Quarterly Journal of the
Royal Meteorological Society RMetS

Estimating model error covariances using particle filters

Mengbin Zhu[1] | Peter J. van Leeuwen[2,3] | Weimin Zhang[1]

[1]Academy of Ocean Science and Engineering, National University of Defense Technology, Changsha, China

[2]Department of Meteorology, University of Reading, Reading, UK

[3]National Centre for Earth Observation, University of Reading, Reading, UK

Correspondence
Peter J. van Leeuwen, Department of Meteorology, University of Reading, Reading RG6 6BB, UK.
Email: p.j.vanleeuwen@reading.ac.uk

A method is presented for estimating the error covariance of the errors in the model equations in observation space. Estimating model errors in this systematic way opens up the possibility to use data assimilation for systematic model improvement at the level of the model equations, which would be a huge step forward. This model error covariance is perhaps the hardest covariance matrix to estimate. It represents how the missing physics and errors in parametrizations manifest themselves at the scales the model can resolve.

A new element is that we use an efficient particle filter to avoid the need to estimate the error covariance of the state as well, which most other data assimilation methods do require. Starting from a reasonable first estimate, the method generates new estimates iteratively during the data assimilation run, and the method is shown to converge to the correct model error matrix. We also investigate the influence of the accuracy of the observation error covariance on the estimation of the model error covariance and show that, when the observation errors are known, the model error covariance can be estimated well, but, as expected and perhaps unavoidably, the diagonal elements are estimated too low when the observation errors are estimated too high, and vice versa.

KEYWORDS

model error covariance, non-degeneracy, particle filter; localization

1 | INTRODUCTION

Linear and linearized data assimilation (DA) rely on prescribing or accurately estimating covariance matrices, related to the (near-)Gaussian assumptions on the probability densities in Bayes Theorem. For example, standard 3D- and 4D-Var stand or fall with an accurately prescribed state covariance **B**, and an accurately prescribed observation error covariance **R**. Especially for the former, numerous person-years are typically needed to tune the prior state covariance matrix. Ensemble Kalman Filters and Smoothers need an accurate **R**, and their prior covariance at each observation time is the sample covariance from the ensemble (Evensen, 1994; Evensen and van Leeuwen, 2000). Because of the small ensemble size that can be afforded in numerical weather prediction, that matrix is rank deficient, underestimates the covariances, and has large Monte Carlo errors, resulting in spurious long-range correlations. *Ad hoc* methods like inflation and localization are

needed to generate useful prior covariances from the sample covariance matrix (Anderson, 2007; 2009). (Of course, for high-dimensional systems the covariance is never calculated explicitly, but the issues and partial solutions mentioned above are still relevant.) Several issues arise when using inflation and localization, for instance how to localize for an observation that is an integral along a line. In so-called hybrid methods that combine variational and Ensemble Kalman Filter methods, the estimation problems do not disappear, and in iterative ensemble smoothers like 4DEnsVar one has to estimate space–time covariances from an ensemble, leading to further problems with localization in time (Liu *et al.*, 2008; Buehner *et al.*, 2010).

Several methods to estimate state covariances have been developed, and Bannister (2008a; 2008b) gives a good overview of what has been done. The so-called Desroziers diagnostic (Desroziers *et al.*, 2005) has become very popular because of its relative simplicity. It uses the statistics

of innovations and analysis minus observations to estimate either the prior covariance $\mathbf{B}$ in observation space, or the observation error covariance $\mathbf{R}$, or sometimes both, exploring relations from linear estimation theory. Estimation of $\mathbf{R}$ is very popular nowadays because it has become clear that correlations between observation errors need to be taken into account to extract most information from the observations and to obtain the best analysis (Stewart *et al.*, 2013; Weston *et al.*, 2014). Another issue is that the error covariances $\mathbf{B}$ and $\mathbf{R}$ should have different characteristics, such as length-scales, to be able to estimate both matrices together.

Recently the Desroziers method has also been used to estimate the model error covariance $\mathbf{Q}$. Todling (2015a; 2015b) gives an alternative method to diagnose model error, which requires two overlapping DA systems, one sequential filter and one fixed lag-1 smoother. The model error diagnosed by Todling (2015a; 2015b) comes from the differences between two model states. Todling stresses the problem that estimating $\mathbf{Q}$ while $\mathbf{R}$ and $\mathbf{B}$ are not known precisely can lead to attributing structures to $\mathbf{Q}$ that belong in $\mathbf{R}$ and especially $\mathbf{B}$, and vice versa.

A major advantage of particle filters is that the prior state error covariance does not play a role, so $\mathbf{B}$ does not have to be prescribed or estimated. So, under the perfect model assumption, the only matrix to be estimated is $\mathbf{R}$. No studies have been performed in this direction to our knowledge. One of the reasons is that standard particle filters are degenerate when the number of observations is large, meaning that the weight that each particle obtains during the assimilation varies enormously over the particles, with one particle having a much larger weight than all the others. This then means that the ensemble effectively has collapsed onto that one particle.

