

What the collapse of the ensemble Kalman filter tells us about particle filters

Matthias Morzfeld, Daniel Hodyss & Chris Snyder

To cite this article: Matthias Morzfeld, Daniel Hodyss & Chris Snyder (2017) What the collapse of the ensemble Kalman filter tells us about particle filters, Tellus A: Dynamic Meteorology and Oceanography, 69:1, 1283809, DOI: [10.1080/16000870.2017.1283809](https://doi.org/10.1080/16000870.2017.1283809)

To link to this article: <https://doi.org/10.1080/16000870.2017.1283809>



© 2017 Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Mar 2017.



Submit your article to this journal [↗](#)



Article views: 1182



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 9 View citing articles [↗](#)

What the collapse of the ensemble Kalman filter tells us about particle filters

By MATTHIAS MORZFELD^{1*}, DANIEL HODYSS² and CHRIS SNYDER³, ¹*Department of Mathematics, University of Arizona, Tucson, AZ, USA;* ²*Marine Meteorology Division, Naval Research Laboratory, Monterey, CA, USA;* ³*National Center for Atmospheric Research, Boulder, CO, USA*

(Manuscript Received 31 May 2016; in final form 19 December 2016)

ABSTRACT

The ensemble Kalman filter (EnKF) is a reliable data assimilation tool for high-dimensional meteorological problems. On the other hand, the EnKF can be interpreted as a particle filter, and particle filters (PF) collapse in high-dimensional problems. We explain that these seemingly contradictory statements offer insights about how PF function in certain high-dimensional problems, and in particular support recent efforts in meteorology to ‘localize’ particle filters, i.e. to restrict the influence of an observation to its neighbourhood.

Keywords: ensemble Kalman filter, particle filter, collapse of particle filters

1. Introduction

Many applications in geophysics require that one updates the state of a numerical model by information from sparse and noisy observations, see, e.g. van Leeuwen (2009), Bocquet et al. (2010), Fournier et al. (2010) and van Leeuwen et al. (2015) for recent reviews in geophysics. Particle filters (PF) are sequential Monte Carlo algorithms that solve such a ‘data assimilation’ problem as follows (Doucet et al., 2001; Arulampalam et al., 2002). One draws samples from a proposal density, and then attaches weights to the samples such that the weighted samples form an empirical estimate of a posterior density. The samples are often called ‘particles’ in the PF literature, or ‘ensemble members’ in the geophysics literature. The empirical estimate approximates the posterior density in the sense that weighted averages over the samples converge to expected values with respect to the posterior density. The PFs in the literature differ in their choice of proposal density, and therefore in their weights. Several particle filters can be found, e.g. in Zaritskii and Shimelevich (1975), Gordon et al. (1993), Liu and Chen (1995), Doucet (1998), Doucet et al. (2000), Chorin and Tu (2009), Weare (2009), Chorin et al. (2010), Morzfeld et al. (2012) and Vanden-Eijnden and Weare (2012).

The variance of the weights determines the efficiency of a PF. If this variance is large, then a PF ‘collapses’, i.e. the weights of all but one sample are near zero. The collapse of particle filters has been described in Snyder et al. (2008); Bickel et al. (2008);

Bengtsson et al. (2008); Snyder (2011); Chorin and Morzfeld (2013); Snyder et al. (2015). It is argued that an effective dimension defines the ensemble size required to prevent collapse. While there are various definitions of effective dimensions, we explain in Section 2.4 that the huge state dimension of global weather models (hundreds of millions of state variables) and the vast number of observations (millions) cause any effective dimension to be large, which in turn implies that particle filters would require huge ensemble sizes in meteorological problems.

It is interesting however that the ensemble Kalman filter (EnKF, Tippet et al., 2003; Evensen, 2006), used routinely with ensembles of 50 to a few hundred members in meteorological problems, can in fact be interpreted as a particle filter. Specifically, as shown by Papadakis et al. (2010), one can view the EnKF ensemble members as draws from a proposal density and attach weights to each ensemble member, interpreting the EnKF as a particle filter. It follows from the optimality result of Snyder et al. (2015) that the particle filter interpretation of the EnKF collapses in the same way as other particle filters. The goal of this paper is to clarify this situation: *Why does the EnKF do so well in meteorological problems in which its particle filter interpretation collapses?* Resolving this apparent contradiction reveals the significance of a certain type of particle filter collapse, and suggests improvements of particle filters for meteorological applications.

*Corresponding author. e-mail: mmo@math.arizona.edu

2. Preliminaries

We consider linear Gaussian data assimilation problems because we can make our points in this simplified setting. Moreover, linear problems are standard in the context of the collapse of particle filters and have been used by [Snyder et al. \(2008\)](#), [Bickel et al. \(2008\)](#), [Bengtsson et al. \(2008\)](#), [Snyder \(2011\)](#), [Chorin and Morzfeld \(2013\)](#), [Snyder et al. \(2015\)](#). Specifically, we consider

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

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

where the state, \mathbf{x} , is a real n_x -dimensional vector, the linear model, \mathbf{M} , is an $n_x \times n_x$ matrix, the observation, \mathbf{y} , is a real n_y -dimensional vector, \mathbf{H} is a $n_y \times n_x$ matrix and where $k = 1, 2, 3, \dots$ is discrete time. We further assume that the initial condition \mathbf{x}_0 is Gaussian distributed, and that no observations are available at time $k = 0$. Throughout this paper, we use subscripts for time indexing and superscripts for indexing samples of random variables. For example, \mathbf{x}_k^j is the j th sample of the state \mathbf{x} at time k . We write $\mathbf{y}_{1:k}$ for a set of vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_k\}$. The random variables $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$ in Equations (1) and (2) represent model errors and errors in the observations, respectively, and we assume that $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}_k$ are independent identically distributed Gaussians with mean zero, which we write as $\boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $\boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. The covariances \mathbf{Q} and \mathbf{R} are $n_x \times n_x$, respectively, $n_y \times n_y$ symmetric positive definite matrices. We exclude data assimilation problems with a deterministic model for which $\mathbf{Q} = \mathbf{0}$, but many of our results hold for data assimilation problems with deterministic models as well by simply setting $\mathbf{Q} = \mathbf{0}$ in equations. We explain which results generalize in Section 6.

2.1. Filtering and smoothing densities

The goal of data assimilation is to update the state of the model (1) in view of the noisy observations in (2). Model and observations jointly define the conditional random variable $\mathbf{x}_{0:k}|\mathbf{y}_{1:k}$, which describes the state trajectory up to time k , given the observations up to time k . This random variable is specified by the ‘smoothing density’

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k}). \quad (3)$$

The ‘filtering density’ is a marginal of the smoothing density,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \int \dots \int p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k}) d\mathbf{x}_0 \dots d\mathbf{x}_{k-1}. \quad (4)$$

and describes the state at observation time, conditioned on all observations up to now, but it does not describe how the system

got to that state. In practice, the filtering density may be more important. Forecast models, for example, require initial conditions sampled from the filtering density; samples from the smoothing density, i.e. descriptions of past weather, are not required. On the other hand, smoothing algorithms have recently received more and more attention, especially in the context of re-analyses.

2.2. The localized EnKF

The EnKF uses a Monte Carlo step within the Kalman filter formalism, and recursively approximates the filtering density by an ensemble ([Kalman, 1960](#); [Kalman and Bucy, 1961](#); [Tippett et al., 2003](#); [Evensen, 2006](#)). Specifically, let \mathbf{x}_{k-1}^j , $j = 1, \dots, N_e$, be samples of the filtering density $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ at time $k-1$. The collection of N_e samples is called the ensemble, and each sample is an ensemble member; N_e is the ensemble size. The ensemble is evolved to time k using the model (1) to generate the forecast ensemble

$$\hat{\mathbf{x}}_k^j = \mathbf{M}\mathbf{x}_{k-1}^j + \boldsymbol{\eta}_k^j,$$

where $\boldsymbol{\eta}_k^j$ is a sample of $\boldsymbol{\eta}_k$. Let

$$\mathbf{P}^f = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{x}_k^j - \bar{\mathbf{x}}_k)(\mathbf{x}_k^j - \bar{\mathbf{x}}_k)^T, \quad (5)$$

be the forecast covariance, where $\bar{\mathbf{x}}_k = (\sum_{j=1}^{N_e} \mathbf{x}_k^j) / N_e$ is the forecast mean. The forecast covariance is usually localized and inflated, see, e.g. [Houtekamer and Mitchell \(2001\)](#) and [Hamill et al. \(2001\)](#). During localization, one decreases long-range correlations in the error covariances, typically by Schur (entrywise) multiplication of the forecast covariance with a specified $n_x \times n_x$ localization matrix ρ :

$$\mathbf{P}^f \leftarrow \rho \circ \mathbf{P}^f.$$

In atmospheric applications, localization matrices as described by [Gaspari and Cohn \(1999\)](#) are commonly used. During inflation, one increases the localized forecast covariance, e.g., by multiplying the forecast covariances by a number slightly larger than one, e.g. $\mathbf{P}^f \leftarrow (1 + \alpha)\mathbf{P}^f$, $\alpha > 0$. Combined, localization and inflation are known to be key factors in making the EnKF a practical tool for meteorological data assimilation, see, e.g. [Houtekamer et al. \(2005\)](#), [Wang et al. \(2007\)](#) and [Raeder et al. \(2012\)](#). Here, ‘practical’ means that an EnKF with a small ensemble ($N_e = 50-100$) yields a small mean square error (MSE), and a comparable ensemble ‘spread’, defined by the trace of the posterior covariance divided by the system dimension n_x (see Section 4).

