boldsymbol{R}_t),
\end{aligned}
$$

with ϕ the probability density function of the Gaussian distribution and $p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{\theta}_{\boldsymbol{Q},t-1})$ a transition kernel which describes the evolution of the stochastic parameters (e.g., a random walk).

The goal is to estimate $\boldsymbol{x}_t$ and $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ at each time t knowing the observations $\boldsymbol{y}_{1:t} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_t)$. This relies on

$$
p(\boldsymbol{x}_t, \boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t}) = p(\boldsymbol{x}_t|\boldsymbol{\theta}_{\boldsymbol{Q},t}, \boldsymbol{y}_{1:t})\; p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t}), \tag{2}
$$

where the conditional distribution $p(\boldsymbol{x}_t|\boldsymbol{\theta}_{\boldsymbol{Q},t}, \boldsymbol{y}_{1:t})$ is estimated using the EnKF and $p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t})$ using the PF. The equations related to each filter are detailed below.

EnKF for $p(\boldsymbol{x}_t|\boldsymbol{\theta}_{\boldsymbol{Q},t}, \boldsymbol{y}_{1:t})$:

For the forecast step at time t, N_{memb} state estimates, called the forecasted members, are generated by applying the dynamical model to the analyzed members at time $t-1$ denoted $\boldsymbol{x}_{t-1}^{a,i}$ and adding a random perturbation depending on $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ to obtain for $i \in \{1, ..., N_{memb}\}$

$$
\boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) = M(\boldsymbol{x}_{t-1}^{a,i}) + \boldsymbol{\eta}_t^i \quad \text{with } \boldsymbol{\eta}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t})). \tag{3}
$$

Then the empirical mean and covariance matrix of the forecasted members are computed

$$
\boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}) = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}), \tag{4}
$$

$$
\boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}) = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} \left(\boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) - \boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t})\right)\left(\boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) - \boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t})\right)^\top. \tag{5}
$$

For the analysis step, each forecasted member is corrected using the available observation $\boldsymbol{y}_t$ as in Burgers et al. (1998) to generate the analyzed members for $i \in \{1, ..., N_{memb}\}$

$$\boldsymbol{x}_t^{a,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) = \boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) + \boldsymbol{K}_t(\boldsymbol{\theta}_{\boldsymbol{Q},t}) \left(\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t^i - \boldsymbol{H}\boldsymbol{x}_t^{f,i}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) \right) \tag{6}$$

with

$$\boldsymbol{K}_t(\boldsymbol{\theta}_{\boldsymbol{Q},t}) = \boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t})\boldsymbol{H}^\top \left(\boldsymbol{H}\boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t})\boldsymbol{H}^\top + \boldsymbol{R}_t \right)^{-1} \tag{7}$$

and

$$\boldsymbol{\varepsilon}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t). \tag{8}$$

PF for $p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t})$:

According to Bayes' theorem we have

$$p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t-1})\ p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{\boldsymbol{Q},t}). \tag{9}$$

Using a sequential importance resampling (SIR) scheme, the conditional distribution $p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t-1})$ is approximated by a discrete distribution

$$p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t-1}) \approx \frac{1}{N_{\boldsymbol{\theta}}} \sum_{j=1}^{N_{\boldsymbol{\theta}}} \delta_{\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}}(\boldsymbol{\theta}_{\boldsymbol{Q},t})$$

where $\delta(\cdot)$ denotes the Dirac distribution. The forecasted particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}$, for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$, are generated from the conditional distribution $p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{\theta}_{\boldsymbol{Q},t-1}^{a,j})$ where $\boldsymbol{\theta}_{\boldsymbol{Q},t-1}^{a,j}$ denotes the j-th analyzed particle at $t-1$.

The previous EnKF scheme allows to approximate the likelihood in Eq. (9) using

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{\boldsymbol{Q},t}) = \int p(\boldsymbol{y}_t|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}_{\boldsymbol{Q},t}) d\boldsymbol{x}_t.$$

Finally, based on Rubin (1988), Eq. (9) can be approximated by the discrete distribution

$$p(\boldsymbol{\theta}_{\boldsymbol{Q},t}|\boldsymbol{y}_{1:t}) \approx \sum_{j=1}^{N_{\boldsymbol{\theta}}} \gamma_t^j \delta_{\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}}(\boldsymbol{\theta}_{\boldsymbol{Q},t}) \tag{10}$$

where for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$

$$\gamma_t^j = \frac{\phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}), \boldsymbol{H}\boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j})\boldsymbol{H}^\top + \boldsymbol{R}_t)}{\sum_{k=1}^{N_{\boldsymbol{\theta}}} \phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,k}), \boldsymbol{H}\boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,k})\boldsymbol{H}^\top + \boldsymbol{R}_t)}. \tag{11}$$

PF-EnKF algorithm:

Now following Eq. (2), the PF and EnKF need to be combined. There are a number of possible approaches to this, where a compromise between computational cost and accuracy must be found. In this work, a combination with low additional computational cost is favoured, meaning that the number of runs of the dynamical model and the number of analysis computations should not be increased compared to the usual EnKF.

For the forecast step at time t, the previous members $\boldsymbol{x}_{t-1}^{a,i}$ are propagated by the dynamical model to obtain the propagated members

$$\boldsymbol{x}_t^{p,i} = M(\boldsymbol{x}_{t-1}^{a,i}) \quad \text{for } i \in \{1, ..., N_{memb}\},$$

with their empirical mean and covariance matrix

$$\boldsymbol{x}_t^p = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{p,i}$$

$$\boldsymbol{P}_t^p = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} (\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)(\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)^\top.$$

The PF is now performed to weight the forecasted particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}$. To do this, Eq. (11) is used but instead of computing the empirical mean $\boldsymbol{x}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j})$ and covariance matrix $\boldsymbol{P}_t^f(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j})$ of the forecasted members, we use the empirical mean $\boldsymbol{x}_t^p$ and covariance matrix $\boldsymbol{P}_t^p$ of the propagated members $\boldsymbol{x}_t^{p,i}$ with the theoretical mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j})$ of the

random perturbations $\boldsymbol{\eta}_t^j$ to obtain as in Stroud et al. (2018)

$$\boldsymbol{x}_t^f = \boldsymbol{x}_t^p, \tag{12}$$

$$\boldsymbol{P}_t^{f,j} = \boldsymbol{P}_t^p + \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}) \quad \text{for } j \in \{1, ..., N_{\boldsymbol{\theta}}\}. \tag{13}$$

The weight of each forecasted particle $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}$ is thus given for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$ by

$$\gamma_t^j = \frac{\phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^f, \boldsymbol{H}\boldsymbol{P}_t^{f,j}\boldsymbol{H}^\top + \boldsymbol{R}_t)}{\sum_{k=1}^{N_{\boldsymbol{\theta}}} \phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^f, \boldsymbol{H}\boldsymbol{P}_t^{f,k}\boldsymbol{H}^\top + \boldsymbol{R}_t)}. \tag{14}$$

The numerical experiments that we performed show that the estimates of the mean and covariance matrix of the forecasted members in Eqs. (12)-(13) improve the performance of the algorithm compared to using the empirical mean and covariance matrix of the forecasted members. Indeed, it permits to reduce the impact of the sampling error on the likelihood used to weight each particle, this allows to better identify the relevant particles.

The forecasted particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}$ are then resampled using the discrete distribution in Eq. (10) with the weights from Eq. (14) to give the analyzed particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{a,j}$ for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$.

Then Eqs. (3)-(8) of the EnKF are used to estimate the true state. A first strategy would consist in replacing the vector of unknown stochastic parameters $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ by each analyzed particle $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{a,j}$ for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$, as it has been done by Stroud et al. (2018) or Frei and Künsch (2012) with their Algorithm 1 for $\boldsymbol{R}$. However it leads to compute $N_{\boldsymbol{\theta}}$ Kalman gain matrices which is computationally not suitable for high-dimensional applications. Another strategy is to replace $\boldsymbol{\theta}_{\boldsymbol{Q},t}$ by the mean of the analyzed particles given by

$$\boldsymbol{\theta}_{\boldsymbol{Q},t}^a = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{j=1}^{N_{\boldsymbol{\theta}}} \boldsymbol{\theta}_{\boldsymbol{Q},t}^{a,j},$$

allowing to compute a single Kalman gain matrix as in Algorithm 2 of Frei and Künsch (2012) for $\boldsymbol{R}$. This saves computational resources compared to the first strategy. The numerical experiments that we performed (not shown) indicate that both strategies give similar results.

