An Efficient State-Parameter Filtering Scheme Combining Ensemble Kalman and Particle Filters

Boujemaa Ait-El-Fquih*, and Ibrahim Hoteit†

King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Corresponding author address: Boujemaa Ait-El-Fquih, King Abdullah University of Science and Technology (KAUST), Division of Physical Science and Engineering, 23955-6900 Thuwal, Saudi Arabia.

E-mail: boujemaa.aitelfquih@kaust.edu.sa

†Current affiliation: King Abdullah University of Science and Technology (KAUST), Division of Physical Science and Engineering, 23955-6900 Thuwal, Saudi Arabia.

ABSTRACT

This work addresses the state-parameter filtering problem for dynamical systems with relatively large-dimensional state and low-dimensional parameters' vector. We propose a Bayesian filtering algorithm combining the strengths of the particle filter (PF) and the ensemble Kalman filter (EnKF). At each assimilation cycle of the proposed EnKF-PF, the PF is first used to sample the parameters' ensemble followed by the EnKF to compute the state ensemble conditional on the resulting parameters' ensemble. The proposed scheme is expected to be more efficient than the traditional state augmentation techniques, which suffer from the curse of dimensionality and inconsistency that is particularly pronounced when the state is a strongly nonlinear function of the parameters. In the new scheme, the EnKF and PF interact via their ensembles' members, in contrast with the recently introduced two-stage EnKF-PF (TS-EnKF-PF), which exchanges point estimates between EnKF and PF while requiring almost double the computational load. Numerical experiments are conducted with the Lorenz-96 model to assess the behavior of the proposed filter and to evaluate its performances against the joint PF, joint EnKF and TS-EnKF-PF. Numerical results suggest that the EnKF-PF performs best in all tested scenarios. It was further found to be more robust, successfully estimating both state and parameters in different sensitivity experiments.

1. Introduction

Data assimilation combines data and dynamical models to provide the best possible estimate of the state of the underlying system. This approach is often implemented sequentially, infering the state given the available observations and the knowledge of the system dynamics. This is known as the so-called *Bayesian filtering* problem. In a Bayesian framework, the posterior probability density function (pdf) of the state, also called analysis pdf, is needed to compute any type of state estimate (and higher moments), as for instance the posterior mean (PM), which minimizes the mean-squared error (MSE) (van Trees 1968). In a filtering algorithm, the analysis pdf is computed recursively following two steps (Künsch 2001): a forecast (or propagation) step that integrates the previous analysis pdf with the dynamical model to obtain the forecast pdf and an analysis (or update) step to update the forecast pdf with the incoming observation. One can also reverse the order of the propagation-update steps without "violating" the Bayesian formulation of the filtering problem, which results in another generic algorithm involving the one- (or two-) step-ahead smoothing pdf of the state in addition to the forecast and analysis pdfs (Desbouvries and Ait-El-Fquih 2008; Desbouvries et al. 2011; Ait-El-Fquih et al. 2016).

In practice, however, the optimal solution of the generic filtering algorithm, in the sense of MSE minimization, is generally not available except for linear-Gaussian systems for which this algorithm reduces to the famous Kalman filter (KF) (Kalman 1960; Jazwinski 1970; Anderson and Moore 1979). A number of suboptimal numerical methods have therefore been proposed, as for instance, the well-known sequential Monte Carlo (or particle filter, PF) method (Doucet et al. 2001a). The PF algorithm recursively computes an approximation of the analysis distribution as a random sample of state vectors, called particles. The forecast step propagates the particles forward

with the dynamical model to approximate the forecast distribution. These forecast particles are then weighted in the update step based on their relative likelihoods (Gordon et al. 1993).

The PF suffers from the weights' degeneracy problem in which all particles' weights, except very few, become almost zero after few assimilation cycles, which severly limits the filter's performance (Liu and Chen 1998; Doucet et al. 2001a; Snyder et al. 2008; van Leeuwen 2009). This phenomena is in part due to the fact that in the update step, the PF only uses the incoming observations to update the weights and not the states (Hoteit et al. 2008; van Leeuwen 2009; Hoteit et al. 2012). One could drastically increase the number of particles to avoid this phenomenon, but this would certainly lead to a prohibitive computational cost. The most standard solution to mitigate this phenomenon is to resample the particles by duplicating those with large weights and abandoning those with low weights (Rubin 1988; Gordon et al. 1993; Liu and Chen 1998; Doucet et al. 2001a). Although the PFs were shown to perform well in a number of low-dimensional systems (Kivman 2003; Subramanian et al. 2012), they nevertheless remain inefficient in large-dimensional systems due to the exorbitant number of particles needed to efficiently sample the state space (curse of dimensionality); in certain situations, the number of particles needed scales exponentially with the system dimension (Crisan and Doucet 2002; Snyder et al. 2008).

Despite the promising PF-type schemes that have recently been proposed to cope with the curse of dimensionality (see e.g. Spiller et al. (2008); Husz et al. (2011); Morzfelda et al. (2012); Ades and van Leeuwen (2013); Djuric and Bugallo (2013); Ait-El-Fquih and Hoteit (2015); Septier and Peters (2015); Ait-El-Fquih and Hoteit (2016)), the so-called ensemble KF (EnKF) (Evensen 1994, 2006; Hoteit et al. 2015) is still most commonly used for data assimilation into large-scale geophysical problems. The EnKF shares the same forecast step with the PF, but assumes the joint state-observation forecast pdf to be Gaussian at the update step. Indeed, it updates the particles (often called "ensemble members") following a KF-like correction based on stochastically per-

turbed observations. The KF-like corrections keep the particles close to the observations, which helps mitigating the risk of degeneracy (Kivman 2003; Hoteit et al. 2008). This leads to remarkable performances even when the filter is implemented with small ensembles, especially when equipped with localization and inflation techniques (Anderson and Anderson 1999; Hamill and Snyder 2000).

With more studies demonstrating the efficiency of the EnKF in various applications, other EnKF-like algorithms were introduced. These include derivation of deterministic EnKF variants, which were shown to be more robust with limited-size ensembles (e.g., Pham (2001); Bishop et al. (2001); Hoteit et al. (2002); Tippett et al. (2003); Hunt et al. (2007); Hoteit et al. (2015)); use of the less restrictive Gaussian mixture assumption (e.g., Hoteit et al. (2008); Stordal et al. (2011); Hoteit et al. (2012); Frei and Künsch (2013); Liu et al. (2015)); development of ensemble Kalman smoothing algorithms (e.g., Evensen and van Leeuwen (2000); Dunne et al. (2007)); and extending the EnKF to the framework of state-parameter filtering problems (e.g., Moradkhani et al. (2005); Annan et al. (2005); Chen and Zhang (2006); Aksoy et al. (2006); Bellsky et al. (2014); Rasmussen et al. (2015); Gharamti et al. (2014, 2015); Ait-El-Fquih et al. (2016)).

The state-parameter filtering problem consists of filtering systems where the state model is not perfectly known, and depends on a set of unknown (static) parameters. In such a problem, the estimation of the state requires that of the parameters. The EnKF-like methods that have been introduced to address this problem follow either the so-called joint or dual approach. The joint approach consists in gathering the state and parameters in the same vector and applying the standard EnKF to the resulting augmented system to simultaneously estimate the state and the parameters (Chen and Zhang 2006; Gharamti et al. 2015). In the dual approach, the state and the parameters are treated separately, starting with an update of the parameters followed by an update of the state (Moradkhani et al. 2005). The separation of the update steps was shown to provide more consis-

tent estimates of the parameters (see e.g., Wen and Chen (2007)). Despite its heuristic framework (Hendricks Franssen and Kinzelbach 2008), the dual EnKF was found to be more efficient than the Joint-EnKF at the cost of increased computational burden (Gharamti et al. 2015). Recently, Ait-El-Fquih et al. (2016) followed the concept of one-step-ahead smoothing in a fully Bayesian filtering framework to derive a new dual-like EnKF that performs better than the standard dual EnKF, while requiring almost no increase in the computational load.

While EnKFs are commonly used in large-scale (state and parameters) applications, we focus here on the situation in which the number of parameters to be estimated is small, so a PF does not require a prohibitive number of particles to sample the parameters space (provided that the number of observations is not too large[1]). We thus introduce a new filtering scheme that combines the PF, to sample the parameters' ensemble, with an EnKF, to sample the state ensemble conditionally on the parameters' ensemble. The use of a PF to estimate the parameters is intended to deal with the non-Gaussian character of the parameters posterior pdf (Kivman 2003). The benefit of the PF is expected to be more pronounced when the nonlinearity between the state and the parameters is pronounced, case in which the Joint-EnKF was shown to provide spurious cross-covariances between the state and the parameters, leading to poor performances (Carrassi and Vannitsem 2011; Santitissadeekorn and Jones 2015).

Among the proposed works that combine the EnKF and PF (see e.g., Hoteit et al. (2008, 2012); Frei and Künsch (2012); Slivinski et al. (2015); Liu et al. (2015); Zhang et al. (2016)), only that of Santitissadeekorn and Jones (2015) relates to our study, as it deals with dynamical systems in which the state model is large and depends on few unknown parameters. In that work, the

[1]For a narrow likelihood (i.e., that peaks in a very small portion of the observations space), the curse of dimensionality may also occur even if the parameters space is small, unless the amount of sampled particles is very large. As discussed in Bengtsson et al. (2008) (see also Snyder et al. (2008); van Leeuwen (2009)), such a scenario occurs especially when the data involve a large number of conditionally independent and Gaussian observations.

authors introduced a two-stage EnKF-PF scheme that exchanges estimates (ensemble means) of parameters and state between the PF and the EnKF components. In contrast, our scheme exchanges the ensembles, representing the approximate posterior distributions, instead of only their means, which is expected to lead to better performances while requiring only half the computational load. The remainder of this paper is organized as follows. Section 2 states the problem and reviews the joint PF and EnKF. Section 3 derives the new scheme combining EnKF with PF and discusses the main differences with the algorithm of Santitissadeekorn and Jones (2015). Section 4 presents results of various numerical experiments with the Lorenz-96 model, and Section 5 concludes the work and discusses future directions.

2. Problem statement

Consider a discrete-time state-parameter dynamical system:

$$\begin{cases} \mathbf{x}_n &= \mathbf{f}_{n-1}\left(\mathbf{x}_{n-1}, \boldsymbol{\theta}\right) + \mathbf{u}_{n-1} \\ \mathbf{y}_n &= \mathbf{h}_n(\mathbf{x}_n) + \mathbf{v}_n \end{cases}, \tag{1}$$

in which $\mathbf{x}_n \in \mathbb{R}^{n_\mathbf{x}}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}$ and $\mathbf{y}_n \in \mathbb{R}^{n_\mathbf{y}}$ denote the system state, the parameters' vector and the observation at time t_n, of dimensions $n_\mathbf{x}$, $n_{\boldsymbol{\theta}}$ and $n_\mathbf{y}$, respectively. $\mathbf{f}_n$ is the dynamical operator integrating the state from time t_n to t_{n+1}, and $\mathbf{h}_n$ is the observational operator at time t_n. The noise processes, $\mathbf{u} = \{\mathbf{u}_n\}_{n\in\mathbb{N}}$ and $\mathbf{v} = \{\mathbf{v}_n\}_{n\in\mathbb{N}}$, are assumed to be independent, jointly independent and independent of $\mathbf{x}_0$ and $\boldsymbol{\theta}$. Throughout the paper, $\mathbf{y}_{0:n} \overset{\text{def}}{=} \{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_n\}$, and $p(\mathbf{x}_n)$ and $p(\mathbf{x}_n|\mathbf{y}_{0:l})$ stand for the prior pdf of $\mathbf{x}_n$ and the posterior pdf of $\mathbf{x}_n$ given $\mathbf{y}_{0:l}$, respectively. All other pdfs used are defined in a similar way.

The state-parameter filtering problem aims at estimating at each time, t_n, the state, $\mathbf{x}_n$, and the parameters vector, $\boldsymbol{\theta}$, from the history of the observations, $\mathbf{y}_{0:n}$. Let for a random variable $\mathbf{u}$ and a realization $\mathbf{v}$ of another random variable, $\mathbb{E}[\mathbf{u}]$ and $\mathbb{E}[\mathbf{u}|\mathbf{v}]$ denote the expected values of $\mathbf{u}$ with

respect to (w.r.t.) $p(\mathbf{u})$ and $p(\mathbf{u}|\mathbf{v})$, respectively. A standard solution of this problem is the PM:

$$\mathbb{E}[\mathbf{x}_n|\mathbf{y}_{0:n}] = \int \mathbf{x}_n p(\mathbf{x}_n|\mathbf{y}_{0:n})\, d\mathbf{x}_n, \tag{2}$$

$$\mathbb{E}[\boldsymbol{\theta}|\mathbf{y}_{0:n}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}_{0:n})\, d\boldsymbol{\theta}, \tag{3}$$

which minimizes the MSE. The evaluation of (2) and (3) requires the knowledge of the so-called analysis pdfs, $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ and $p(\boldsymbol{\theta}|\mathbf{y}_{0:n})$. One classical way to compute these pdfs starts by computing the analysis pdf $p(\mathbf{z}_n|\mathbf{y}_{0:n})$ of the augmented state $\mathbf{z}_n \overset{\text{def}}{=} [\mathbf{x}_n^T, \boldsymbol{\theta}_n^T]^T$, where $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1}$, followed by a marginalization. By virtue of the factorization,

$$p(\mathbf{z}_{0:N}, \mathbf{y}_{0:N}) = p(\mathbf{z}_0) \prod_{n=1}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}) \times \prod_{n=0}^{N} p(\mathbf{y}_n|\mathbf{z}_n), \tag{4}$$

which, in turn, follows from the hidden Markov chain character of the pairwise process, $(\mathbf{z}, \mathbf{y}) = \{\mathbf{z}_n, \mathbf{y}_n\}_{n\in\mathbb{N}}$, it is possible to *recursively* compute $p(\mathbf{z}_n|\mathbf{y}_{0:n})$ (see e.g., Ait-El-Fquih and Desbouvries (2006); Ait-El-Fquih and Desbouvries (2011) and references therein). Two steps are performed to compute $p(\mathbf{z}_n|\mathbf{y}_{0:n})$ from $p(\mathbf{z}_{n-1}|\mathbf{y}_{0:n-1})$:

- *Forecast step.* The joint transition pdf, $p(\mathbf{z}_n|\mathbf{z}_{n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1}, \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n|\boldsymbol{\theta}_{n-1})$, is used to compute the joint forecast pdf, $p(\mathbf{z}_n|\mathbf{y}_{0:n-1})$, following a marginalization formula (Chapman-Kolmogorov equation (Jazwinski 1970)):

$$p(\mathbf{z}_n|\mathbf{y}_{0:n-1}) = \int p(\mathbf{z}_n|\mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}|\mathbf{y}_{0:n-1}) d\mathbf{z}_{n-1}. \tag{5}$$

- *Analysis step.* The likelihood, $p(\mathbf{y}_n|\mathbf{z}_n) = p(\mathbf{y}_n|\mathbf{x}_n)$, is combined with the joint forecast pdf, following the Bayes' rule, to obtain:

$$p(\mathbf{z}_n|\mathbf{y}_{0:n}) \propto p(\mathbf{y}_n, \mathbf{z}_n|\mathbf{y}_{0:n-1}), \tag{6}$$

$$= p(\mathbf{y}_n|\mathbf{z}_n) p(\mathbf{z}_n|\mathbf{y}_{0:n-1}). \tag{7}$$

Before we focus on the practical derivation of this generic algorithm, it is worth mentioning that the PF is inevitably vulnerable to the "sample attrition" issue when dealing with (fixed) parameters.