The localized and inflated forecast covariance defines the Kalman gain

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (6)$$

which is used to compute the analysis ensemble

$$\mathbf{x}_k^j = \hat{\mathbf{x}}_k^j + \mathbf{K} (\hat{\mathbf{y}}_k^j - \mathbf{H} \hat{\mathbf{x}}_k^j), \quad (7)$$

where $\hat{\mathbf{y}}_k^j = \mathbf{y}_k + \varepsilon_k^j$, is a perturbed observation and where ε_k^j is a sample of ε_k . The sample mean and sample covariance of the analysis ensemble approximate the mean and covariance of the filtering density $p(\mathbf{x}_k | \mathbf{y}_{1:k})$. This is the ‘perturbed observations’ implementation of the localized EnKF. Other implementations of the EnKF are also available, see, e.g. [Anderson \(2001\)](#), [Bishop et al. \(2001\)](#) and [Tippett et al. \(2003\)](#).

2.3. Particle filters

Particle filters are sequential importance sampling methods that construct an empirical estimate of the smoothing density (3), see, e.g. [Doucet et al. \(2001\)](#). The empirical estimate consists of a collection of weighted samples, such that expected values of functions of $\mathbf{x}_{0:k} | \mathbf{y}_{1:k}$ can be approximated by weighted averaging over the samples, see, e.g. [Kalos and Whitlock \(1986\)](#) and [Chorin and Hald \(2013\)](#). In particular, particle filters make use of a recursion of the smoothing density,

$$p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) = p(\mathbf{x}_{0:k-1} | \mathbf{y}_{1:k-1}) \frac{p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{y}_k | \mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})}, \quad (8)$$

and draw samples from a proposal density which is a similar product of functions,

$$\pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) \propto \pi_0(\mathbf{x}_0) \prod_{l=1}^k \pi_l(\mathbf{x}_l | \mathbf{y}_{1:l}, \mathbf{x}_{0:l-1}). \quad (9)$$

Thus, only the most recent product needs to be updated when a new observation is collected. At each step k , the update $\pi_k(\mathbf{x}_k | \mathbf{y}_{1:k}, \mathbf{x}_{0:k-1})$ is used to propose samples \mathbf{x}_k^j , which are appended to the samples of $\mathbf{x}_{0:k-1}^j$ to extend the trajectory to time k . The weights satisfy the recursion

$$w_k \propto \frac{p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k})}{\pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k})} \propto w_{k-1} \frac{p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{y}_k | \mathbf{x}_k)}{\pi_k(\mathbf{x}_k | \mathbf{y}_{1:k}, \mathbf{x}_{0:k-1})}, \quad (10)$$

and account for the fact that one draws samples sequentially from the proposal density $\pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k})$, rather than directly from

the smoothing density. After each step, the weights are normalized such that their sum is one, and the ensemble is ‘re-sampled’ ([Arulampalam et al., 2002](#)), i.e. ensemble members with a small weight are deleted, and ensemble members with a large weight are repeated. Various choices for the updates $\pi_k(\mathbf{x}_k | \mathbf{y}_{1:k}, \mathbf{x}_{0:k-1})$ lead to the various particle filters developed over the past years, each with different weights, e.g. [Gordon et al. \(1993\)](#), [Doucet \(1998\)](#), [Chorin and Tu \(2009\)](#), [Weare \(2009\)](#), [Chorin et al. \(2010\)](#), [Morzfeld et al. \(2012\)](#) and [Vanden-Eijnden and Weare \(2012\)](#).

More specifically, the ‘standard particle filter’ ([Gordon et al. \(1993\)](#)) uses the proposal density $\pi_k = p(\mathbf{x}_k | \mathbf{x}_{k-1})$, i.e. the ensemble is generated by running the numerical model (1). The weights are the ratio of target and proposal density and, thus, are proportional to the likelihood, $w_k \propto w_{k-1} p(\mathbf{y}_k | \mathbf{x}_k)$. Note that the standard particle filter does not make use of the most recent observation when generating the ensemble. A particle filter with proposal density $\pi_k = p(\mathbf{x}_k | \mathbf{y}_k, \mathbf{x}_{k-1})$ however does make use of the most recent observation when generating the ensemble. The weights of this particle filter are given by $w_k \propto w_{k-1} p(\mathbf{y}_k | \mathbf{x}_{k-1})$, and it can be shown that these weights have minimum variance over all particle filters with a proposal density of the form (9) ([Snyder et al. \(2015\)](#)). For that reason, this particle filter is called the ‘optimal particle filter’ (see, e.g. [Zaritskii and Shimelevich, 1975](#); [Liu and Chen, 1995](#); [Doucet et al., 2000](#)). There are, however, more general particle filters that make use of more complicated proposal densities than the ones described by Equation (9). For example, one can use ‘tempering’, i.e. techniques that ensure a more gradual transition from a proposal density to a posterior density (see, e.g. [Beskos et al., 2014](#) and [Kantas et al., 2014](#)). Such filters may have lower variance weights than the ‘optimal’ particle filter described above. For the remainder of this paper, however, we only consider ‘classical’ particle filters with proposal densities of the form (9).

2.4. The collapse of particle filters and effective dimensions

A particle filter collapses if only one ensemble member carries a significant weight. The collapse occurs when the variance of the weights is large, and avoiding the collapse can require huge ensemble sizes. Indeed, it is often said that particle filters require ensemble sizes that are exponential in the ‘dimension’. However, the situation is more delicate. It was argued by [Chorin and Morzfeld \(2013\)](#) that there are several mechanisms that can cause particle filters to collapse, not merely a large state dimension. These mechanisms are described by the theory of the collapse of particle filters and effective dimensions: when the effective dimension is large, an even larger ensemble size is needed to avoid collapse. We briefly review this theory here because it is needed to identify the cause for the collapse of particle filters in meteorological problems.

The collapse of the ‘standard particle filter’ with proposal density $\pi_k = p(\mathbf{x}_k | \mathbf{x}_{k-1})$ is described by [Snyder et al. \(2008\)](#), [Bickel et al. \(2008\)](#) and [Bengtsson et al. \(2008\)](#). Specifically, in [Snyder et al. \(2008\)](#), it is shown that the collapse of this particle filter depends on the variance

$$\tau^2 = \text{var}(\log w) = \sum_{j=1}^{n_y} \left(\frac{1}{2} \lambda_j^2 + \lambda_j (\mathbf{y}_k)_j^2 \right),$$

where $(\mathbf{y}_k)_j$ is the j th element of the observation \mathbf{y}_k , and where λ_j are the eigenvalues of the covariance matrix in observation space, normalized by the observation covariance matrix, $\text{cov}(\mathbf{R}^{-1/2} \mathbf{H} \mathbf{x}_k)$. Asymptotic statements about the collapse of the standard particle filter can be made for large state dimensions and large number of observations ($n_x, n_y \gg 1$), when $\sum \lambda_j \gg 1$.

In this case, collapse occurs if $(\log N_e)/\tau^2 \rightarrow 0$, i.e. the ensemble size must be exponential in τ^2 to avoid the collapse. Similar arguments hold for the optimal particle filter, with a slightly different definition of τ^2 ([Snyder, 2011](#); [Snyder et al., 2015](#)). The exponential scaling of the ensemble size N_e with τ^2 suggests that the ensemble’s size required to avoid collapse may be too large to be feasible in numerical weather prediction, where n_y is so large – typically millions – that even moderate eigenvalues λ_j lead to large τ^2 .

Effective dimensions are quantities related to τ^2 , which describe how difficult it is to solve a given data assimilation problem by a particle filter. There are several definitions of effective dimensions in the literature, some of which are reviewed by [Agapiou et al. \(2016\)](#). For example, in [Bickel et al. \(2008\)](#), an effective dimension is defined by

$$\tau_{\text{spf}} = \text{trace}(\text{cov}(\mathbf{R}^{-1/2} \mathbf{H} \mathbf{x}_k)) = \sum_{j=1}^{n_y} \lambda_j. \quad (11)$$

Note that this effective dimension is a property of the data assimilation problem *and* the standard particle filter algorithm. In [Chorin and Morzfeld \(2013\)](#), effective dimensions for the standard and optimal particle filter are defined by Frobenius norms, rather than traces of covariances. The Frobenius norm of a matrix is the square root of the sum of the squares of the singular values. For the symmetric covariance matrices we consider, this is the same as the square root of the sum of the squares of the eigenvalues. For example, an effective dimension of the standard particle filter is defined by

$$\hat{\tau}_{\text{spf}}^2 = \|\text{cov}(\mathbf{R}^{-1/2} \mathbf{H} \mathbf{x}_k)\|_F^2 = \sum_{j=1}^{n_y} \lambda_j^2.$$

All effective dimensions above depend on the covariance of \mathbf{x}_k and on the data assimilation algorithm one uses. In [Chorin and Morzfeld \(2013\)](#), the covariance of \mathbf{x}_k is calculated during steady state of a linear and Gaussian data assimilation problem, so that the effective dimension does not vary over time. Note that the trace of the prior and posterior covariance matrices often oscillates about a steady-state value in non-linear problems as well. Indeed, the Frobenius norm, or the trace, of a (steady state) posterior covariance matrix \mathbf{P} is a natural candidate for an effective dimension of a data assimilation problem, independently of the particle filter one applies to solve it:

$$\tau_{\text{sys}}^2 = \|\mathbf{P}\|_F^2 = \sum_{j=1}^{n_x} \alpha_j^2. \quad (12)$$

Here, α_j are the eigenvalues of \mathbf{P} , and the Frobenius norm depends on all aspects of the data assimilation problem, including the number of observations, the number of state variables, as well as the covariances that define model and observation error (\mathbf{Q} and \mathbf{R}).