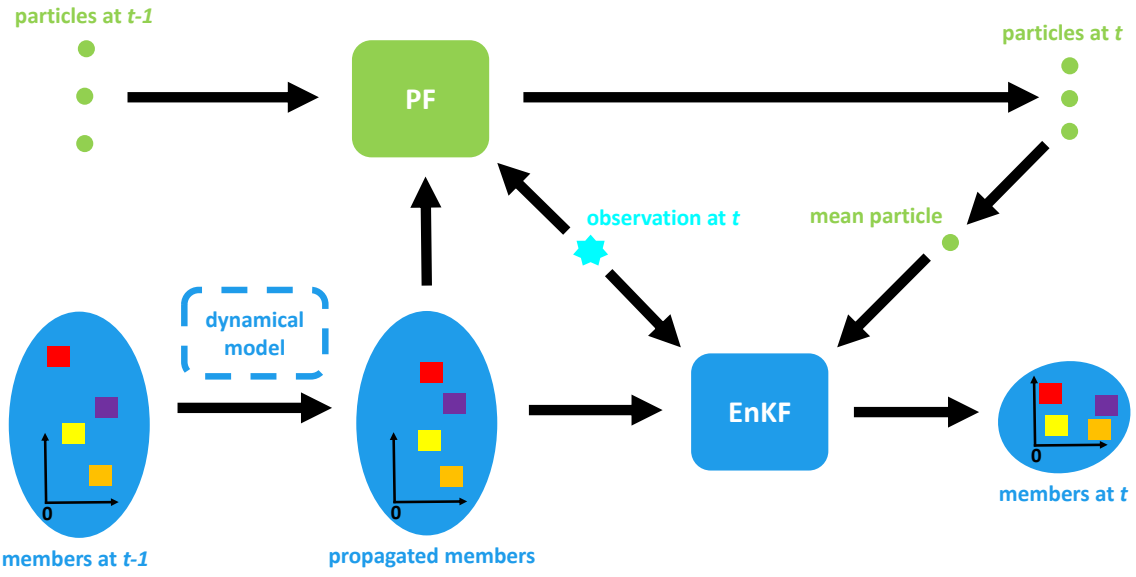

FIG. 1: Global scheme of the PF-EnKF cycle at time t. The PF is dedicated to the stochastic parameters estimation whereas the EnKF provides the state estimation.

The EnKF is thus run for the mean particle $\boldsymbol{\theta}^a_{\boldsymbol{Q},t}$ giving for $i \in \{1,\dots,N_{memb}\}$ the forecasted members

$$\boldsymbol{x}_t^{f,i} = \boldsymbol{x}_t^{p,i} + \boldsymbol{\eta}_t^i \quad \text{with } \boldsymbol{\eta}_t^i \sim \mathcal{N}(0, \boldsymbol{Q}(\boldsymbol{\theta}^a_{\boldsymbol{Q},t}))$$

and the analyzed members

$$\boldsymbol{x}_t^{a,i} = \boldsymbol{x}_t^{f,i} + \boldsymbol{K}_t(\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t^i - \boldsymbol{H}\boldsymbol{x}_t^{f,i}) \quad \text{with } \boldsymbol{\varepsilon}_t^i \sim \mathcal{N}(0, \boldsymbol{R}_t),$$

where

$$\boldsymbol{K}_t = \boldsymbol{P}_t^f \boldsymbol{H}^\top (\boldsymbol{H}\boldsymbol{P}_t^f \boldsymbol{H}^\top + \boldsymbol{R}_t)^{-1} \quad \text{with } \boldsymbol{P}_t^f = \boldsymbol{P}_t^p + \boldsymbol{Q}(\boldsymbol{\theta}^a_{\boldsymbol{Q},t}). \tag{15}$$

The PF-EnKF algorithm for estimating the model error is detailed in Algorithm 1 and can be easily implemented from an existing EnKF with few extra computational cost. A cycle at time t of our generic method is schematized in Figure 1.

Algorithm 1 PF-EnKF for estimating $\boldsymbol{Q}_t$

Require: $\boldsymbol{Q}(\cdot)$, $\boldsymbol{R}_t$ and $\boldsymbol{\theta}_{\boldsymbol{Q},0}$

for $i = 1, \ldots, N_{memb}$ **do**
  $\boldsymbol{\eta}_1^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},0}))$
  $\boldsymbol{x}_1^{a,i} = \boldsymbol{x}_0 + \boldsymbol{\eta}_1^i$
end for
for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
  $\boldsymbol{\theta}_{\boldsymbol{Q},1}^{a,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{Q},1} | \boldsymbol{\theta}_{\boldsymbol{Q},0})$
end for
for $t = 2, \ldots, T$ **do**
  for $i = 1, \ldots, N_{memb}$ **do**
    $\boldsymbol{x}_t^{p,i} = M(\boldsymbol{x}_{t-1}^{a,i})$
  end for
  $\boldsymbol{x}_t^p = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{p,i}$
  $\boldsymbol{P}_t^p = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} (\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)(\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)^\top$
  for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
    $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{Q},t} | \boldsymbol{\theta}_{\boldsymbol{Q},t-1}^{a,j})$
    $\boldsymbol{P}_t^{f,j} = \boldsymbol{P}_t^p + \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j})$
  end for
  for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
    $\gamma_t^j = \frac{\phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^p, \boldsymbol{H}\boldsymbol{P}_t^{f,j}\boldsymbol{H}^\top + \boldsymbol{R}_t)}{\sum_{k=1}^{N_{\boldsymbol{\theta}}} \phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^p, \boldsymbol{H}\boldsymbol{P}_t^{f,k}\boldsymbol{H}^\top + \boldsymbol{R}_t)}$
  end for
  resample the forecasted particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{f,j}$ using the weights γ_t^j to obtain the analyzed particles $\boldsymbol{\theta}_{\boldsymbol{Q},t}^{a,j}$
  $\boldsymbol{\theta}_{\boldsymbol{Q},t}^a = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{j=1}^{N_{\boldsymbol{\theta}}} \boldsymbol{\theta}_{\boldsymbol{Q},t}^{a,j}$
  $\boldsymbol{P}_t^f = \boldsymbol{P}_t^p + \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t}^a)$
  $\boldsymbol{K}_t = \boldsymbol{P}_t^f \boldsymbol{H}^\top (\boldsymbol{H}\boldsymbol{P}_t^f \boldsymbol{H}^\top + \boldsymbol{R}_t)^{-1}$
  for $i = 1, \ldots, N_{memb}$ **do**
    $\boldsymbol{\eta}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},t}^a))$

27: $\boldsymbol{x}_t^{f,i} = \boldsymbol{x}_t^{p,i} + \boldsymbol{\eta}_t^i$

28: $\boldsymbol{\varepsilon}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t)$

29: $\boldsymbol{x}_t^{a,i} = \boldsymbol{x}_t^{f,i} + \boldsymbol{K}_t(\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t^i - \boldsymbol{H}\boldsymbol{x}_t^{f,i})$

30: **end for**

31: **end for**

b. PF-EnKF for inflation and localization

In our approach the inflation and localization of the covariance matrix of the forecasted members $\boldsymbol{P}_t^f$ are led jointly by the Schur product $\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t}) \circ \boldsymbol{P}_t^f$ which is the element-wise multiplication between $\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t})$ and $\boldsymbol{P}_t^f$. $\boldsymbol{\theta}_{\boldsymbol{L},t} = (\lambda_{\boldsymbol{L},t}, l_{\boldsymbol{L},t})$ is the couple of unknown stochastic parameters and $\boldsymbol{L}(\cdot)$ is a n by n positive semi-definite matrix that represents the decay of correlations in the physical space. Each element of $\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t})$ is computed using the correlation function of Gaspari and Cohn (1999), denoted GC(·), that depends on the localization parameter $l_{\boldsymbol{L},t} \succ 0$ and inflated by $\lambda_{\boldsymbol{L},t} \succ 0$, so that for $(k, k') \in \{1, ..., n\}^2$

$$\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t})[k, k'] = \lambda_{\boldsymbol{L},t}\mathrm{GC}(l_{\boldsymbol{L},t})[k, k'].$$

The correlation function GC(·) is a polynomial approximation of a Gaussian density but with compact support whose radius is equal to $2l_{\boldsymbol{L},t}$. A lower value of $l_{\boldsymbol{L},t}$ corresponds to a stronger localization. The goal is to estimate $\boldsymbol{\theta}_{\boldsymbol{L},t}$ using the same methodology as in Section a. The resulting PF-EnKF algorithm jointly estimates time-dependent inflation and localization parameters and is detailed in Algorithm 2. To show the relevancy of using the likelihood as a criterion for the estimation of inflation and localization parameters, an illustrative example is given in Appendix B.

Algorithm 2 PF-EnKF for inflation and localization

Require: $\boldsymbol{Q}_t, \boldsymbol{R}_t, \boldsymbol{L}(\cdot)$ and $\boldsymbol{\theta}_{\boldsymbol{L},0}$

1: **for** $i = 1, \ldots, N_{memb}$ **do**

2: $\boldsymbol{\eta}_1^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_1)$

3: $\boldsymbol{x}_1^{a,i} = \boldsymbol{x}_0 + \boldsymbol{\eta}_1^i$

4: **end for**

5: **for** $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**

$\boldsymbol{\theta}_{\boldsymbol{L},1}^{a,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{L},1}|\boldsymbol{\theta}_{\boldsymbol{L},0})$

end for

for $t = 2,\ldots,T$ **do**

  for $i = 1,\ldots,N_{memb}$ **do**

    $\boldsymbol{\eta}_t^i \sim \mathcal{N}(0,\boldsymbol{Q}_t)$

    $\boldsymbol{x}_t^{f,i} = M(\boldsymbol{x}_{t-1}^{a,i}) + \boldsymbol{\eta}_t^i$

  end for

  $\boldsymbol{x}_t^f = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{f,i}$

  $\boldsymbol{P}_t^f = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} (\boldsymbol{x}_t^{f,i} - \boldsymbol{x}_t^f)(\boldsymbol{x}_t^{f,i} - \boldsymbol{x}_t^f)^\top$

  for $j = 1,\ldots,N_{\boldsymbol{\theta}}$ **do**

    $\boldsymbol{\theta}_{\boldsymbol{L},t}^{f,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{L},t}|\boldsymbol{\theta}_{\boldsymbol{L},t-1}^{a,j})$

