



Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters*

Item Type	Article
Authors	Hoteit, Ibrahim; Luo, Xiaodong; Pham, Dinh-Tuan
Citation	Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters* 2012, 140 (2):528 Monthly Weather Review
Eprint version	Publisher's Version/PDF
DOI	10.1175/2011MWR3640.1
Publisher	American Meteorological Society
Journal	Monthly Weather Review
Rights	<p>© Copyright 2012 American Meteorological Society (AMS). Permission to use figures, tables, and brief excerpts from this work in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is determined to be "fair use" under Section 107 of the U.S. Copyright Act September 2010 Page 2 or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the AMS's permission. Republication, systematic reproduction, posting in electronic form, such as on a web site or in a searchable database, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. Additional details are provided in the AMS Copyright Policy, available on the AMS Web site located at (http://www.ametsoc.org/) or from the AMS at 617-227-2425 or copyrights@ametsoc.org.</p>

Download date	23/08/2022 14:20:59
Link to Item	http://hdl.handle.net/10754/552775

Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters*

IBRAHIM HOTEIT AND XIAODONG LUO

King Abdullah University of Sciences and Technology, Thuwal, Saudi Arabia

DINH-TUAN PHAM

Centre National de la Recherche Scientifique, Grenoble, France

(Manuscript received 27 September 2010, in final form 26 July 2011)

ABSTRACT

This paper investigates an approximation scheme of the optimal nonlinear Bayesian filter based on the Gaussian mixture representation of the state probability distribution function. The resulting filter is similar to the particle filter, but is different from it in that the standard weight-type correction in the particle filter is complemented by the Kalman-type correction with the associated covariance matrices in the Gaussian mixture. The authors show that this filter is an algorithm in between the Kalman filter and the particle filter, and therefore is referred to as the particle Kalman filter (PKF).

In the PKF, the solution of a nonlinear filtering problem is expressed as the weighted average of an “ensemble of Kalman filters” operating in parallel. Running an ensemble of Kalman filters is, however, computationally prohibitive for realistic atmospheric and oceanic data assimilation problems. For this reason, the authors consider the construction of the PKF through an “ensemble” of ensemble Kalman filters (EnKFs) instead, and call the implementation the particle EnKF (PEnKF). It is shown that different types of the EnKFs can be considered as special cases of the PEnKF. Similar to the situation in the particle filter, the authors also introduce a resampling step to the PEnKF in order to reduce the risk of weights collapse and improve the performance of the filter. Numerical experiments with the strongly nonlinear Lorenz-96 model are presented and discussed.

1. Introduction

Estimating the state of the atmosphere and the ocean has long been one of the main goals of modern science. Data assimilation, which consists of combining data and dynamical models to determine the best possible estimate of the state of a system, is now recognized as the best approach to tackle this problem (Ghil and Malanotte-Rizzoli 1991). The strongly nonlinear character of the atmospheric and oceanic models, combined with their important computational burden, makes data assimilation in these systems quite challenging.

Based on the Bayesian estimation theory, the optimal solution of the nonlinear data assimilation problem can be obtained from the optimal nonlinear filter (ONF; Doucet et al. 2001). This involves the estimation of the conditional probability distribution function (pdf; not necessarily Gaussian) of the system state given all available measurements up to the estimation time. Knowledge of the state pdf allows determining different estimates of the state, such as the minimum variance estimate or the maximum a posteriori estimate (Todling 1999). The ONF recursively operates as a succession of a correction (or analysis) step at measurement times to correct the state (predictive) pdf using the Bayes' rule, and a prediction step to propagate the state (analysis) pdf to the time of the next available observation. Although conceptually simple, the numerical implementation of the optimal nonlinear filter can be computationally prohibitive, even for systems with few dimensions (Doucet et al. 2001). Its use with atmospheric and oceanic data assimilation problems is therefore not possible because of the huge dimension of these systems.

* Supplemental information related to this paper is available at the Journals Online Web site: <http://dx.doi.org/10.1175/2011MWR3640.s1>.

Corresponding author address: Ibrahim Hoteit, Earth Sciences and Engineering/Applied Mathematics and Computational Science, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
E-mail: ibrahim.hoteit@kaust.edu.sa

In recent years, two approximation schemes of the ONF have attracted the attention of researchers for their potentials to tackle nonlinear and non-Gaussian data assimilation problems. One is based on the point-mass representation (mixture of Dirac functions) of the state pdf, and leads to the celebrated particle filter (PF; Doucet et al. 2001; Pham 2001; Nakano et al. 2007; Van Leeuwen 2003, 2009). The other is based on the Gaussian mixture representation of the state pdf, and results in a filter that is in between the Kalman filter and the particle filter (Anderson and Anderson 1999; Bengtsson et al. 2003; Chen and Liu 2000; Hoteit et al. 2008; Luo et al. 2010; Sorenson and Alspach 1971), as to be shown later. For this reason, we refer to this filter as the particle Kalman filter (PKF).

In terms of computational efficiency, the particle filter needs to generate large samples for a good approximation of the state pdf. In certain circumstances, in order to avoid weights collapse, the number of samples needs to scale exponentially with the dimension of the system in assimilation (Bengtsson et al. 2008), which may be infeasible for high-dimensional systems (Snyder et al. 2008). On the other hand, in some comparison studies (Han and Li 2008; Nakano et al. 2007), it has been reported that the ensemble Kalman filter (EnKF) and its variants (Anderson 2001; Bishop et al. 2001; Burgers et al. 1998; Evensen 1994; Evensen and van Leeuwen 1996; Houtekamer and Mitchell 1998; Whitaker and Hamill 2002) can achieve lower estimation errors than the particle filter given a small ensemble size. In this paper we confine ourselves to the PKF and make performance comparison only between the PKF and the EnKF.

Using a Gaussian mixture representation of the state pdf, the resulting PKF consists of an ensemble of parallel nonlinear Kalman filters (Hoteit et al. 2008; Luo et al. 2010). Different variants of the Kalman filter (KF), including the extended Kalman filter (Chen and Liu 2000; Sorenson and Alspach 1971), the reduced-rank Kalman filter (Hoteit et al. 2008; Luo et al. 2010), the EnKF (Anderson and Anderson 1999; Bengtsson et al. 2003), can be used to construct the PKF. The focus of this paper is to investigate the PKF that is constructed by an ensemble of parallel EnKFs. Common to all the implementations of the PKF, the mixture of normal distributions (MON)—a more general pdf representation than the single Gaussian pdf approximation in the EnKF—can be used to tackle nonlinearity and non-Gaussianity in data assimilation. On the other hand, choosing the EnKF to construct the PKF is based on the consideration of computational efficiency, since the EnKF itself is a very efficient algorithm for data assimilation in high-dimensional systems. In this regard, this work is very similar to the earlier works of Anderson and Anderson (1999) and Bengtsson et al.

(2003), but is different from them mainly in the following aspect.

In Anderson and Anderson (1999) and Bengtsson et al. (2003), the PKF was constructed without a resampling step. As a result, the PKF may suffer from weights collapse as in the particle filter. To overcome this problem, Bengtsson et al. (2003) considered a hybrid of the EnKF and the PKF, which, however, involves the computation of the inverses of sample covariance matrices in the “global to local” adjustments. In doing so, it is not only computationally intensive, but also encounters singularities in computing the inverses when the ensemble size is smaller than the system dimension, such that the sample covariances themselves are rank deficient. Therefore, it is not clear how the hybrid scheme in Bengtsson et al. (2003) can be applied to the scenario with the ensemble size smaller than the system dimension. For the implementation of the PKF scheme in this work, we introduce a resampling step similar to those in Musso et al. (2001) and Stavropoulos and Titterton (2001) to tackle the collapse of the weights. Our experience shows that, with this resampling step, the PKF becomes much more stable and can conduct data assimilation in the small ensemble scenario, as to be demonstrated through the numerical experiments presented in this work.

As may be of particular interest for the ensemble filtering community, we will show that different EnKFs can be considered as special cases of the PEnKF following our implementation. This point of view allows for a better understanding of the EnKFs’ behaviors and/or their differences.

The paper is organized as follows. The optimal nonlinear filter is first described in section 2. The PKF and its ensemble implementation are discussed in section 3. Results of numerical experiments with the Lorenz-96 model are presented in section 4. A summary of the main results and a general discussion on the potential of the PEnKF for tackling realistic atmospheric and oceanic data assimilation problems concludes the paper in section 5.

2. The optimal nonlinear filter

Starting from a random initial condition with a known probability density function, the optimal nonlinear filter provides the conditional density function of the system state given all available measurements up to the estimation time. To describe the algorithm of the optimal nonlinear filter, consider the nonlinear stochastic discrete-time dynamical system:

$$\mathbf{x}_k = M_k(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k, \quad (1)$$

$$\mathbf{y}_k = H_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k, \quad (2)$$

where \mathbf{x}_k is the state vector (to be estimated) of dimension n ; \mathbf{y}_k is the observation vector of dimension p ; M_k and H_k are two continuously differentiable maps from \mathbb{R}^n to \mathbb{R}^n and from \mathbb{R}^n to \mathbb{R}^p , respectively, representing the transition and the observational operators; and $\boldsymbol{\eta}_k$ and $\boldsymbol{\epsilon}_k$ denote the dynamical and the observational noise, respectively. We assume that $\boldsymbol{\eta}_k$ and $\boldsymbol{\epsilon}_k$ are Gaussian with zero mean and nonsingular covariance matrices \mathbf{Q}_k and \mathbf{R}_k , respectively, and are independent of the system state at any time instant. Under this setting, the dynamical system in Eq. (1) is Markovian.