It was pointed out in [Chorin et al. \(2016\)](#) that none of the effective dimensions are in fact ‘dimensions’. The effective dimensions above rather describe ‘effective variance’, and one can define more effective dimensions, for example, by

$$\ell_{\text{sys}} = \min \ell \in \mathbb{Z} : \sum_{j=1}^{\ell} \alpha_j^2 \geq (1 - \kappa) \sum_{j=1}^{n_x} \alpha_j^2, \quad (13)$$

where κ is a predetermined small parameter ([Chorin et al., 2016](#)). Similar ideas were used in the context of ‘partial noise’ by [Morzfeld and Chorin \(2012\)](#). However, it is not yet clear which effective dimension will ultimately be most useful. Nonetheless, all effective dimensions have in common that they describe that ‘if the effective dimension is large, an even larger ensemble size is needed to avoid collapse’.

2.5. The collapse of particle filters in meteorology

We are interested in the collapse of particle filters as it may occur in *meteorological problems*. Meteorological problems are characterized by

- (a) large state dimension n_x (hundreds of millions), and large observation dimension n_y (millions);
- (b) large spatial domains, in the sense that there are many correlation lengths of a prior density contained within the overall system domain;
- (c) the matrix \mathbf{H} is ‘local’, in the sense that each element of the vector $\mathbf{H} \mathbf{x}$ depends only on ‘a few’ elements of \mathbf{x} .

These conditions are satisfied in a canonical linear example, where particle filter collapse occurs and where the required ensemble size scales exponentially with dimension (see, e.g. Snyder et al., 2008; Bickel et al., 2008; Bengtsson et al., 2008; Snyder, 2011; Chorin and Morzfeld, 2013; Snyder et al., 2015, and also Section 4). More generally, the above criteria imply that all effective dimensions are large. Since even the optimal particle filter requires large ensemble sizes if an effective dimension is large, one can expect that *all* particle filters collapse in meteorological problems unless the ensemble size is huge (see Chorin and Morzfeld, 2013; Snyder et al., 2015).

To understand why the above conditions imply large effective dimensions, recall that the Frobenius norm of a matrix is equal to the square root of the sum of the squares of all of its elements, $\|\mathbf{P}\|^2 = \sum_{i,j} P_{i,j}^2$. Thus, if n_x is huge, the effective dimension τ_{sys} is large unless each entry of \mathbf{P} is tiny.

A large τ_{sys} implies that the effective dimensions τ_{spf} and $\hat{\tau}_{\text{spf}}$ are also large (see Chorin and Morzfeld, 2013). The same arguments apply to effective dimensions defined for the optimal particle filter (Chorin and Morzfeld, 2013; Snyder, 2011; Snyder et al., 2015). The effective dimension ℓ_{sys} is large since, by property (b), far away variables de-correlate, so that the overall system can be thought of as a collection of loosely coupled lower dimensional ‘sub-systems’. Since the correlation lengths are small compared to the overall system domain, the overall system contains a large number of subsystems. Thus, the effective dimension ℓ_{sys} is large. In fact, it increases with the number of loosely coupled subsystems.

We emphasize that the above statements about the collapse of particle filters in meteorology only apply to particle filters of a certain kind (see Section 2.3). The collapse may be avoided, even for meteorological problems, by a next generation of particle filters that make use of more general approach, e.g. using techniques to transition more smoothly from a proposal density to a posterior density. Indeed, some of these algorithms have already been proven to scale algebraically, rather than exponentially with (effective) dimension (see, e.g. Beskos et al., 2014; Kantas et al., 2014).

3. The collapse of the particle filter interpretation of the EnKF

The EnKF performs well in high-dimensional meteorological applications in which particle filters are expected to collapse. On the other hand, it was shown by Papadakis et al. (2010) that the EnKF can be interpreted as a particle filter (PF-EnKF), and, thus, can be expected to collapse under the same conditions as other particle filters. We study this apparent contradiction and in particular the mechanisms that cause the collapse of the PF-EnKF in meteorological problems. This approach allows us to combine what we know about the performance of EnKF in *real* meteorological problems with what we know about the collapse of particle filters. In this context, we first consider the

PF-EnKF as a particle filter for the smoothing density, and then make connections to the filtering density using ideas similar to marginal particle filters introduced by Klaas et al. (2005).

3.1. Collapse of the PF-EnKF for the smoothing density

It was pointed out by Papadakis et al. (2010) that the EnKF ensemble members can be viewed as draws from the proposal density

$$\pi_{k,\text{EnKF}}^j(\mathbf{x}_k | \mathbf{x}_{k-1}^j, \mathbf{y}_k) = \mathcal{N}(\mu_k^j, \Sigma_k), \quad (14)$$

where

$$\begin{aligned} \mu_k^j &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{M}\mathbf{x}_{k-1}^j + \mathbf{K}\mathbf{y}_k, \\ \Sigma_k &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{Q}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T, \end{aligned}$$

with the Kalman gain \mathbf{K} defined in (6). This proposal density is of the form (9) so that the EnKF can be interpreted as a particle filter. A related construction using Kernel density estimation at each step is discussed in Khlalil et al. (2015). The weights that can be attached to the EnKF ensemble are the ratio of the smoothing and proposal densities

$$w_{k,\text{EnKF}}^j \propto w_{k-1,\text{EnKF}}^j \frac{p(\mathbf{x}_k^j | \mathbf{x}_{k-1}^j) p(\mathbf{y}_k | \mathbf{x}_k^j)}{\pi_{k,\text{EnKF}}^j(\mathbf{x}_k | \mathbf{x}_{k-1}^j, \mathbf{y}_k)}. \quad (15)$$

The theory for the collapse of particle filters reviewed in Section 2.4 applies to these weights, and their variance governs the collapse of the PF-EnKF. This variance is at least as large as the corresponding variance of the optimal PF (Snyder et al., 2015). Thus, the PF-EnKF collapses whenever the optimal PF collapses. As a specific example, we will demonstrate in Section 4 that the ensemble size required to avoid the collapse of the PF-EnKF scales exponentially with the dimension n_x in the canonical system studied in Snyder et al. (2008), Bickel et al. (2008), Bengtsson et al. (2008), Snyder (2011), Chorin and Morzfeld (2013) and Snyder et al. (2015).

While covariance localization and inflation of EnKF will modify the mean and covariance of the proposal density (14), these mechanisms will not avoid the collapse when sampling from the smoothing density. The reason is that, even with tuned localization and inflation, the proposal density defined by the EnKF ensemble cannot improve on the performance of the optimal proposal and must collapse in the same way as the optimal particle filter. Thus, using even a localized EnKF that is operating correctly by tracking the truth to within the ensemble variance to define a proposal density cannot prevent the collapse for small ensemble sizes in meteorological problems.

3.2. Collapse of the PF-EnKF for the filtering density

The particle filter interpretation of the EnKF defines a sequential importance sampler for the smoothing density. However, the EnKF ensemble can also be viewed as draws from a proposal density for the filtering density. Indeed, the idea of designing particle filters directly for the filtering density, rather than the smoothing density, was explored by [Klaas et al. \(2005\)](#) via ‘marginal particle filters’. We connect these ideas to the EnKF and study the conditions for the collapse of the particle filter interpretation of the EnKF for the filtering density. The reason is that it may be possible that the PF-EnKF collapses as a sampler for the smoothing density, but performs well as a sampler for the filtering density, and that sampling the filtering density is sufficient in meteorology. Below we argue that this is not the case, i.e. the collapse of the PF-EnKF occurs, under the same conditions, for filtering and smoothing densities, unless the ensemble size is huge.