    $\boldsymbol{P}_t^{f,j} = \boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t}^{f,j}) \circ \boldsymbol{P}_t^f$

  end for

  for $j = 1,\ldots,N_{\boldsymbol{\theta}}$ **do**

    $\gamma_t^j = \frac{\phi(\boldsymbol{y}_t;\boldsymbol{H}\boldsymbol{x}_t^f,\boldsymbol{H}\boldsymbol{P}_t^{f,j}\boldsymbol{H}^\top+\boldsymbol{R}_t)}{\sum_{k=1}^{N_{\boldsymbol{\theta}}} \phi(\boldsymbol{y}_t;\boldsymbol{H}\boldsymbol{x}_t^f,\boldsymbol{H}\boldsymbol{P}_t^{f,k}\boldsymbol{H}^\top+\boldsymbol{R}_t)}$

  end for

  resample the forecasted particles $\boldsymbol{\theta}_{\boldsymbol{L},t}^{f,j}$ using the weights γ_t^j to obtain the analyzed particles $\boldsymbol{\theta}_{\boldsymbol{L},t}^{a,j}$

  $\boldsymbol{\theta}_{\boldsymbol{L},t}^a = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{j=1}^{N_{\boldsymbol{\theta}}} \boldsymbol{\theta}_{\boldsymbol{L},t}^{a,j}$

  $\overline{\boldsymbol{P}_t^f} = \boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},\boldsymbol{t}}^{\boldsymbol{a}}) \circ \boldsymbol{P}_t^f$

  $\boldsymbol{K}_t = \overline{\boldsymbol{P}_t^f}\boldsymbol{H}^\top(\boldsymbol{H}\overline{\boldsymbol{P}_t^f}\boldsymbol{H}^\top + \boldsymbol{R}_t)^{-1}$

  for $i = 1,\ldots,N_{memb}$ **do**

    $\boldsymbol{\varepsilon}_t^i \sim \mathcal{N}(0,\boldsymbol{R}_t)$

    $\boldsymbol{x}_t^{a,i} = \boldsymbol{x}_t^{f,i} + \boldsymbol{K}_t(\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t^i - \boldsymbol{H}\boldsymbol{x}_t^{f,i})$

  end for

end for

3. Numerical results

The PF-EnKF is tested with the Lorenz-96 dynamical model for estimating $\boldsymbol{Q}_t$, $\boldsymbol{R}_t$ and then both inflation and localization parameters. It is compared to the EnKF that uses the true $\boldsymbol{Q}_t$ in the first case study, then with Algorithm 2 of Frei and Künsch (2012) for estimating $\boldsymbol{R}_t$ and to the

combination of state-of-the-art inflation and localization methods in the third case study.

The Lorenz-96 model (see Lorenz 1996) is a one-dimensional model, representing the evolution of a meteorological quantity in n sectors of a latitude circle. It is defined for $k \in \{1, ..., n = 40\}$ by the ordinary differential equation

$$\frac{\partial \boldsymbol{x}_{t,k}}{\partial t} = (\boldsymbol{x}_{t,k+1} - \boldsymbol{x}_{t,k-2})\boldsymbol{x}_{t,k-1} - \boldsymbol{x}_{t,k} + F$$

with periodic boundary conditions. The time step is 0.05 and the usual value $F = 8$ is chosen to have a chaotic behavior.

The true state $\boldsymbol{x}_t$ is generated following Eq. (1a) for $t \in \{1, ..., T = 500\}$ with for $(k, k') \in \{1, ..., n\}^2$

$$\boldsymbol{Q}_t[k, k'] = \lambda_{\boldsymbol{Q},t}^2 \, \mathrm{e}^{-\frac{d_{\boldsymbol{x}}(k,k')^2}{l_{\boldsymbol{Q},t}^2}}.$$

$d_{\boldsymbol{x}}(k, k') = \min(|k - k'|, n - |k - k'|)$ is the distance between two grid points on the circle in the state space, $\lambda_{\boldsymbol{Q},t}$ and $l_{\boldsymbol{Q},t}$ are respectively the variance and spatial scale parameters of $\boldsymbol{Q}_t$. The values on the diagonal of $\boldsymbol{Q}_t$ are equal to $\lambda_{\boldsymbol{Q},t}^2$ and $l_{\boldsymbol{Q},t}$ controls the speed of decrease of the correlation with the distance.

The initial true state is $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. One grid point out of two is observed ($p = 20$) and the observation $\boldsymbol{y}_t$ is generated following Eq. (1b) with for $(k, k') \in \{1, ..., p\}^2$

$$\boldsymbol{R}_t[k, k'] = \lambda_{\boldsymbol{R},t}^2 \, \mathrm{e}^{-\frac{d_{\boldsymbol{y}}(k,k')^2}{l_{\boldsymbol{R},t}^2}}. \tag{16}$$

$d_{\boldsymbol{y}}(k, k') = \min(|k - k'|, p - |k - k'|)$ is the distance between two observed grid points on the circle in the observation space. The vectors of stochastic parameters of $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are respectively $\boldsymbol{\theta}_{\boldsymbol{Q},t} = (\lambda_{\boldsymbol{Q},t}, l_{\boldsymbol{Q},t})$ and $\boldsymbol{\theta}_{\boldsymbol{R},t} = (\lambda_{\boldsymbol{R},t}, l_{\boldsymbol{R},t})$.

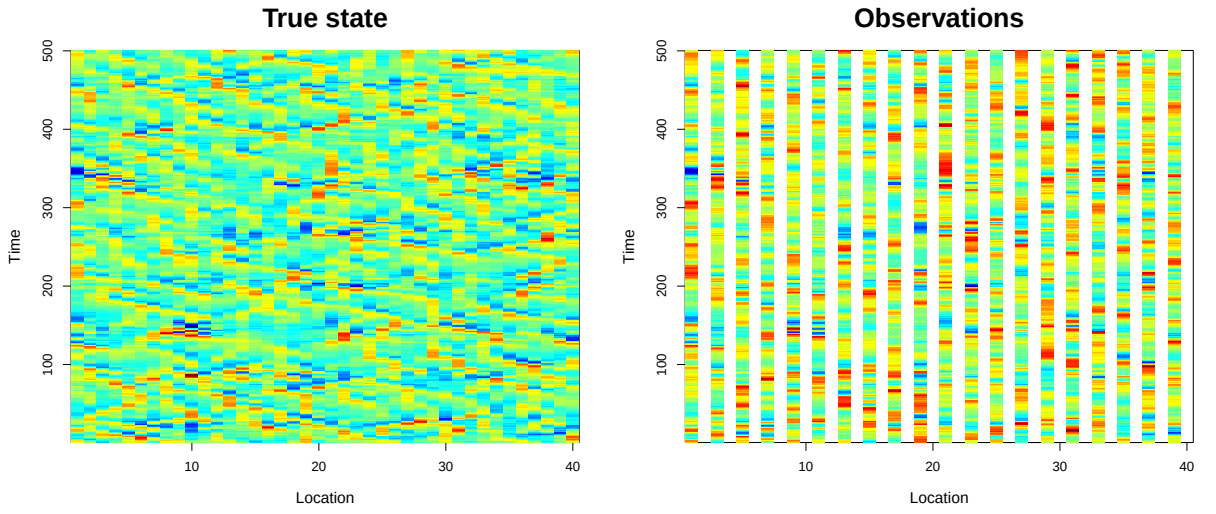

FIG. 2: True spatio-temporal evolution of the system (on the left) with the partial and noisy observation data (on the right).

a. Results for the estimation of the model error

The aim is to retrieve $\boldsymbol{x}_t$ and the parameters of $\boldsymbol{Q}_t$ for $t \in \{1,...,T\}$

$$\lambda_{\boldsymbol{Q},t} = 1 + 0.5 \sin\left(\frac{t}{10}\right),$$
$$l_{\boldsymbol{Q},t} = \sqrt{3 + 2\cos\left(\frac{t}{20}\right)},$$

assuming that $\boldsymbol{R}_t = 0.1\boldsymbol{I}$ is known. The temporal evolution of $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ is shown in Figure 2.

The PF-EnKF algorithm detailed in Section a is run with the following implementation. The number of members is $N_{memb} = 100$ and the number of particles is $N_{\boldsymbol{\theta}} = 100$. At $t = 1$, the members are initialized using a Gaussian distribution with mean $\boldsymbol{x}_0$ and covariance matrix $\boldsymbol{Q}(\boldsymbol{\theta}_{\boldsymbol{Q},0})$ where $\boldsymbol{\theta}_{\boldsymbol{Q},0} = (0.5,0.5)$. The initial particles $\boldsymbol{\theta}^{a,j}_{\boldsymbol{Q},1} = (\lambda^{a,j}_{\boldsymbol{Q},1}, l^{a,j}_{\boldsymbol{Q},1})$ are simulated using independent uniform distributions on the intervals $]0,1]\times]0,1]$ for $j \in \{1,...,N_{\boldsymbol{\theta}}\}$. For $t \geq 2$, the forecasted particles are simulated using the random walk

$$\boldsymbol{\theta}^{f,j}_{\boldsymbol{Q},t} = \max(\boldsymbol{\theta}^{a,j}_{\boldsymbol{Q},t-1} + \tilde{\boldsymbol{\theta}}^{j}_{\boldsymbol{Q},t}, \boldsymbol{\theta}_{min})$$

which is bounded by $\boldsymbol{\theta}_{min} = (10^{-4}, 10^{-4})$ to avoid negative values for the parameters. The additive Gaussian white noise $\tilde{\boldsymbol{\theta}}^j_{\boldsymbol{Q},t}$ avoids the degeneracy problem of the PF when the particles concentrate on a single particle over time. The covariance matrix of $\tilde{\boldsymbol{\theta}}^j_{\boldsymbol{Q},t}$ is

$$\boldsymbol{C} = \begin{pmatrix} \sigma^2_{\lambda,\boldsymbol{Q}} & 0 \\ 0 & \sigma^2_{l,\boldsymbol{Q}} \end{pmatrix}$$

and depends on hyperparameters set to $\sigma_{\lambda,\boldsymbol{Q}} = 0.1$ and $\sigma_{l,\boldsymbol{Q}} = 0.1$. Appendix C shows that our algorithm is not very sensitive to the choice of these hyperparameters.