Three solutions have been presented in the literature: localized particle filters, like the Local Ensemble Transform Particle Filter by Reich (2013), and the localized particle filter by Poterjoy (2016), methods that combine Particle Filters with Ensemble Kalman Filters (Lei and Bickel, 2011; Frei and Künsch, 2013), and methods that explore the proposal density freedom (Chorin and Tu, 2009; van Leeuwen, 2010; Zhu *et al.*, 2016). Reich (2013) estimates a local transformation matrix via optimal transportation to transform the weighted prior ensemble into an equal-weighted posterior ensemble, while Poterjoy (2016) employs a very delicate smoothing procedure to ensure smooth resampled particles. A problem with particle filters that use localization is that the localization radius has to be taken very small to avoid this weight collapse, as the variance of the weights scales with the number of independent observations within the localization area. The Poterjoy (2016) scheme avoids this issue by setting a minimal weight for each particle, and this issue has not been discussed yet for the other schemes. Furthermore, issues with localization as mentioned above for ensemble Kalman filters play a role here too, and also inflation has to be applied. The second method needs to estimate sample covariance matrices for the EnKF part of the algorithm, again with the same issues

as above. The third method needs model errors to be able to move the particles in state space different from the deterministic model equations. Hence for particle filters practical for high-dimensional geophysical problems, we need localization or we need to estimate $\mathbf{Q}$.

Interestingly, there is a huge advantage having to estimate $\mathbf{Q}$ over $\mathbf{B}$. $\mathbf{B}$ is just a statistical quantity that tells us something about the accuracy of a best estimate, while $\mathbf{Q}$ contains information about how and where our model is wrong. So estimating $\mathbf{Q}$ gives us a direct route to improving our model, one of the holy grails of DA, in which very little progress has been made in the last 30 years.

This article is organized as follows. The new scheme for estimating the model error covariance is described in section 2; other variants are described in Appendix 4, and details on convergence are presented in Appendix 4. The method is tested on the 1000-dimensional Lorenz-96 model and discussed further in section 3. A summary and conclusions are included in section 4.

2 | ESTIMATING THE MODEL ERROR COVARIANCE

In the Desroziers diagnostic, one calculates the innovation and/or the analysis minus observation statistics, or the statistics of the combination of these two. In a particle filter there is no unique best forecast or analysis. One could choose the ensemble mean, but that is not always the natural choice, for instance the probability density functions (pdfs) can be bimodal with modes of similar magnitude. So, instead, it is natural to explore the innovations of all particles. The model equations are denoted as

$$\mathbf{x}^n = f(\mathbf{x}^{n-1}) + \beta^n, \tag{1}$$

in which the superscript denotes the time index, $f(..)$ denotes the nonlinear deterministic model, and β^n denotes the stochastic term representing the missing physics in the model, taken as additive noise drawn from $N(0, \mathbf{Q})$.

To simplify the notation we introduce $f_i = f(x_i^{n-1})$ with i the particle index, $\bar{f}$ is the ensemble mean of the f_i, and $f_t = f(\mathbf{x}_t^{n-1})$ in which the subscript t denotes the true state. We then find

$$
\begin{aligned}
\mathbf{y}^n &- H(f_i) \\
&= \mathbf{y}^n - H(\mathbf{x}_t^n) + H(\mathbf{x}_t^n) - H(f_i) \\
&= \epsilon_o^n + H(\mathbf{x}_t - f_t + f_t) - H(f_i) \\
&\approx \epsilon_o^n + \tilde{H}(\mathbf{x}_t - f_t) + H(f_t) - H(f_i) \\
&= \epsilon_o^n + \tilde{H}\epsilon_q^n + H(f_t) - H(\bar{f}) + H(\bar{f}) - H(f_i),
\end{aligned} \tag{2}
$$

in which $\epsilon_o^n = \mathbf{y}^n - H(\mathbf{x}_t^n)$, $\epsilon_q^n = \mathbf{x}_t - f_t$. This allows us to write

$$\mathbf{y}^n - H(f_i) \approx \epsilon_o^n + \tilde{\mathbf{H}}\epsilon_q^n + v_t^n + v_i^n, \tag{3}$$

in which the approximation is due to the Taylor series expansion of H at point f_t and $\tilde{\mathbf{H}}$ is the derivative of H to its

argument at that point. If H is linear, the above is true without approximation taking $\tilde{\mathbf{H}} = H$.

From now on we will denote $H(..)$ (note the parentheses) as the full nonlinear H, and $H..$ (without parentheses) for the derivative of $H(..)$ at the true value. Furthermore,

$$v_i^n = H(\bar{f}) - H(f_i) \tag{4}$$

and similarly for $v_t^n = H(f_t) - H(\bar{f})$.