As before, we interpret the EnKF ensemble members as draws from a suitable proposal density $\pi_{\text{EnKF}}(\mathbf{x}_k)$. We wish to compute the weights,

$$w \propto \frac{p(\mathbf{x}_k | \mathbf{y}_{1:k})}{\pi_{\text{EnKF}}(\mathbf{x}_k)}. \quad (16)$$

such that the weighted ensemble approximates the filtering density, not the smoothing density as before. Note that the EnKF proposal density for the smoothing density in (14) is conditioned on the previous analysis state but the EnKF proposal density for the filtering density in (16) is not. Thus, the EnKF proposal density for the filtering density is not the same as the EnKF proposal density for the smoothing density. The weights (16), are difficult to compute in general because the filtering density and the EnKF proposal density are not known up to a multiplicative constant (see the Appendix and literature about marginal particle filters, e.g. [Klaas et al. \(2005\)](#), for more detail). However, for the linear Gaussian problems we consider, the filtering density is also Gaussian and we write $p(\mathbf{x}_k | \mathbf{y}_{1:k}) \sim \mathcal{N}(\mu_k, \mathbf{P}_k)$. The proposal density generated by the EnKF ensemble is the Gaussian,

$$\pi_{\text{EnKF}}(\mathbf{x}_k) = \mathcal{N}(\mu_{k,\text{EnKF}}, \mathbf{P}_{k,\text{EnKF}}), \quad (17)$$

where $\mu_{k,\text{EnKF}}$ is the mean of the analysis ensemble and where

$$\begin{aligned} \mathbf{P}_{k,\text{EnKF}} = & (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{M}\mathbf{P}_{k-1,\text{EnKF}}\mathbf{M}^T(\mathbf{I} - \mathbf{K}\mathbf{H})^T \\ & + (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{Q}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T. \end{aligned} \quad (18)$$

Under our assumptions, the weights are

$$w \propto \frac{\exp\left(-\frac{1}{2}(\mathbf{x}_k^j - \mu_k)^T \mathbf{P}_k^{-1}(\mathbf{x}_k^j - \mu_k)\right)}{\exp\left(-\frac{1}{2}(\mathbf{x}_k^j - \mu_{k,\text{EnKF}})^T \mathbf{P}_{k,\text{EnKF}}^{-1}(\mathbf{x}_k^j - \mu_{k,\text{EnKF}})\right)}. \quad (19)$$

Asymptotically, for large ensembles, the sample mean converges to the posterior mean and the sample covariance converges to the posterior covariance, so that the variance of the PF-EnKF weights for the filtering density goes to zero asymptotically (see, e.g. [Burgers et al., 1998](#)).

We are interested in the collapse of the particle filter interpretation of a localized EnKF with small ensemble sizes, and study this situation by making simplifying assumptions. Specifically, we neglect errors in the ensemble mean, i.e. we set $\mu_{k,\text{EnKF}} = \mu_k$, and assume that errors in the ensemble covariance matrix are small, i.e. $\mathbf{P}_{k,\text{EnKF}} = (1 + \beta)\mathbf{P}_k$, where $\beta \geq 0$ is a small number. One could also consider perturbations of the form $(1 - \beta)\mathbf{P}_k$, but, in the context of importance sampling, perturbations as above are more favourable, and we wish to study a ‘best-case-scenario’ for EnKF. The reason why $\beta > 0$ is favourable is that importance sampling requires that the support of the proposal density contains the support of the target density, i.e. one wants to use proposal densities with variances larger than those of the target densities.

While drastic, the above simplifications do lead to useful expansions that can accurately describe the behaviour of the weights of a localized EnKF with small ensembles (see also Section 4.3). Those simplifications require that the EnKF mean and covariance are close to correct. In typical large geophysical problems, this can be achieved for small ensemble size ($N_e \approx 100$) only when using covariance localization. Without localization, ensemble sizes substantially larger than n_x would be necessary. Thus, the subsequent arguments in this section hold only for a localized EnKF and an EnKF with large ensembles. We further assume, for simplicity, that the posterior covariance \mathbf{P}_k is full rank, so that all its eigenvalues are positive. While the rank of \mathbf{P}_k may not be equal to the state dimension n_x , conditions (a)–(c) in Section 2.5 imply that the rank of \mathbf{P}_k is large (we discuss what happens when some eigenvalues are zero below). In the context of the localized EnKF, our above assumption that the EnKF covariance is close to correct implies that $\mathbf{P}_{k,\text{EnKF}}$ has the same rank as the posterior covariance.

Under our assumptions, the weights in Equation (19) become

$$w \propto \exp\left(\frac{1}{2}\mathbf{s}^T \mathbf{s}\right), \quad \mathbf{s} = \sqrt{\frac{\beta}{1 + \beta}} \left(\mathbf{P}_k^{-1/2}(\mathbf{x}_k^j - \mu_k)\right), \quad (20)$$

where $\mathbf{x}_k^j \sim \mathcal{N}(\mu_k, \mathbf{P}_{k,\text{EnKF}})$, $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \beta\mathbf{I})$.

The variance of the above weights can be computed from the expected values

$$\begin{aligned} E[w] &= \int_{-\infty}^{\infty} \exp(-\mathbf{s}^T \mathbf{s}/2) \frac{\exp(-\mathbf{s}^T \mathbf{s}/(2\beta))}{(2\pi\beta)^{n/2}} d\mathbf{s} \\ &= \left(\frac{1}{1 + \beta}\right)^{n_x/2}, \\ E[w^2] &= \int_{-\infty}^{\infty} \exp(-\mathbf{s}^T \mathbf{s}) \frac{\exp(-\mathbf{s}^T \mathbf{s}/(2\beta))}{(2\pi\beta)^{n/2}} d\mathbf{s} \end{aligned}$$

$$= \left(\frac{1}{1 + 2\beta} \right)^{n_x/2}.$$

Indeed, one can define an ‘effective number of samples’ based on the variance of the weights (Arulampalam et al., 2002; Owen, 2013)

$$N_{e,\text{eff}} = \frac{N_e}{G}, \quad G = \frac{\text{var}(w)}{E(w)^2} - 1 = \frac{E(w^2)}{E(w)^2}. \quad (21)$$

Thus, the collapse occurs when $N_{e,\text{eff}} = 1$, i.e. if $G = N_e$, and the ‘quality measure’ G can be used to estimate the ensemble size required to avoid collapse. Under our assumptions,

$$G(\beta) = \left(\frac{1 + \beta}{\sqrt{1 + 2\beta}} \right)^{n_x} \approx 1 + \frac{\beta^2 n_x}{2} + O(\beta^3).$$

The approximation is obtained by expanding $G(\beta)$ around $\beta = 0$, i.e. for small errors β in the covariance matrix. Indeed, if we interpret $\beta \propto 1/\sqrt{N_e}$ as sampling error, then G can remain constant as dimension n_x increases, provided that the ensemble size increases linearly with the dimension. Thus, the collapse of the PF-EnKF for the filtering density does not occur for ensemble sizes proportional to the dimension n_x .

A small G however is not a guarantee that a sampling scheme is ‘working well’. In particular, G may be small if the variance of the proposal density is chosen too small compared to the target density. In this case, one may find that the weights are well distributed; however, other indicators, such as MSE, will show that one is sampling too narrowly within state space.

We can relax the assumption that all eigenvalues of the posterior covariance \mathbf{P}_k are positive by making use of the effective dimension ℓ_{sys} in Equation (13). If some eigenvalues are zero, or if some are close to zero, then one can substitute the $\ell_{\text{sys}} \leq n_x$ dimensional approximation of this covariance matrix \mathbf{P}_k into the above formulas. Using the pseudo-inverse of this matrix in (20), one obtains the same expression for G as above with ℓ_{sys} replacing n_x . We argued above that while ℓ_{sys} may be smaller than n_x in meteorological problems, ℓ_{sys} can be expected to be large in these problems (see property (b) in Section 2.5). A large ℓ_{sys} implies that the collapse of the particle filter interpretation of the EnKF can only be avoided by large ensemble sizes, in particular much larger than the 50–100 ensemble members used in practice.

4. Performance of the EnKF during the collapse of its particle filter interpretation

We have shown that the collapse of the PF-EnKF for the filtering density can be prevented by ensemble sizes that scale linearly with the dimension n_x , or, more generally, with the effective dimension ℓ_{sys} in (13). We further argued that the large dimensions

n_x and n_y in meteorological problems cause *all* particle filters for the smoothing density, including the particle filter interpretation of a *localized* EnKF, to collapse. On the other hand, localized EnKFs with small ensemble sizes (50 to several hundred) are used in meteorological problems with hundreds of millions of state variables and millions of observations and produce a MSE comparable to the normalized trace of the posterior covariance matrix, see, e.g. Houtekamer et al. (2005), Wang et al. (2007) and Raeder et al. (2012). Thus, it seems to happen that the EnKF ensemble is ‘good’, i.e. there is sufficient spread in the ensemble and MSE is small, while at the same time, the weights are ‘bad’, i.e. all but one weight are near zero. We now demonstrate that this situation indeed occurs: localized EnKFs with small ensemble sizes yield a small MSE while their particle filter interpretation collapses.

4.1. Scaling of MSE with dimension

The MSE is defined by

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n ((\bar{\mathbf{x}}_k)_i - (\mathbf{x}_k^t)_i)^2, \quad \bar{\mathbf{x}}_k = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{x}_k^j. \quad (22)$$

Here, \mathbf{x}_k^t is the ‘true’ state of the model at time k , $(\mathbf{x}_k^t)_i$, $i = 1, \dots, n$, are its n_x components and $\bar{\mathbf{x}}_k$ is the sample mean at time k , computed from the EnKF analysis ensemble members \mathbf{x}_k^j , $j = 1, \dots, N_e$. We obtain useful scaling equations under the same assumptions as above, i.e. we consider a Gaussian filtering density $\mathcal{N}(\mu_k, \mathbf{P}_k)$ and assume that the EnKF ensemble members are draws from $\mathcal{N}(\mu_k, (1 + \beta)\mathbf{P}_k)$, where $\beta > 0$. As indicated above, these assumptions simplify the problem, but describe a best-case scenario for the EnKF, and in particular model the localized EnKF (see Section 3.2). Indeed, the scalings below are not valid for unlocalized EnKF, unless N_e is sufficiently large, and, thus, larger than ever used in practice.