Figure 3 compares the temporal evolutions of the true stochastic parameters with the ones of the estimates obtained by the PF-EnKF. It indicates that our method is able to track the time-evolution of the true stochastic parameters, even if the initial parameter value is ill-chosen. The algorithm provides an accurate estimate for the variance parameter $\lambda_{\boldsymbol{Q},t}$, whereas the estimate of the spatial scale parameter $l_{\boldsymbol{Q},t}$ seems to be delayed in time compared with the true parameter evolution. It suggests that the likelihood is more sensitive to the value of $\lambda_{\boldsymbol{Q},t}$ compared to $l_{\boldsymbol{Q},t}$. This appears to be linked to the observation density, since the time delay is significantly reduced when the number of observed grid points is increased to $p = 40$ instead of $p = 20$ (see Figure 4).

Focusing now on the state estimation, our algorithm is compared to the EnKF with true $\boldsymbol{Q}_t$. This latter uses the theoretical covariance matrix of the forecasted members $\boldsymbol{P}^f_t = \boldsymbol{P}^p_t + \boldsymbol{Q}_t$, instead of the empirical one, to reduce the impact of the sampling error as our algorithm within Eq. (15). The experiment is repeated 10 times with the same true states and observations for the two algorithms, only the stochastic parts of the EnKF and PF vary from a trial to another. Their global root mean square error (RMSE) and coverage probability are computed for each experiment. The global RMSE is given by

$$\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{nN_{memb}}\sum_{i=1}^{N_{memb}}(\boldsymbol{x}^{a,i}_t - \boldsymbol{x}_t)^\top(\boldsymbol{x}^{a,i}_t - \boldsymbol{x}_t)}.$$

The coverage probability is defined as the empirical probability to have $\boldsymbol{x}_{t,k}$, with $k \in \{1, ..., n\}$, falling in the 95% confidence interval obtained by the algorithm. The global RMSE provides

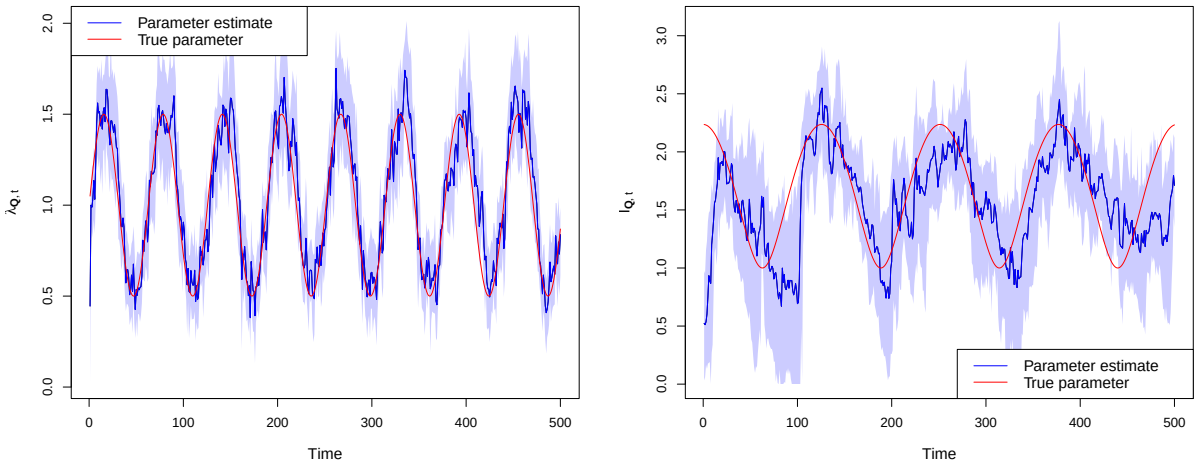

FIG. 3: Estimation of the variance parameter $\lambda_{\boldsymbol{Q},t}$ and spatial scale parameter $l_{\boldsymbol{Q},t}$ with the PF-EnKF. For each panel, the 95% confidence interval is plotted using the quantiles of the particle distribution.

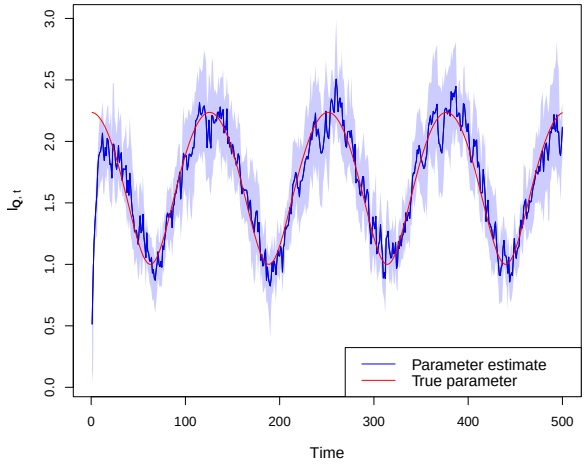

FIG. 4: Estimation of the spatial scale parameter $l_{\boldsymbol{Q},t}$ with the PF-EnKF when $p = n$ (the state is fully observed).

informations on the ensemble mean and the coverage probability on the ensemble spread. The results shown in Table 1 are quite similar for the two algorithms. As expected, the global RMSE of the PF-EnKF is larger than the one of the EnKF with true $\boldsymbol{Q}_t$, but the difference is small, indicating that the PF-EnKF is able to relevantly estimate both the state and stochastic parameters. The coverage probability of each algorithm is close to the optimal value 0.95, showing that they

Algorithm	Global RMSE	Coverage probability
PF-EnKF	1.19 ± 0.03	0.95 ± 0.01
EnKF with $\boldsymbol{Q}_t$	1.09 ± 0.01	0.94 ± 0.01

TABLE 1: Comparison between the PF-EnKF with estimated $\boldsymbol{Q}_t$ and the EnKF with true $\boldsymbol{Q}_t$. Each result corresponds to the average over the 10 experiments ± the standard deviation.

provide a good estimation of the uncertainty on the reconstructed state. The standard deviations computed for the global RMSE and coverage probability of our algorithm show its stability. Note that the numerical experiments that we performed show that increasing the number of particles $N_{\boldsymbol{\theta}}$ does not lead to better results. The PF-EnKF is thus able to retrieve the true state, estimating $\boldsymbol{Q}_t$ with few extra computational cost compared to the EnKF with true $\boldsymbol{Q}_t$.

b. Results for the estimation of the observation error

We obtained similar results as in Section a when using the PF-EnKF algorithm detailed in Appendix A for estimating time-varying parameters of $\boldsymbol{R}_t$ (results not shown). Here, to allow the comparison with Algorithm 2 of Frei and Künsch (2012), we consider static parameters of $\boldsymbol{R}_t$ to retrieve for $t \in \{1, ..., T\}$

$$\lambda_{\boldsymbol{R},t} = 2,$$
$$l_{\boldsymbol{R},t} = \sqrt{2}.$$

Moreover, the number of observed grid points $p = 10$ is lower than in Section a, in order to assess the efficiency of each algorithm when the observation density is low.

The PF-EnKF algorithm detailed in Appendix A is run with the following implementation. At $t = 1$, for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$, the initial particles $\boldsymbol{\theta}^{a,j}_{\boldsymbol{R},1} = (\lambda^{a,j}_{\boldsymbol{R},1}, l^{a,j}_{\boldsymbol{R},1})$ are simulated using independent uniform distributions on the intervals $]0, 2\lambda_{\boldsymbol{R},0}] \times]0, 2l_{\boldsymbol{R},0}]$ with $\boldsymbol{\theta}_{\boldsymbol{R},0} = (\lambda_{\boldsymbol{R},0}, l_{\boldsymbol{R},0}) = (0.05, 0.05)$. For $t \geq 2$, the forecasted particles $\boldsymbol{\theta}^{f,j}_{\boldsymbol{R},t}$ are simulated as in Section a using the bounded random walk for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$

$$\boldsymbol{\theta}^{f,j}_{\boldsymbol{R},t} = \max(\boldsymbol{\theta}^{a,j}_{\boldsymbol{R},t-1} + \tilde{\boldsymbol{\theta}}^{j}_{\boldsymbol{R},t}, \boldsymbol{\theta}_{min})$$

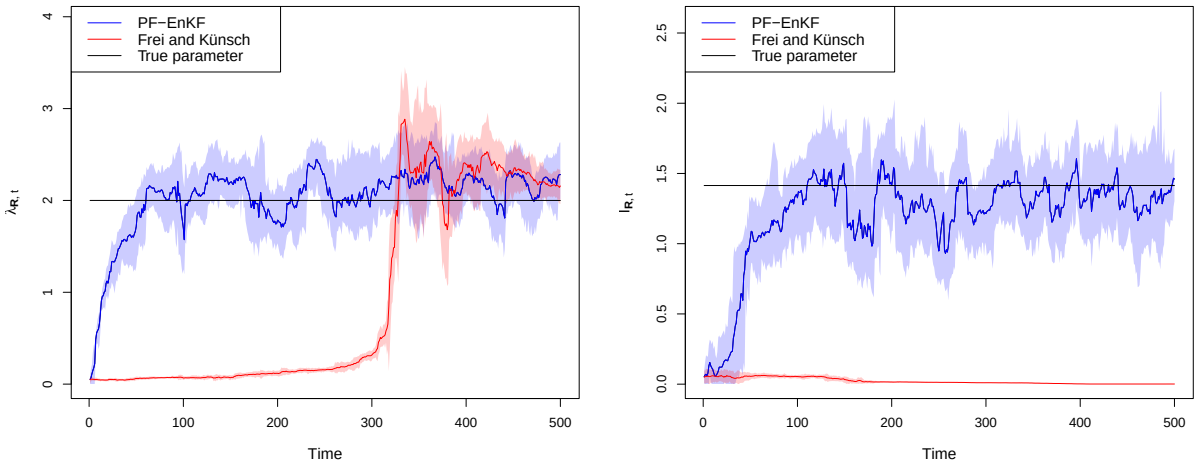

FIG. 5: Estimation of the variance parameter $\lambda_{\boldsymbol{R},t}$ and spatial scale parameter $l_{\boldsymbol{R},t}$ with the PF-EnKF and Algorithm 2 of Frei and Künsch (2012). For each panel, the 95% confidence interval of each algorithm is plotted.

with the additive noise $\tilde{\boldsymbol{\theta}}^j_{\boldsymbol{R},t} \sim \mathcal{N}(\mathbf{0}, 0.0025\boldsymbol{I})$.