The optimal nonlinear filter recursively operates with a succession of prediction and correction steps as summarized below. The reader is referred to Doucet et al. (2001) for an extensive description of the filter. To

simplify the notation, $\mathbf{y}_{1:k}$ is defined as a shorthand for the set of all observations $\mathbf{y}_1, \dots, \mathbf{y}_k$ up to and including time t_k . Let $p_k^f(\cdot | \mathbf{y}_{1:k-1})$ be the conditional (predictive) pdf of \mathbf{x}_k given $\mathbf{y}_{1:k-1}$ and $p_k^a(\cdot | \mathbf{y}_{1:k})$ be the conditional (analysis) pdf of \mathbf{x}_k given $\mathbf{y}_{1:k}$, both determined at time t_k . The filter steps are described as follows.

- *Prediction step:* Given the analysis pdf $p_{k-1}^a(\cdot | \mathbf{y}_{1:k-1})$ at time t_{k-1} , the predictive pdf $p_k^f(\cdot | \mathbf{y}_{1:k-1})$ is obtained by integrating $p_{k-1}^a(\cdot | \mathbf{y}_{1:k-1})$ with the model in Eq. (1) to the time of the next available observation t_k . Under the assumptions made on the model noise $\boldsymbol{\eta}_k$, the likelihood function for the state vector \mathbf{x}_{k-1} to transit to \mathbf{x}_k at the next time instant is described by the Gaussian pdf $N[\mathbf{x}_k : M_k(\mathbf{x}_{k-1}), \mathbf{Q}_k]$, where $N(\mathbf{x} : \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Thus,

$$p_k^f(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int_{\mathbb{R}^n} N[\mathbf{x}_k : M_k(\mathbf{x}_{k-1}), \mathbf{Q}_k] p_{k-1}^a(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}. \quad (3)$$

- *Correction step:* After a new observation \mathbf{y}_k has been made, the analysis pdf $p_k^a(\cdot | \mathbf{y}_{1:k})$ at time t_k is updated from $p_k^f(\cdot | \mathbf{y}_{1:k-1})$ using Bayes' rule:

$$p_k^a(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{1}{b_k} p_k^f(\mathbf{x}_k | \mathbf{y}_{1:k-1}) N[\mathbf{y}_k : H_k(\mathbf{x}_k), \mathbf{R}_k]. \quad (4)$$

The analysis pdf is therefore obtained by multiplying the predictive pdf by the observation likelihood function $N[\mathbf{y}_k : H_k(\mathbf{x}_k), \mathbf{R}_k]$, and then being normalized by $b_k = \int_{\mathbb{R}^n} p_k^f(\mathbf{x}_k | \mathbf{y}_{1:k-1}) N[\mathbf{y}_k : H_k(\mathbf{x}_k), \mathbf{R}_k] d\mathbf{x}_k$.

While the expressions of the state pdfs can be obtained conceptually, determining the exact values of them at each point of the state space is practically infeasible in high-dimensional systems (Doucet et al. 2001). For instance, the determination of the predictive pdf requires the evaluation of the model $M_k(\mathbf{x})$ for a prohibitively large number of \mathbf{x} , given that one single evaluation might already be computationally very expensive in realistic atmospheric and oceanic applications.

3. The particle ensemble Kalman filter

a. Particle Kalman filtering and its ensemble implementation

Given N independent samples $\mathbf{x}^1, \dots, \mathbf{x}^N$ from a (multivariate) density p , an estimator \hat{p} of p can be obtained by

the kernel density estimation method (Silverman 1986), in the form of a mixture of N Gaussian pdfs:

$$\hat{p}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N N(\mathbf{x} : \mathbf{x}^i, \mathbf{P}), \quad (5)$$

where \mathbf{P} is a positive definite matrix. Inspired from this estimator, the PKF approximates the conditional state pdfs in the optimal nonlinear filter by mixtures of N Gaussian densities of the following form:

$$p_k^s(\mathbf{x}_k | \mathbf{y}_{1:k}) = \sum_{i=1}^N w_k^i N(\mathbf{x}_k : \mathbf{x}_k^{s,i}, \mathbf{P}_k^{s,i}). \quad (6)$$

The subscript s replaces a at the analysis time and f at the prediction time. The parameters of the mixture are the weights w_k^i , the centers of the distributions $\mathbf{x}_k^{s,i}$, and the covariance matrices $\mathbf{P}_k^{s,i}$. In particular, if $N = 1$, $p_k^s(\mathbf{x}_k | \mathbf{y}_{1:k})$ reduces to a single Gaussian pdf, so that the PKF reduces to the KF or its variants trivially (a nontrivial simplification will also be discussed below). Consequently, the KF and its variants can be considered special cases of the PKF.

Two special cases of Eq. (6) may be of particular interest. In the first case, $\mathbf{P}_k^{s,i} \rightarrow \mathbf{0}$, such that the Gaussian pdfs $N(\mathbf{x}_k : \mathbf{x}_k^{s,i}, \mathbf{P}_k^{s,i})$ tend to a set of Dirac functions $\delta(\mathbf{x}_k^{s,i})$, with the mass points at $\mathbf{x}_k^{s,i}$. In this case, the Gaussian mixture Eq. (6) reduces to the Monte Carlo approximation used in the particle filter (Doucet et al. 2001). In

the second case, all Gaussian pdfs $N(\mathbf{x}_k : \mathbf{x}_k^{s,i}, \mathbf{P}_k^{s,i})$ have (almost) identical centers and covariances, such that the Gaussian mixture Eq. (6) tends to a (single) Gaussian approximation, an assumption often used in various nonlinear Kalman filters (including the EnKF). In this sense, the PKF can be considered as a filter in between the Kalman filter and the particle filter (Hoteit et al. 2008; Luo et al. 2010).

The main procedures of the PKF are summarized as follows. Without loss of generality, suppose that at time instant $k - 1$, the analysis pdf, after a resampling step, is given by $\tilde{p}_{k-1}(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) = \sum_{i=1}^N \tilde{w}_{k-1}^i N(\mathbf{x}_{k-1} : \boldsymbol{\theta}_{k-1}^i, \boldsymbol{\Phi}_{k-1}^i)$. Then by applying Eq. (3) at the prediction step, one obtains the background pdf, in terms of a new MON:

$$p_k^f(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \approx \sum_{i=1}^N \tilde{w}_{k-1}^i N(\mathbf{x}_k : \hat{\mathbf{x}}_k^{f,i}, \hat{\mathbf{P}}_k^{f,i}), \quad (7)$$

where $\hat{\mathbf{x}}_k^{f,i}$ and $\hat{\mathbf{P}}_k^{f,i}$ are the propagations of the mean $\boldsymbol{\theta}_{k-1}^i$ and the covariance $\boldsymbol{\Phi}_{k-1}^i$ of the Gaussian component $N(\mathbf{x}_{k-1} : \boldsymbol{\theta}_{k-1}^i, \boldsymbol{\Phi}_{k-1}^i)$ through the system model Eq. (1), respectively.

Given an incoming observation \mathbf{y}_k , one applies Eq. (4) to update $p_k^f(\mathbf{x} | \mathbf{y}_{1:k})$ to the analysis pdf, also in the form of an MON:

$$p_k^a(\mathbf{x}_k | \mathbf{y}_{1:k}) = \sum_{i=1}^N w_k^i N(\mathbf{x}_k : \hat{\mathbf{x}}_k^{a,i}, \hat{\mathbf{P}}_k^{a,i}), \quad (8)$$

where $\hat{\mathbf{x}}_k^{a,i}$ and $\hat{\mathbf{P}}_k^{a,i}$ are updated from $\hat{\mathbf{x}}_k^{f,i}$ and $\hat{\mathbf{P}}_k^{f,i}$ through the Kalman filter or its variants, and the new weights:

$$w_k^i = \frac{\tilde{w}_{k-1}^i N[\mathbf{y}_k : H_k(\hat{\mathbf{x}}_k^{f,i}), \boldsymbol{\Sigma}_k^i]}{\sum_{j=1}^N \tilde{w}_{k-1}^j N[\mathbf{y}_k : H_k(\hat{\mathbf{x}}_k^{f,j}), \boldsymbol{\Sigma}_k^j]}, \quad (9)$$

where $\boldsymbol{\Sigma}_k^i$ is the innovation matrix. If evaluated through the extended Kalman filter, $\boldsymbol{\Sigma}_k^i = \mathbf{H}_k^i \hat{\mathbf{P}}_k^{f,i} (\mathbf{H}_k^i)^T + \mathbf{R}_k$, with \mathbf{H}_k^i being the gradient of H_k evaluated at $\hat{\mathbf{x}}_k^{f,i}$. Alternatively, if evaluated in the context of the EnKF, $\boldsymbol{\Sigma}_k^i$ can be expressed as the covariance of the projected background ensemble onto the observation space plus the observation covariance \mathbf{R}_k (Evensen 1994; Whitaker and Hamill 2002). Finally, a resampling step can be introduced to improve the performance of the PKF (Hoteit et al. 2008; Luo et al. 2010), so that the analysis pdf becomes $\tilde{p}_k(\mathbf{x}_k | \mathbf{y}_{1:k}) = \sum_{i=1}^N \tilde{w}_k^i N(\mathbf{x}_k : \boldsymbol{\theta}_k^i, \boldsymbol{\Phi}_k^i)$. Such a resampling algorithm is presented in the next section.

The PKF correction step can be interpreted as composed of two types of corrections: a *Kalman-type correction* used to update $\hat{\mathbf{x}}_k^{f,i}$ and $\hat{\mathbf{P}}_k^{f,i}$ to $\hat{\mathbf{x}}_k^{a,i}$ and $\hat{\mathbf{P}}_k^{a,i}$, and a *particle-type correction* used to update the weights \tilde{w}_{k-1}^i

to w_k^i . In the PKF, the Kalman correction reduces the risk of the weights' collapse by allocating the estimates $\hat{\mathbf{x}}_k^{f,i}$ (whose projections onto the observation space far away from the observation \mathbf{y}_k) relatively more weights than in the particle filter (Hoteit et al. 2008; Van Leeuwen 2009). Indeed, Eq. (9) has the same form as in the PF (Doucet et al. 2001), but uses the innovation matrices $\boldsymbol{\Sigma}_k^i$ to normalize the model-data misfit, rather than \mathbf{R}_k . As $\boldsymbol{\Sigma}_k^i$ are always greater than \mathbf{R}_k , the estimates that are close to the observation will receive relatively less weights than in the PF, while those far from the observation will receive relatively more weights. This means that the support of the local predictive pdf and the observation likelihood function will be more coherent than in the PF. Resampling will therefore be needed less often, so that Monte Carlo fluctuations are reduced.