Under our assumptions, the statistics of the ensemble mean are

$$\bar{\mathbf{x}}_k \sim \mathcal{N}(\mu_k, \frac{1 + \beta}{N_e} \mathbf{P}_k).$$

If the true state is a sample of the filtering density, one can combine the above expression for the sample mean with the expression (22) for the MSE to obtain the statistics of the MSE:

$$\begin{aligned} \text{MSE} &\sim \mathcal{N}\left(\frac{1 + \beta + N_e}{N_e} \left(\frac{\text{tr}(\mathbf{P}_k)}{n_x}\right), \right. \\ &\quad \left. 2 \left(\frac{(1 + \beta + N_e)}{N_e} \frac{\|\mathbf{P}_k\|_F}{n_x}\right)^2\right). \end{aligned} \quad (23)$$

If we assume that the errors in the covariance matrix are due to sampling error, we can set $\beta = O(1/\sqrt{N_e})$, so that the mean of the normalized MSE is

$$n_x \cdot \frac{E(\text{MSE})}{\text{tr}(\mathbf{P}_k)} = 1 + O(N_e^{-1}) + O(N_e^{-3/2}). \quad (24)$$

Thus, even with small ensemble sizes, and even if the dimensions n_x and n_y are huge, one can have that the MSE is on the order of the ‘spread’, defined by the trace of the posterior covariance matrix \mathbf{P}_k divided by n_x . At the same time, the particle filter interpretation of the EnKF collapses for small, or even moderate ensemble sizes (see Section 3.2).

4.2. Local error and MSE vs. global error and collapse of the PF-EnKF

Note that MSE describes the error one can expect on average per dimension. Thus, MSE assesses the performance of the EnKF locally, rather than globally. This assessment of performance is useful in practice. For example, errors are usually examined locally, e.g. one might state that the data assimilation is ‘working poorly in the tropics, but well in mid-latitudes’. Similarly, an EnKF user routinely asserts that data assimilation has succeeded if average errors at stations around the globe are small.

However, when the performance of the PF-EnKF is assessed by the variance of its weights, then one does the opposite. The variance of the weights is due to differences between the posterior density and the proposal density generated by the EnKF ensemble. If the proposal and posterior densities are effectively high-dimensional (with large l_{sys} and τ_{sys} , see Sections 2.4 and 2.5), the differences, or errors, may be small in each dimension, corresponding to small λ_j , but the errors add up to a large overall error during the weight calculation. This leads to the collapse of the PF-EnKF, and happens similarly for filtering or smoothing densities.

We have shown that it can indeed occur that localized EnKFs with small ensemble sizes can produce a small MSE while at the same time their particle filter interpretation collapses. The success of the EnKF in meteorology, where MSE is the gold standard, thus suggests that assessing errors locally is sufficient for these problems. There is no need for keeping a ‘global’ variance of the weights small: the EnKF can work well even when this variance is huge, as is illustrated by the collapse of the particle filter interpretation of the EnKF. Indeed, this collapse may occur regularly and without significant practical consequences. However, it is important to note that the particle filter interpretation of the EnKF performs poorly no matter how we evaluate its performance – by MSE or the variance of the weights. As we will describe below, this problem is caused by the fact that the PF-EnKF is not using localization during the weight calculation, which is equivalent to its ‘update’, or analysis. More specifically, localization *is* used for generating the ensemble, but it is *not* used when calculating the weights (see Section 5).

4.3. Numerical example

We illustrate that the localized EnKF with small ensemble sizes N_e can yield a small MSE while at the same time its particle

filter interpretation collapses by the canonical linear example usually studied in the context of filter collapse (Snyder et al., 2008; Bickel et al., 2008; Bengtsson et al., 2008; Snyder, 2011; Chorin and Morzfeld, 2013; Snyder et al., 2015). The example is defined by

$$\mathbf{M} = \mathbf{I}, \mathbf{H} = \mathbf{I}, \mathbf{Q} = \mathbf{I}, \mathbf{R} = \mathbf{I}, \quad (25)$$

and all effective dimensions reviewed in Section 2.4 increase with dimension $n = n_x = n_y$. Thus, for large dimension n , any particle filter will collapse unless the ensemble size is huge. However, the localized EnKF with small ensemble sizes yields a small MSE even when its particle filter interpretation collapses. Here, we demonstrate this scenario by computing the quality measure G in (21) for various n , which is equivalent to computing the ensemble size required to avoid collapse. Specifically, if G depends exponentially on dimension n , then the required ensemble size to avoid collapse also depends exponentially on dimension. Since all effective dimensions increase with dimension n , the scaling of G also illustrates how the required ensemble size scales with effective dimensions.

Computing G for one time step, from $k - 1$ to k , requires that one knows the density of the state at $k - 1$. We assume that this density is Gaussian with mean $\mu_{k-1} = \mathbf{0}$ and covariance $\mathbf{P}_{k-1} = \mathbf{I}$. We thus have

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{x}_{k-1}^j) &= \mathcal{N}(\mathbf{x}_k^j | \mathbf{I}), & p(\mathbf{y}_k | \mathbf{x}_k^j) &= \mathcal{N}(\mathbf{x}_k^j | \mathbf{I}), \\ p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) &= \mathcal{N}(\mathbf{0} | \mathbf{I}). \end{aligned}$$

We consider the following algorithms:

- (1) an unlocalized/uninflated EnKF and its PF interpretation for the filtering density, $N_e = 50$,
- (2) a localized/inflated EnKF and its PF interpretation for the filtering density, $N_e = 50$,
- (3) a localized/inflated EnKF and its PF interpretation for the filtering density, $N_e = n$,
- (4) a localized/inflated EnKF and its PF interpretation for the smoothing density, $N_e = 50$,
- (5) an optimal PF for the smoothing density with $N_e = 50$.

For each algorithm, we approximate G in (21) from an ensemble of size $N_e = 10^6$ and repeat the calculation 100 times to average out errors in our approximations of G . We do this for problems with dimensions between $n = 5$ and $n = 500$.

4.3.1. Implementation, localization and inflation of the EnKF. We localize the forecast covariance by the identity matrix, i.e. we set all off-diagonal elements of the EnKF forecast covariance matrix to zero. Note that this is a ‘perfect’ localization for this diagonal example and therefore we are depicting a best-case

scenario for the localized EnKF and its particle filter interpretation. Prior inflation is inversely proportional to the ensemble size, i.e. we multiply the forecast covariance matrix by $(1 + 1/N_e)$. However, we tested our results with and without inflation and found that, for this example, the impact of inflation is insignificant compared with localization.

The weights of the PF-EnKF for the smoothing density are computed by (15). Here, we assume that $\mathbf{x}_{k-1} \sim p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$, i.e. the PF-EnKF starts off with samples from the correct density, so that $w^{k-1} = 1/N_e$. The weights of the PF-EnKF for the filtering density are computed by (16). The proposal density is given by (17), and its covariance is localized by the identity matrix as before. The filtering density is Gaussian, $p(\mathbf{x}_k | \mathbf{y}_{1:k}) \sim \mathcal{N}(\mu_k, \mathbf{P}_k)$, and its mean and covariance are computed by the Kalman filter formulas.

4.3.2. Results. The left panels of Fig. 1 show the values of G for the algorithms for each experiment (dots) along with the log-linear least-squares approximations and one standard deviation confidence intervals. The right panels show the scaling of MSE with dimension, specifically the mean and two standard deviation error bars of the MSE based on the 100 experiments we perform. We observe the expected exponential scaling of G for the particle filter interpretation of the localized EnKF for the smoothing density. The optimal PF is also characterized by an exponential scaling of G , however slower the rate (see also Snyder et al. (2015)). The particle filter interpretations of localized or unlocalized EnKFs for the filtering density are also characterized by an exponentially increasing G if the ensemble size is fixed at $N_e = 50$. However, the collapse of the particle filter interpretation of the localized EnKF for the filtering density can be avoided if the ensemble size is proportional to dimension, as described by our theoretical considerations above. Specifically, we observe that the PF-EnKF for the filtering density produces the same G independently of n when $N_e = n$.

In the right panels of Fig. 1, we observe that the localized EnKFs give a small MSE while the unlocalized EnKF and the optimal PF cause a large MSE, which is consistent with what is reported in practice in large-dimensional problems. Specifically, the localized EnKF with $N_e = 50$ gives a small MSE even when its particle filter interpretation collapses. Also shown is the mean and variance of the MSE we obtain without sampling error, i.e. by an optimal algorithm. The localized EnKFs and the optimal sampling algorithm give almost identical MSE. In summary, our numerical experiments confirm our scaling of MSE with dimension and illustrate that a localized EnKF with a small ensemble size indeed can yield a small MSE, while at the same time its particle filter interpretation collapses.