The PF-EnKF is compared with Algorithm 2 of Frei and Künsch (2012) which uses kernel resampling with shrinkage for updating the particles. Both algorithms are run with $N_{memb} = 100$, $N_{\boldsymbol{\theta}} = 100$ and $\boldsymbol{Q}_t = \boldsymbol{I}$. Their estimates for the parameters of $\boldsymbol{R}_t$ are shown on Figure 5. The initial parameters' values are close to 0 as a consequence of the small values chosen for $\boldsymbol{\theta}_{\boldsymbol{R},0}$. This leads to a poor estimation of $\boldsymbol{\theta}_{\boldsymbol{R},t}$ in the first iterations of both algorithms. Then, after about 50 time steps, the PF-EnKF is able to correct this bad initial estimation and identify parameters' values close to the true values. Algorithm 2 of Frei and Künsch (2012) is able to retrieve the variance parameter but with an important delay compared with the PF-EnKF. However, it is not able to retrieve the spatial scale parameter and provides estimates close to 0 even after 500 iterations. Figure 6 shows the results obtained when using more realistic values for $\boldsymbol{\theta}_{\boldsymbol{R},0}$, in this case Algorithm 2 of Frei and Künsch (2012) seems to perform as well as the PF-EnKF. This experiment highlights that our algorithm is less sensitive to the initialization and as a consequence may be more suitable for estimating the parameters of $\boldsymbol{R}_t$ in an adaptive way.

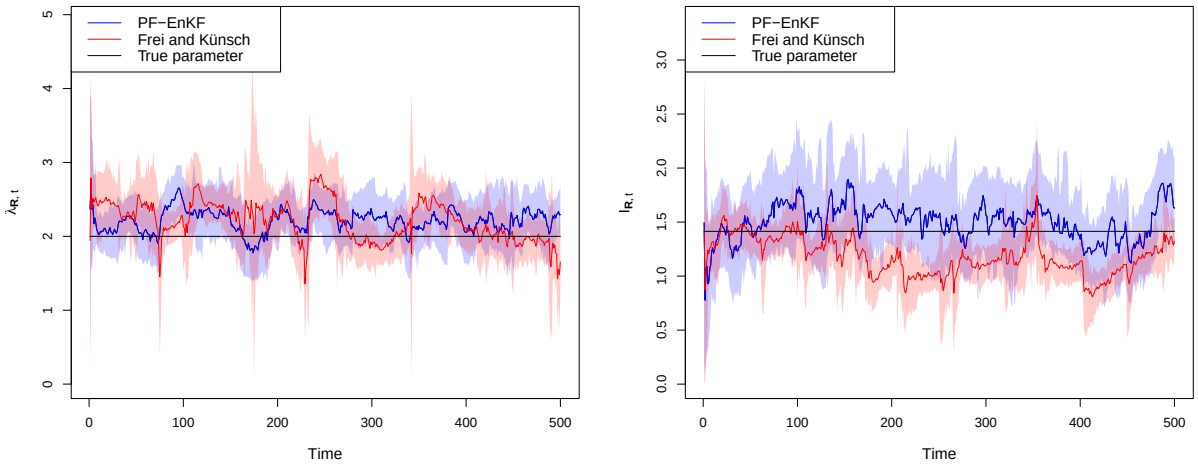

FIG. 6: As in Figure 5 but with $\boldsymbol{\theta}_{\boldsymbol{R},0} = (2,1.5)$.

Algorithm	Global RMSE	Coverage probability
PF-EnKF with good initialization	4.68 ± 0.04	0.95 ± 0.01
PF-EnKF with bad initialization	4.69 ± 0.04	0.92 ± 0.01
Frei and Künsch with good initialization	4.70 ± 0.06	0.95 ± 0.01
Frei and Künsch with bad initialization	4.74 ± 0.06	0.25 ± 0.22
EnKF with $\boldsymbol{R}_t$	4.68 ± 0.06	0.94 ± 0.01

TABLE 2: Comparison between the PF-EnKF, Algorithm 2 of Frei and Künsch (2012) and the EnKF with true $\boldsymbol{R}_t$, for good and bad initializations. Each result corresponds to the average over the 10 experiments ± the standard deviation.

Focusing now on the state estimation as in Section a, our algorithm is compared with Algorithm 2 of Frei and Künsch (2012) and the EnKF with true $\boldsymbol{R}_t$. The PF-EnKF and Algorithm 2 of Frei and Künsch (2012) are run with a good initialization: $\boldsymbol{\theta}_{\boldsymbol{R},0} = (2,1.5)$, and with a bad initialization: $\boldsymbol{\theta}_{\boldsymbol{R},0} = (0.05,0.05)$. The global RMSE and coverage probability of each algorithm are shown in Table 2. When the initialization is good, the algorithms have similar results. However, when the initialization is bad, the global RMSE of Algorithm 2 of Frei and Künsch (2012) is slightly higher and its coverage probability is low because $\boldsymbol{R}_t$ is most of the time underestimated (see Figure 5), whereas the results of the PF-EnKF remain close to the ones of the EnKF with true $\boldsymbol{R}_t$. Moreover, the standard deviations computed for the global RMSE and coverage probability of our algorithm show its stability. The PF-EnKF is thus able to retrieve the true state estimating $\boldsymbol{R}_t$ more efficiently than Algorithm 2 of Frei and Künsch (2012), which is more sensitive to the inialization.

c. Results for the PF-EnKF with inflation and localization

The PF-EnKF algorithm detailed in Section b is used with $N_{\boldsymbol{\theta}} = 100$ and $N_{memb} = 10$ which is smaller than in Sections a and b to have an important sampling error to treat. To do this, we perform the simultaneous inflation and localization of $\boldsymbol{P}_t^f$ with $\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t})$ whose inflation parameter $\lambda_{\boldsymbol{L},t}$ and localization parameter $l_{\boldsymbol{L},t}$ are estimated by the PF-EnKF. The error covariance matrices $\boldsymbol{Q}_t = \boldsymbol{I}$ and $\boldsymbol{R}_t = \boldsymbol{I}$ are assumed to be known. At $t = 1$, for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$, the initial particles $\boldsymbol{\theta}_{\boldsymbol{L},1}^{a,j} = (\lambda_{\boldsymbol{L},1}^{a,j}, l_{\boldsymbol{L},1}^{a,j})$ are simulated using independent uniform distributions on the intervals $]0, 2\lambda_{\boldsymbol{L},0}]\times]0, 2l_{\boldsymbol{L},0}]$ with $\boldsymbol{\theta}_{\boldsymbol{L},0} = (\lambda_{\boldsymbol{L},0}, l_{\boldsymbol{L},0}) = (0.5, 2.5)$. For $t \geq 2$, the forecasted particles $\boldsymbol{\theta}_{\boldsymbol{L},t}^{f,j}$ are simulated as in Sections a and b using the bounded random walk for $j \in \{1, ..., N_{\boldsymbol{\theta}}\}$

$$\boldsymbol{\theta}_{\boldsymbol{L},t}^{f,j} = \max(\boldsymbol{\theta}_{\boldsymbol{L},t-1}^{a,j} + \tilde{\boldsymbol{\theta}}_{\boldsymbol{L},t}^{j}, \boldsymbol{\theta}_{min})$$

with the additive noise $\tilde{\boldsymbol{\theta}}_{\boldsymbol{L},t}^{j} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$ where

$$\boldsymbol{C} = \begin{pmatrix} 0.001 & 0 \\ 0 & 0.1 \end{pmatrix}.$$

The goal is to retrieve $\boldsymbol{x}_t$ for $t \in \{1, ..., T\}$ and to estimate the inflation and localization parameters. The PF-EnKF for inflation and localization is compared to the combination of Desroziers-based inflation and grid search for localization in the EnKF (which also uses $N_{memb} = 10$). More precisely, the time-varying inflation parameter λ_t is estimated using the innovation statistics of Desroziers et al. (2005) with a temporal smoothing to obtain for $t \geq 2$, $\lambda_2 = 1$ and $v_{min} = 10^{-4}$

$$\lambda_{t+1} = \max(\rho\tilde{\lambda}_t + (1-\rho)\lambda_t, v_{min})$$
$$\text{with } \tilde{\lambda}_t = \frac{(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_t^f)^\top(\boldsymbol{y}_t - \boldsymbol{H}\boldsymbol{x}_t^f) - \mathrm{Tr}(\boldsymbol{R}_t)}{\mathrm{Tr}(\boldsymbol{H}\boldsymbol{P}_t^f\boldsymbol{H}^\top)}.$$

The smoothing parameter is set to $\rho = 0.05$. For the localization, assuming that the localization parameter l in GC(l) is constant in time, a grid search is performed using the global RMSE on the true state, although such an approach is not possible in practical applications since the true state is unknown. The resulting estimate is hereafter referred to as "optimal localization".