This may result in loss of diversity of particles, and eventually one ending up with identical particles (see e.g., West and Liu (2001); Frei and Künsch (2012); Santitissadeekorn and Jones (2015)). One popular way to cope with this is to impose some artificial dynamics on the parameters:

$$\boldsymbol{\theta}_n = \mathbf{g}_{n-1}(\boldsymbol{\theta}_{n-1}) + \mathbf{w}_{n-1}, \tag{8}$$

where $\{\mathbf{w}_n\}_{n\in\mathbb{N}}$ is a zero mean Gaussian process assumed to be independent and independent of $\mathbf{z}_0$, $\{\mathbf{u}_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{v}_n\}_{n\in\mathbb{N}}$. The reader may consult West and Liu (2001) and Frei and Künsch (2012) and references therein for a comprehensive literature about the choice of the operator, $\mathbf{g}_{n-1}(.)$, and the covariance, $\mathbf{W}_{n-1}$, of the noise, $\mathbf{w}_{n-1}$. We also assume $\mathbf{z}_0$, $\mathbf{u}_n$ and $\mathbf{v}_n$ to be Gaussian with $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{U}_n)$ and $\mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_n)$. The transition pdfs, $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \boldsymbol{\theta}_n)$ and $p(\boldsymbol{\theta}_n|\boldsymbol{\theta}_{n-1})$, and the likelihood, $p(\mathbf{y}_n|\mathbf{x}_n)$, are therefore Gaussian with

$$p(\mathbf{x}_n|\mathbf{x}_{n-1}, \boldsymbol{\theta}_n) \overset{(1)}{=} \mathcal{N}_{\mathbf{x}_n}\left(\mathbf{f}_{n-1}(\mathbf{x}_{n-1}, \boldsymbol{\theta}_n), \mathbf{U}_{n-1}\right), \tag{9}$$

$$p(\boldsymbol{\theta}_n|\boldsymbol{\theta}_{n-1}) \overset{(8)}{=} \mathcal{N}_{\boldsymbol{\theta}_n}\left(\mathbf{g}_{n-1}(\boldsymbol{\theta}_{n-1}), \mathbf{W}_{n-1}\right), \tag{10}$$

$$p(\mathbf{y}_n|\mathbf{x}_n) \overset{(1)}{=} \mathcal{N}_{\mathbf{y}_n}\left(\mathbf{h}_n(\mathbf{x}_n), \mathbf{V}_n\right), \tag{11}$$

where $\mathcal{N}_{\mathbf{x}}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian pdf of argument $\mathbf{x}$ and parameters $(\mathbf{m}, \mathbf{C})$.

Joint state-parameters filtering with PF and EnKF

The main idea is to transform the state-parameter system $(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})$ into a classical state-space system with an augmented state $\mathbf{z} = (\mathbf{x}, \boldsymbol{\theta})$ and then applying the standard PF (Gordon et al. 1993) and EnKF (Evensen 1994) to the augmented system. The resulting Joint-PF and Joint-EnKF are Monte Carlo implementations of the generic algorithms ((5),(7)) and ((5),(6)), respectively. They share the same forecast step (that follows from the Markov property (5)) and differ in their analysis steps (that follow from the Bayes' formulas (7) and (6), respectively).

Assume that one has an independently and identically distributed (i.i.d.) (analysis) random ensemble of members (or particles), $\{\mathbf{z}_{n-1}^{a,(m)} \overset{\text{def}}{=} (\mathbf{x}_{n-1}^{a,(m)}, \boldsymbol{\theta}_{n-1}^{a,(m)})\}_{m=1}^{M}$, from $p(\mathbf{z}_{n-1}|\mathbf{y}_{0:n-1})$, and wants to sample an i.i.d. (forecast) random ensemble, $\{\mathbf{z}_n^{f,(m)}\}_{m=1}^{M}$ from $p(\mathbf{z}_n|\mathbf{y}_{0:n-1})$. This can be achieved by applying the hierarchical sampling property 1 (see Appendix A) on eq. (5):

$$\boldsymbol{\theta}_n^{f,(m)} = \mathbf{g}_{n-1}(\boldsymbol{\theta}_{n-1}^{a,(m)}) + \mathbf{w}_{n-1}^{(m)}, \tag{12}$$

$$\mathbf{x}_n^{f,(m)} = \mathbf{f}_{n-1}(\mathbf{x}_{n-1}^{a,(m)}, \boldsymbol{\theta}_n^{f,(m)}) + \mathbf{u}_{n-1}^{(m)}; \tag{13}$$

$\mathbf{w}_{n-1}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{n-1})$ and $\mathbf{u}_{n-1}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{U}_{n-1})$. In other words, the forecast members are obtained by integrating the previous analysis members forward with the state-parameter model.

The analysis step of the Joint-PF involves a weighting and then a resampling of the forecast particles. The weight of each augmented particle, $\mathbf{z}_n^{f,(m)}$, is computed by applying the Rubin's sampling importance resampling (SIR) mechanism (recalled in Property 2 , Appendix A) on (7):

$$\lambda_n^{(m)} \propto \mathcal{N}_{\mathbf{y}_n}(\mathbf{h}_n(\mathbf{x}_n^{f(m)}), \mathbf{V}_n); \qquad \sum_{m=1}^{M} \lambda_n^{(m)} = 1. \tag{14}$$

The analysis estimates of state and parameters are computed from $\{\mathbf{z}_n^{f,(m)}, \lambda_n^{(m)}\}_{m=1}^{M}$, which approximates the analysis distributions in (2) and (3). The ensemble $\{\mathbf{z}_n^{f,(m)}, \lambda_n^{(m)}\}_{m=1}^{M}$ is then sampled with replacement to obtain the (empirical) posterior $\{\mathbf{z}_n^{a,(m)}, \frac{1}{M}\}_{m=1}^{M}$ before we proceed to the next assimilation cycle.

The Joint-PF was shown to provide satisfactory performances in many applications involving low-dimensional systems, but remains impractical in large dimensions. Indeed, when an insufficient number of particles are used, regions of high probabilities are unlikely to be sampled in the forecast step. The analysis step cannot remedy this because the analysis particles are obtained from a (discrete) sampling of the forecast ensemble.

The Joint-EnKF avoids this issue thanks to the (continuous) sampling character of its analysis step. The incoming observations are indeed used to update the forecast particles (or ensemble

members) using a KF-like analysis step, as in (15), which results in analysis members that are not only different from the forecast members, but also more likely to belong to regions of high probabilities. Let, hereafter, for any ensemble $\{\mathbf{u}^{(m)}\}_{m=1}^{M}$, $\hat{\mathbf{u}}$ and $\mathbf{P}_{\mathbf{u}}$ denote its empirical mean and covariance, respectively; and $\mathbf{P}_{\mathbf{u},\mathbf{v}}$ the cross-covariance between $\{\mathbf{u}^{(m)}\}_{m=1}^{M}$ and $\{\mathbf{v}^{(m)}\}_{m=1}^{M}$. Assuming $p(\mathbf{z}_n, \mathbf{y}_n|\mathbf{y}_{0:n-1})$ is Gaussian, the analysis members are computed by applying the conditional sampling property 3 (Appendix A) to eq. (6):

$$\underbrace{\begin{bmatrix}\mathbf{x}_n^{a,(m)}\\ \boldsymbol{\theta}_n^{a,(m)}\end{bmatrix}}_{\mathbf{z}_n^{a,(m)}} = \underbrace{\begin{bmatrix}\mathbf{x}_n^{f,(m)}\\ \boldsymbol{\theta}_n^{f,(m)}\end{bmatrix}}_{\mathbf{z}_n^{f,(m)}} + \underbrace{\begin{bmatrix}\mathbf{P}_{\mathbf{x}_n^f,\mathbf{y}_n^f}\\ \mathbf{P}_{\boldsymbol{\theta}_n^f,\mathbf{y}_n^f}\end{bmatrix}}_{\mathbf{P}_{\mathbf{z}_n^f,\mathbf{y}_n^f}} \mathbf{P}_{\mathbf{y}_n^f}^{-1}\left(\mathbf{y}_n - \mathbf{y}_n^{f,(m)}\right), \tag{15}$$

where $\mathbf{y}_n^{f,(m)}$ is an observation forecast member that is computed by sampling from the Gaussian $\mathcal{N}(\mathbf{h}_n(\mathbf{x}_n^{f,(m)}), \mathbf{V}_n)$. The KF-like update of the parameters is based on the sampled cross-covariance, $\mathbf{P}_{\boldsymbol{\theta}_n^f,\mathbf{y}_n^f}$. The later may however not be well estimated when the parameters and state variables are strongly nonlinearly related, which may severely limit the performance of the Joint-EnKF (Kivman 2003). In the subsequent section, we introduce an alternative scheme that uses the PF to sample the parameters and the EnKF to sample the state.

3. The EnKF-PF scheme

a. The generic algorithm

The forecast step of the proposed scheme is identical to that of the Joint-PF and the Joint-EnKF. In the analysis step, instead of using a truly augmented (i.e., joint) approach to compute $p(\mathbf{z}_n|\mathbf{y}_{0:n})$ from $p(\mathbf{z}_n|\mathbf{y}_{0:n-1})$, we adopt a conditional strategy that involves an analysis step that separately updates $p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n})$ and $p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n})$, from which one then obtains:

$$p(\mathbf{z}_n|\mathbf{y}_{0:n}) = p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n}) \times p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n}). \tag{16}$$

The idea of using the conditional (or marginalization) strategy (16) has been already used in systems with particular structures, including those for which a part of the augmented state (here $\mathbf{x}_n$) involves linear-Gaussian models (Doucet et al. 2000, 2001b; Schön et al. 2005; Guimaraes et al. 2010) or models with finite (discrete) state spaces (Ghahramani and Jordan 1997; Doucet et al. 2000). In these cases, $p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n})$ is analytically tractable either using the KF or the hidden Markov model (HMM) algorithm (depending on whether the state is continuous or discrete). This enables one to marginalize out $\mathbf{x}_n$ from the joint posterior distribution and only focus on estimating, using the PF, $p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n})$, which belongs to a space of reduced dimension. More related to our work, this conditional technique, also known as Rao-Blackwellisation, has been recently adopted by Santitissadeekorn and Jones (2015) in the context of state-parameter filtering, yet with a nonlinear (conditional) state model, in order to apply the EnKF to estimate the state. This allows the authors to derive a two-stage EnKF-PF scheme that exchanges estimates (ensemble means) of parameters and state between the PF and the EnKF components. Here, we propose an alternative scheme that exchanges the ensembles instead of only their means. A thorough discussion about the differences between these two approaches is given in Subsection c below.

1) Parameters' analysis step

Since the parameters will be updated based on the PF, one chooses the form (7) for the generic analysis step:

$$p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n}) \propto p(\mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}). \tag{17}$$

The conditional likelihood, $p(\mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$, is not known. As will be made clearer in next section, this pdf needs to be assumed Gaussian in the (EnKF) state update step, as it is used to sample the observations forecast ensemble. While we exploit this Gaussian assumption here, we still need to compute the first two moments of this likelihood. To that end, we propose to approximate these

moments based on a linear minimum MSE (LMMSE) optimization criteria[2] (Anderson and Moore 1979; Kailath et al. 2000). Given $\mathbf{y}_{0:n-1}$, one readily obtains a LMMSE-based estimation of the first two moments of $p(\mathbf{y}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ as:

$$\check{\mathbf{y}}_n(\boldsymbol{\theta}_n) = \mathbb{E}[\mathbf{y}_n|\mathbf{y}_{0:n-1}] + \mathrm{Cov}[\mathbf{y}_n,\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]\mathrm{Cov}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]^{-1}(\boldsymbol{\theta}_n - \mathbb{E}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]), \quad (18)$$

$$\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n} = \mathrm{Cov}[\mathbf{y}_n|\mathbf{y}_{0:n-1}] - \mathrm{Cov}[\mathbf{y}_n,\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]\mathrm{Cov}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]^{-1}\mathrm{Cov}[\boldsymbol{\theta}_n,\mathbf{y}_n|\mathbf{y}_{0:n-1}]. \quad (19)$$

2) (CONDITIONAL) STATE ANALYSIS STEP

Similarly to (6), one has:

$$p(\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n}) \propto p(\mathbf{y}_n,\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1}). \quad (20)$$

The use of the EnKF to "sample" (20) requires $p(\mathbf{y}_n,\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ to be Gaussian. On the other hand, the EnKF-like sampling of (20) relies on sampling the marginals, $p(\mathbf{y}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ and $p(\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$, and thus a beforehand computation of their first two moments. A LMMSE estimate of those of $p(\mathbf{y}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ has been already considered in (18)-(19). Those of $p(\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ are similarly approximated as

$$\check{\mathbf{x}}_n(\boldsymbol{\theta}_n) = \mathbb{E}[\mathbf{x}_n|\mathbf{y}_{0:n-1}] + \mathrm{Cov}[\mathbf{x}_n,\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]\mathrm{Cov}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]^{-1}(\boldsymbol{\theta}_n - \mathbb{E}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]), \quad (21)$$

$$\check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n} = \mathrm{Cov}[\mathbf{x}_n|\mathbf{y}_{0:n-1}] - \mathrm{Cov}[\mathbf{x}_n,\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]\mathrm{Cov}[\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1}]^{-1}\mathrm{Cov}[\boldsymbol{\theta}_n,\mathbf{x}_n|\mathbf{y}_{0:n-1}]. \quad (22)$$

Note that in the linear observational operator case (i.e., $\mathbf{h}_n(\mathbf{x}_n) = \mathbf{H}_n\mathbf{x}_n$ in system (1)), only $p(\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$ needs to be assumed Gaussian since in such a case the pdf $p(\mathbf{y}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})$, which is given by

$$p(\mathbf{y}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1}) = \int p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\boldsymbol{\theta}_n,\mathbf{y}_{0:n-1})d\mathbf{x}_n, \quad (23)$$

[2]For two random variables $\mathbf{u}$ and $\mathbf{v}$, the LMMSE estimator of $\mathbf{u}$ given (a realization of) $\mathbf{v}$ and the associated error covariance are respectively given by $\mathbb{E}[\mathbf{u}] + \mathrm{Cov}[\mathbf{u},\mathbf{v}]\mathrm{Cov}[\mathbf{v}]^{-1}(\mathbf{v} - \mathbb{E}[\mathbf{v}])$ and $\mathrm{Cov}[\mathbf{u}] - \mathrm{Cov}[\mathbf{u},\mathbf{v}]\mathrm{Cov}[\mathbf{v}]^{-1}\mathrm{Cov}[\mathbf{v},\mathbf{u}]$; where $\mathrm{Cov}[\mathbf{u}]$, $\mathrm{Cov}[\mathbf{v}]$ and $\mathrm{Cov}[\mathbf{u},\mathbf{v}]$ denote the covariances of $\mathbf{u}$ and $\mathbf{v}$, and the cross-covariance between $\mathbf{u}$ and $\mathbf{v}$, respectively.

is also Gaussian with parameters exactly equal to the LMMSE estimates (18) and (19), which, in turn, are equal to $\mathbf{H}_n\check{\mathbf{x}}_n(\boldsymbol{\theta}_n)$ and $\mathbf{H}_n\check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n}\mathbf{H}_n^T+\mathbf{V}_n$, respectively.

b. Practical implementation

We resort to classical sampling properties, recalled in the Appendix, to derive a Monte Carlo implementation of the generic algorithm (5), (17) and (20). As stated previously, the forecast step of the proposed scheme is the same as those of the PF and EnKF, given by (12)-(13). In the analysis step and using (16), each joint analysis member $\mathbf{z}_n^{a,(m)} \sim p(\mathbf{z}_n|\mathbf{y}_{0:n})$ is computed following a conditioning strategy that requires first computing $\boldsymbol{\theta}_n^{a,(m)} \sim p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n})$ using the PF update and then $\mathbf{x}_n^{a,(m)} \sim p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)},\mathbf{y}_{0:n})$ using an EnKF update. The proposed EnKF-PF algorithm is summarized below (the proof is given in Appendix B).