The main issue with the PKF is the prohibitive computational burden associated with running an ensemble of KFs, knowing that running a KF or an extended KF in high-dimensional systems is already a challenge. To reduce computational cost, we use an ensemble of EnKFs, rather than the KF or the extended KF, to construct the PKF. We refer to this approach as the particle ensemble Kalman filter (PEnKF). In the PEnKF, the (analysis) ensembles representing the Gaussian components are propagated forward in time to obtain a set of background ensembles at the next assimilation cycle. Then for each background ensemble, a stochastic or deterministic EnKF is used to update the background ensemble to its analysis counterpart. This amounts to simultaneously running a weighted ensemble of EnKFs, and the final state estimate is the weighted average of all the EnKFs solutions.

b. A resampling algorithm

We adopt a resampling algorithm that combines those in Hoteit et al. (2008), Luo et al. (2010), and Pham (2001). The main idea is as follows: given a MON, we first employ an information-theoretic criterion used in Hoteit et al. (2008) and Pham (2001) to check if it needs to conduct resampling. If there is such a need, we then reapproximate the MON by a new MON, based on the criterion that the mean and covariance of the new MON match those of the original MON as far as possible (Luo et al. 2010).

More concretely, let $p(\mathbf{x})$ be the pdf of the n -dimensional random vector \mathbf{x} , expressed in terms of an MON with N Gaussian pdfs so that

$$p(\mathbf{x}) = \sum_{i=1}^N w_i N(\mathbf{x} : \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (10)$$

where w_i are the set of normalized weights of the Gaussian pdfs $N(\mathbf{x} : \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$, satisfying $w_i \geq 0$ for $i = 1, \dots, N$ and $\sum_{i=1}^N w_i = 1$. To decide

whether to conduct resampling or not, the entropy E_w of the weights w_i is computed, which reads as follows (Hoteit et al. 2008; Pham 2001):

$$E_w = -\sum_{i=1}^N w_i \log w_i. \quad (11)$$

Ideally, when the distribution of the weights w_i is uniform, which yields the maximum weight entropy $E_w^u = \log N$, there is no need to conduct resampling. Thus, as a criterion, if E_w is within a certain distance d to E_w^u ,

$$E_w^u - E_w = \log N + \sum_{i=1}^N w_i \log w_i \leq d, \quad (12)$$

where d is a user-defined threshold, then we choose not to conduct resampling. In this work we set the threshold $d = 0.25$ following Hoteit et al. (2008).

In case that there is a need to conduct resampling, we follow the procedure similar to that in Luo et al. (2010). Here the idea is to treat resampling as a pdf approximation problem, in which we seek a new MON:

$$\tilde{p}(\mathbf{x}) = \frac{1}{q} \sum_{i=1}^q N(\mathbf{x} : \boldsymbol{\theta}_i, \boldsymbol{\Phi}_i), \quad (13)$$

with q equally weighted Gaussian pdfs, to approximate the original $p(\mathbf{x})$ in Eq. (10). In approximation, we require that the mean and covariance of $\tilde{p}(\mathbf{x})$ be as close as possible to those of $p(\mathbf{x})$. To this end, we need to choose proper values of $\boldsymbol{\theta}_i$ and $\boldsymbol{\Phi}_i$ in order to achieve this objective.

The means and covariances of $p(\mathbf{x})$ and $\tilde{p}(\mathbf{x})$, denoted by $\bar{\mathbf{x}}$ and $\bar{\mathbf{P}}$, and $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{P}}$, respectively, are given by

$$\bar{\mathbf{x}} = \sum_{i=1}^N w_i \boldsymbol{\mu}_i, \quad \text{and} \quad \bar{\mathbf{P}} = \sum_{i=1}^N w_i [\boldsymbol{\Sigma}_i + (\boldsymbol{\mu}_i - \bar{\mathbf{x}})(\boldsymbol{\mu}_i - \bar{\mathbf{x}})^T], \quad (14a)$$

$$\tilde{\mathbf{x}} = \frac{1}{q} \sum_{i=1}^q \boldsymbol{\theta}_i, \quad \text{and} \quad \tilde{\mathbf{P}} = \frac{1}{q} \sum_{i=1}^q [\boldsymbol{\Phi}_i + (\boldsymbol{\theta}_i - \tilde{\mathbf{x}})(\boldsymbol{\theta}_i - \tilde{\mathbf{x}})^T]. \quad (14b)$$

Thus, our objective is equivalent to balancing the above equation such that

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}}, \quad \text{and} \quad \tilde{\mathbf{P}} \approx \bar{\mathbf{P}}. \quad (15)$$

In the trivial case with $q = N = 1$, Eq. (15) can be satisfied by letting $\boldsymbol{\theta}_1 = \boldsymbol{\mu}_1$ and $\boldsymbol{\Phi}_1 = \boldsymbol{\Sigma}_1$, and the PEnKF reduces to an EnKF. In nontrivial cases, for simplicity in solving Eq. (15) and reducing computational cost (as to be shown later), one may choose the covariances $\boldsymbol{\Phi}_i$ to be identical, say $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}$, for $i = 1, \dots, q$, so that

$$\frac{1}{q} \sum_{i=1}^q \boldsymbol{\theta}_i = \bar{\mathbf{x}}, \quad \text{and} \quad \boldsymbol{\Phi} + \frac{1}{q} \sum_{i=1}^q (\boldsymbol{\theta}_i - \bar{\mathbf{x}})(\boldsymbol{\theta}_i - \bar{\mathbf{x}})^T \approx \bar{\mathbf{P}}. \quad (16)$$

When an EnKF is used to construct the PKF, one needs to represent the solution of Eq. (16) in terms of some ensembles $\{\mathbf{X}_{\text{en}}^i, i = 1, \dots, q\}$, where \mathbf{X}_{en}^i is a matrix containing the (analysis) ensemble of the i th Gaussian component in Eq. (13), with mean $\boldsymbol{\theta}_i$ and covariance $\boldsymbol{\Phi}$. For simplicity, we assume that \mathbf{X}_{en}^i are all of dimension $n \times m$, with the ensemble size m for each i . Similar results can be easily obtained in the case with nonuniform ensemble sizes.

We then define a constant c , hereafter called the *fraction coefficient*, which satisfies $0 \leq c \leq 1$. We let $\boldsymbol{\Phi} \approx c^2 \bar{\mathbf{P}}$, so that Eq. (16) is reduced to

$$\frac{1}{q} \sum_{i=1}^q \boldsymbol{\theta}_i = \bar{\mathbf{x}}, \quad \text{and} \quad \frac{1}{q} \sum_{i=1}^q (\boldsymbol{\theta}_i - \bar{\mathbf{x}})(\boldsymbol{\theta}_i - \bar{\mathbf{x}})^T \approx (1 - c^2) \bar{\mathbf{P}}. \quad (17)$$

In other words, the centers $\{\boldsymbol{\theta}_i, i = 1, \dots, q\}$ can be generated as a set of state vectors whose sample mean and covariance are $\bar{\mathbf{x}}$ and $(1 - c^2) \bar{\mathbf{P}}$, respectively. After obtaining $\boldsymbol{\theta}_i$, one can generate the corresponding ensembles \mathbf{X}_{en}^i , with the sample means and covariances being $\boldsymbol{\theta}_i$ and $\boldsymbol{\Phi} \approx c^2 \bar{\mathbf{P}}$, respectively. (How $\boldsymbol{\theta}_i$ and \mathbf{X}_{en}^i can be generated is discussed in more detail in the supplemental material.)

From the above discussion, we see that c is a coefficient that sets how to divide $\bar{\mathbf{P}}$ among $\boldsymbol{\Phi}$ and $1/q \sum_{i=1}^q (\boldsymbol{\theta}_i - \bar{\mathbf{x}})(\boldsymbol{\theta}_i - \bar{\mathbf{x}})^T$, so that the constraints in Eq. (16) are satisfied. When $c \rightarrow 0$, we have $\boldsymbol{\Phi} \rightarrow \mathbf{0}$ so that $\tilde{p}(\mathbf{x})$ in Eq. (13) approaches the Monte Carlo approximation in the particle filter, with the mass points equal to $\boldsymbol{\theta}_i$. On the other hand, when $c \rightarrow 1$, we have $1/q \sum_{i=1}^q (\boldsymbol{\theta}_i - \bar{\mathbf{x}})(\boldsymbol{\theta}_i - \bar{\mathbf{x}})^T \rightarrow \mathbf{0}$, so that all $\boldsymbol{\theta}_i$ approach $\bar{\mathbf{x}}$ and $\boldsymbol{\Phi}$ approaches $\bar{\mathbf{P}}$. As a result, $\tilde{p}(\mathbf{x})$ in Eq. (13) approaches the Gaussian pdf $N(\mathbf{x} : \bar{\mathbf{x}}, \bar{\mathbf{P}})$, which is essentially the assumption used in the EnKF. In this sense, when equipped with the resampling algorithm, the PEnKF is a filter in between the particle filter and the EnKF, with an adjustable parameter c that influences its behavior.

We note that, when $c \rightarrow 0$ and under the constraint of matching the first two moments, our resampling scheme is quite similar to the posterior Gaussian resampling strategy used in the Gaussian particle filter (Kotecha and Djurić 2003; Xiong et al. 2006), in which one generates particles from a Gaussian distribution with mean and covariance equal to those of the posterior pdf of the system states. As a result, there is no guarantee that higher-order moments of the new MON match those of the original MON in our resampling scheme. If matching higher-order moments is a concern, one may adopt alternative criteria,

for instance, the one that aims to minimize the distance (in certain metric) between the new MON and the original one. The resampling procedure is then recast as an optimization problem, in which one looks for the parameters (i.e., means and covariances) of the new MON that satisfy the chosen criterion as close as possible. In principle, this type of parameter estimation problem may be solved by the expectation-maximization (EM) algorithm (Redner and Walker 1984; Smith 2007), but in practice, it is often computationally very intensive to implement, due to the slow convergence rate of the EM algorithm and the high dimensionality of the parameter space in constructing the new MON. We, therefore, do not consider this type of sophisticated resampling strategy in the present study.