Algorithm	Global RMSE	Coverage probability
PF-EnKF	2.26 ± 0.06	0.88 ± 0.01
Desroziers-based inflation and optimal localization	2.17 ± 0.04	0.87 ± 0.01

TABLE 3: Comparison between the PF-EnKF for inflation and localization and the combination of the Desroziers-based inflation with the optimal localization when $N_{memb} = 10$.

In the first instance, both approaches are compared with respect to their ability to retrieve the true state by repeating 10 times the experiment with the same true states and observations (as for Sections a and b). The global RMSE and coverage probability shown in Table 3 are almost the same between the two algorithms. The localization has a significant effect on the global RMSE (by reducing the sampling error impact) and the inflation on the coverage probability (by adjusting the ensemble spread). The concurrent method, based on Desroziers inflation and optimal localization, is favoured in comparison with the PF-EnKF because it assumes that the true state is known for calibrating the localization parameter, whereas our algorithm only uses the observations. However, the PF-EnKF has a lower computational cost than the concurrent method because the grid search used for the localization parameter is very expensive. Moreover, good results were also obtained by our method with a lower computational cost when a smaller number of particles $N_{\boldsymbol{\theta}} = 10$ was used. The standard deviations computed for the global RMSE and coverage probability of the PF-EnKF show, as for the concurrent method, that our algorithm is stable. The PF-EnKF with inflation and localization seems thus to efficiently retrieve the true state when a small ensemble size N_{memb} is used. This is also shown in Figure 7 where the estimate of the first state variable $\boldsymbol{x}_{t,1}$ is close to the truth for both algorithms.

In the second instance, the temporal evolutions of the estimates of the localization and inflation parameters are shown for both methods on Figure 8. For the inflation parameter, the estimate of the PF-EnKF is quite close to the Desroziers-based inflation. Note that the PF-EnKF can be seen as a likelihood-based method whereas the Desroziers-based inflation is a moment-based method as explained by Tandeo et al. (2020). For the localization parameter, the estimate of the PF-EnKF varies around the optimal value $l = 1$ obtained by grid search.

A sensitivity analysis is performed to further investigate the properties of the estimate of the localization parameter. Figure 9 shows the distribution of the values taken over time by the estimate

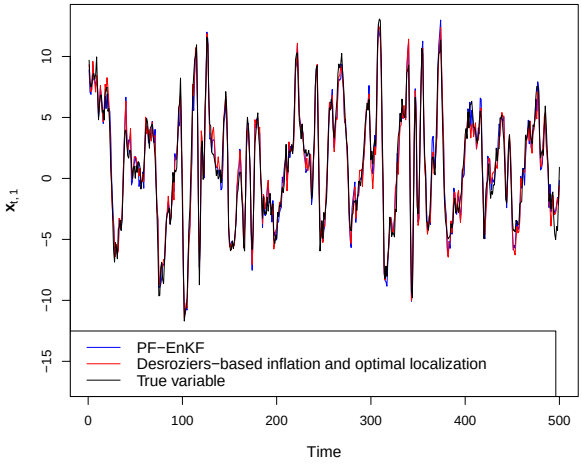

FIG. 7: Temporal evolution of the estimates of $\boldsymbol{x}_{t,1}$ obtained with the PF-EnKF with inflation and localization and the Desroziers-based inflation with optimal localization.

of the localization parameter of the PF-EnKF for different ensemble sizes N_{memb}. These values are mostly close to the optimal localization and increase, as expected, when the ensemble size is larger. In the same manner but without inflation to better see the localization sensitivity, the left panel of Figure 10 shows that our localization parameter estimate depends on the number of observed grid points p, allowing to handle a time-dependent observation operator $\boldsymbol{H}_t$. The same experiment is then led with a time-varying covariance matrix $\boldsymbol{R}_t$ and $p = 40$. The right panel of Figure 10 shows that our localization parameter estimate also depends on the observation error. All these results are in accordance with the ones of Ying et al. (2018) who illustrated that the optimal value of the localization parameter notably depends on the ensemble size, the observation density (given by p) and the observation error.

The PF-EnKF is thus able to relevantly estimate both time-dependent inflation and localization parameters.

4. Conclusion and perspectives

A new combination of the EnKF with the PF has been developed in this paper to simultaneously estimate the latent true state and time-dependent stochastic parameters of a data assimilation system. To achieve this, the PF generates the particles that estimate the vector of stochastic parameters, then the mean particle is used within the EnKF to generate the members that estimate the state.

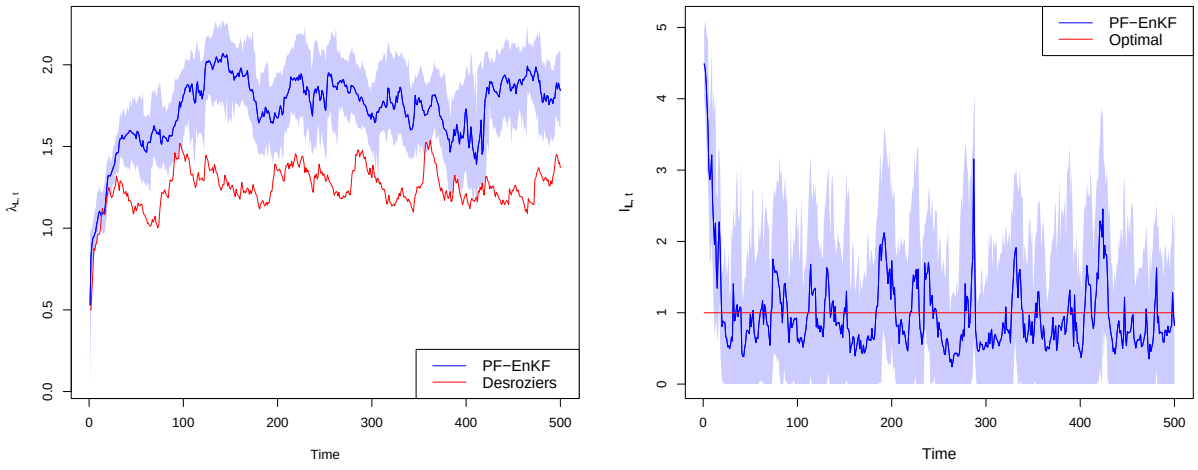

FIG. 8: Comparison of the estimates of the inflation (on the left) and localization (on the right) parameters of the PF-EnKF with the Desroziers-based inflation and the optimal localization. For the PF-EnKF, the 95% confidence interval is plotted for each parameter.

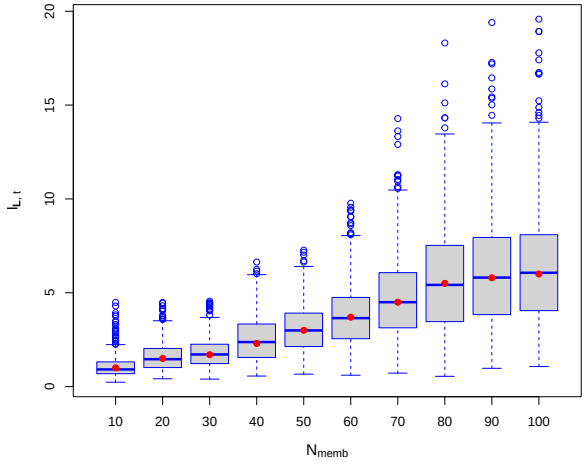

FIG. 9: Comparison of the values of the estimate of the PF-EnKF localization parameter (blue boxplot) with the optimal localization (red point) for different ensemble sizes N_{memb}.

The generic PF-EnKF algorithm, easy to implement, has been detailed for estimating $\boldsymbol{Q}_t$, $\boldsymbol{R}_t$, or both time-dependent inflation and localization parameters. The numerical results have shown that our algorithm is able to stably retrieve the latent state and stochastic parameters related to $\boldsymbol{Q}_t$ or $\boldsymbol{R}_t$ with a reasonable computational cost. The experiments have also shown that the PF-EnKF provides good state estimates even when a small ensemble size is used, thanks to its ability to

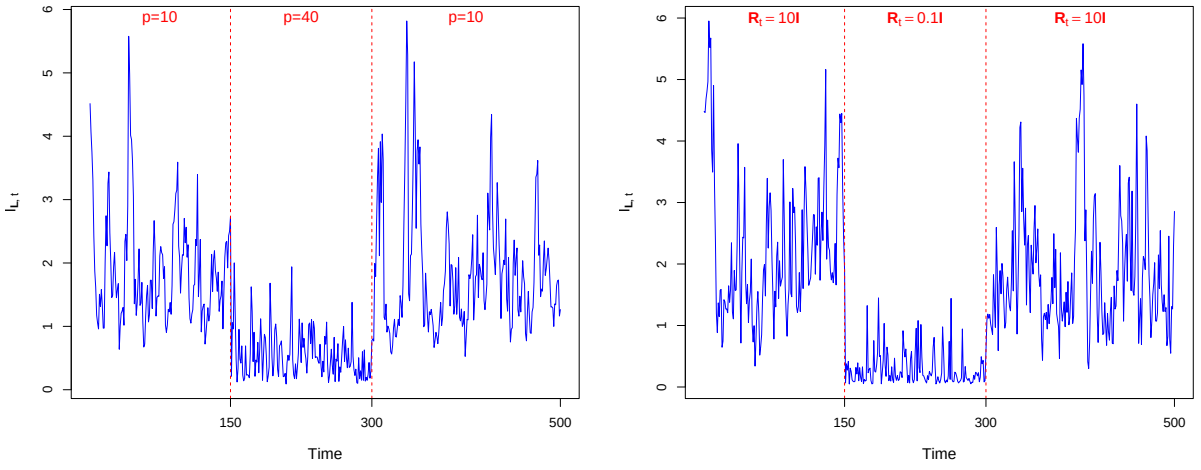

FIG. 10: Temporal evolution of the estimate of the PF-EnKF localization parameter according to the number of observed grid points p (on the left) and according to $\boldsymbol{R}_t$ (on the right).

estimate time-dependent inflation and localization parameters with a competitive computational cost compared to existing methods. Also, the localization parameter estimate obtained by our method has turned out to depend on the ensemble size, the observation density and the observation error.