• *Joint forecast of state and parameters.*

1. The joint forecast ensemble, $\{\mathbf{x}_n^{f,(m)},\boldsymbol{\theta}_n^{f,(m)}\}_{m=1}^M$, is first computed using (12)-(13). This step is initialized at time t_0 by sampling from a given joint prior, $p(\mathbf{x}_0,\boldsymbol{\theta}_0)$.

2. The observation forecast ensemble is obtained as $\mathbf{y}_n^{f,(m)}=\boldsymbol{\eta}_n^{(m)}+\mathbf{v}_n^{(m)}$, with $\boldsymbol{\eta}_n^{(m)}=\mathbf{h}_n(\mathbf{x}_n^{f,(m)})$ and $\mathbf{v}_n^{(m)} \sim \mathcal{N}(\mathbf{0},\mathbf{V}_n)$.

3. Compute the empirical means, $\widehat{\mathbf{x}}_n^f$, $\widehat{\boldsymbol{\theta}}_n^f$ and $\widehat{\boldsymbol{\eta}}_n^f$, and the forecast perturbed matrices, $\mathbf{S}_{\mathbf{x}_n^f}$, $\mathbf{S}_{\boldsymbol{\theta}_n^f}$ and $\mathbf{S}_{\boldsymbol{\eta}_n^f}$, from which the (cross-)covariances, $\mathbf{P}_{\mathbf{x}_n^f,\boldsymbol{\theta}_n^f}=\mathbf{S}_{\mathbf{x}_n^f}\mathbf{S}_{\boldsymbol{\theta}_n^f}^T$, $\mathbf{P}_{\mathbf{x}_n^f,\boldsymbol{\eta}_n^f}=\mathbf{S}_{\mathbf{x}_n^f}\mathbf{S}_{\boldsymbol{\eta}_n^f}^T$, $\mathbf{P}_{\boldsymbol{\eta}_n^f,\boldsymbol{\theta}_n^f}=\mathbf{S}_{\boldsymbol{\eta}_n^f}\mathbf{S}_{\boldsymbol{\theta}_n^f}^T$, $\mathbf{P}_{\boldsymbol{\theta}_n^f}=\mathbf{S}_{\boldsymbol{\theta}_n^f}\mathbf{S}_{\boldsymbol{\theta}_n^f}^T$, $\mathbf{P}_{\boldsymbol{\eta}_n^f}=\mathbf{S}_{\boldsymbol{\eta}_n^f}\mathbf{S}_{\boldsymbol{\eta}_n^f}^T$, $\mathbf{P}_{\boldsymbol{\eta}_n|\boldsymbol{\theta}_n}=\mathbf{P}_{\boldsymbol{\eta}_n^f}-\mathbf{K}_{\boldsymbol{\eta}_n^f,\boldsymbol{\theta}_n^f}\mathbf{P}_{\boldsymbol{\theta}_n^f,\boldsymbol{\eta}_n^f}$, $\mathbf{P}_{\mathbf{x}_n,\mathbf{y}_n|\boldsymbol{\theta}}=\mathbf{P}_{\mathbf{x}_n^f,\boldsymbol{\eta}_n^f}-\mathbf{K}_{\mathbf{x}_n^f,\boldsymbol{\theta}_n^f}\mathbf{P}_{\boldsymbol{\theta}_n^f,\boldsymbol{\eta}_n^f}$, and "the gains", $\mathbf{K}_{\mathbf{x}_n^f,\boldsymbol{\theta}_n^f}=\mathbf{P}_{\mathbf{x}_n^f,\boldsymbol{\theta}_n^f}\mathbf{P}_{\boldsymbol{\theta}_n^f}^{-1}$, $\mathbf{K}_{\boldsymbol{\eta}_n^f,\boldsymbol{\theta}_n^f}=\mathbf{P}_{\boldsymbol{\eta}_n^f,\boldsymbol{\theta}_n^f}\mathbf{P}_{\boldsymbol{\theta}_n^f}^{-1}$, are then evaluated[3].

• *Parameters' analysis (PF-like update).*

[3]Note that the $n_{\boldsymbol{\theta}}\times n_{\boldsymbol{\theta}}$ matrix, $\mathbf{P}_{\boldsymbol{\theta}_n^f}$, whose inverse is needed in the expression of "the gains" is positive-definite, thus invertible, as the number of particles, M, is supposed to be larger than the number of parameters $n_{\boldsymbol{\theta}}$.

1. The normalized weights, $w_n^{(m)} \propto \mathcal{N}_{\mathbf{y}_n}(\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)}), \check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n})$, are computed based on

$$\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)}) \approx \widehat{\boldsymbol{\eta}}_n^f + \mathbf{K}_{\boldsymbol{\eta}_n^f, \boldsymbol{\theta}_n^f}(\boldsymbol{\theta}_n^{f,(m)} - \widehat{\boldsymbol{\theta}}_n^f), \tag{24}$$

$$\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n} \approx \mathbf{P}_{\boldsymbol{\eta}_n|\boldsymbol{\theta}_n} + \mathbf{V}_n. \tag{25}$$

An approximation of the PM analysis estimate (3) and the associated error covariance are then given by $\widehat{\boldsymbol{\theta}}_n^a = \sum_{m=1}^M w_n^{(m)} \boldsymbol{\theta}_n^{f,(m)}$ and $\sum_{m=1}^M w_n^{(m)} (\boldsymbol{\theta}_n^{f,(m)} - \widehat{\boldsymbol{\theta}}_n^a)(\boldsymbol{\theta}_n^{f,(m)} - \widehat{\boldsymbol{\theta}}_n^a)^T$, respectively.

2. The weighted ensemble, $\{\boldsymbol{\theta}_n^{f,(m)}, w_n^{f,(m)}\}_{m=1}^M$, is then sampled with replacement to obtain an analysis ensemble with equal weights, $\{\boldsymbol{\theta}_n^{a,(m)}, \frac{1}{M}\}_{m=1}^M$.

• *State analysis (EnKF-like update).*

1. The means (for all m) of the conditional forecast pdf, $p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$, are estimated as,

$$\check{\mathbf{x}}_n(\boldsymbol{\theta}_n^{a,(m)}) \approx \widehat{\mathbf{x}}_n^f + \mathbf{K}_{\mathbf{x}_n^f, \boldsymbol{\theta}_n^f}(\boldsymbol{\theta}_n^{a,(m)} - \widehat{\boldsymbol{\theta}}_n^f). \tag{26}$$

2. A singular value decomposition (SVD) of the $M \times M$ symmetric matrix $\mathbb{I}_M - \mathbf{S}_{\boldsymbol{\theta}_n^f}^T \mathbf{P}_{\boldsymbol{\theta}_n^f}^{-1} \mathbf{S}_{\boldsymbol{\theta}_n^f}$ is performed leading to $M \times M$ orthogonal $\mathscr{U}_n$ and diagonal Σ_n matrices.

3. Samples $\xi_{\mathbf{x}_n}^{(m)}$ are then generated from $p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$ as

$$\xi_{\mathbf{x}_n}^{(m)} \approx \check{\mathbf{x}}_n(\boldsymbol{\theta}_n^{a,(m)}) + \mathbf{S}_{\mathbf{x}_n^f} \mathscr{U}_n (\Sigma_n)^{\frac{1}{2}} \mathbf{r}_n^{(m)}, \tag{27}$$

where $(\Sigma_n)^{\frac{1}{2}}$ denotes the square-root of the diagonal matrix Σ_n and $\mathbf{r}_n^{(m)} \sim \mathscr{N}(\mathbf{0}, \mathbb{I}_M)$.

4. These are then updated based on the observation, $\mathbf{y}_n$, following a KF-like correction, leading to the state analysis ensemble of interest, $\{\mathbf{x}_n^{a,(m)}\}_{m=1}^M$, as,

$$\xi_{\mathbf{y}_n}^{(m)} = \mathbf{h}_n(\xi_{\mathbf{x}_n}^{(m)}) + \mathbf{v}_n^{(m)}; \quad \mathbf{v}_n^{(m)} \sim \mathscr{N}(\mathbf{0}, \mathbf{V}_n), \tag{28}$$

$$\mathbf{x}_n^{a,(m)} = \xi_{\mathbf{x}_n}^{(m)} + \mathbf{P}_{\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}} \mathbf{P}_{\mathbf{y}_n|\boldsymbol{\theta}}^{-1} (\mathbf{y}_n - \xi_{\mathbf{y}_n}^{(m)}), \tag{29}$$

and whose mean is taken as an approximation of the PM estimate (2).

A schematic illustration of these steps is given in Fig. 1. Once the forecast members $\boldsymbol{\theta}_n^{f,(m)}$ and $\mathbf{x}_n^{f,(m)}$ are generated, they are updated, successively, based on the observation, $\mathbf{y}_n$. The update of the parameters is performed following a PF-like mechanism which consists of a succession of a weighting step evaluating a (normalized) weight, $w_n^{(m)}$, for each member $\boldsymbol{\theta}_n^{f,(m)}$, based on the density, $p(\mathbf{y}_n|\boldsymbol{\theta}_n^{f,(m)}, \mathbf{y}_{0:n-1})$, which is assumed to be Gaussian with moments computed based on LMMSE criteria; and a resampling step leading to the (equi-weighted) analysis members of interest, $\boldsymbol{\theta}_n^{a,(m)}$. The update of the state is then performed by first computing the moments of the assumed Gaussian density, $p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$, based on the LMMSE criteria, then sampling this density before applying an EnKF-like update to the resulting samples to obtain the state analysis members of interest, $\mathbf{x}_n^{a,(m)}$. Here, the joint pdf $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ needs to be Gaussian, not only its marginals as above. Finally, one notes that in the particular case of a linear observation operator, only $p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ needs to be Gaussian (as this implies that both $p(\mathbf{y}_n|\boldsymbol{\theta}_n^{f,(m)}, \mathbf{y}_{0:n-1})$ and $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ are Gaussian).

c. Discussion

The proposed algorithm combines the PF to sample $\boldsymbol{\theta}_n^{a,(m)} \sim p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n})$, with the EnKF to sample $\mathbf{x}_n^{a,(m)} \sim p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n})$. The $\mathbf{x}_n^{a,(m)}$ are taken as samples of the state analysis pdf, $p(\mathbf{x}_n|\mathbf{y}_{0:n})$. The algorithm has been derived following a Bayesian formulation, while assuming $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ to be Gaussian with moments computed based on the LMMSE criteria. As it is well known, the Gaussian assumption of the joint state and observation forecast pdf is a standard assumption in the context of ensemble Kalman filtering. We further exploited this Gaussianity, along with the LMMSE criteria, in the PF stage to efficiently compute the likeli-

hood, $p(\mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$, which, in turn, is used to evaluate the particles' weights. Now, assuming[4] $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ is Gaussian implies that its marginals $p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ and $p(\mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ are also Gaussian, but not the forecast / analysis pdfs of interest, $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$, $p(\mathbf{x}_n|\mathbf{y}_{0:n})$, $p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n-1})$, nor $p(\boldsymbol{\theta}_n|\mathbf{y}_{0:n})$, in contrast with the Joint-EnKF in which all these pdfs need to be Gaussian as the derivation of this filter relies on the Gaussian assumption over $p(\mathbf{x}_n, \boldsymbol{\theta}_n, \mathbf{y}_n|\mathbf{y}_{0:n-1})$.

Although the proposed EnKF-PF involves the cross-covariances between the state (resp. observations) and parameters' forecast pdfs as in (21)-(22) (resp. (18)-(19)), these are indeed not used in practice, as in the Joint-EnKF, to generate the analysis members of interest, $\mathbf{x}_n^{a,(m)}$ (resp. $\boldsymbol{\theta}_n^{a,(m)}$). Instead they are only used to compute the means of $p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$ in the EnKF stage as in (26) (resp. the weights in the PF stage as in (24)), which are then used to generate these members. The resulting members are therefore not restricted to approximate Gaussian distributions as in the Joint-EnKF. While $\boldsymbol{\theta}_n^{a,(m)}$ are computed by the free-Gaussian sampling PF, the non-Gaussian character of $\mathbf{x}_n^{a,(m)}$ stems from the fact that these members are computed based on an EnKF update but conditionally on $\boldsymbol{\theta}_n^{a,(m)}$ (this can be seen from eq. (29) which involves $\boldsymbol{\theta}_n^{a,(m)}$ through $\boldsymbol{\xi}_{\mathbf{y}_n}^{(m)}$ and (27)-(28)). On the other hand, the proposed EnKF-PF requires as many model integrations as the Joint-EnKF (as they share the same forecast step). The analysis steps of these algorithms involve operations with linear complexity in $n_{\mathbf{x}}$. Thus, although the analysis step of the proposed scheme involves a few more operations than that of the Joint-EnKF, the computational complexities of these schemes should be approximately of the same order in relatively large-dimensional state applications.

[4] It should be stressed that eqs. (21)-(22) (resp. (18)-(19)) that approximate the moments of the assumed Gaussian pdf $p(\mathbf{x}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ (resp. $p(\mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$) under the LMMSE optimization criteria, have the same form as the *true* moments if $p(\mathbf{x}_n, \boldsymbol{\theta}_n|\mathbf{y}_{0:n-1})$ (resp. $p(\mathbf{y}_n, \boldsymbol{\theta}_n|\mathbf{y}_{0:n-1})$) is Gaussian. However, in our study, these joint pdfs are not required to be Gaussian as the proposed algorithm uses only their first two moments, that are empirically estimated from the members, and still valid if they are not Gaussian.