For the purpose of pdf reapproximation, it is clear that the MON is not the only choice. Several other alternatives are available in the context of kernel density estimation (KDE; Silverman 1986), and in principle all of these can be used for pdf reapproximation. For instance, KDE is adopted at the resampling step in the regularized particle filter (RPF; Musso et al. 2001; Stavropoulos and Titterton 2001) to construct a continuous pdf from the particles before resampling. The new particles are then sampled from the continuous pdf. In this respect, the PEnKF is somehow similar to the RPF, especially if the Gaussian kernel is adopted in the RPF for density estimation. However, there also exist several differences, which we summarize as follows:

- The RPF first constructs a continuous pdf, and then draws a set of new particles with equal weights from the resulting pdf. In contrast, the PEnKF aims to directly approximate a MON by a new MON with equal weights.
- In the RPF, various kernels can be adopted for the purpose of constructing the continuous pdf. In the PEnKF, we are confined to use the MON, since we seek to build the PEnKF consisting of a set of parallel EnKFs.
- In the present work, the pdf reapproximation criterion used in the PEnKF only captures the first two moments of the underlying pdf. In contrast, the KDE used in the RPF could in principle yield a more comprehensive pdf estimate, provided that there are sufficient particles. In certain situations, however, the number of required particles may also suffer from the “curse of dimensionality” (Silverman 1986, see chapter 4).

c. Outline of the PEnKF algorithm

To facilitate the comprehension of the PEnKF, here we provide an outline of the main steps of its algorithm. To avoid distraction, we will discuss the initialization of the PEnKF in the next section. Throughout this paper, we assume that the number q of Gaussian components at the resampling step and the number N of Gaussian

components at the prediction and correction steps are time invariant. This implies the choice $q = N$.

Without loss of generality, we also assume that at time instant $k - 1$, the posterior pdf $p_{k-1}^a(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ is reapproximated, through the resampling step, by a mixture model:

$$\tilde{p}_{k-1}(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) = \sum_{i=1}^q \tilde{w}_{k-1}^i N(\mathbf{x}_{k-1} : \boldsymbol{\theta}_{k-1,i}, \boldsymbol{\Phi}_{k-1}).$$

Moreover, the reapproximated analysis ensembles $\{\mathbf{X}_{\text{approx}}^{k-1,i}, i = 1, \dots, q\}$ representing the Gaussian components $N(\mathbf{x}_{k-1} : \boldsymbol{\theta}_{k-1,i}, \boldsymbol{\Phi}_{k-1})$ are also generated. The procedures at the next assimilation cycle are outlined as follows.

- *Prediction step:* For $i = 1, \dots, q$, propagate the ensembles $\mathbf{X}_{\text{approx}}^{k-1,i}$ forward through Eq. (1) to obtain the corresponding background ensembles $\mathbf{X}_{\text{bg}}^{k,i}$ at instant k . Accordingly, the background pdf becomes

$$p_k^b(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \sum_{i=1}^q \tilde{w}_{k-1}^i N(\mathbf{x}_k : \hat{\mathbf{x}}_{k,i}^b, \hat{\mathbf{P}}_{k,i}^b),$$

with $\hat{\mathbf{x}}_{k,i}^b$ and $\hat{\mathbf{P}}_{k,i}^b$ being the sample mean and covariance of the ensemble $\mathbf{X}_{\text{bg}}^{k,i}$, respectively.

- *Correction step:* With an incoming observation \mathbf{y}_k , for each background ensemble $\mathbf{X}_{\text{bg}}^{k,i}, i = 1, \dots, q$, apply an EnKF to obtain the analysis mean $\hat{\mathbf{x}}_{k,i}^a$ and the analysis ensemble $\mathbf{X}_{\text{ana}}^{k,i}$. During the correction, covariance inflation and localization [cf. section 4b(2)] can be conducted on the EnKF. In addition, update the associated weights \tilde{w}_{k-1}^i to w_k^i according to Eq. (9). After the corrections, the analysis pdf becomes

$$p_k^a(\mathbf{x}_k | \mathbf{y}_{1:k}) = \sum_{i=1}^q w_k^i N(\mathbf{x}_k : \hat{\mathbf{x}}_{k,i}^a, \hat{\mathbf{P}}_{k,i}^a),$$

where w_k^i are computed according to Eq. (9) in the context of the EnKF, and $\hat{\mathbf{P}}_{k,i}^a$ are the sample covariances of $\mathbf{X}_{\text{ana}}^{k,i}$.

- *Resampling step:* Use the criterion in Eq. (12) to determine whether to conduct resampling or not.
 - 1) If there is no need for resampling, then assign $\tilde{p}_k(\mathbf{x}_k | \mathbf{y}_{1:k}) = p_k^a(\mathbf{x}_k | \mathbf{y}_{1:k})$, and $\mathbf{X}_{\text{approx}}^{k,i} = \mathbf{X}_{\text{ana}}^{k,i}$ for $i = 1, \dots, q$;
 - 2) Otherwise, $\tilde{p}_k(\mathbf{x}_k | \mathbf{y}_{1:k}) = (1/q) \sum_{i=1}^q N(\mathbf{x}_k : \boldsymbol{\theta}_{k,i}, \boldsymbol{\Phi}_k)$, where parameters $\boldsymbol{\theta}_{k,i}$ and $\boldsymbol{\Phi}_k$ are computed following the method in section 3b, and the associated weights become $1/q$. The ensembles $\mathbf{X}_{\text{approx}}^{k,i}$ are produced accordingly.

4. Numerical experiments

a. Experiments' design

In the present work, we focus on two different implementations of the PEnKF: the first is based on the

stochastic EnKF (SEnKF) of Evensen (1994) and the second is based on the ensemble transform Kalman filter (ETKF) of Bishop et al. (2001). These two implementations are referred to as the PSEnKF and the PETKF, respectively.

The strongly nonlinear 40-dimensional system model created by Lorenz and Emanuel (1998) (hereafter the LE98 model) was chosen as the test bed to evaluate and study the performance of these two filters. This model mimics the time evolution of a scalar atmospheric quantity. It is governed by the following set of equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + 8, \quad i = 1, \dots, 40, \quad (18)$$

where the nonlinear quadratic terms simulate advection and the linear term represents dissipation. Boundary conditions are cyclic (i.e., we define $x_{-1} = x_{39}$, $x_0 = x_{40}$, and $x_{41} = x_1$). The model was numerically integrated using the Runge–Kutta fourth-order scheme from time $t = 0$ to $t = 35$ with a constant time step $\Delta t = 0.05$ (which corresponds to 6 h in real time). To eliminate the impact of transition states, the model trajectory between times $t = 0$ and $t = 25$ was discarded. The assimilation experiments were carried out during the period $t = 25.05$ to $t = 35$ where the model trajectory was considered to be the “truth.” Reference states were then sampled from the true trajectory and a filter performance is evaluated by how well it is able to estimate the reference states using a perturbed model and assimilating a set of (perturbed) observations that was extracted from the reference states.

In this work we consider two scenarios: one with a linear observation operator and the other with a nonlinear operator. The concrete forms of these two observational operators will be given in the relevant sections below.

The time-averaged root-mean-square error (rmse) is used to evaluate the performance of a filter. Given a set of n -dimensional state vectors $\{\mathbf{x}_k; \mathbf{x}_k = (x_{k,1}, \dots, x_{k,n})^T, k = 0, \dots, k_{\max}\}$, with k_{\max} being the maximum time index ($k_{\max} = 199$ in our experiments), then the rmse \hat{e} is defined as

$$\hat{e} = \frac{1}{k_{\max} + 1} \sum_{k=0}^{k_{\max}} \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{x}}_{k,i}^a - \mathbf{x}_{k,i})^2}, \quad (19)$$

where $\hat{\mathbf{x}}_k^a = (\hat{\mathbf{x}}_{k,1}^a, \dots, \hat{\mathbf{x}}_{k,n}^a)^T$ is the analysis state of \mathbf{x}_k .

A possible problem in directly using \hat{e} as the performance measure is that \hat{e} itself may depend on some intrinsic parameters of the filters (e.g., the covariance inflation factor and localization length scale, which will be discussed later). This may lead to inconsistent conclusions at different parameter values. To avoid this problem,

we adopted the following strategy: we relate a filter’s best possible performance to the minimum rmse \hat{e}_{\min} , which is the minimum value of \hat{e} that the filter can achieve within the chosen ranges of the filter’s intrinsic parameters. In performance comparison, if the minimum rmse \hat{e}_{\min}^A of filter A is less than the minimum rmse \hat{e}_{\min}^B of filter B , filter A is said to perform better than filter B .

b. Implementation details

1) FILTER INITIALIZATION

To initialize the PEnKF, we first estimate the mean and covariance of the LE98 model over some time interval following Hoteit et al. (2008). These statistics are then used to produce the pdf $p_0^f(\mathbf{x}_0)$ of the background at the first assimilation cycle as a MON.