In a future work, the PF-EnKF could be used for estimating more complex parametric forms for $\boldsymbol{Q}_t$ or $\boldsymbol{R}_t$, with a reasonable number of parameters to estimate (because of the curse of dimensionality). Moreover it may be applied to data-driven data assimilation, for instance when unresolved equations of the dynamical model are emulated using machine learning as in Brajard et al. (2021). Our algorithm could also be applied to more realistic settings. An intermediate step may consist in performing numerical experiments using an intermediate complexity atmospheric general circulation model such as SPEEDY (see Molteni 2003).

Acknowledgments. We thank the two anonymous Reviewers and the Editor in Chief for their helpful and constructive comments on the earlier version of this paper.

APPENDIX A

PF-EnKF for estimating the observation error

The aim is to retrieve the true state $\boldsymbol{x}_t$ and the vector of stochastic parameters $\boldsymbol{\theta}_{\boldsymbol{R},t}$ related to the observation error, assuming that $\boldsymbol{Q}_t$ is known. For this purpose, the same methodology as in Section a is applied for the observation error and detailed in Algorithm 3. The latter is different from Algorithm 2 of Frei and Künsch (2012) where the EnKF is used before the PF in the cycle. Moreover, their EnKF works with the estimate of $\boldsymbol{R}$ obtained at the previous cycle and their PF is based on a ratio of likelihoods.

Algorithm 3 PF-EnKF for estimating $\boldsymbol{R}_t$

Require: $\boldsymbol{Q}_t$, $\boldsymbol{R}(\cdot)$ and $\boldsymbol{\theta}_{\boldsymbol{R},0}$

for $i = 1, \ldots, N_{memb}$ **do**
  $\boldsymbol{\eta}_1^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_1)$
  $\boldsymbol{x}_1^{a,i} = \boldsymbol{x}_0 + \boldsymbol{\eta}_1^i$
end for
for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
  $\boldsymbol{\theta}_{\boldsymbol{R},1}^{a,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{R},1} | \boldsymbol{\theta}_{\boldsymbol{R},0})$
end for
for $t = 2, \ldots, T$ **do**
  for $i = 1, \ldots, N_{memb}$ **do**
    $\boldsymbol{x}_t^{p,i} = M(\boldsymbol{x}_{t-1}^{a,i})$
    $\boldsymbol{\eta}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$
    $\boldsymbol{x}_t^{f,i} = \boldsymbol{x}_t^{p,i} + \boldsymbol{\eta}_t^i$
  end for
  $\boldsymbol{x}_t^p = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{p,i}$
  $\boldsymbol{P}_t^p = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} (\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)(\boldsymbol{x}_t^{p,i} - \boldsymbol{x}_t^p)^\top$
  $\boldsymbol{P}_t^f = \boldsymbol{P}_t^p + \boldsymbol{Q}_t$
  for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
    $\boldsymbol{\theta}_{\boldsymbol{R},t}^{f,j} \sim p(\boldsymbol{\theta}_{\boldsymbol{R},t} | \boldsymbol{\theta}_{\boldsymbol{R},t-1}^{a,j})$
  end for
  for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
    $\gamma_t^j = \frac{\phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^p, \boldsymbol{H}\boldsymbol{P}_t^f\boldsymbol{H}^\top + \boldsymbol{R}(\boldsymbol{\theta}_{\boldsymbol{R},t}^{f,j}))}{\sum_{k=1}^{N_{\boldsymbol{\theta}}} \phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^p, \boldsymbol{H}\boldsymbol{P}_t^f\boldsymbol{H}^\top + \boldsymbol{R}(\boldsymbol{\theta}_{\boldsymbol{R},t}^{f,k}))}$

end for

resample the forecasted particles $\boldsymbol{\theta}_{\boldsymbol{R},t}^{f,j}$ using the weights γ_t^j to obtain the analyzed particles $\boldsymbol{\theta}_{\boldsymbol{R},t}^{a,j}$

$\boldsymbol{\theta}_{\boldsymbol{R},t}^{a} = \frac{1}{N_{\boldsymbol{\theta}}} \sum_{j=1}^{N_{\boldsymbol{\theta}}} \boldsymbol{\theta}_{\boldsymbol{R},t}^{a,j}$

$\boldsymbol{K}_t = \boldsymbol{P}_t^f \boldsymbol{H}^\top \left(\boldsymbol{H} \boldsymbol{P}_t^f \boldsymbol{H}^\top + \boldsymbol{R}(\boldsymbol{\theta}_{\boldsymbol{R},t}^{a}) \right)^{-1}$

for $i = 1, \ldots, N_{memb}$ **do**

$\boldsymbol{\varepsilon}_t^i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\theta}_{\boldsymbol{R},t}^{a}))$

$\boldsymbol{x}_t^{a,i} = \boldsymbol{x}_t^{f,i} + \boldsymbol{K}_t(\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t^i - \boldsymbol{H}\boldsymbol{x}_t^{f,i})$

end for

end for

APPENDIX B

Likelihood as a criterion for the joint estimation of inflation and localization parameters: an illustrative example

The stochastic parameters are estimated using the likelihood in Eq. (20) of Algorithm 2. To assess the relevance of using the likelihood as a criterion, a simple study is led by comparing the use of the likelihood to the use of the root mean square error (RMSE) for estimating both inflation and localization parameters. In this illustrative example the true state is supposed to be known in order to be able to compute the RMSE, although it is not the case for real applications.

We focus on the following state-space model for $t \in \{1, ..., T = 10^4\}$:

$$\begin{cases} \boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{P}), \\ \boldsymbol{y}_t \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{x}_t, \boldsymbol{R}). \end{cases}$$

The true state $\boldsymbol{x}_t$ is in $\mathbb{R}^n$ and its $n = 20$ components correspond to equally spaced grid points on a circle. It is a Gaussian white noise with covariance matrix for $(k, k') \in \{1, ..., n\}^2$

$$\boldsymbol{P}[k, k'] = \alpha \, \mathrm{e}^{-\frac{d(k,k')}{\beta}}$$

where $d(k,k') = \min(|k-k'|, n-|k-k'|)$ is the distance between two grid points on the circle, $\alpha = 2$ and $\beta = 5$. One component of $\boldsymbol{x}_t$ out of two is observed such as $\boldsymbol{y}_t$ is in $\mathbb{R}^{10}$ and a Gaussian white noise with covariance matrix $\boldsymbol{R} = \boldsymbol{I}$ is added to $\boldsymbol{y}_t$.

In our experiment summarized in Algorithm 4, we consider that α is unknown and $\widehat{\boldsymbol{P}}$ is used instead of $\boldsymbol{P}$ to generate the members that estimate $\boldsymbol{x}_t$ with for $(k,k') \in \{1,...,n\}^2$

$$\widehat{\boldsymbol{P}}[k,k'] = \mathrm{e}^{-\frac{d(k,k')}{\beta}}.$$

Covariance inflation is led to compensate the misspecification of $\widehat{\boldsymbol{P}}$. Moreover, the ensemble size $N_{memb} = 20$ is small, so covariance localization is used to reduce the sampling error impact. The covariance inflation and localization are practiced jointly following the strategy shown in Section b through $\boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L},t}) \circ \boldsymbol{P}_t^f$. At each time t the same grids of possible values for the unknown stochastic parameters are used: $\{1, 1.25, ..., 3\}$ for the inflation parameter $\lambda_{\boldsymbol{L},t}$ and $\{2, 3, ..., 10\}$ for the localization parameter $l_{\boldsymbol{L},t}$. Then the log-likelihood and RMSE related to each couple of possible values $\boldsymbol{\theta}_{\boldsymbol{L}}^j$ are computed in Eqs. (11)-(12) of Algorithm 4 at every time t. Finally, for each couple $\boldsymbol{\theta}_{\boldsymbol{L}}^j$, the temporal means of the log-likelihood and RMSE are computed, allowing to obtain Figure B1. The best couple that maximizes the likelihood is close to the best couple that minimizes the RMSE. The likelihood is able to identify similar optimal values for the localization and inflation parameters compared with the ones obtained using the RMSE, while it seems to be less sensitive to the parameters' values. It is not surprising since it is based on partial and noisy observations, contrary to the RMSE that requires the true state. This suggests that the likelihood is a relevant criterion for estimating both inflation and localization parameters.