Assuming all terms in the particle innovations are independent, we can now form

$$\mathbf{C} = \frac{1}{N-1} \sum_i \{\mathbf{y}^n - H(f_i)\}\{\mathbf{y}^n - H(f_i)\}^{\mathrm{T}}$$

$$= \mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}} + \frac{1}{N-1} \sum_i v_t^n v_t^{n\mathrm{T}} + \frac{1}{N-1} \sum_i v_i^n v_i^{n\mathrm{T}}, \tag{5}$$

in which we use the linearized $\mathbf{H}$ in the last line exploring the notation explained above. We now note that, for a well-performing DA system, on average the distance of an ensemble member to the ensemble mean is equal to the distance between the truth and the ensemble mean, so the ensemble variance is equal to the square of the RMSE, so that the last two terms are approximately equal. This allows us to approximate

$$\mathbf{C} = \mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}} + 2\mathbf{V} \tag{6}$$

with

$$\mathbf{V} = \frac{1}{N-1} \sum_i v_i^n v_i^{N\mathrm{T}}$$

$$= \frac{1}{N-1} \sum_i \{H(\bar{f}) - H(f_i)\}\{H(\bar{f}) - H(f_i)\}^{\mathrm{T}}. \tag{7}$$

This then allows us to find an estimate for $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$ as:

$$\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}} = \mathbf{C} - \mathbf{R} - 2\mathbf{V}. \tag{8}$$

As with the Desroziers diagnostics, we could also try to use the statistics of $(\mathbf{y}^n - H\mathbf{x}_i^n)$, i.e. the difference between observation at time n and the ensemble members at time n. In the appendix we show that, at least for the particle filter we use here, these methods do not reveal new information, but they rather complicate the computation of $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$.

The ensemble estimate of C and V, and hence $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$ will be of very low rank because of the limited ensemble size, typically of order 10 to 250. This leads to a large sampling error. To this end, we apply averaging over both space and time of the estimated $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$, as described below.

An important point is that the scheme described above only updates $\mathbf{Q}$ in observation space, so $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$, and the rest of $\mathbf{Q}$ is untouched. This is a shortcoming of the estimation scheme, similar to estimating $\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ using innovations in linear estimation theory. A partial solution is to identify parts of the state space with similar physics and copy $\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}$ from observed parts of the state space to unobserved parts with similar physics. For parts of the state that have no observed counterpart defined in this way we would have to rely on interpolation, with all problems attached to that. We have no full solution to this problem. In the following we assume we can

estimate $\mathbf{Q}$ completely, and in the numerical experiments we assume that we observe the whole state space.

2.1 | Temporal smoothing

The temporal averaging is described by the following algorithm.

Algorithm 1: Estimating $\mathbf{Q}$

1:   $\mathbf{Q}0 \leftarrow s\mathbf{B}$
2:   **for** $k \leftarrow 1, 150$ **do**
3:      $\mathbf{Q}_e \leftarrow Estimate\_Q \ (using \ \mathbf{Q} = \mathbf{Q}0 \ in \ the \ DA)$
4:      $\mathbf{Q}_m \leftarrow \frac{k-1}{k}\mathbf{Q}_m + \frac{1}{k}\mathbf{Q}_e$
5:   **end for**
6:   **for** $k \leftarrow 151, 250$ **do**
7:      $\mathbf{Q}_e(k) \leftarrow Estimate\_Q(updating \ \mathbf{Q} = \mathbf{Q}_m \ in \ the \ DA)$
8:      $\mathbf{Q}_m \leftarrow \frac{k-1}{k}\mathbf{Q}_m + \frac{1}{k}\mathbf{Q}_e$
9:   **end for**

In the algorithm, $\mathbf{Q}_e(k)$ is the estimated model error covariance in time step k and $\mathbf{Q}_m$ is the average of the $\mathbf{Q}_e$ matrices in the past k time steps. s is a scalar which is multiplied by $\mathbf{B}$ to find an initial value of $\mathbf{Q}0$. This $\mathbf{B}$ matrix comes from the prior at the start of the DA procedure, and plays no role in the rest of the algorithm or the particle filter. If better initial estimates of $\mathbf{Q}0$ are available, they should be used, of course. The algorithm is split in two parts. The first part uses the first-guess model error covariance $\mathbf{Q}0$ as the current value in the first 150 time steps of the DA. So, we do estimate $\mathbf{Q}$ at each observation step, but we do not use that raw estimate yet in the DA. The second part uses the average of the estimated model error covariances as the current estimate of model error covariance in the DA. The reason for this two-stage algorithm is that, because of the relatively small ensemble size, the instantaneous estimates for $\mathbf{Q}$ have large sampling error, and the temporal averaging will increase this sample size. The initial averaging is taken here over 150 time steps in the pseudocode and the total running time of the whole scheme is 250 time steps, but that can (and should) be tested for each DA system separately.