Concretely, the LE98 model was first integrated for a long period (between $t = 0$ and $t = 1000$) starting from an initial state that has been drawn at random. The trajectory that falls between $t = 50.05$ and $t = 1000$ was used to estimate the mean $\hat{\mathbf{x}}_{\text{ds}}$ and covariance $\hat{\mathbf{P}}_{\text{ds}}$ of the dynamical system. To initialize $p_0^f(\mathbf{x}_0)$ as a mixture of N Gaussian distributions,

$$p_0^f(\mathbf{x}_0) = \frac{1}{N} \sum_{i=1}^N N(\mathbf{x}_0; \mathbf{x}_0^{f,i}, \mathbf{P}_{\text{com}}), \quad (20)$$

where $\mathbf{x}_0^{f,i}$ are the means, and \mathbf{P}_{com} is the common covariance matrix of the Gaussian distributions in the mixture, we draw N samples $\mathbf{x}_0^{f,i}$ from the Gaussian distribution $N(\mathbf{x}_b; \hat{\mathbf{x}}_{\text{ds}}, \hat{\mathbf{P}}_{\text{ds}})$, and set $\mathbf{P}_{\text{com}} = \hat{\mathbf{P}}_{\text{ds}}$. If $\hat{\mathbf{x}}_0^f = 1/N \sum_{i=1}^N \mathbf{x}_0^{f,i}$ denotes the sample mean of $\mathbf{x}_0^{f,i}$, then the covariance \mathbf{P}_0^f of $p_0^f(\mathbf{x}_0)$ is given by

$$\mathbf{P}_0^f = \hat{\mathbf{P}}_{\text{ds}} + \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_0^{f,i} - \hat{\mathbf{x}}_0^f)(\mathbf{x}_0^{f,i} - \hat{\mathbf{x}}_0^f)^T, \quad (21)$$

which is always larger than $\hat{\mathbf{P}}_{\text{ds}}$. The rationale behind this choice is not far from the covariance inflation technique (Anderson and Anderson 1999; Whitaker and Hamill 2002). In practice, a data assimilation system is often subject to various errors, such as poorly known model and observational errors, sampling errors, etc. In such circumstances, an inflated background covariance would allocate more weights to the incoming observation when updating the background to the analysis, making the filter more robust (Jazwinski 1970; Simon 2006).

2) COVARIANCE INFLATION AND LOCALIZATION

Covariance inflation (Anderson and Anderson 1999; Whitaker and Hamill 2002) and localization (Hamill et al. 2001) are two popular techniques that are used to improve the stability and performance of the EnKF

(Hamill et al. 2009; Van Leeuwen 2009), especially in the small ensemble scenario. In our experiments, these two techniques are implemented for each EnKF in the PEnKF.

More concretely, to introduce covariance inflation to the i th EnKF at instant k , we multiply the analysis covariance $\hat{\mathbf{P}}_{k,i}^a$ (before the resampling step) by a factor $(1 + \delta)^2$, where the scalar $\delta \geq 0$, called *covariance inflation factor*, is introduced as an intrinsic parameter of the EnKF. On the other hand, we follow the method in Hamill et al. (2001) to conduct covariance localization on the background covariance and its projection onto the observation space, with the tapering function (for smoothing out spuriously large values in covariance matrices) being the fifth-order function defined in Eq. (4.10) of Gaspari and Cohn (1999). In doing so, another intrinsic scalar parameter $l_c > 0$, called the *length scale* (Hamill et al. 2001), is introduced to the EnKF. Roughly speaking, l_c is a parameter that determines the critical distance beyond which the tapering function becomes zero.

c. Experiments results with a linear observation operator

In the first scenario, we let the (synthetic) observations be generated every day (four model time steps) from the reference states using the following linear observation system:

$$\mathbf{y}_k = (x_{k,1}, x_{k,3}, \dots, x_{k,39})^T + \mathbf{v}_k, \quad (22)$$

where only the odd state variables $x_{k,i}$ ($i = 1, 3, \dots, 39$) of the system state $\mathbf{x}_k \equiv (x_{k,1}, \dots, x_{k,40})^T$ at time index k are observed. The observation noise \mathbf{v}_k follows the 20-dimensional Gaussian distribution $N(\mathbf{v}_k; \mathbf{0}, \mathbf{I}_{20})$ with \mathbf{I}_{20} being the 20×20 identity matrix.

1) EFFECT OF THE NUMBER OF GAUSSIAN DISTRIBUTIONS

In the first experiment we examine the effect of the number of Gaussian distributions on the performance of the PSEnKF and the PETKF. The experiment settings are as follows.

We initialize the pdf $p_0^f(\mathbf{x}_0)$ with N Gaussian pdfs. In our experiments we let N take values between 1 and 60. Since it is costly to carry out the computation for each integer in this interval, we choose to let N increase from 1 to 10, with an even increment of 1 each time, and then increase it from 15 to 60, with a larger increment of 5 each time, as N becomes larger. For convenience, we denote this choice by $N \in \{1: 1: 10, 15: 5: 60\}$, where the notation $v_{\min}: v_{\text{inc}}: v_{\max}$ represents a set of values

increasing from v_{\min} to v_{\max} , with an even increment of v_{inc} each time. If there is a need to conduct resampling, we reapproximate the analysis MON by a new MON with equal weights and with the same number of normal distributions. In doing so, we introduce a new parameter (i.e., the fraction coefficient c defined in section 3b) to the PSEnKF/PETKF. To examine its effect on the performance of the filter, we let $c \in \{0.05: 0.1: 0.95\}$. The ensemble size is set to $m = 20$ in each SEEnKF/ETKF, which is relatively small compared to the system dimension 40. In this case, it is customary to conduct covariance inflation (Anderson and Anderson 1999; Whitaker and Hamill 2002) and localization (Hamill et al. 2001) to improve the robustness and performance of the filters (Hamill et al. 2009; Van Leeuwen 2009). The impacts of covariance inflation and localization on the performance of the EnKF have been examined in many works (e.g., see Whitaker and Hamill 2002). In our experiments we let the covariance inflation factor $\delta = 0.02$. We follow the settings in Luo et al. (2010, their section 7.2.3) to conduct covariance localization and choose the length scale $l_c = 50$. To reduce statistical fluctuations, we repeat the experiments 20 times, each time with a randomly drawn initial background ensemble, but with the same true trajectory and the corresponding observations. The same repetition setting is adopted in all the other experiments.

Figure 1 plots the rms errors of both the PSEnKF and PETKF as functions of the fraction coefficient c and the number N of Gaussian pdfs. First, we examine how the rmse of the PSEnKF changes with c when N is fixed. In Fig. 1a, if N is relatively small (say $N < 40$), the rmse tends to decrease as c increases. For larger N (say $N = 55$), the rmse of the filter exhibits the bell-shaped behavior: at the beginning it increases when c grows from 0; after c becomes relatively large (i.e., $c = 0.4$), further increasing c reduces the rmse instead. Next, we examine the behavior of the rmse of the PSEnKF with respect to N when c is fixed. When c is relatively small (i.e., $c = 0.1$), the rmse exhibits the U-turn behavior: at the beginning it intends to decrease as N grows; after N becomes relatively large (i.e., $N = 45$), further increasing N increases the rmse instead. When c is larger, say, $c = 0.6$, the rmse appears less sensitive to the change of N . However, for even larger values of c , say, $c = 0.9$, the rmse appears to monotonically decrease with N .

The behavior of the PETKF (cf. Fig. 1b) with respect to the changes of N and c is similar to that of the PSEnKF. Therefore we do not repeat its description here.

To examine the minimum rms errors \hat{e}_{\min} of the PSEnKF and the PETKF within the tested values of c and N , we plot \hat{e}_{\min} of both filters as functions of N in Fig. 2.

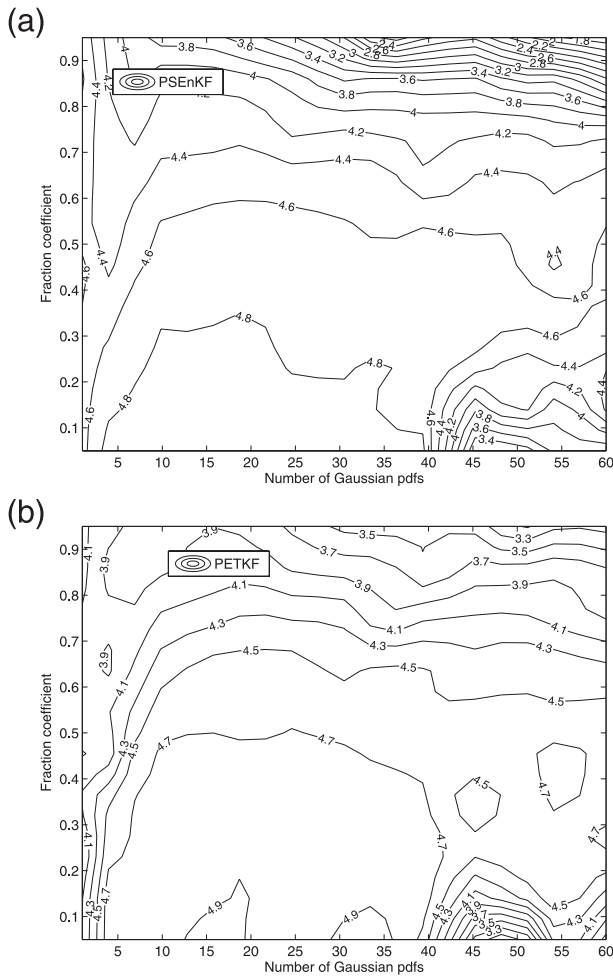


FIG. 1. RMS errors (over 20 experiments) of the (a) stochastic EnKF- and (b) ETKF-based PENKFs (with a fixed ensemble size of 20 in each ensemble filter) as functions of the fraction coefficient and the number of Gaussian pdfs in the MON.

The \hat{e}_{\min} of both filters tends to decrease as the number N of Gaussian distributions increases, though there also exhibit certain local minima. The PSEnKF achieves its lowest \hat{e}_{\min} at $N = 60$, while the PETKF at $N = 50$. As N grows, both the PSEnKF and the PETKF tend to have lower \hat{e}_{\min} than their corresponding base filters, the SEnKF and the ETKF (corresponding to the PSEnKF and the PETKF with $N = 1$, as discussed in section 3b), respectively. This confirms the benefit of accuracy improvement by using the PEnKF instead of an EnKF. A comparison between the PSEnKF and the PETKF shows that the PETKF performs better than the PSEnKF when the number N of Gaussian distributions is relatively small (i.e., $N \leq 7$). However, as N becomes larger, the PSEnKF outperforms its ETKF-based counterpart. Similar phenomena can also be observed in other experiments, as will be seen later.