5. The importance of localization for particle filters

We argued in Section 2.5 that the collapse of particle filters in meteorology is caused by extremely large state dimensions

n_x and very large numbers of observations n_y , which in turn imply large effective dimensions. On the other hand, we expect that meteorological problems are characterized by correlation lengths that are short compared to the overall domain size, and that the observation operator \mathbf{H} is ‘local’ (see Section 2.5). These assumptions imply that forecast and posterior covariances are sparse, but the PF-EnKF does not utilize this sparsity during the weight calculation, but it *does* use it when generating the ensemble. If one could make use of the sparsity of the data assimilation problem during weight calculation, then one may be able to avoid this type of filter collapse of the PF-EnKF, even for small ensemble sizes. The situation is perhaps analogous to numerical linear algebra, where computations with large dense matrices are expensive, but large sparse matrices can be dealt with efficiently. We now explain in more detail that ‘localization for particle filters’ means exactly that: making use of sparsity of forecast and posterior covariances. We start by discussing a diagonal problem, for which it is straightforward to define what ‘localization for particle filters’ is, and then move onto discuss more realistic cases.

5.1. Localizing particle filters for diagonal problems

In the diagonal example of Section 4.3, exploiting sparsity of covariance matrices means to make use of the fact that the system is diagonal. Since a diagonal problem uncouples into independent scalar sub-problems, one can apply scalar PF-EnKFs to each sub-problem and also compute the weights for each sub-problem independently. The ensemble size required for each scalar sub-problem is independent of dimension, so that the catastrophic scaling of the required ensemble size of the PF-EnKF with dimension disappears. In fact, this strategy of decoupling the problem applies to any particle filter (for smoothing or filtering densities) so that the scaling of the required ensemble size with dimension disappears for every particle filter (see also Rebescini and van Handel, 2015).

We can connect the above localization of particle filters by decoupling to covariance localization in EnKF as follows. If the EnKF is localized by the identity matrix, which is optimal for this diagonal problem, then covariance localization is equivalent to applying n scalar EnKFs to each sub-problem. Thus, for diagonal examples, there is a one-to-one analogy between covariance localization of EnKF, and a localization of the particle filter interpretation of EnKF by decoupling of the problem.

5.2. Localizing particle filters for coupled problems

While a localization of particle filters by decoupling diagonal problems is rather intuitive, the situation is less clear in ‘real’ meteorological problems, which are not diagonal. In this context, it is important to realize that making the weights vary over space is not enough to define a localization strategy. The weights must be local and, in addition, must be used in a way to modify

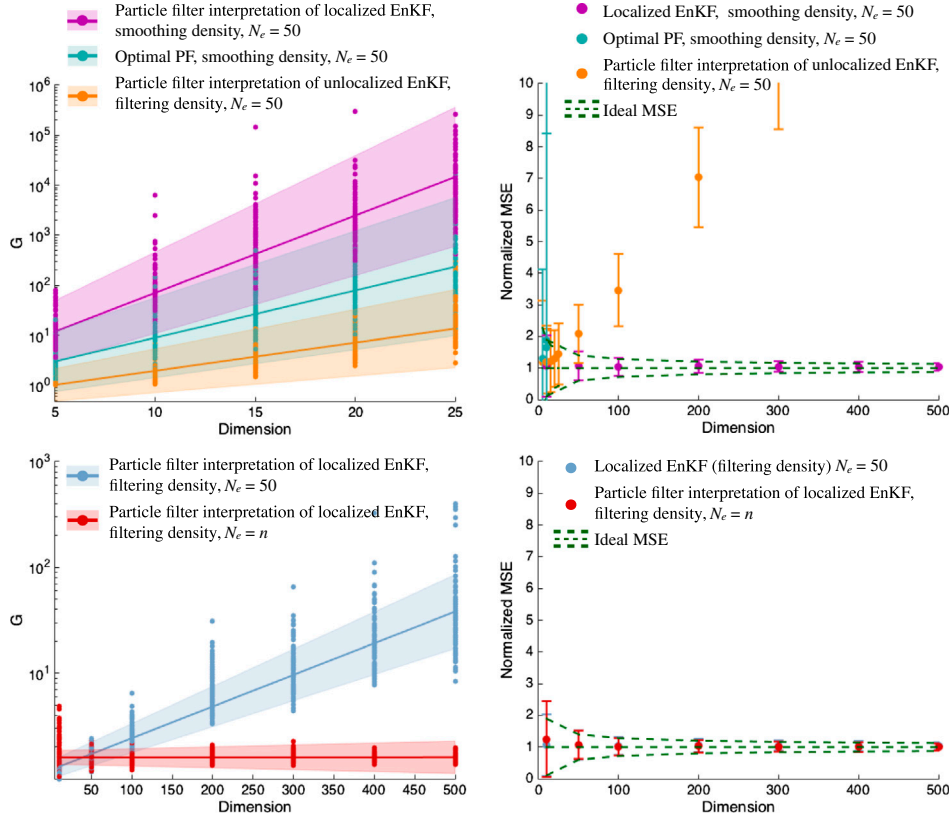


Fig. 1. Left: quality measure G as a function of the dimension for the various methods. Shown are the values of G for each experiment (dots) along with log-linear least-squares fits (straight lines) and one standard deviation confidence intervals (shaded regions). Right: normalized MSE as a function of the dimension for the various methods. Shown are the mean (dots) and confidence intervals (error bars) computed from 100 numerical experiments. Also shown (dashed lines) is the mean and two standard deviation confidence intervals of the ideal MSE (dashed green).

the ensemble locally. In the diagonal example, this is done by resampling each variable separately. In more general situations, this step may not be as straightforward and introduces bias, as highlighted in the mathematically rigorous paper by [Rebeschini and van Handel \(2015\)](#). However, the same is true for covariance localization of EnKF. We argued that covariance localization of the EnKF and localization of particle filters by (approximate) decoupling require the same conditions and we argued that these conditions are satisfied in meteorological problems (see Section 2.5). The impressive performance of localized EnKFs in meteorology suggests that the bias introduced by the covariance localization is small. Thus, if particle filters can be localized by making use of the same properties, then the bias introduced by localization of particle filters can also be expected to be small.

Indeed, several localization strategies for particle filters have been published recently ([Reich, 2013](#); [Poterjoy, 2015](#); [Poterjoy and Anderson, 2016](#); [Penny and Miyoshi, 2015](#); [Rebeschini and van Handel, 2015](#); [Tödter and Ahrens, 2015](#)). Earlier work includes the study by [Bengtsson et al. \(2003\)](#), where localization of mixture filters is considered, as well as the paper by [Lei and Bickel \(2011\)](#). All of these strategies yield good results in

numerical simulations and in one way or another exploit sparsity of the data assimilation problem by restricting the effects of an observation to its neighbourhood. Our study supports these efforts and connects localization of particle filters to real meteorological problems. Specifically, we draw attention to the fact that the successful performance of the localized EnKF in numerical weather prediction in conjunction with the collapse of its particle filter interpretation on the same problem indicates that localization schemes for particle filters may prevent particle filter collapse in meteorology at the cost of a small bias.

6. Discussion

Localizing particle filters by decoupling as described in the context of a diagonal problem applies similarly to block diagonal problems – one may apply independent particle filters to each independent block, which removes any dependence of the performance of the particle filter on dimension or the number of the blocks considered. A meteorological problem is unlikely to be block diagonal, but it may be similar to a large number of loosely coupled sub-problems (see also Section 2). This makes the local-

ization of particle filters less straightforward, as described above, but one can anticipate that localized particle filters can only be successful if the effective dimensions of each sub-problem are small enough to prevent the collapse of particle filters. Our study does not provide any estimate for the effective dimension of each sub-problem, or the number of sub-problems or the degree to which the various blocks are ‘coupled’. Thus, we cannot provide ‘proof’ that localization will indeed make particle filters effective in meteorology. Our study merely indicates that sparsity of the data assimilation problem can be exploited by particle filters and that such strategies may prevent filter collapse in meteorology, provided that each sub-problem is manageable.

There are more analogies between covariance localization of the EnKF and localizing particle filters by (approximate) decoupling. Recall that covariance localization of the EnKF reduces or deletes long-range correlations by Schur products (see Section 2.2). The localization thus restricts the influence of an observation to its neighbourhood. The ‘localized’ weight calculation we describe for diagonal problems does precisely that. Moreover, unlocalized EnKFs with small ensemble sizes fail in high-dimensional problems, and localization makes them effective. Similarly, particle filters collapse in high-dimensional problems if the ensemble size is small; however, a localization by decoupling may make them effective.

Finally, a localization of particle filters is only useful if local errors such as MSE are of greater importance than the global errors measured by the variance of the weights. We showed in Section 4 that the MSE of a (localized) EnKF with a small ensemble can be small even when global errors, measured by the weights of the PF-EnKF, are huge. The success of EnKFs in meteorology suggests that it is sufficient to evaluate errors locally.