In terms of algorithmic comparison with the recently introduced two-stage EnKF-PF (TS-EnKF-PF) by Santitissadeekorn and Jones (2015), we note the following aspects:

- As the proposed scheme, the TS-EnKF-PF inherently relies on the Gaussian assumption over $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$. However, instead of using the LMMSE criteria to compute the first two moments, the authors approximate $p(\mathbf{x}_n|\theta_n, \mathbf{y}_{0:n-1})$ with the forecast pdf $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$, whose moments can be estimated from the ensemble members. Such an assumption further enables to compute the mean of the pdf $p(\mathbf{y}_n|\theta_n, \mathbf{y}_{0:n-1})$, which serves to evaluate the particles' weights in the PF stage, and the observation forecast members in the EnKF stage (see eqs. (9) and (11) in Santitissadeekorn and Jones (2015)). As for the covariance of $p(\mathbf{y}_n|\theta_n, \mathbf{y}_{0:n-1})$, it is set to be equal to the covariance of $p(\mathbf{y}_n|\mathbf{x}_n)$, $\mathbf{V}_n$. However, based on (23), this covariance is actually equal to $\mathbf{C}_n \overset{\text{def}}{=} \text{cov}[\mathbf{h}_n(\mathbf{x}_n)|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1}] + \mathbf{V}_n$, which therefore suggests that $\text{cov}[\mathbf{h}_n(\mathbf{x}_n)|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1}]$ is neglected in Santitissadeekorn and Jones (2015). This may strongly underestimates the true covariance, $\mathbf{C}_n$, when it is much larger than $\mathbf{V}_n$. Such a problem could be mitigated in the proposed approach in which the particles' weights are computed based on $p(\mathbf{y}_n|\boldsymbol{\theta}_n^{f,(m)}, \mathbf{y}_{0:n-1}) = \mathscr{N}_{\mathbf{y}_n}(\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)}), \check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n})$, where $\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)})$ and $\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n}$ are given in (24) and (25), respectively. Indeed, $\mathbf{C}_n$ is approximated by $\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n} = \mathbf{P}_{\boldsymbol{\eta}_n|\boldsymbol{\theta}_n} + \mathbf{V}_n$, where $\mathbf{P}_{\boldsymbol{\eta}_n|\boldsymbol{\theta}_n}$ is the LMMSE-based approximation of $\text{cov}[\mathbf{h}_n(\mathbf{x}_n)|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1}]$, which is not assumed to be zero as in the TS-EnKF-PF, but evaluated from the ensembles.
- In the EnKF-PF, the information between the PF and EnKF is exchanged through the ensembles' members (see e.g., Fig. 1). In the TS-EnKF-PF, the exchange only carried through the ensembles' means, which should result in loss of performance, especially when multimodal posterior distributions are involved. As discussed in Santitissadeekorn and Jones (2015), "feeding" the EnKF with the parameters' mean only may divert the background ensemble of the EnKF from high-probability regions, and likewise when the state mean "feeds" the PF.

- Because of the independent two-stage formulation of the TS-EnKF-PF, the dynamical model needs to be integrated in both PF and EnKF stages, each of them involving a different state forecast ensemble. In the proposed scheme, the model is run only in the (shared) forecast stage, to obtain a state forecast ensemble that is used in both PF and EnKF update stages. Therefore, the proposed scheme can be roughly half computationally less demanding than the TS-EnKF-PF when the same ensemble size is used in both PF and EnKF.

4. Numerical experiments

Numerical experiments are performed with the strongly nonlinear Lorenz-96 (L96) (Lorenz and Emanuel 1998) to assess the behavior of the proposed EnKF-PF and to evaluate its performance against the Joint-PF, the Joint-EnKF and the TS-EnKF-PF. The L96 model, which describes the time evolution of an atmospheric quantity, solves the following set of differential equations:

$$\frac{dx(j,t)}{dt} = x(j-1,t)[x(j+1,t) - x(j-2,t)] - x(j,t) + F(j), \tag{30}$$

where $x(j,t)$, $j = 1,2,\cdots,n_{\mathbf{x}}$, with, unless otherwise stated, $n_{\mathbf{x}} = 40$, denotes the j^{th} element of the state at time t. Boundary conditions are cyclic, (i.e., $x(-1,t) = x(n_{\mathbf{x}}-1,t)$, $x(0,t) = x(n_{\mathbf{x}},t)$, and $x(1,t) = x(n_{\mathbf{x}}+1,t)$). The constant $F(j)$, which represents dissipation, is commonly equal to 8 for all j; such a value has been shown to exhibit a chaotic behavior of the model (Karimi and Paul 2010). Here, we follow Santitissadeekorn and Jones (2015) and assume it unknown and a function of two parameters, $\boldsymbol{\theta} = [\theta(1), \theta(2)]^T$, through the equation:

$$F(j) = \theta(1)\sin\left(\frac{2\pi}{\theta(2)}j\right) + 8. \tag{31}$$

We consider for now a perfect model scenario (i.e., the noise $\mathbf{u}_n$ in model (1) vanishes), so the uncertainty in $\boldsymbol{\theta}$ is the only source of model error. The imperfect model scenario will be considered later.

We use the standard fourth-order Runge-Kutta method to numerically integrate the model (30) with a time step size $\delta_m = 0.05$, equivalent to six hours in real time. The reference (true) state trajectory is built based on the reference parameters, with values $\theta^*(1) = 2$ and $\theta^*(2) = 40$. The initial (reference) state is set to F, based on the reference parameters. The model is then integrated for 36000 time steps (corresponding to 24.6575 years in real time). The first 30000 time steps of the resulting trajectory is discarded (as a spin-up period), and the remaining 6000 time steps are considered as the reference states. The observations are obtained by perturbing the corresponding states with a standard Gaussian noise (i.e., $\mathbf{V}_n = \mathbb{I}_{n_\mathbf{y}}$).

We follow Santitissadeekorn and Jones (2015) and choose for all filters involving the PF (i.e., all except the Joint-EnKF), the "parameters' kernel smoothing" model (see eq. (8)) of West and Liu (2001) in which $\mathbf{g}_{n-1}(\boldsymbol{\theta}_{n-1}) = \alpha\boldsymbol{\theta}_{n-1} + (1-\alpha)\widehat{\boldsymbol{\theta}}^a_{n-1}$ and $\mathbf{W}_{n-1} = (1-\alpha^2)\mathbf{P}_{\boldsymbol{\theta}^a_{n-1}}$, with $\alpha = 0.9$. This model has been shown to be more efficient than the random walk model (Gordon et al. 1993), for which $\mathbf{g}_{n-1}(\boldsymbol{\theta}_{n-1}) = \boldsymbol{\theta}_{n-1}$ and $\mathbf{W}_{n-1} = \mathbf{P}_{\boldsymbol{\theta}^a_{n-1}}$. This was explained by the ability of this model to push the new set of particles towards the mean, which eliminates "the overdispersion" issue, the main limitation of the random walk model.

In all filters, the initial forecast ensembles of the parameters $\theta(1)$ and $\theta(2)$ are independently generated according to the (prior) Gaussian distributions $\mathcal{N}(4,\sqrt{2})$ and $\mathcal{N}(60,3)$, respectively (Santitissadeekorn and Jones 2015). The initial forecast state ensemble is generated from a Gaussian distribution centred around the mean of the reference states with an identity covariance, following Hamill and Whitaker (2010) and Song et al. (2010).

In all experiments, the Joint-PF and the PF components of the EnKF-PF and TS-EnKF-PF use the residual resampling strategy (Liu and Chen 1998; van Leeuwen 2009), which is well-known to improve upon the simple probabilistic one used in the bootstrap PF of Gordon et al. (1993). On the other hand, the EnKF was implemented with the covariance inflation and localization

techniques to enhance the filter's robustness and improve its performance with small ensembles. Covariance inflation mitigates for the underestimation of the sample error variances that result from the use of a small ensemble, among other neglected uncertainties and filtering approximations (Anderson and Anderson 1999; Hoteit et al. 2002). Covariance localization tackles the rank-deficiency and spuriously large cross-correlations between distant state variables in the ensemble covariance matrix (Hamill and Whitaker 2010). Here, localization is applied using the fifth-order correlation function given in (Gaspari and Cohn 1999, eq. (4.10)). In all experiments, we follow Santitissadeekorn and Jones (2015) and use a localization length-scale $\ell = 2$ and an inflation factor $\alpha = 1.2$ for which the Joint-EnKF provided the best performances[5].

In the first set of experiments, we compare the filters' behavior with $M = 100$ ensemble members in a setting where the data are assimilated every four model time steps, which is equivalent to one day in real time (i.e., the observations time step is $\delta_o = 4\delta_m$). In all the experiments, the results are averaged over 30 independent simulations, each time with a randomly generated initial ensembles and observation noise. In terms of computational cost, the EnKF and PF components of the TS-EnKF-PF use the same ensemble size; this means that the latter is more computationally demanding than the Joint-EnKF and the proposed EnKF-PF.

Figure 2 plots the time evolution of the analysis estimates of the marginal parameters and associated 95% confidence intervals (bounded by $\pm 1.96\times$ standard deviation). As expected, the Joint-PF seems to suffer from the high dimensionality of the system[6]. The Joint-EnKF outperforms the Joint-PF, suggesting the relevance of using a KF-like update on the forecast members.

[5]More precisely, Santitissadeekorn and Jones (2015) used an inflation $\alpha = 1$, suggesting that the Joint-EnKF results are not too sensitive to α when $\ell \lesssim 4$. Our tests reveal that the Joint-EnKF provides the best estimates with $\alpha = 1.2$, especially when implemented with not too large ensembles.

[6]Other experiments that have been conducted using alternative initial priors of the parameters with supports including the true values $\theta^*(1)$ and $\theta^*(2)$, suggested an improvement of the Joint-PF in tracking these true values, while remaining less accurate than the three other filters.

Its performance nevertheless saturates after about 100 assimilation cycles and the contribution of the observations becomes less and less significant. The associated standard deviations are very small (close to zero), meaning a very narrow confidence interval with bounds coinciding with the estimates and far from containing the true values of parameters. The TS-EnKF-PF estimates very well both parameters compared to the joint schemes (even though $\theta(2)$ requires about 500 assimilation cycles to be well estimated). Furthermore, the true values of $\theta(2)$ and $\theta(1)$ fall within the TS-EnKF-PF confidence intervals, in contrast with the Joint-PF and the Joint-EnKF. The proposed EnKF-PF provides the most accurate estimates (in terms of reaching the true values) for both parameters, further suggesting confidence intervals that include the true values of these parameters[7].

Figure 3 plots the bias of the analysis estimates of the first four components of the state and the parameters (i.e., the error between the reference states / parameters and the average of their estimates over 30 different set of observations). Overall, the proposed EnKF-PF and, to a lesser extent, the TS-EnKF-PF suggest the lowest biases, followed by the Joint-EnKF, which, in turn, clearly outperforms the Joint-PF.

In the next set of experiments, the filters' performances are evaluated based on the root MSE (RMSE) misfits between the reference augmented states $z(j,n)$ and the filter (analysis) estimates $\hat{z}(j,n)$ averaged over all variables and over the last $N_n = 100$ assimilation cycles

$$\text{RMSE} \;=\; \frac{1}{N_n}\sum_{n=1}^{N_n}\sqrt{\frac{1}{n_{\mathbf{x}}+n_{\boldsymbol{\theta}}}\sum_{j=1}^{n_{\mathbf{x}}+n_{\boldsymbol{\theta}}}\left(z(j,n)-\hat{z}(j,n)\right)^2}. \tag{32}$$

The RMSEs of the parameters' vector and the actual state are evaluated similarly. The relative estimation error is also considered for the (scalar) marginal parameters. Similarly to above, the presented RMSEs / relative errors are averaged over 30 independent repetitions.

[7] As commonly done, we tolerate here some flexibility in the notion of "accuracy" and limit it to the closeness to the reference state/parameters instead of its true PM estimate, which is not accessible. We further provide the (empirical) confidence intervals (or uncertainties) only as a way to assess whether the range of values, which is likely to include the estimates, includes the reference state/parameters.

a. Sensitivity to the ensemble size

We now study the sensitivity of the four filtering algorithms to the ensemble size, assimilating again the data every $\delta_o = 4\delta_m$. Figure 4 plots the RMSE (resp. the relative error), as a function of the ensemble size, of the analysis estimates of the full joint state-parameters' vector, $\mathbf{z}_n = (\mathbf{x}_n, \boldsymbol{\theta})$, the state, $\mathbf{x}_n$, and the parameters' vector, $\boldsymbol{\theta} = (\theta(1), \theta(2))$ (resp. the marginal parameters $\theta(1)$ and $\theta(2)$). As can be seen, the "blended" schemes that combine PF with EnKF roughly lead to comparable performances, and both clearly outperform the Joint-EnKF, especially for not too large ensembles (i.e., $M < 200$). Indeed, one can notice that the Joint-EnKF requires more than 150 members to reach the accuracy of the blended schemes using only 50 members. On the other hand, although the Joint-EnKF with $M \geq 200$ members performs slightly better than the blended schemes in estimating the marginal parameter, $\theta(1)$, it, nevertheless, provides similar performances in estimating the joint parameters' vector, $\boldsymbol{\theta}$. Recalling

$$p(\underbrace{\theta(1), \theta(2)}_{\boldsymbol{\theta}} | \mathbf{y}_{0:n}) = \underbrace{p(\theta(2)|\theta(1), \mathbf{y}_{0:n})}_{term\,1} \underbrace{p(\theta(1)|\mathbf{y}_{0:n})}_{term\,2},$$

this can be explained by the fact that, compared to the Joint-EnKF, the EnKF-PF and TS-EnKF-PF generate more representative analysis members of $\theta(2)$ given those of $\theta(1)$ (i.e., they more efficiently sample $p(\theta(2)|\theta(1)^{a,(m)}, \mathbf{y}_{0:n})$ in *term 1*), which compensates for the effect of (slightly) less accurately estimating $\theta(1)$, which, in turn, may originate from less efficiently sampling $p(\theta(1)|\mathbf{y}_{0:n})$ (in *term 2*). Furthermore, the linear relationship between $\theta(1)$ and $\mathbf{y}_n$ (which comes from (30)-(31) and the linear observation model) may explain the reasonable estimation of $\theta(1)$ by the Joint-EnKF.

b. Sensitivity to the frequency of observations

We now fix the ensemble size to $M = 100$ and test the filters' performances with different lengths of assimilation windows, $f_o = \frac{\delta_o}{\delta_m}$ (i.e., the times at which data are assimilated). Figure 5 displays the RMSE (resp. the relative error), as a function of f_o, of the analysis estimates of the state-parameters' vector, $\mathbf{z}_n$, the state, $\mathbf{x}_n$ and the parameters' vector, $\boldsymbol{\theta}$ (resp. the marginal parameters $\theta(1)$ and $\theta(2)$). Again, overall, the TS-EnKF-PF and the EnKF-PF suggest better performances than the Joint-EnKF, in particular for estimating the full state-parameters' vector. The two former filters lead to comparable results when the assimilation period of the data does not exceed 2 days. However, when the data are assimilated less frequently, the performance of TS-EnKF-PF degrades increasingly with f_o, while the proposed EnKF-PF is shown to be more robust to changes in f_o. One can also see from Sub-Figures (b) and (c) that the Joint-EnKF performs almost as well as the blended schemes in estimating the state, but it is clearly less accurate when it comes to the parameters' vector. The fact that the Joint-EnKF suggests RMSEs closer to those of the hybrid schemes for the state (than for the parameters' vector), is also consistent with the results of Figures 4 and 6 which outline the impact of the ensemble size and the observation noise, respectively. Furthermore, while the Joint-EnKF performs poorly with $\boldsymbol{\theta}$, it still estimates $\theta(1)$ as good as the hybrid algorithms. Based on

$$p(\underbrace{\theta(1),\theta(2)}_{\boldsymbol{\theta}}|\mathbf{y}_{0:n}) = \underbrace{p(\theta(1)|\theta(2),\mathbf{y}_{0:n})}_{term\,1}\underbrace{p(\theta(2)|\mathbf{y}_{0:n})}_{term\,2},$$

the (relative) poor performance of the Joint-EnKF in estimating $\boldsymbol{\theta}$ is not only due to the filter's inability to well estimate the cross-covariance of the posterior marginal pdfs (which is involved in *term 1*), but also to its inability to efficiently sample the posterior $p(\theta(2)|\mathbf{y}_{0:n})$ (*term 2*). The strong nonlinear relationship between $\theta(1)$ and $\theta(2)$, and between $\theta(2)$ and $\mathbf{y}_n$ (see (30)-(31)),

may explain the poor estimation of the cross-covariance in *term 1* and the inefficient sampling in *term 2*, respectively.