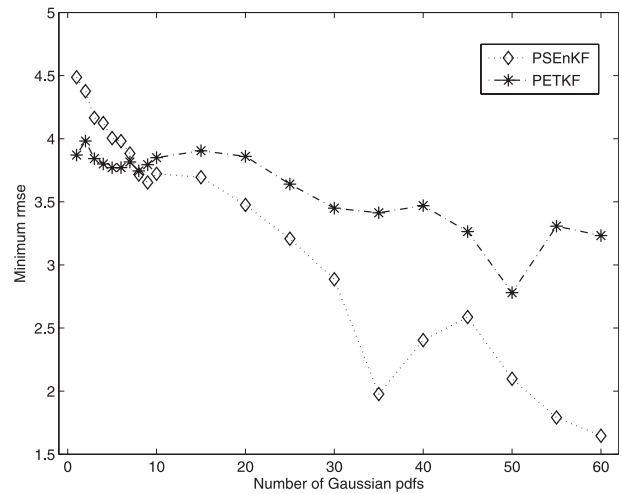


FIG. 2. Minimum rms errors \hat{e}_{\min} (over 20 experiments) of the stochastic EnKF- and ETKF-based PENKFs (with a fixed ensemble size of 20 in each ensemble filter) as a function of the number of Gaussian pdfs in the MON.

2) EFFECT OF THE ENSEMBLE SIZE

In the second experiment we examine the effect of the ensemble size of each SEnKF/ETKF in the PEnKF, on the performance of the PSEnKF/PETKF. For reference, we also examine the performance of the SEnKF and the ETKF under various ensemble sizes. The experiment settings are as follows. For the PSEnKF and the PETKF, we let the ensemble size m of each EnKF take values from the set $\{20, 40, 80, 100, 200, 400, 800, 1000\}$. For a single SEnKF/ETKF, we let $m \in \{20, 40, 80, 100, 200, 400, 600, 800, 1000\}$, with two more values at 60 and 600.

In the PSEnKF and the PETKF, we also vary the fraction coefficient c such that $c \in \{0.05: 0.1: 0.95\}$. We fix the number N of Gaussian pdfs (i.e., the number of ensemble filters) to be 3. To conduct covariance inflation, we let the inflation factor $\delta = 0.02$. We choose to conduct covariance localization, and set the length scale $l_c = 50$, only if the ensemble size m is not larger than the dimension 40 of the LE98 model. No covariance localization was conducted if $m > 40$. Our experience shows that for $m > 40$, the benefit of conducting localization is not significant even if the length scale l_c is properly chosen, while an improper value of l_c is more likely to deteriorate the filter performance. To reduce statistical fluctuations, the experiments are again repeated 20 times.

In Fig. 3 we show the rms errors of the SEnKF and the ETKF as functions of the ensemble size m . The rmse of the ETKF exhibits a U-turn behavior. The rmse of the ETKF monotonically decreases as long as $m < 100$. Beyond that, the rmse monotonically increases instead as m increases. On the other hand, the SEnKF exhibits a different behavior. Its rmse decreases for $m \leq 200$, and

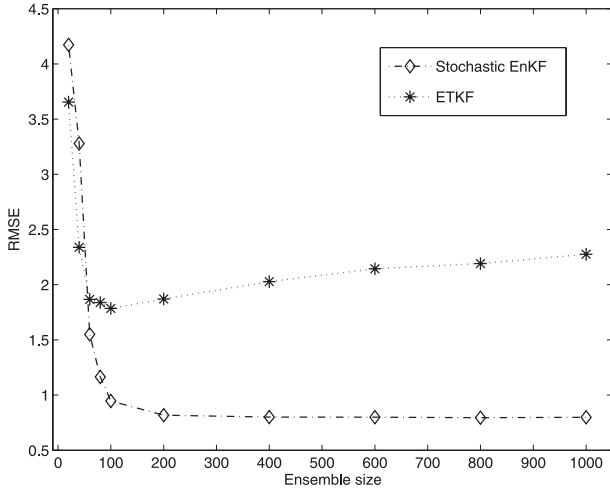


FIG. 3. RMS errors (over 20 experiments) of the stochastic EnKF and the ETKF as functions of the ensemble size.

then reaches a plateau where the rmse remains almost unchanged as m further increases.

Figure 4 plots the rms errors of the PSEnKF and the PETKF as functions of the fraction coefficient c , and the ensemble size m in the SEnKF and the ETKF used to construct the corresponding PEnKFs. The rmse errors, as functions of the ensemble size m (with fixed c), are consistent with our observations in Fig. 3. On the other hand, for both PEnKFs, their rms errors tend to decrease as the fraction coefficient c increases.

Per analogy to the first experiment, Fig. 5 plots the minimum rms errors \hat{e}_{\min} of the PSEnKF and the PETKF within the tested fraction coefficient c and the ensemble size m . A comparison between Figs. 5 and 3 shows that, the minimum rms errors \hat{e}_{\min} of the PEnKFs behave very similarly to the rms errors of their corresponding EnKFs in Fig. 3. Moreover, the values of \hat{e}_{\min} in Fig. 5 tends to be lower than the corresponding rms errors in Fig. 3, indicating the benefit of accuracy improvement in using the PEnKFs. Again, a comparison between the PSEnKF and the PETKF shows that the PETKF performs better than the PSEnKF when the ensemble size m is relatively small (i.e., $m \leq 40$). However, as m becomes larger, the PSEnKF outperforms the PETKF.

d. Experiments results with a nonlinear observation operator

In the second scenario, we introduce nonlinearity to the observation system. To this end, we let the observations be generated by the following nonlinear process:

$$\mathbf{y}_k = 0.05(x_{k,1}^2, \dots, x_{k,39}^2)^T + \mathbf{v}_k \quad (23)$$

for every four model time steps. In Eq. (23), again only the odd state variables $x_{k,i}$ ($i = 1, 3, \dots, 39$) of the system

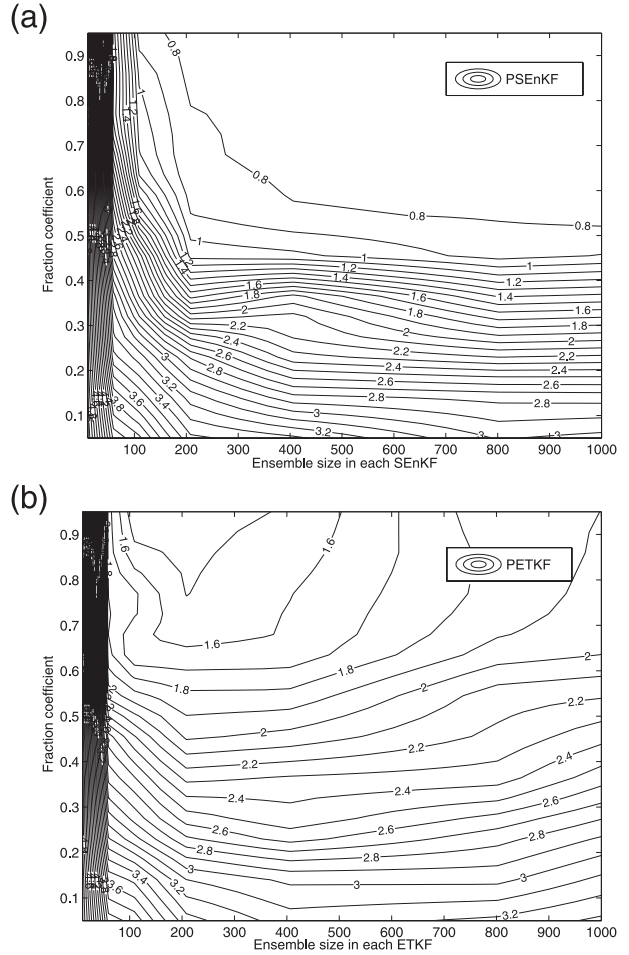


FIG. 4. RMS errors (over 20 experiments) of the (a) stochastic EnKF- and (b) ETKF-based PEnKFs (with a fixed number of Gaussian pdfs of 3 in each PKF) as functions of the fraction coefficient and the ensemble size of the ensemble filter.

state $\mathbf{x}_k \equiv (x_{k,1}, \dots, x_{k,40})^T$ at time index k are observed. The observation noise \mathbf{v}_k also follows the 20-dimensional Gaussian distribution $N(\mathbf{v}_k; \mathbf{0}, \mathbf{I}_{20})$. We conduct the same experiments as those in the case of linear observation operator.

1) EFFECT OF THE NUMBER OF GAUSSIAN DISTRIBUTIONS

We first examine the effect of the number of Gaussian distributions. The experiment settings are the same as those in section 4c(1). Concretely, For either the PSEnKF or the PETKF, the number of Gaussian distributions $N \in \{1: 10, 15: 5: 60\}$ and the fraction coefficient $c \in \{0.05: 0.1: 0.95\}$. For each individual SEnKF/ETKF in the PEnKF, the ensemble size $m = 20$, the covariance inflation factor $\delta = 0.02$, and the length scale $l_c = 50$ for covariance localization. As before, the experiments are repeated 20 times to reduce statistical fluctuations.

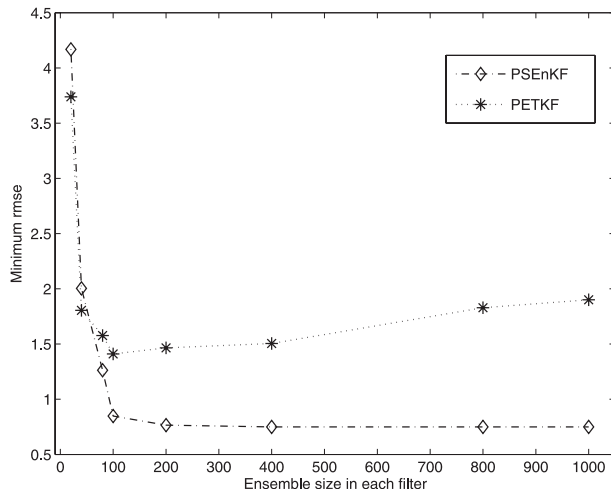


FIG. 5. Minimum rms errors \hat{e}_{\min} (over 20 experiments) of the stochastic EnKF- and ETKF-based PEnKFs (with a fixed number of Gaussian pdfs of 3 in each PKF) as a function of the ensemble size in each ensemble filter.