6.1. Extensions of our results

Our results concerning the PF-EnKF for the filtering density in Section 3.2, as well as the results concerning the MSE in Section 4 generalize to data assimilation problems with a ‘perfect’ model for which $\mathbf{Q} = \mathbf{0}$. However, the results concerning the smoothing density cannot be carried over to data assimilation problems with deterministic models by setting $\mathbf{Q} = \mathbf{0}$. The reason is that the smoothing density $p(\mathbf{x}_{1:k}|\mathbf{y}_{1:k})$ no longer exists. In the limit $\mathbf{Q} \rightarrow \mathbf{0}$, one can consider either the filtering density $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, or the density $p(\mathbf{x}_0|\mathbf{y}_{1:k})$. Using the formulas and setting $\mathbf{Q} = \mathbf{0}$, one considers methods for the latter density, i.e. the density that describes initial conditions conditioned on data collected at later times. The formulas for the optimal PF with $\mathbf{Q} = \mathbf{0}$ lead to a sampling algorithm for $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, but this algorithm is not optimal. Other optimal algorithms can be obtained, as linear examples show.

Throughout our paper, we assume that the data assimilation problem is linear and Gaussian. In NWP practice, this is not the case. In fact, EnKF can struggle in non-linear problems

(see, e.g. Hodyss, 2011; Hodyss and Campbell, 2013) and in these situations computing weights and using the particle filter interpretation of the EnKF may be particularly helpful. While our results do not directly extend to non-linear problems, the results we collect for linear problems indicate that weight calculations for the EnKF must be localized for non-linear problems as well or else the weights may not be useful in meteorology.

6.2. EnKF stability in time

Work towards a stability theory for the EnKF is partially done (Kelly et al., 2014; Kelly et al., 2016a) and more is underway ((Kelly et al., 2016b); Kelly et al., 2016c; Majda and Tong, 2016). The goal of these papers is to understand the EnKF and its approximations better by studying stability of its estimates in time. For example, Kelly et al. (2014) interpret the EnKF as a method for approximating the mean of the filtering density, and the expected distance of an ensemble member to ‘the truth’ is analysed over time. It is shown that this distance is bounded as time evolves and can become small, for large t , if the forecast covariances are inflated. We took a different perspective and address scaling of the computational requirements of the particle filter interpretation of the EnKF with dimension. We considered this scaling during a single time step, as is customary in the particle filter literature (Snyder et al., 2008; Bickel et al., 2008; Bengtsson et al., 2008; Snyder, 2011; Chorin and Morzfeld, 2013; Snyder et al., 2015). We thus neglected accumulation of variance of the weights over time due to the sequential approach, and our study complements recent papers about stability of the EnKF, as well as earlier results by Moral (2004), Chopin (2004) and Künsch (2005) that address accumulation of variances over time.

7. Summary

We summarize the main results we have collected and derived.

- (1) We reviewed and extended the literature about the collapse of particle filters which is caused by a large ‘effective dimension’. There are several mechanisms that can cause a large effective dimension, and, thus, lead to particle filter collapse. We describe how particle filter collapse occurs in meteorology due to huge numbers of state variables and observations, and a ‘local’ structure of the system (see Section 2). Moreover, while there are several definitions of effective dimension in the literature, we argue that every effective dimension can be expected to be large in meteorology.
- (2) We made explicit connections between the collapse of particle filters and the EnKF. Our study and our assumptions are specific to the *localized* EnKF and are violated by the EnKF without localization. Specifically, we showed that the particle filter interpretation of the localized EnKF for the filtering density collapses unless the

ensemble size increases linearly with dimension. This means in particular that using a localized EnKF as the proposal density does not prevent particle filter collapse in meteorology where dimension is large.

- (3) We have demonstrated that a localized EnKF with a small ensemble size can produce a small MSE even while its particle filter interpretation collapses. The MSE is small because it is a local measure of error, and the covariance localization of the EnKF can keep local errors small. The collapse of the particle filter interpretation of the EnKF on the other hand is caused by a global weight calculation, during which small errors at various locations add up, rather than average out. We argued that the success of the EnKF indicates that a ‘global’ assessment of error may not be required in meteorology and that assessing errors locally is sufficient.
- (4) We have described analogies between covariance localization of EnKF and ‘localization’ of particle filters. We first consider a diagonal problem and show that localization of a particle filter by decoupling the problem independent sub-problems is the equivalent to localizing the EnKF by a diagonal matrix. Relaxing the assumption of diagonality provides hints about how localization for particle filters may be implemented in more realistic, coupled problems. In particular, covariance localization for the EnKF and a localization of particle filters rely on the same assumptions of short correlation lengths within a large system domain, and local assessment of error. The impressive performance of the localized EnKF in meteorological problems indicates that these assumptions are indeed valid in meteorology, which implies that localization of particle filters may be feasible in meteorology without introducing large bias. Our collection of results thus corroborates the applicability of recently proposed localization schemes for particle filters.

Acknowledgements

We thank Prof. Alexandre J. Chorin of UC Berkeley and Berkeley National Laboratory for interesting discussion and encouragement.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

MM was supported by the Office of Naval Research N00173-17-2-C003 [grant number N00173-17-2-C003]; by an Alfred P. Sloan Research Fellowship, and by the National Science Foundation [grant number DMS-1619630]. DH gratefully acknowledges support from the Office of Naval Research [PE-0601153N].

The work upon which this material is based upon was started while MM was a postdoc, supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program [contract number DE-AC02005CH11231]; and by the National Science Foundation [grant number DMS-1419044].

References

- Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D. and Stuart, A. (2016). Importance sampling: Computational complexity and intrinsic dimension. submitted, pre-print available on <https://arxiv.org>.
- Anderson, J. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903.
- Arulampalam, M., Maskell, S., Gordon, N. and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50**(2), 174–188.
- Bengtsson, T., Bickel, P. and Li, B. (2008). Curse of dimensionality revisited: the collapse of importance sampling in very large scale systems., *IMS Collections: Probability and Statistics: Essays in Honor of David A. Freedman* **2**, 316–334.
- Bengtsson, T., Snyder, C. and Nychka, D. (2003). Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.* **108**, 8775.
- Beskos, A., Crisan, D. and Jasra, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* **24**(4), 1396–1445.
- Bickel, P., Li, B. and Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions., *Pushing Limits Contemp. Stat.: Contrib. Honor Jayanta K. Ghosh* **3**, 318–329.
- Bishop, C., Etherton, B. and Majumdar, S. (2001). Adaptive sampling with the ensemble transform Kalman filter. part I: Theoretical aspects. *Monthly Weather Review* **129**, 420–436.
- Bocquet, M., Pires, C. and Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Weather Rev.* **138**, 2997–3023.
- Burgers, G., van Leeuwen, P. J. and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**, 1719–1724.
- Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statistics* **32**(6), 2385–2411.
- Chorin, A. and Hald, O. (2013). *Stochastic Tools in Mathematics and Science*, 3rd ed., Springer, New York.
- Chorin, A., Lu, F., Miller, R., Morzfeld, M. and Tu, X. (2016). Sampling, feasibility, and priors in Bayesian estimation, *Discrete Continuous Dyn. Syst.* **36**, 4227–4246.
- Chorin, A. and Morzfeld, M. (2013). Conditions for successful data assimilation. *J. Geophys. Res. – Atmos.* **118**, 11522–11533.
- Chorin, A., Morzfeld, M. and Tu, X. (2010). Implicit particle filters for data assimilation. *Commun. Appl. Math. Comput. Sci.* **5**(2), 221–240.
- Chorin, A. and Tu, X. (2009). Implicit sampling for particle filters. *Proc. Nat. Acad. Sci.* **106**(41), 17249–17254.
- Doucet, A. (1998). On sequential Monte Carlo methods for Bayesian filtering. *Technical Report*. Cambridge: Department of Engineering, University Cambridge.
- Doucet, A., de Freitas, N. and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.

- Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**, 197–208.
- Evensen, G. (2006). *Data Assimilation: The Ensemble Kalman Filter*. Springer, New York.
- Fournier, A., Hulot, G., Jault, D., Kuang, W., Tangborn, W., co-authors. (2010). An introduction to data assimilation and predictability in geomagnetism, *Space Sci. Rev.* **155**, 247–291.
- Gaspari, G. and Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**, 723–757.
- Gordon, N., Salmond, D. and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proc. Radar Signal Process.* **140**(2), 107–113.
- Hamill, T. M., Whitaker, J. and Snyder, C. (2001). Distance-dependent filtering of background covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**, 2776–2790.
- Hodyss, D. (2011). Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation. *Mon. Weather Rev.* **139**, 3571–3588.
- Hodyss, D. and Campbell, W. (2013). Square root and perturbed observation ensemble generation techniques in Kalman and quadratic ensemble filtering algorithms. *Mon. Weather Rev.* **141**, 2561–2573.
- Houtekamer, P. and Mitchell, H. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**, 123–136.
- Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., co-authors. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Weather Rev.* **133**, 604–620.
- Kalman, R. (1960). A new approach to linear filtering and prediction theory. *Trans. ASME-J. Basic Eng.* **82**(Series D), 35–48.
- Kalman, R. and Bucy, R. (1961). New results in linear filtering and prediction theory. *ASME J. Basic Eng. Ser. D* **83**, 95–108.
- Kalos, M. and Whitlock, P. (1986). *Monte Carlo Methods*, 1st ed., Vol. 1. John Wiley & Sons, New York.
- Kantas, N., Beskos, A. and Jasra, A. (2014). Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier-Stokes equations, *SIAM/ASA. J. Uncertainty Quantification* **2**, 464–489.
- Kelly, D., Law, K. and Stuart, A. (2014). Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity* **27**, 2579–2603.
- Kelly, D., Majda, A. and Tong, X. (2016a). Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proc. Nat. Acad. Sci.* **112**(34), 10589–10594.
- Kelly, D., Majda, A. and Tong, X. (2016b). Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *Commun. Math. Sci.*
- Kelly, D., Majda, A. and Tong, X. (2016c). Nonlinear stability and ergodicity of ensemble based kalman filters. *Nonlinearity* **29**(2), 657–691.
- Khlalil, M., Sarkar, A., Adhikari, S. and Poirel, D. (2015). The estimation of time-invariant parameters of noisy nonlinear oscillatory systems. *J. Sound Vib.* **344**, 81–100.
- Klaas, M., de Freitas, N. and Doucet, A. (2005). *Towards practical N^2 Monte Carlo: The marginal particle filter*, Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI-05), Arlington, VA. pp. 308–315.
- Künsch, H. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statistics* **33**(5), 1983–2021.
- Lei, J. and Bickel, P. (2011). A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Weather Rev.* **139**, 3964–3973.
- Liu, J. and Chen, R. (1995). Blind deconvolution via sequential imputations. *J. Am. Stat. Assoc.* **90**(430), 567–576.
- Majda, A. & Tong, X. (2016). Robustness and accuracy of finite ensemble Kalman filters in large dimensions. arXiv, p. 1606.09321.
- Moral, P. D. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.
- Morzfeld, M. and Chorin, A. (2012). Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation, *Nonlinear Process. Geophys.* **19**, 365–382.
- Morzfeld, M., Tu, X., Atkins, E. and Chorin, A. (2012). A random map implementation of implicit filters. *J. Comput. Phys.* **231**, 2049–2066.
- Owen, A. B. (2013). *Monte Carlo Theory, Methods and Examples*. Online at: <http://statweb.stanford.edu/~owen/mc/>
- Papadakis, N., Memin, E., Cuzol, A. and Gengembre, N. (2010). Data assimilation with the weighted ensemble Kalman filter. *Tellus* **62A**, 673–697.
- Penny, S. and Miyoshi, T. (2015). A local particle filter for high dimensional geophysical systems, *Nonlinear Process. Geophys.* **2**, 1631–1658.
- Poterjoy, J. (2015). A localized particle filter for high-dimensional nonlinear systems. *Mon. Weather Rev.* **144**, 59–76.
- Poterjoy, J. and Anderson, J. (2016). Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Weather Rev.* **144**, 2007–2020.
- Raeder, K., Anderson, J. L., Collins, N., Hoar, T. J., Kay, J. E. and co-authors. (2012). DART/CAM: An ensemble data assimilation system for CESM atmospheric models. *J. Clim.* **25**, 6304–6317.
- Rebeschini, P. and van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.* **25**(5), 2809–2866.
- Reich, S. (2013). A nonparametric ensemble transform method for Bayesian inference. *Mon. Weather Rev.* **35**(4), 1337–1367.
- Snyder, C. (2011). Particle filters, the "optimal" proposal and high-dimensional systems. *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, Reading, UK, 1–10.
- Snyder, C., Bengtsson, T., Bickel, P. and Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**(12), 4629–4640.
- Snyder, C., Bengtsson, T. and Morzfeld, M. (2015). *Performance bounds for particle filters using the optimal proposal*. *Mon. Weather Rev.* **143**(11), 4750–4761.
- Tippett, M., Anderson, J., Bishop, C., Hamill, T. and Whitaker, J. (2003). Ensemble square root filters. *Mon. Weather Rev.* **131**, 1485–1490.
- Tödter, J. and Ahrens, B. (2015). A second-order exact ensemble square root filter for nonlinear data assimilation. *Mon. Weather Rev.* **143**, 1337–1367.
- van Leeuwen, P. (2009). Particle filtering in geophysical systems. *Mon. Weather Rev.* **137**, 4089–4144.
- van Leeuwen, P., Cheng, Y. and Reich, S. (2015). *Nonlinear Data Assimilation*. Springer.

- Vanden-Eijnden, E. and Weare, J. (2012). Data assimilation in the low noise regime with application to the Kuroshio. *Mon. Weather Rev.* **141**, 1822–1841.
- Wang, X., Hamill, T. M., Whitaker, J. S. and Bishop, C. H. (2007). A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Mon. Weather Rev.* **135**, 1055–1076.
- Weare, J. (2009). Particle filtering with path sampling and an application to a bimodal ocean current model. *J. Comput. Phys.* **228**, 4312–4331.
- Zaritskii, V. and Shimelevich, L. (1975). Monte Carlo technique in problems of optimal data processing. *Autom Remote Control* **12**, 95–103.

Appendix 1. Weights for the particle filter interpretation of the EnKF for the filtering density of non-linear problems

The weights of the PF-EnKF when sampling the filtering density are given in (16) and can be re-written as

$$w \propto \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1})}{\pi_{\text{EnKF}}(\mathbf{x}_k)}.$$

Evaluation of these weights requires integration since

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}, \quad (\text{A1})$$

and evaluating this integral is difficult unless the problem is linear and Gaussian (see above). Similarly, evaluating the density of the analysis ensemble at time k also requires integration since

$$\pi_{\text{EnKF}}(\mathbf{x}_k) = \int \pi_{k,\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}, \quad (\text{A2})$$

where $\pi_{k,\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k)$ is not Gaussian. However, a basic requirement for weighted sampling (and many other sampling schemes, e.g. Markov Chain Monte Carlo) is that the target and proposal densities be known up to a multiplicative constant. This requirement is not met here so that exact weights for the PF-EnKF when sampling the filtering density are out of reach.

Using ideas of Klaas et al. (2005) for marginal particle filters, one can develop approximate weights. These approximations rely on Monte Carlo for evaluating the integral (A1) and become exact as the ensemble size N_e goes to infinity. We demonstrate here how to use these ideas. Let \mathbf{x}_{k-1}^j , $j = 1, \dots, N_e$, be the analysis ensemble at time $k-1$, which defines an approximation to the filtering density at time $k-1$:

$$p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \approx \frac{1}{N_e} \sum_{j=1}^{N_e} \delta(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^j), \quad (\text{A3})$$

which can be used in the integral (A1) to obtain

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \approx \frac{1}{N_e} \sum_{j=1}^{N_e} p(\mathbf{x}_k | \mathbf{x}_{k-1}^j).$$

The target density can thus be approximated by

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \approx p(\mathbf{y}_k | \mathbf{x}_k) \left(\frac{1}{N_e} \sum_{j=1}^{N_e} p(\mathbf{x}_k | \mathbf{x}_{k-1}^j) \right). \quad (\text{A4})$$

Similarly, we obtain by Monte Carlo approximation of the integral (A2)

$$\pi_{\text{EnKF}}(\mathbf{x}_k) \approx \frac{1}{N_e} \sum_{j=1}^{N_e} \pi_{\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}^j, \mathbf{y}_k).$$

With these approximations, the PF-EnKF weights with respect to the marginal posterior are

$$w_{\text{EnKF}} \approx \hat{w}_{\text{EnKF}} \propto \frac{p(\mathbf{y}_k | \mathbf{x}_k) \sum_{j=1}^{N_e} p(\mathbf{x}_k | \mathbf{x}_{k-1}^j)}{\sum_{j=1}^{N_e} \pi_{\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}^j, \mathbf{y}_k)}. \quad (\text{A5})$$

However, the approximate weights are difficult to compute unless the model's time stepper is in sync with the observations. The difficulties can be illustrated by considering a problem with one additional model step between two consecutive observations. In this case, the proposal density can be written as

$$\pi_{\text{EnKF}}(\mathbf{x}_k) = \int \int \pi_{k,\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) p(\mathbf{x}_{k-1} | \mathbf{x}_{k-2}) \times p(\mathbf{x}_{k-2} | \mathbf{y}_{1:k-2}) d\mathbf{x}_{k-1} d\mathbf{x}_{k-2}.$$

We can use the analysis ensemble at time $k-2$ to approximate $p(\mathbf{x}_{k-2} | \mathbf{y}_{1:k-2})$ as before, but we then must perform the integration over \mathbf{x}_{k-1} (assuming there is no observation available at that time):

$$\pi_{\text{EnKF}}(\mathbf{x}_k) \approx \frac{1}{N_e} \sum_{j=1}^{N_e} \int \pi_{k,\text{EnKF}}(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) \times p(\mathbf{x}_{k-1} | \mathbf{x}_{k-2}^j) d\mathbf{x}_{k-1}.$$

This integration is difficult in general, which complicates the computation of the weights for the PF-EnKF for the filtering density. In fact, these weights may be impractical because the observations are usually not frequently available (at least not at every step of the model). These results can be extended to the case of deterministic models by putting $p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \delta(\mathbf{x}_k - \mathbf{x}_{k-1})$, where $\delta(\cdot)$ is the delta distribution.