c. Sensitivity to the observation noise

The filters are now tested in more challenging scenarios by studying their sensitivity to larger observation noise (i.e., to different variances $\sigma^2 \geq 1$). We fix the ensemble size to $M = 100$ and the assimilation window to $f_o = 4$. Figure 6 displays the RMSE (resp. the relative error), as a function of σ^2, of the analysis estimates of the joint state-parameters' vector $\mathbf{z}_n$, the state $\mathbf{x}_n$ and the parameters' vector $\boldsymbol{\theta}$ (resp. the marginal parameters $\theta(1)$ and $\theta(2)$). The Joint-EnKF, EnKF-PF and, to a lesser extent, TS-EnKF-PF, are not very sensitive to the increase of the noise in the data (especially for $\sigma^2 \geq 2$). Furthermore, for all tested values of σ^2, the proposed EnKF-PF exhibits the best behavior in estimating the joint state-parameters' vector and its marginal (state and parameters) components. For instance, for $\sigma^2 = 5$, which is the noisiest scenario, EnKF-PF is more accurate (in terms of RMSE of estimating $\mathbf{z}_n$) than TS-EnKF-PF with $\sigma^2 = 2$ and Joint-EnKF with $\sigma^2 = 1$. On the other hand, while the TS-EnKF-PF and Joint-EnKF estimate the state almost as accurately as the proposed scheme, the former performs clearly better than the second for both the joint parameters' vector and its marginal $\theta(2)$. As for $\theta(1)$, however, the TS-EnKF-PF performs worse than the Joint-EnKF, which, in turn, compares well with the EnKF-PF. This however does not prevent the TS-EnKF-PF from more accurately computing the analysis members of $\theta(2)$ given those of $\theta(1)$, providing joint estimates of these parameters (i.e. of $\boldsymbol{\theta}$) that are more accurate than those provided by the Joint-EnKF.

d. Case of stochastic state model and nonlinear observation operator

We conduct here the same experiment as above in a scenario where the L96 model is stochastic given the parameters. We therefore include, as in system (1), an additive Gaussian noise with zero

mean and covariance $\mathbf{U} = 0.1\mathbb{I}_{40}$ (see e.g., Stordal et al. (2011); Shen and Tang (2015)). The RMSE and relative errors of the analysis estimates of the state and parameters are displayed in Figure 7. As one can see, the errors are slightly larger than those in Figure 6, but these were obtained in the perfect model case, while roughly suggesting a similar global behavior. More specifically, the proposed filter exhibits the best behavior, especially for strongly noisy scenarios ($\sigma^2 \geq 2$), followed by the TS-EnKF-PF when the state-parameters' vector, $\mathbf{z}$, the parameters' vector, $\boldsymbol{\theta}$, and the marginal parameter, $\theta(2)$, are concerned, or by the Joint-EnKF when the state, $\mathbf{x}$, and the marginal parameter, $\theta(1)$, are concerned.

We further consider a more challenging scenario which includes, as above, a noise in the L96 model, and a strongly nonlinear observation operator $\mathbf{h}_n(\mathbf{x}_n) = 5\tanh(\mathbf{x}_n)$ (Shen and Tang 2015). The averaged estimation errors are plotted in Figure 8. As expected, these errors are generally larger than those of Figure 7. Furthermore, in contrast with the linear observation operator case, the Joint-EnKF outperforms the TS-EnKF-PF in all cases, even for $\mathbf{z}$, $\boldsymbol{\theta}$ and $\theta(2)$. The behavior of the TS-EnKF-PF is probably due to the strong linearity of the observation operator, which may lead to more pronounced multimodal distributions, case for which passing only the mean of the parameters' posterior distribution to one EnKF may result in the filter divergence as the background ensemble in the EnKF may diverge from a high probability region and likewise for passing only the mean of the state posterior distribution to the PF (Santitissadeekorn and Jones 2015, p. 2031). The performance of the TS-EnKF-PF improves for mildly nonlinear observation operator, as can be seen in Figure 9 corresponding to $\mathbf{h}_n(\mathbf{x}_n) = 5\tanh(\frac{\mathbf{x}_n}{10})$ (Shen and Tang 2015).

5. Conclusion

We proposed a Bayesian filtering algorithm for state-parameter dynamical systems of large dimension and depending on few unknown parameters. The new scheme combines the PF to sam-

ple the parameters' ensembles with the EnKF to compute the state ensembles conditional on the parameters' ensembles. In the case of (relatively) small ensembles, the proposed EnKF-PF is expected to perform better than the Joint-PF, as it uses the EnKF for the large-dimensional component (state) to avoid the curse of dimensionality, and the Joint-EnKF, as it uses the PF for the low-dimensional component (parameters) to avoid inconsistency issues related to the limitation of the Gaussian framework. Compared to a recently introduced two-stage EnKF-PF (TS-EnKF-PF) scheme, in which the PF and EnKF exchange information through their ensembles means, the EnKF-PF exchanges the EnKF and PF ensembles.

We tested the proposed EnKF-PF with the nonlinear Lorenz-96 model in which the dissipation factor is parametrized by two parameters, $\boldsymbol{\theta} = (\theta(1), \theta(2))$. We conducted various sensitivity experiments to evaluate the behavior and robustness of the proposed scheme and to compare its performance against those of the Joint-PF, Joint-EnKF and TS-EnKF-PF. In the linear observation operator scenario, the results suggest that the EnKF-PF always outperforms the other filters, being more robust and successfully estimating both state and parameters under different scenarios. While the Joint-PF leads to the worst performances for state estimation, the Joint-EnKF and the TS-EnKF-PF often exhibited performances that are similar to that of the proposed EnKF-PF. For parameters estimation, however, the proposed scheme always provided the best results, exhibiting robust behavior under challenging scenarios, where data are very noisy or assimilated less frequently. Regarding the TS-EnKF-PF and the Joint-EnKF, while the former provides less reliable estimates of the marginal parameter $\theta(1)$ when data are not frequently assimilated, it still nevertheless provides better estimates of the joint parameters' vector, $\boldsymbol{\theta}$. However, in the case of a strongly nonlinear observation operator, the Joint-EnKF becomes more accurate that the TS-EnKF-PF.

In the proposed EnKF-PF, the number of particles / members must be the same in both PF and EnKF, as these communicate via their ensembles, and not only through the means, as in the TS-

EnKF-PF, or the means and cross-covariances, as in the divided EnKF scheme of Luo and Hoteit (2014). The EnKF-PF could however be extended to derive a new two-stage scheme for which the EnKF and PF exchange information via their means, or any other statistics of the ensembles. In this case, one may consider using different ensemble sizes in the EnKF and PF, as in the TS-EnKF-PF and divided EnKF. It could also be generalized to the case of Gaussian-mixture EnKFs, instead of EnKF, and to the unsupervised case for which one or more of the system hyper-parameters is unknown, as for instance the observation noise covariance. Such extensions will be considered in future studies.

APPENDIX A

Some useful random sampling results

Property 1 (Hierarchical sampling (Robert 2007)). Assuming that one can sample from $p(\mathbf{x}_1)$ and $p(\mathbf{x}_2|\mathbf{x}_1)$, then a sample, $\mathbf{x}_2^*$, from $p(\mathbf{x}_2)$ can be obtained by drawing $\mathbf{x}_1^*$ from $p(\mathbf{x}_1)$ and then $\mathbf{x}_2^*$ from $p(\mathbf{x}_2|\mathbf{x}_1^*)$.

Property 2 (Rubin's SIR (Rubin 1988)). Assuming that one has sampled a set, $\{\mathbf{x}^{(s)}\}_{s=1}^S$, of independent and identical (i.i.d.) realizations from a prior, $p(\mathbf{x})$, then, one can asymptotically draw an i.i.d. set, $\{\mathbf{x}^{*(s)}\}_{s=1}^S$, from the posterior, $p(\mathbf{x}|\mathbf{y})$, in two steps:

1. *Weighting.* A normalized weight, $\omega^{(s)} \propto p(\mathbf{y}|\mathbf{x}^{(s)})$, is first assigned to each sample $\mathbf{x}^{(s)}$.
2. *Resampling.* Each sample, $\mathbf{x}^{*(s)}$, is then computed by sampling from the probability mass function, $\sum_{\ell=1}^S \omega^{(\ell)} \delta(\mathbf{x}-\mathbf{x}^{(\ell)})$, where $\delta(.)$ is the Kronecker delta.

Property 3 (Conditional sampling (Hoffman and Ribak 1991)). Consider a Gaussian pdf, $p(\mathbf{x},\mathbf{y})$, with $\mathbf{P}_{xy}$ and $\mathbf{P}_y$ denoting the cross-covariance of $\mathbf{x}$ and $\mathbf{y}$ and the covariance of $\mathbf{y}$, respectively. Then, a sample, $\mathbf{x}^*$, from $p(\mathbf{x}|\mathbf{y})$, can obtained as, $\mathbf{x}^* = \widetilde{\mathbf{x}} + \mathbf{P}_{xy}\mathbf{P}_y^{-1}[\mathbf{y}-\widetilde{\mathbf{y}}]$, where $(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}}) \sim p(\mathbf{x},\mathbf{y})$.

APPENDIX B

The EnKF-PF update steps

a. PF-like update of the parameters

Starting from a forecast ensemble, $\{\boldsymbol{\theta}_n^{f,(m)}\}_{m=1}^M$, an analysis ensemble, $\{\boldsymbol{\theta}_n^{a,(m)}\}_{m=1}^M$, is computed following a PF-like strategy based on Property 2 and (17). The normalized weights of $\boldsymbol{\theta}_n^{f,(m)}$ are first computed as $w_n^{(m)} \propto p(\mathbf{y}_n|\boldsymbol{\theta}_n^{f,(m)}, \mathbf{y}_{0:n-1})$. The forecast ensemble is then resampled according to the resulting weights. As shown in Section 3.a.1, $p(\mathbf{y}_n|\boldsymbol{\theta}_n^{f,(m)}, \mathbf{y}_{0:n-1})$ is approximated by a Gaussian density with parameters given by $\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)})$ and $\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n}$ in (18)-(19), with $\boldsymbol{\theta}_n = \boldsymbol{\theta}_n^{f,(m)}$. Approximating expectations and (cross)-covariances in (18)-(19) using the forecast ensembles $\{\boldsymbol{\theta}_n^{f,(m)}\}_{m=1}^M$ and $\{\mathbf{y}_n^{f,(m)}\}_{m=1}^M$, one eventually obtains (24) and (25) as approximations of $\check{\mathbf{y}}_n(\boldsymbol{\theta}_n^{f,(m)})$ and $\check{\mathbf{P}}_{\mathbf{y}_n|\boldsymbol{\theta}_n}$, respectively.

b. EnKF-like update of the state

The state analysis ensemble, $\{\mathbf{x}_n^{a,(m)}\}_{m=1}^M$, is computed using an EnKF-like update based on (20) and Property 3, by assuming that $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n, \mathbf{y}_{0:n-1})$ is Gaussian. Indeed, for each $m = 1, \cdots, M$, given $\xi_{\mathbf{x}_n}^{(m)} \sim p(\mathbf{x}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$ and $\xi_{\mathbf{y}_n}^{(m)} \sim p(\mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$, one obtains $\mathbf{x}_n^{a,(m)}$ through an update of $\xi_{\mathbf{x}_n}^{(m)}$ based on $\mathbf{y}_n$, following a KF-like correction with a gain, $\mathbf{K}_n = \mathrm{cov}[\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1}] \left(\mathrm{cov}[\mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1}]\right)^{-1}$:

$$\mathbf{x}_n^{a,(m)} = \xi_{\mathbf{x}_n}^{(m)} + \mathbf{K}_n(\mathbf{y}_n - \xi_{\mathbf{y}_n}^{(m)}). \tag{B1}$$

An ensemble estimation of the (cross)-covariances in the expression of the gain, $\mathbf{K}_n$, can be obtained from the LMMSE-based estimation of the covariance of $p(\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1})$: $\mathrm{cov}[\mathbf{x}_n, \mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1}] \approx \mathbf{P}_{\mathbf{x}_n^f, \boldsymbol{\eta}_n^f} - \mathbf{K}_{\mathbf{x}_n^f, \boldsymbol{\theta}_n^f} \mathbf{P}_{\boldsymbol{\theta}_n^f, \boldsymbol{\eta}_n^f}$ and $\mathrm{cov}[\mathbf{y}_n|\boldsymbol{\theta}_n^{a,(m)}, \mathbf{y}_{0:n-1}] \approx \mathbf{P}_{\boldsymbol{\eta}_n^f} - \mathbf{K}_{\boldsymbol{\eta}_n^f, \boldsymbol{\theta}_n^f} \mathbf{P}_{\boldsymbol{\theta}_n^f, \boldsymbol{\eta}_n^f} + \mathbf{V}_n$. These are then inserted in (B1) to obtain the analysis update (29).