Figure 6 plots the rms errors of both the PSEnKF and the PETKF as functions of the fraction coefficient c and the number N of Gaussian pdfs. When c and N changes, both the PSEnKF and the PETKF behave very similar to their counterparts in the linear case. The rms errors of the filters tend to decrease as N increases, meaning that the PSEnKF/PETKF with $N > 1$ in general performs better than the stochastic EnKF/ETKF (corresponding to the case $N = 1$ in the PEnKF), consistent with the results obtained in the linear observer case.

We also examine the minimum rms errors \hat{e}_{\min} of the PSEnKF and the PETKF within the tested values of c and N . Figure 7 plots \hat{e}_{\min} as functions of N . For the PSEnKF, the lowest \hat{e}_{\min} is achieved at $N = 50$. And for the PETKF, its \hat{e}_{\min} tends to decrease within the tested range of N , and achieves its minimum at $N = 60$. The PEnKF with more than one Gaussian distributions ($N > 1$) performs better than the corresponding EnKF ($N = 1$). In addition, a comparison between the PSEnKF and the PETKF again shows that the PETKF performs better than the PSEnKF when the number N of Gaussian distributions is relatively small, but tends to become worse as N increases.

A comparison between Figs. 2 and 7 suggests that the rmse of a filter (e.g., the PSEnKF at $N = 2$) with a nonlinear observer may sometimes be lower than that of the same filter with a linear observer.¹ (This can be also

¹ The result of comparison should also depend on the filter in use, its configuration, the system in assimilation, and so on, and therefore may change from case to case.

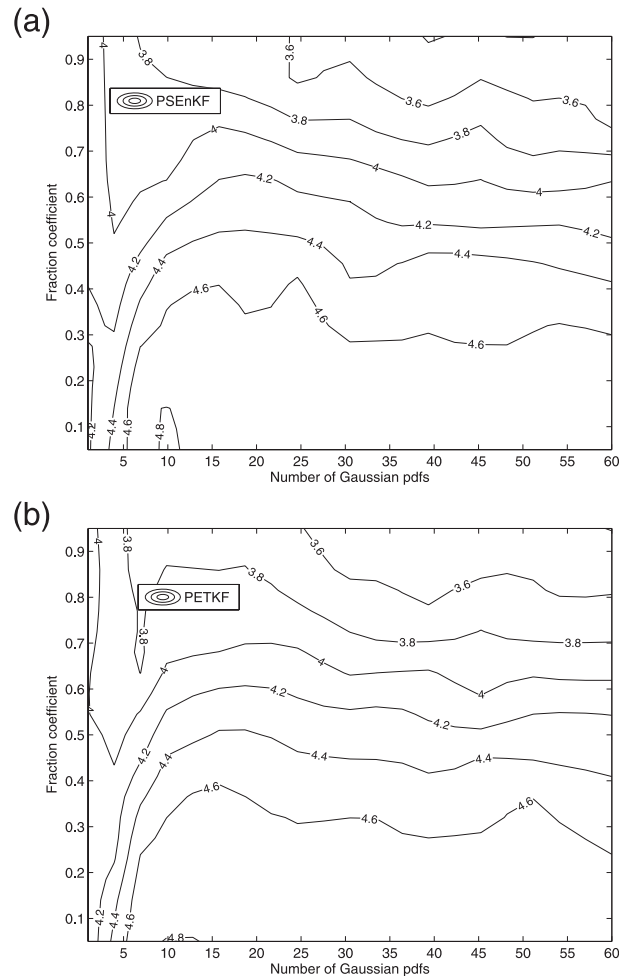


FIG. 6. RMS errors (over 20 experiments) of the (a) stochastic EnKF- and (b) ETKF-based PEnKFs (with a fixed ensemble size of 20 in each ensemble filter) as functions of the fraction coefficient and the number of Gaussian pdfs in the MON.

noticed in the results of Figs. 3 and 5 compared with those from Figs. 8 and 10, respectively, with different ensemble sizes.) This seemingly counterintuitive result may happen because in such situations, the effect of sampling error due to the relatively small ensemble size dominates the effect of nonlinearity in the observation system. However, as the number N of Gaussian distributions increases, the effect of nonlinearity becomes more prominent so that the filters with linear observation system performs better than those with the nonlinear observation system.

2) EFFECT OF THE ENSEMBLE SIZE

In the second experiment we examine the effect of the ensemble size in each ensemble filter on the performance of the corresponding PEnKF. For reference, we also examine the performance of the SEnKF and the ETKF

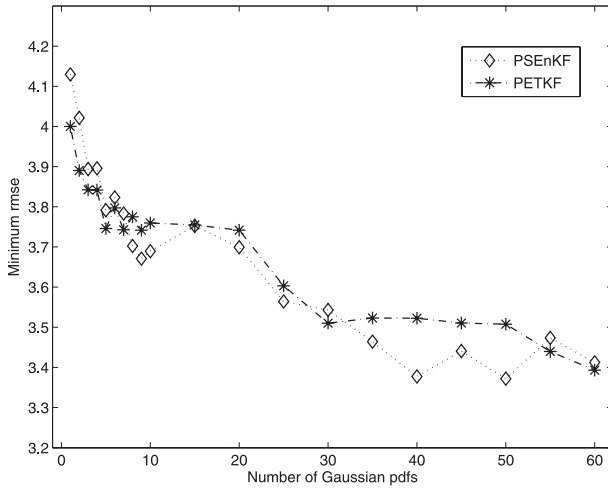


FIG. 7. Minimum rms errors \hat{e}_{\min} (over 20 experiments) of the stochastic EnKF- and ETKF-based PEnKFs (with a fixed ensemble size of 20 in each ensemble filter) as a function of the number of Gaussian pdfs in the MON.

under various ensemble sizes. The experiment settings are the same as those in section 4c(2). In the PSEnKF and PETKF, we choose the fraction coefficient $c \in \{0.05: 0.1: 0.95\}$. We also choose the number of ensemble filters in each PEnKF to be 3. For each individual EnKF in the corresponding PEnKF, we let the ensemble size m take values from the set $\{20, 40, 80, 100, 200, 400, 800, 1000\}$, and for the experiments on the single EnKF, we let $m \in \{20, 40, 60, 80, 100, 200, 400, 600, 800, 1000\}$. To conduct covariance inflation and localization in each individual EnKF, we choose the inflation factor $\delta = 0.02$, and the length scale $l_c = 50$. As in section 4c(2), the covariance localization is conducted only if the ensemble size m is no larger than the dimension 40.

Figure 8 shows the rms errors of the SEnKF and the ETKF as functions of the ensemble size m . For both filters, their rms errors decreases as the ensemble size m increases. The ETKF performs better than the SEnKF in the small sample scenario with $m = 20$. But as m increases, the SEnKF outperforms the ETKF. In particular, divergence in the ETKF occurs if $m > 400$, which did not happen in the linear observer case (cf. Fig. 3). On the other hand, the rmse of the SEnKF appears to reach a plateau for $m > 400$, similar to the linear observer case. Comparing Fig. 8 with Fig. 3, it is easy to see that, except for the stochastic EnKF at $m = 20$, the presence of nonlinearity in the observer deteriorates the performance of the ensemble filters.

Figure 9 plots the rms errors of the PSEnKF and the PETKF as functions of the fraction coefficient c , and the ensemble size m in the corresponding SEnKF and the ETKF, respectively. In the PSEnKF (cf. Fig. 9a),

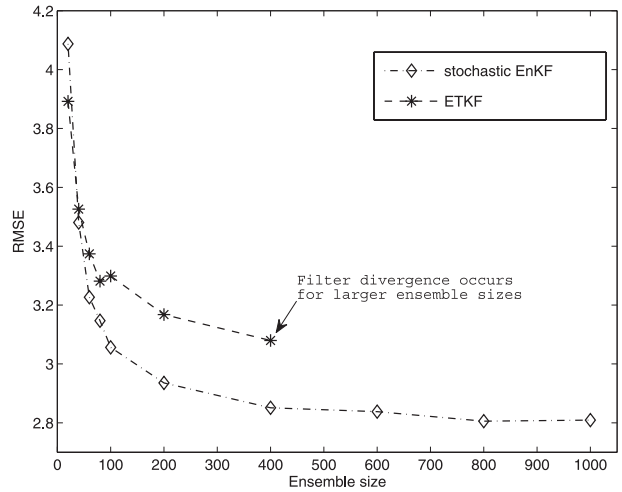


FIG. 8. RMS errors (over 20 experiments) of the stochastic EnKF and the ETKF as a function of the ensemble size.

the rmse tends to decrease as both c and m increases when the ensemble size $m \leq 800$. However, when $m > 800$, the impact of m on the filter performance is not significant, which is consistent with the results in Fig. 8. On the other hand, in the PETKF (cf. Fig. 9b), filter divergence occurs for $m > 200$, which is why we only report its rmse with $m \leq 200$ in Fig. 9b, where the rmse of the PETKF appears to be a monotonically decreasing function of m and c .

In analogy to the first experiment, Fig. 10 plots the minimum rms errors \hat{e}_{\min} of the PSEnKF and the PETKF within the tested fraction coefficient c and ensemble size m . One may observe that, similar to the SEnKF and the ETKF themselves, the \hat{e}_{\min} of both the PSEnKF and the PETKF decrease as m increases. However, for the PETKF, divergence occurs if $m > 200$, rather than $m > 400$ as in Fig. 8, but overall its rmse is closer to that obtained in the PSEnKF. Meanwhile, a comparison between Figs. 8 and 10 shows that the PEnKFs perform better than the corresponding EnKFs. Also, a comparison between Figs. 5 and 10 shows that, except for the PSEnKF at $m = 20$, the nonlinearity in the observer again deteriorates the performance of the ensemble filters.