Algorithm 4 Estimation of the inflation and localization parameters using the likelihood and RMSE

Require: $\widehat{\boldsymbol{P}}$ and $\boldsymbol{R}$

for $t = 1, \ldots, T$ **do**
  for $i = 1, \ldots, N_{memb}$ **do**
    $\boldsymbol{x}_t^{f,i} \sim \mathcal{N}(\boldsymbol{0}, \widehat{\boldsymbol{P}})$
  end for
  $\boldsymbol{x}_t^f = \frac{1}{N_{memb}} \sum_{i=1}^{N_{memb}} \boldsymbol{x}_t^{f,i}$
  $\boldsymbol{P}_t^f = \frac{1}{N_{memb}-1} \sum_{i=1}^{N_{memb}} (\boldsymbol{x}_t^{f,i} - \boldsymbol{x}_t^f)(\boldsymbol{x}_t^{f,i} - \boldsymbol{x}_t^f)^\top$
  for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
    $\boldsymbol{P}_t^{f,j} = \boldsymbol{L}(\boldsymbol{\theta}_{\boldsymbol{L}}^j) \circ \boldsymbol{P}_t^f$
    $\boldsymbol{K}_t^j = \boldsymbol{P}_t^{f,j} \boldsymbol{H}^\top (\boldsymbol{H} \boldsymbol{P}_t^{f,j} \boldsymbol{H}^\top + \boldsymbol{R})^{-1}$
    $\boldsymbol{x}_t^{a,j} = \boldsymbol{x}_t^f + \boldsymbol{K}_t^j (\boldsymbol{y}_t - \boldsymbol{H} \boldsymbol{x}_t^f)$
    $\text{Likelihood}_t^j = \log\left(\phi(\boldsymbol{y}_t; \boldsymbol{H}\boldsymbol{x}_t^f, \boldsymbol{H}\boldsymbol{P}_t^{f,j}\boldsymbol{H}^\top + \boldsymbol{R})\right)$
    $\text{RMSE}_t^j = \sqrt{\frac{1}{n}(\boldsymbol{x}_t^{a,j} - \boldsymbol{x}_t)^\top (\boldsymbol{x}_t^{a,j} - \boldsymbol{x}_t)}$
  end for
end for
for $j = 1, \ldots, N_{\boldsymbol{\theta}}$ **do**
  $\text{Likelihood}^j = \frac{1}{T} \sum_{t=1}^{T} \text{Likelihood}_t^j$
  $\text{RMSE}^j = \frac{1}{T} \sum_{t=1}^{T} \text{RMSE}_t^j$
end for

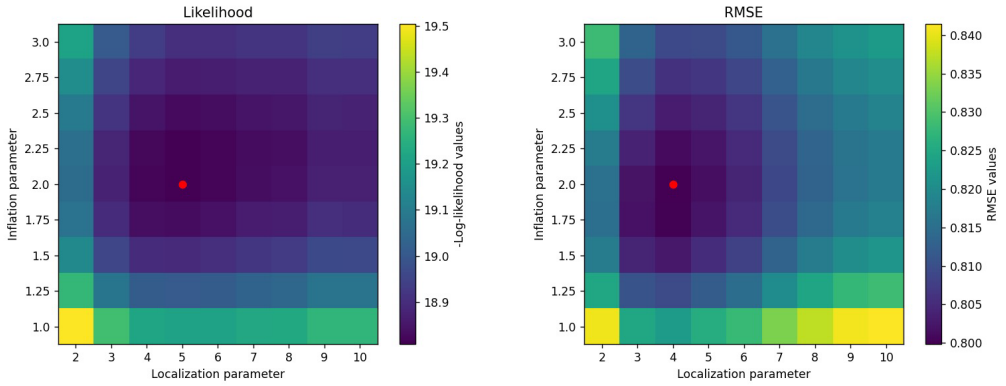

FIG. B1: Comparison of the couples of parameters' values (each represented by a pixel) using the likelihood (on the left) and the RMSE (on the right). The best couple is identified by a red point for each method.

APPENDIX C

RMSE sensitivity to the hyperparameters $\sigma_{\lambda,Q}$ and $\sigma_{l,Q}$ of the PF-EnKF for estimating the model error

The same experiment as in Section a (with the same true states and observations) is repeated for different values of $\sigma_{\lambda,Q}$ first and $\sigma_{l,Q}$ then. For each hyperparameter value, the RMSE with respect to the true state is computed at every time t and the distribution of these RMSE values is represented by a boxplot in Figure C1. This latter shows that the RMSE is not very sensitive to the values of the hyperparameters $\sigma_{\lambda,Q}$ and $\sigma_{l,Q}$.

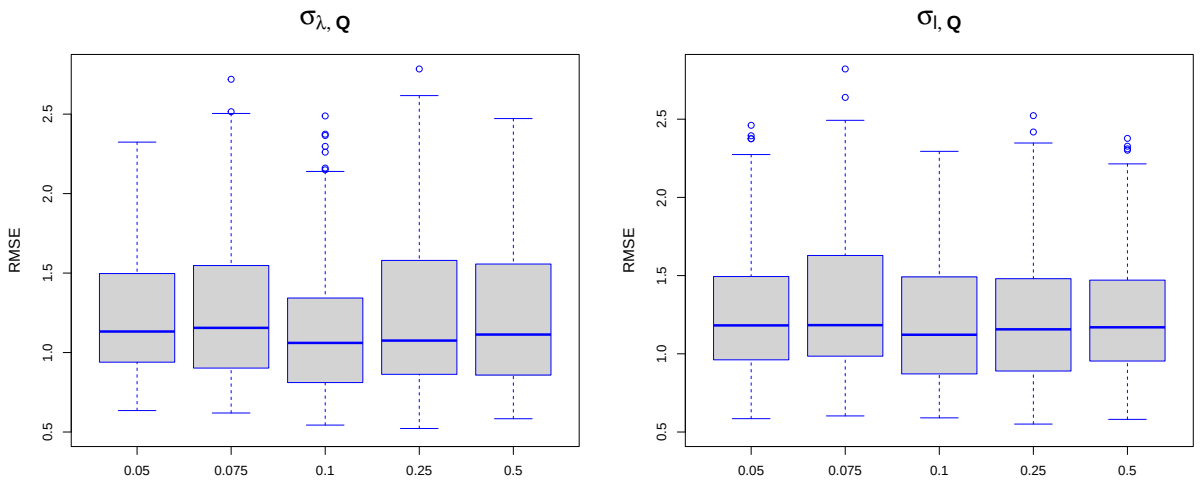

FIG. C1: RMSE values over time of the PF-EnKF when $\boldsymbol{Q}_t$ is estimated, for different values of $\sigma_{\lambda,\boldsymbol{Q}}$ (on the left) and $\sigma_{l,\boldsymbol{Q}}$ (on the right).

References

Ait-El-Fquih, B., and I. Hoteit, 2020: A particle-filter based adaptive inflation scheme for the ensemble kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **146 (727)**, 922–937.

Anderson, J., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic meteorology and oceanography*, **61 (1)**, 72–83.

Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, **129 (12)**, 2884–2903.

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

Anderson, J. L., and S. L. Anderson, 1999: A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly weather review*, **127 (12)**, 2741–2758.

Brajard, J., A. Carrassi, M. Bocquet, and L. Bertino, 2021: Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A*, **379 (2194)**, 20200 086.

Burgers, G., P. J. Van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble kalman filter. *Monthly weather review*, **126 (6)**, 1719–1724.

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

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly weather review*, **123 (4)**, 1128–1145.

DelSole, T., and X. Yang, 2010: State and parameter estimation in stochastic dynamical models. *Physica D: Nonlinear Phenomena*, **239 (18)**, 1781–1788.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **131 (613)**, 3385–3396.

El Gharamti, M., 2018: Enhanced adaptive inflation algorithm for ensemble filters. *Monthly Weather Review*, **146 (2)**, 623–640.

Frei, M., and H. R. Künsch, 2012: Sequential state and observation noise covariance estimation using combined ensemble kalman and particle filters. *Monthly Weather Review*, **140 (5)**, 1476–1495.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757.

Guillot, J., E. Frénod, and P. Ailliot, 2022: Physics informed model error for data assimilation. *Discrete and Continuous Dynamical Systems-Series S*.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129 (1)**, 123–137.

Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144 (713)**, 1257–1278.

Lguensat, R., P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, 2017: The analog data assimilation. *Monthly Weather Review*, **145 (10)**, 4093–4107.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on predictability*, Vol. 1.

Miyoshi, T., E. Kalnay, and H. Li, 2013: Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, **21 (3)**, 387–398.

Molteni, F., 2003: Atmospheric simulations using a gcm with simplified physical parametrizations. i: Model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20**.

Papadakis, N., É. Mémin, A. Cuzol, and N. Gengembre, 2010: Data assimilation with the weighted ensemble kalman filter. *Tellus A: Dynamic Meteorology and Oceanography*, **62 (5)**, 673–697.

Rubin, D. B., 1988: Using the sir algorithm to simulate posterior distributions. *Bayesian statistics*, **3**, 395–402.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle–ensemble kalman filter for lagrangian data assimilation. *Monthly Weather Review*, **143 (1)**, 195–211.

Smith, P., G. Thornhill, S. Dance, A. Lawless, D. Mason, and N. Nichols, 2013: Data assimilation for state and parameter estimation: application to morphodynamic modelling. *Quarterly Journal of the Royal Meteorological Society*, **139 (671)**, 314–327.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, **136 (12)**, 4629–4640.

Stroud, J. R., M. Katzfuss, and C. K. Wikle, 2018: A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. *Monthly weather review*, **146 (1)**, 373–386.

Tandeo, P., P. Ailliot, M. Bocquet, A. Carrassi, T. Miyoshi, M. Pulido, and Y. Zhen, 2020: A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Monthly Weather Review*, **148 (10)**, 3973–3994.

Van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Monthly Weather Review*, **137 (12)**, 4089–4114.

Ying, Y., F. Zhang, and J. L. Anderson, 2018: On the selection of localization radius in ensemble filtering for multiscale quasigeostrophic dynamics. *Monthly Weather Review*, **146 (2)**, 543–560.