Finally, one needs to compute the members $\xi^{(m)}_{\mathbf{x}_n}$ (from which the expression (28) of $\xi^{(m)}_{\mathbf{y}_n}$ is readily obtained by applying Prop. 1 on eq. (23)). To that end, we propose to sample them from the (assumed) Gaussian pdf $p(\mathbf{x}_n|\boldsymbol{\theta}^{a,(m)}_n, \mathbf{y}_{0:n-1})$ with moments $(\check{\mathbf{x}}_n(\boldsymbol{\theta}^{a,(m)}_n), \check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n})$. Using the forecast ensembles, $\{\mathbf{x}^{f,(m)}_n\}_{m=1}^M$ and $\{\boldsymbol{\theta}^{f,(m)}_n\}_{m=1}^M$, the mean $\check{\mathbf{x}}_n(\boldsymbol{\theta}^{a,(m)}_n)$ (in (21)) is approximated as in (26) and the covariance $\check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n}$ (in (22)) is approximated by $\mathbf{P}_{\mathbf{x}^f_n} - \mathbf{P}_{\mathbf{x}^f_n,\boldsymbol{\theta}^f_n}\mathbf{P}^{-1}_{\boldsymbol{\theta}^f_n}\mathbf{P}_{\boldsymbol{\theta}^f_n,\mathbf{x}^f_n}$. This leads to

$$\xi^{(m)}_{\mathbf{x}_n} = \check{\mathbf{x}}_n(\boldsymbol{\theta}^{a,(m)}_n) + \mathbf{C}_{\mathbf{x}_n}\mathbf{r}^{(m)}_n, \tag{B2}$$

where $\mathbf{r}^{(m)}_n \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_M)$ and $(\mathbf{C}_{\mathbf{x}_n})_{n_\mathbf{x}\times M}$ is a square-root of $\check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n}$. To compute a square-root of $\check{\mathbf{P}}_{\mathbf{x}_n|\boldsymbol{\theta}_n} \approx \mathbf{S}_{\mathbf{x}^f_n}\Lambda_n\mathbf{S}^T_{\mathbf{x}^f_n}$ with $\Lambda_n = \mathbb{I}_M - \mathbf{S}^T_{\boldsymbol{\theta}^f_n}\mathbf{P}^{-1}_{\boldsymbol{\theta}^f_n}\mathbf{S}_{\boldsymbol{\theta}^f_n}$, one uses the SVD decomposition to factorize the $M \times M$ symmetric matrix Λ_n as $\mathscr{U}_n\Sigma_n\mathscr{U}_n^T$, where the matrix $\mathscr{U}_n$ is orthogonal and Σ_n diagonal. One then computes $\mathbf{C}_{\mathbf{x}_n} = \mathbf{S}_{\mathbf{x}^f_n}\mathscr{U}_n(\Sigma_n)^{\frac{1}{2}}$, which is finally used in (B2) to obtain (27).

References

Ades, M., and P. van Leeuwen, 2013: An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, **139**, 820–40.

Ait-El-Fquih, B., and F. Desbouvries, 2006: Kalman filtering in triplet Markov chains. *IEEE Transactions on Signal Processing*, **54 (8)**, 2957–63.

Ait-El-Fquih, B., and F. Desbouvries, 2011: Fixed-Interval Kalman Smoothing Algorithms in singular state-space systems. *The Journal of Signal Processing Systems*, **65 (3)**, 469–78.

Ait-El-Fquih, B., M. Gharamti, and I. Hoteit, 2016: A Bayesian Consistent Dual Ensemble Kalman Filter for State-Parameter Estimation in Subsurface Hydrology. *Hydrology and Earth System Sciences*, **20**, 3289–3307.

Ait-El-Fquih, B., and I. Hoteit, 2015: An efficient multiple particle filter based on the variational bayesian approach. *Proceedings of the IEEE International ISSPIT Symposium*, Abu Dhabi, UAE.

Ait-El-Fquih, B., and I. Hoteit, 2016: A variational Bayesian multiple particle filtering scheme for large-dimensional systems. *IEEE Transactions on Signal Processing*, **64 (20)**, 5409–22.

Aksoy, A., F. Zhang, and J. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophysical Research Letters*, **33**, L12 801.

Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering*. Prentice Hall, Englewood Cliffs, New Jersey.

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.

Annan, J. D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005: Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. *Nonlinear Processes Geophys.*, **12**, 363–71.

Bellsky, T., J. Berwald, and L. Mitchell, 2014: Nonglobal parameter estimation using local ensemble Kalman filtering. *Monthly Weather Review*, **142**, 2150–64.

Bengtsson, T., P.Bickel, , and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, **2**, 316–34.

Bishop, C., B. Etherton, and S. Majumdar, 2001: Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Monthly Weather Review*, **129**, 420–36.

Carrassi, A., and S. Vannitsem, 2011: Parameter estimation using a particle method: Inference mixing coefficients from sea-level observations. *Quarterly Journal of the Royal Meteorological Society*, **137**, 435–51, doi:10.1002/qj.762.

Chen, Y., and D. Zhang, 2006: Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, **29**, 1107–1122.

Crisan, D., and A. Doucet, 2002: A survey on convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, **50 (3)**, 736–46.

Desbouvries, F., and B. Ait-El-Fquih, 2008: Direct, prediction-based and smoothing-based particle filter algorithms. *Proceedings of the 4th world conference of the IASC*, Yokohama, Japan.

Desbouvries, F., Y. Petetin, and B. Ait-El-Fquih, 2011: Direct, Prediction- and Smoothing-based Kalman and Particle Filter Algorithms. *Signal Processing*, **91 (8)**, 2064–2077.

Djuric, P., and M. Bugallo, 2013: Particle Filtering for High-Dimensional Systems. *Proceedings of the IEEE International Workshop on CAMSAP*.

Doucet, A., N. de Freitas, and N. Gordon, Eds., 2001a: *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science, Springer Verlag, New York.

Doucet, A., N. D. Freitas, K. P. Murphy, and S. J. Russell, 2000: Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks. *Proceedings of the 16th world Conference on UAI*, Stanford, California, USA, 176–83.

Doucet, A., N. J. Gordon, and V. Krishnamurthy, 2001b: Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, **49 (3)**, 613–24.

Dunne, S., D. Entekhabi, and E. Njoku, 2007: Impact of Multiresolution Active and Passive Microwave Measurements on Soil Moisture Estimation Using the Ensemble Kalman Smoother. *IEEE Transactions on Geoscience and Remote Sensing*, **45 (4)**, 1016–28.

Evensen, G., 1994: Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99 (C5)**, 143–62.

Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. New York: Springer.

Evensen, G., and P. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, **128 (6)**, 1852–67.

Frei, M., and H. Künsch, 2012: Sequential State and Observation Noise Covariance Estimation Using Combined Ensemble Kalman and Particle Filters. *Monthly Weather Review*, **140**, 1476–95, doi:10.1175/MWR-D-10-05088.1.

Frei, M., and H. R. Künsch, 2013: Mixture ensemble Kalman filters. *Computational Statistics and Data Analysis*, **58**, 127–38.

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.

Ghahramani, Z., and M. Jordan, 1997: Factorial Hidden Markov Models. *Machine Learning*, **29**, 245–273.

Gharamti, M., J. Valstar, and I. Hoteit, 2014: An adaptive hybrid EnKF-OI scheme for efficient state-parameter estimation of reactive contaminant transport models. *Advances in Water Resources*, **71**, 1–15.

Gharamti, M. E., B. Ait-El-Fquih, and I. Hoteit, 2015: An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models. *Journal of Hydrology*, **527**, 442–57.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/ non-Gaussian Bayesian state estimation. *IEE Proceedings - F*, **140**, 107–113.

Guimaraes, A., B. Ait-El-Fquih, and F. Desbouvries, 2010: A Fixed-Lag Particle Smoother for Blind SISO Equalization of Time-Varying Channels. *IEEE Transactions on Wireless Communications*, **9 (2)**, 512–16.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Monthly Weather Review*, **128 (8)**, 2905–2919.

Hamill, T. M., and J. S. Whitaker, 2010: What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Monthly Weather Review*, **139**, 117–31.

Hendricks Franssen, H., and W. Kinzelbach, 2008: Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, **44 (9)**.

Hoffman, Y., and E. Ribak, 1991: Constrained realizations of Gaussian fields - a simple algorithm. *The Astrophysical Journal*, **380 (491)**, L5–L8.

Hoteit, I., X. Luo, and D. Pham, 2012: Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters. *Monthly Weather Review*, **140**, 528–42.

Hoteit, I., D. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine Systems*, **36 (1–2)**, 101–127.

Hoteit, I., D.-T. Pham, M. Gharamti, and X. Luo, 2015: Mitigating observation perturbation sampling errors in the stochastic EnKF. *Monthly Weather Review*, **143**, 2918–36.

Hoteit, I., D.-T. Pham, G. Triantafyllou, and G. Korres, 2008: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, **136 (1)**, 317–334.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–26.

Husz, Z. L., A. M. Wallace, and P. R. Green, 2011: Tracking with a hierarchical partitioned particle filter and movement modelling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **41 (6)**, 1571–84.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*, Mathematics in Science and Engineering, Vol. 64. Academic Press, San Diego.

Kailath, T., A. H. Sayed, and B. Hassibi, 2000: *Linear estimation*. Prentice Hall Information and System Sciences Series, Prentice Hall, Upper Saddle River, NJ.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering, Series D*, **82 (1)**, 35–45.

Karimi, A., and M. R. Paul, 2010: Extensive chaos in the Lorenz-96 model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **20 (4)**.

Kivman, G. A., 2003: Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, **10 (3)**, 253–59.

Künsch, H., 2001: State space and hidden markov models. *Complex Stochastic Systems*, O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Eds., CRC Press, chap. 3, 109–173.

Liu, B., B. Ait-El Fquih, and I. Hoteit, 2015: Efficient Kernel-Based Ensemble Gaussian Mixture Filtering. *Monthly Weather Review*, **144**, 781–800.

Liu, J. S., and R. Chen, 1998: Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, **93 (443)**, 1032–44.

Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, **55 (3)**, 399–414.

Luo, X., and I. Hoteit, 2014: Ensemble Kalman filtering with a divided state-space strategy for coupled data assimilation problems. *Monthly Weather Review*, **142**, 4542–58.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser, 2005: Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, **28 (2)**, 135–147.

Morzfelda, M., X. Tub, E. Atkins, and A. Chorina, 2012: A random map implementation of implicit filters. *Journal of Computational Physics*, **231**, 2049–66.

Pham, D., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**, 1194–1207.

Rasmussen, J., H. Madsen, K. H. Jensen, and J. C. Refsgaard, 2015: Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. *Hydrology Earth System Science*, **19**, 2999–3013.

Robert, C., 2007: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Science & Business Media, New York.

Rubin, D., 1988: Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics*, J. Bernardo, M. DeGroot, D. Lindley, and A. Smith, Eds., Vol. 3, Oxford University Press, 395–402.

Santitissadeekorn, N., and C. Jones, 2015: Two-Stage Filtering for Joint State-Parameter Estimation. *Monthly Weather Review*, **143**, 2028–41.

Schön, T., F. Gustafsson, and P.-J. Nordlund, 2005: Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, **53 (7)**, 2279–89.

Septier, F., and G. W. Peters, 2015: *An Overview of Recent Advances in Monte-Carlo Methods for Bayesian Fitlering in High-Dimensional Spaces*, chap. 2 : "Theoretical Aspects of Spatial-Temporal Modeling". G. w. peters and t. matsui ed., SpringerBriefs - JSS Research Series in Statistics.

Shen, Z., and Y. Tang, 2015: A modified ensemble Kalman particle filter for non-Gaussian systems with nonlinear measurement functions. *Journal of Advanced in Modeling Earth Systems*, **7**, 50–66.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation. *Monthly Weather Review*, **143**, 195–211, doi:10.1175/MWR-D-14-00051.1.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles To High-Dimensional Particle Filtering. *Monthly Weather Review*, **136 (12)**, 4629–40.

Song, H., I. Hoteit, B. Cornuelle, and A. Subramanian, 2010: An adaptive approach to mitigate background covariance limitations in the ensemble Kalman filter. *Monthly Weather Review*, **138**, 2825–45.

Spiller, E. T., A. Budhiraja, K. Ide, and C. K. Jones, 2008: Modified particle filter methods for assimilating Lagrangian data into a point-vortex model. *Physica D: Nonlinear Phenomena*, **237 (10)**, 1498–1506.

Stordal, A. S., H. A. Karlsen, G. Naevdal, H. J. Skaug, and B. Valles, 2011: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, **15 (2)**, 293–305.

Subramanian, A., I. Hoteit, B. Cornuelle, and H. Song, 2012: Linear vs. nonlinear filtering with scale selective corrections for balanced dynamics in a simple atmospheric model. *Journal of Atmospheric Sciences*, **69**, 3405–19.

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

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

van Trees, H., 1968: *Detection, Estimation, and Modulation Theory : Part I*. John wiley and Sons, New York.

Wen, X. H., and W. H. Chen, 2007: Real-time reservoir updating using ensemble kalman filter: The confirming approach. *Society of Petroleum Engineering*, **11**, 431–442.

West, M., and J. Liu, 2001: Combined parameter and state estimation in simulation-based filtering. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., Springer, 197–223.

Zhang, H., H.-J. Hendricks-Franssen, X. Han, J. Vrugt, and H. Vereecken, 2016: Joint State and Parameter Estimation of Two Land Surface Models Using the Ensemble Kalman Filter and

774 Particle Filter. *Hydrology and Earth System Sciences Discuss.*, doi:10.5194/hess-2016-42, in
775 review.

LIST OF FIGURES

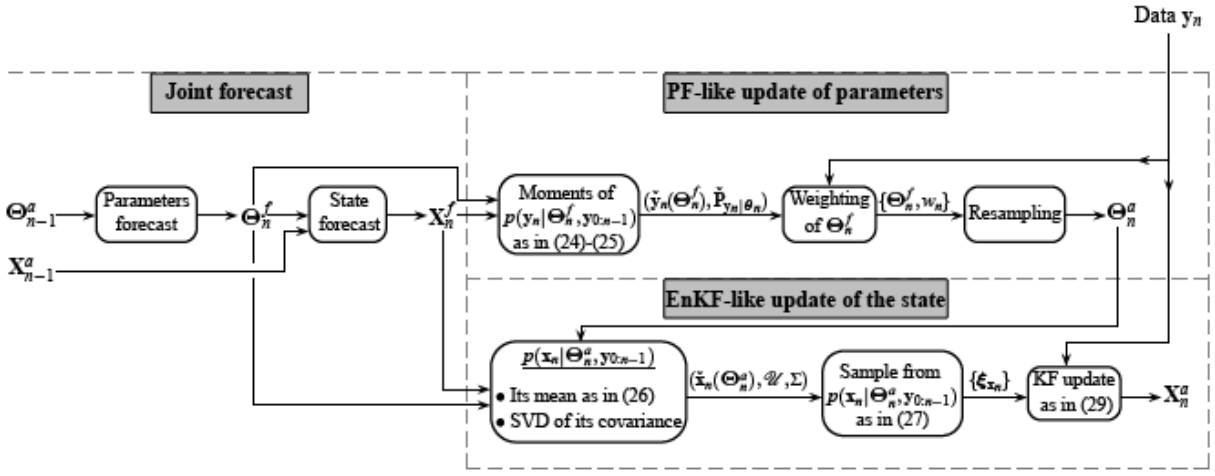


FIG. 1: A schematic illustration of the steps of computing a joint analysis ensemble $(\mathbf{X}^a_n, \Theta^a_n)$ from the previous one $(\mathbf{X}^a_{n-1}, \Theta^a_{n-1})$ using the EnKF-PF.