5. Discussion

This paper presented a discrete solution of the optimal nonlinear filter, called the particle Kalman filter (PKF), based on the Gaussian mixture representation of the state pdf given the observations. The PKF solves the nonlinear Bayesian correction step by complementing the Kalman filter-like correction step of the particles with a particle filter-like correction step of the weights. The PKF simultaneously runs a weighted ensemble of

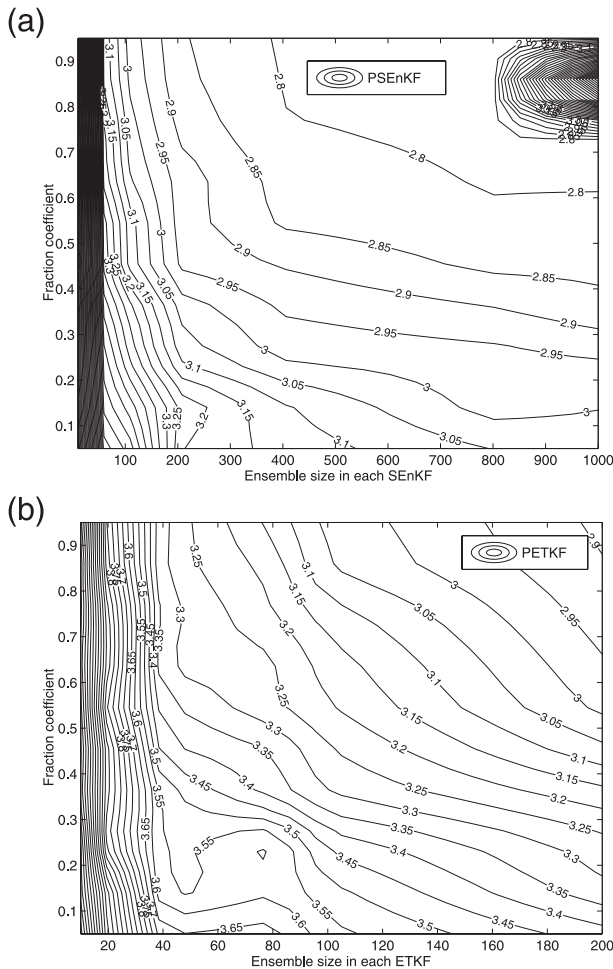


FIG. 9. RMS errors (over 20 experiments) of the (a) stochastic EnKF- and (b) ETKF-based PENKFs (with a fixed number of Gaussian pdfs of 3 in each PKF) as functions of the fraction coefficient and the ensemble size of the ensemble filter. In (b) the ensemble size in each ensemble filter is only up to 200. Divergence occurs in the ETKF-based PKF with ensemble sizes in each ensemble filter larger than 200.

the Kalman filters in parallel. This is far beyond our computing capabilities when dealing with computationally demanding systems, as the atmospheric and oceanic models. Therefore, to reduce computational cost, one may instead consider a low-rank parameterization of the Gaussian mixture covariance matrices of the state pdfs. An efficient way to do that is to resort to the ensemble Kalman filter (EnKF) and use an EnKF-like method to update each component of the Gaussian mixture pdfs. This amounts to running a weighted ensemble of the EnKFs. In this work, the PKF was implemented using the stochastic EnKF and a deterministic EnKF, the ensemble transform Kalman filter (ETKF). We call this type of implementation the particle ensemble Kalman filter (PEnKF).

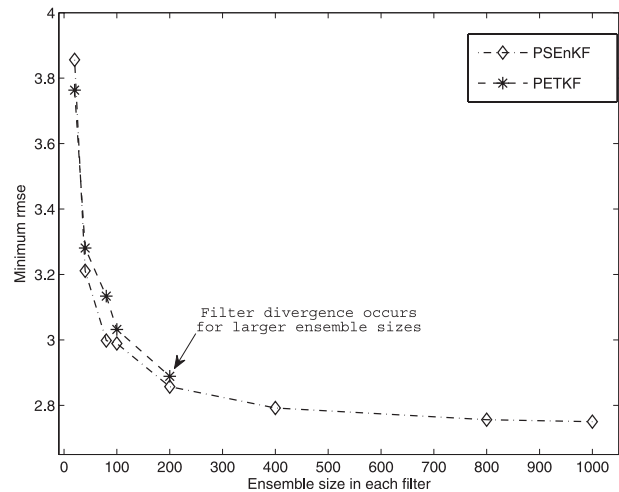


FIG. 10. Minimum rms errors \hat{e}_{\min} (over 20 experiments) of the stochastic EnKF- and ETKF-based PENKFs (with a fixed number of Gaussian pdfs of 3 in each PKF) as a function of the ensemble size in each ensemble filter.

The PEnKF sets a nonlinear Bayesian filtering framework that encompasses the EnKF methods as a special case. As in the EnKF, the Kalman correction in the PEnKF attenuates the degeneracy of the ensemble by allocating the ensemble members far away from the incoming observation relatively more weights than in the particle filter, so that the filter can operate with reasonable size ensembles. To further improve the performance of the PEnKF, we also introduced to the PEnKF a resampling step similar to that used in the regularized particle filter (Musso et al. 2001; Stavropoulos and Titterton 2001).

The stochastic EnKF and ETKF-based PENKFs, called the PSEnKF and the PETKF, respectively, were implemented and their performance was investigated with the strongly nonlinear Lorenz-96 model. These filters were tested with both linear and nonlinear observation operators. Experiments results suggest that the PSEnKF and the PETKF outperform their corresponding EnKFs. It was also found that the ETKF outperforms the stochastic EnKF for small-size ensembles while the stochastic EnKF exhibits better performance for large-size ensembles. We argued that this happens because the EnKF endures less observational sampling errors when the ensemble size is large. Another reason would be the better approximation of the PEnKF distributions provided by the stochastic EnKF compared to the ETKF. This was also true for their PEnKF counterparts. Overall, the conclusions from the numerical results obtained with the linear and nonlinear observation operators were not fundamentally different, except that in general better estimation accuracy was achieved

with the linear observer when the sampling error is not the dominant factor. The results also suggest that the PENKFs could more benefit from the use of more components in the mixture of normals (MON) and larger ensembles in the EnKFs in the nonlinear observations case.

Future work will focus on developing and testing new variants of the PENKF that applies more efficient approximations, in term of computational cost, to update the mixture covariance matrices (Hoteit et al. 2002). Another direction for improvement would also be to work on localizing the correction step of the particle weights (Van Leeuwen 2009). Our final goal is to develop a set of computationally feasible suboptimal PENKFs that can outperform the EnKF methods at reasonable computational cost. As stated by Anderson (2003), developing filters in the context of the optimal nonlinear filtering problem, rather than starting from the Kalman filter, should lead to a more straightforward understanding of their capabilities.

The paper further discussed how the PENKF can also be used as a general framework to simultaneously run several assimilation systems. We believe that this approach provides a framework to merge the solutions of different EnKFs, or to develop hybrid EnKF-variational methods. Work in this direction is under investigation.

Acknowledgments. We thank the three anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, **108**, 8775, doi:10.1029/2002JD002900.
- , P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collect.*, **2**, 316–334.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: On the analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Chen, R., and J. Liu, 2000: Mixture Kalman filters. *J. Roy. Stat. Soc. Series B Stat. Methodol.*, **62** (3), 493–508.
- Doucet, A., N. De Freitas, and N. Gordon, Eds., 2001: *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 612 pp.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Aghulas Current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics*, Vol. 33, Academic Press, 141–266.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- , —, J. L. Anderson, and C. Snyder, 2009: Comments on “Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems.” *J. Atmos. Sci.*, **66**, 3498–3500.
- Han, X., and X. Li, 2008: An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. *Remote Sens. Environ.*, **112**, 1434–1449.
- Hoteit, I., D.-T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in the tropical Pacific. *J. Mar. Syst.*, **36**, 101–127.
- , —, G. Triantafyllou, and G. Korres, 2008: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Mon. Wea. Rev.*, **136**, 317–334.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kotecha, J., and P. Djurić, 2003: Gaussian particle filtering. *IEEE Trans. Signal Process.*, **51** (10), 2592–2601.
- Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.
- Luo, X., I. M. Moroz, and I. Hoteit, 2010: Scaled unscented transform Gaussian sum filter: Theory and application. *Physica D*, **239**, 684–701.
- Musso, C., N. Oudjane, and F. Le Gland, 2001: Improving regularized particle filters. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. De Freitas, and N. Gordon, Eds., Springer-Verlag, 247–271.
- Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlinear Processes Geophys.*, **14**, 395–408.
- Pham, D.-T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Redner, R., and H. Walker, 1984: Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, **26** (2), 195–239.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 176 pp.
- Simon, D., 2006: *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Wiley-Interscience, 552 pp.
- Smith, K. W., 2007: Cluster ensemble Kalman filter. *Tellus*, **59A**, 749–757.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640.
- Sorenson, H. W., and D. L. Alspach, 1971: Recursive Bayesian estimation using Gaussian sums. *Automatica*, **7**, 465–479.

- Stavropoulos, P., and D. M. Titterton, 2001: Improved particle filters and smoothing. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. De Freitas, and N. Gordo, Eds., Springer-Verlag, 295–317.
- Todling, R., 1999: Estimation theory and foundations of atmospheric data assimilation. DAO Office Note 1999-01, 187 pp.
- Van Leeuwen, P. J., 2003: A variance minimizing filter for large-scale applications. *Mon. Wea. Rev.*, **131**, 2071–2084.
- , 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Xiong, X., I. Navon, and B. Uzunoglu, 2006: A note on the particle filter with posterior Gaussian resampling. *Tellus*, **58A**, 456–460.