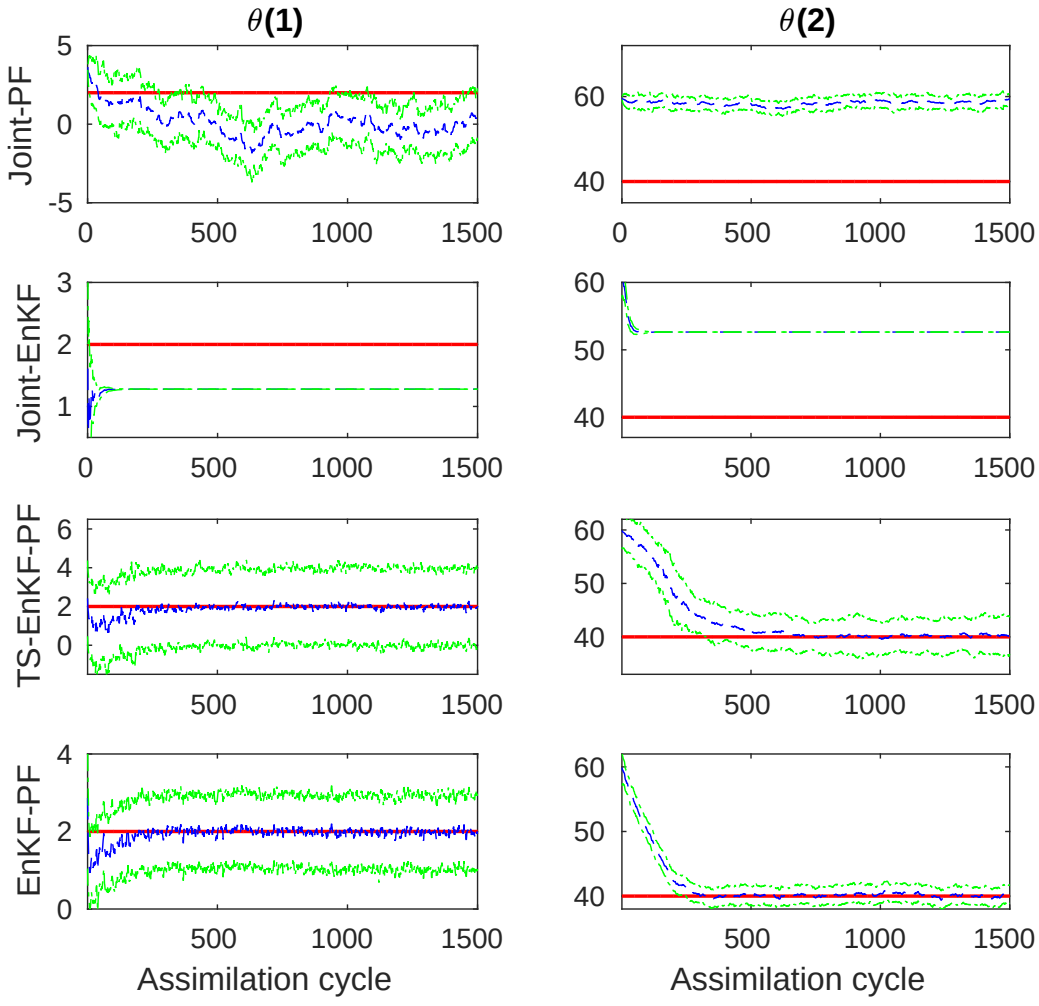
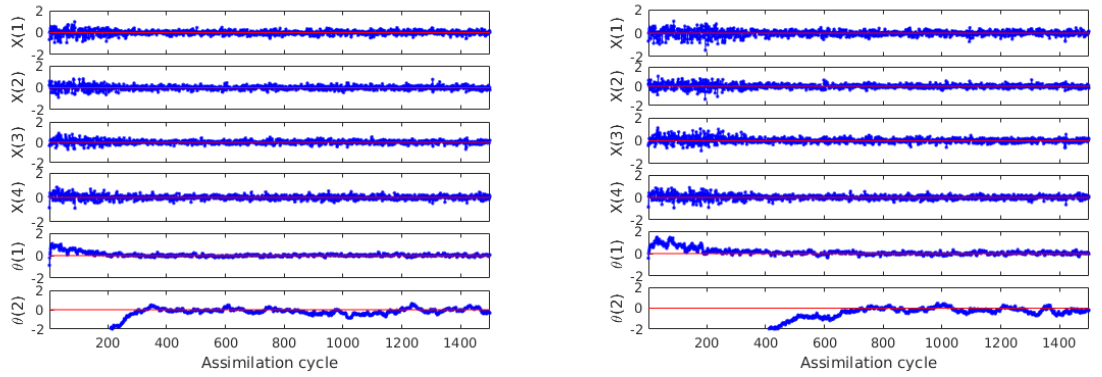
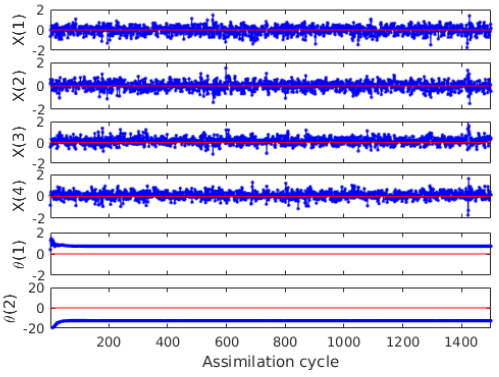


FIG. 2: Time-evolution of the parameters analysis estimates (dashed blue) and associated 95% confidence intervals (green) using the Joint-PF, Joint-EnKF, TS-EnKF-PF and EnKF-PF with 100 members. The true values of parameters are indicated in red. In the last three filters, the localization length scale and inflation factor were set to 2 and 1.2, respectively. All observations were assimilated at every four model time steps (one day in real time).

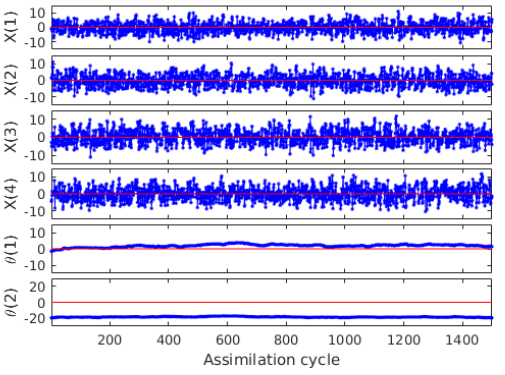


(a) EnKF-PF

(b) TS-EnKF-PF



(c) Joint-EnKF



(d) Joint-PF

FIG. 3: Time-evolution of the bias (dashed blue) of the first four components of the state and the parameters, computed as the error between the true values of these variables and the average (over 30 independent repetitions) of their analysis estimates obtained using the EnKF-PF, TS-EnKF-PF, Joint-EnKF and Joint-PF with 100 members. In the first three filters, the localization length scale and inflation factor were set to 2 and 1.2, respectively. All observations were assimilated at every four model time steps (one day in real time).

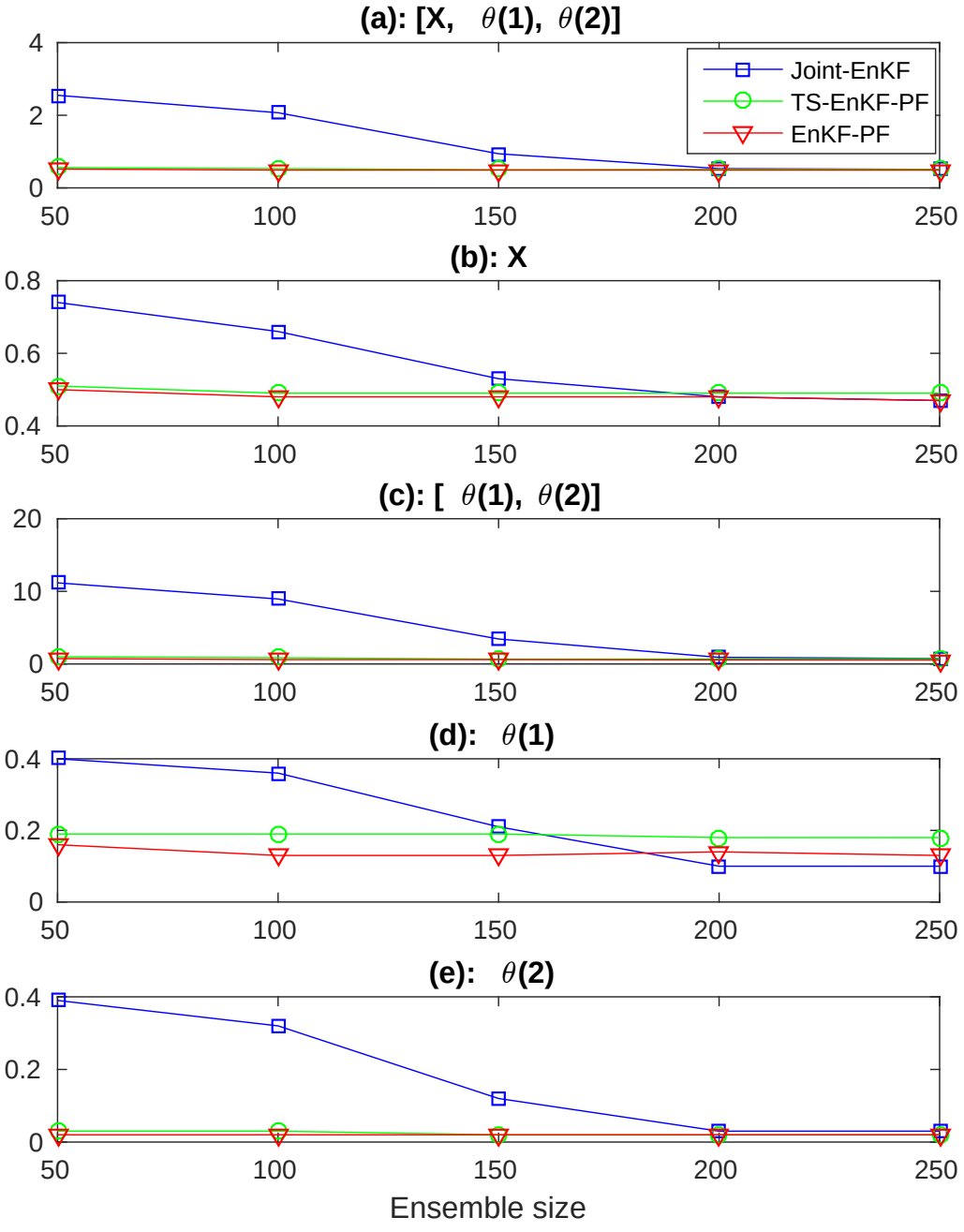


FIG. 4: Time- and variable-averaged RMSE of the analysis estimates of state-parameters' vector (a), state (b) and parameters' vector (c), and time-averaged relative error of the analysis estimates of marginal parameters ((d) and (e)), provided by Joint-EnKF, TS-EnKF-PF and EnKF-PF as a function of ensemble size. In all filters, the localization length scale and inflation factor were set to 2 and 1.2, respectively. All observations were assimilated at every four model time steps (one day in real time).

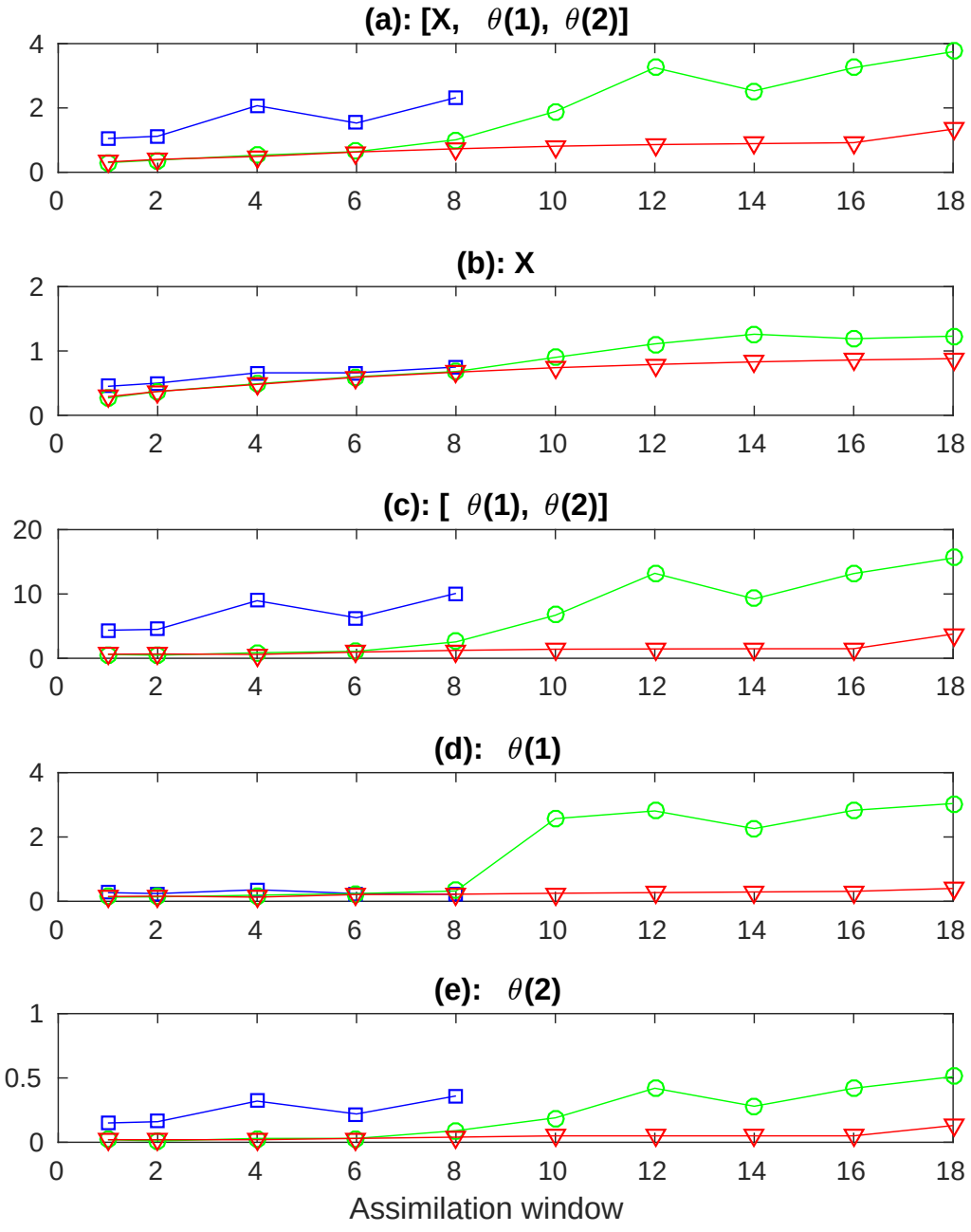


FIG. 5: Time- and variable-averaged RMSE of the analysis estimates of state-parameters' vector (a), state (b) and parameters' vector (c), and time-averaged relative error of the analysis estimates of marginal parameters ((d) and (e)), provided by the Joint-EnKF (blue square), TS-EnKF-PF (green circle) and EnKF-PF (red triangle) as a function of the temporal assimilation period. In all filters, ensemble size, localization length scale and inflation factor were set to 100, 2 and 1.2, respectively.

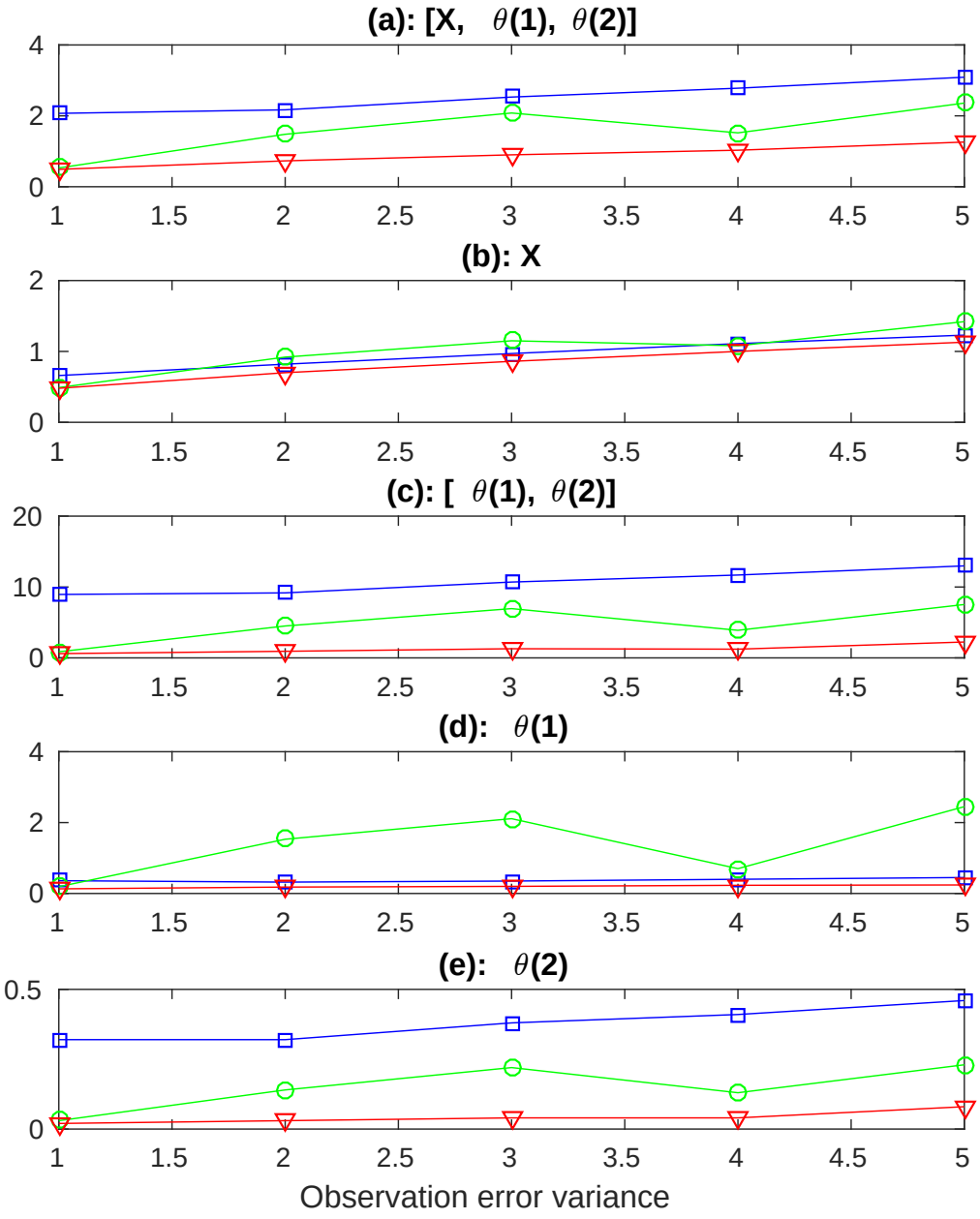


FIG. 6: Time- and variable-averaged RMSE of the analysis estimates of state-parameters' vector (a), state (b) and parameters' vector (c), and time-averaged relative error of the analysis estimates of marginal parameters ((d) and (e)), provided by the Joint-EnKF (blue square), TS-EnKF-PF (green circle) and EnKF-PF (red triangle) as a function of observation error variance. In all filters, ensemble size, localization length scale and inflation factor were set to 100, 2 and 1.2, respectively. All observations were assimilated at every four model time steps (one day in real time).

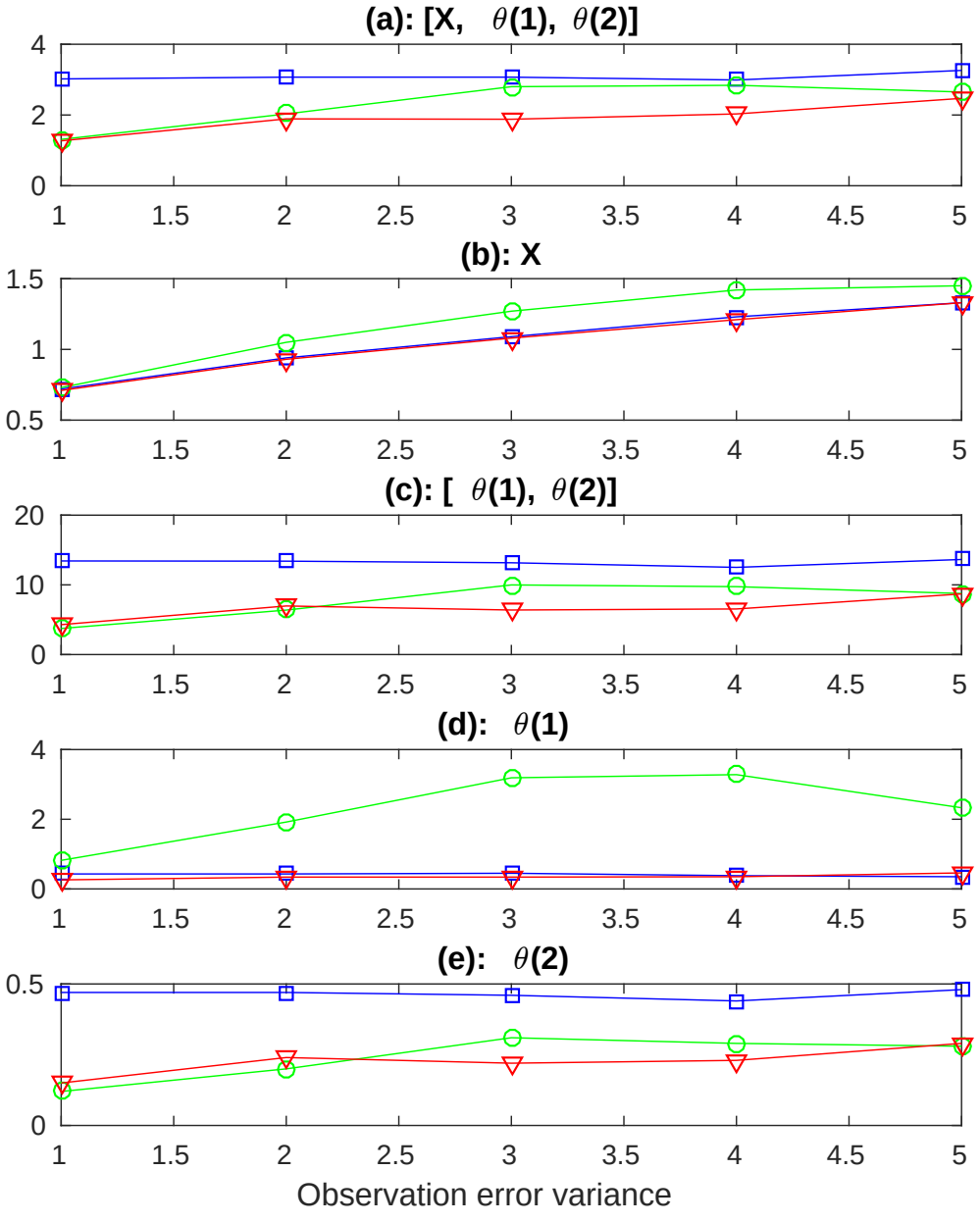


FIG. 7: The same as Fig. 6 in the case of an imperfect state model with an error variance of 0.1.

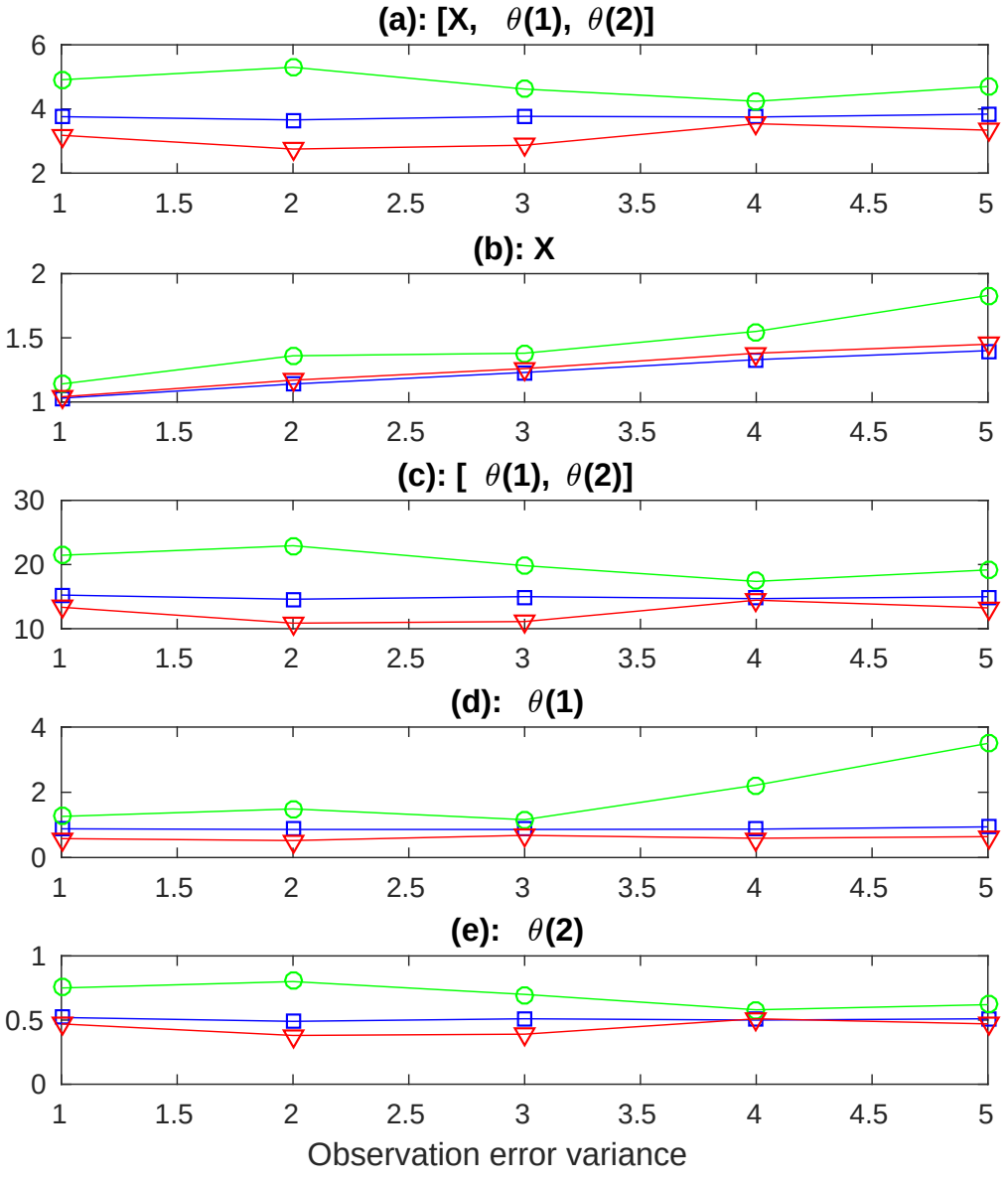


FIG. 8: The same as Fig. 7 in the case of a strongly nonlinear observation operator.

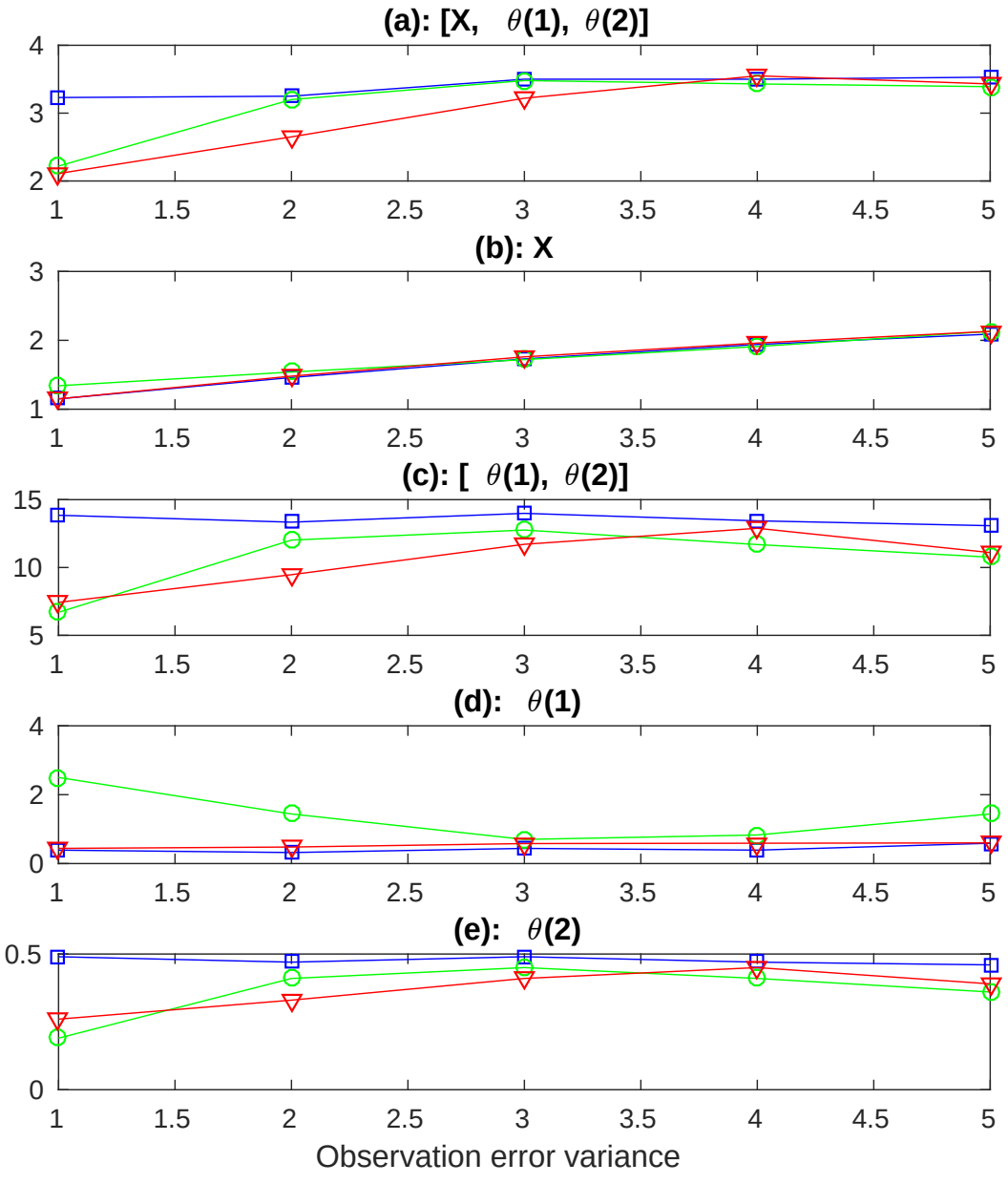


FIG. 9: The same as Fig. 7 in the case of a weakly nonlinear observation operator.