2.2 | Spatial smoothing

Because of the small ensemble size, the temporal filtering will not be enough, or has to be done over an unrealistically long time, to eliminate sampling error even in places where the correlations are high. To this end we smooth the elements over areas that we know have similar physics. In this paper two ways of smoothing are applied, localization to smooth perpendicular to the diagonal, and a simple 10 grid-point moving average over the diagonals of the estimated covariance matrix.

Localization is a way of increasing the rank of the estimated model error covariance matrix and eliminating spurious long-distance correlations. A simple localization strategy has

been implemented here, which is the covariance localization (Houtekamer and Mitchell, 2001). Covariance localization is performed by element-wise (Shur) multiplication of the model error covariance matrix with a predefined correlation matrix (denoted as $\mathbb{L}$ here) representing a decaying function of distance:

$$\mathbf{Q}_s^k = \mathbb{L} \circ \mathbf{Q}_m^k. \tag{9}$$

We use a localization radius of five grid points. Many other localization methods could be considered, tailored to specific applications.

2.3 | Convergence of the method

In the following we discuss the convergence properties of the scheme used to estimate the model error covariance matrix $\mathbf{HQH}^T$. The actual convergence of the method will depend on the particle-filter method used. Here we provide results using the Implicit Equal-Weights Particle Filter (IEWPF) of Zhu *et al.* (2016) as an example of how one could go about such a proof. We first describe that particle filter briefly, followed by the actual proof.

2.3.1 | The implicit equal-weights particle filter

The Implicit Equal-Weight Particle Filter (IEWPF) is described in detail in Zhu *et al.* (2016). It is a particle filter that uses a proposal density as follows. For simplicity we concentrate only on the states at time n and $n-1$, and assume that all observations up to time $n-1$ have been used to find a set of particles at $n-1$ with weights w_i^{n-1}. Then we can write

$$\begin{aligned}
p(\mathbf{x}^n, \mathbf{x}^{n-1} | \mathbf{y}^n) &= \frac{p(\mathbf{y}^n | \mathbf{x}^n)}{p(\mathbf{y}^n)} p(\mathbf{x}^n | \mathbf{x}^{n-1}) p(\mathbf{x}^{n-1}) \\
&= \frac{p(\mathbf{y}^n | \mathbf{x}^n)}{p(\mathbf{y}^n)} p(\mathbf{x}^n | \mathbf{x}^{n-1}) \frac{1}{N} \sum_i \{\delta(\mathbf{x}^{n-1} - \mathbf{x}_i^{n-1})\} \\
&= \frac{p(\mathbf{y}^n | \mathbf{x}^n)}{N p(\mathbf{y}^n)} \\
&\quad \times \sum_i \frac{p(\mathbf{x}^n | \mathbf{x}_i^{n-1})}{q(\mathbf{x}^n | \mathcal{X}^{n-1}, \mathbf{y}^n)} q(\mathbf{x}^n | \mathbf{x}^{n-1}, \mathbf{y}^n) \{\delta(\mathbf{x}^{n-1} - \mathbf{x}_i^{n-1})\}
\end{aligned} \tag{10}$$

in which $q(x^n | \mathcal{X}^{n-1}, y^n)$ is the proposal density that is conditioned on all particles at time $n-1$, indicated by the $\mathcal{X}^{n-1}$. It will be chosen such that the weights of all particles are equal. We note that, no matter what we choose for q, the weights will be given by

$$w_i = \frac{p(\mathbf{y}^n | \mathbf{x}_i^n)}{p(\mathbf{y}^n)} \frac{p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1})}{q(\mathbf{x}_i^n | \mathcal{X}^{n-1}, \mathbf{y}^n)}. \tag{11}$$

Instead of drawing directly from q, which will be complicated because of the equal weight procedure, we draw implicitly from a Gaussian $q(\xi)$, leading to weights

$$w_i = \frac{p(\mathbf{y}^n | \mathbf{x}_i^n)}{p(\mathbf{y}^n)} \frac{p(\mathbf{x}_i^n | \mathbf{x}_i^{n-1})}{q(\xi)} \left\| \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\xi} \right\|. \tag{12}$$

As explained in Zhu *et al.* (2016), equal-weight particles can be constructed for instance by assuming

$$\mathbf{x}_i^n = f(\mathbf{x}_i^{n-1}) + \mathbf{K}\{\mathbf{y}^n - Hf(\mathbf{x}_i^{n-1})\} + \alpha_i^{1/2} \mathbf{P}^{1/2} \xi_i^n, \tag{13}$$

where
$\mathbf{K} = \mathbf{Q}\mathbf{H}^T(\mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R})^{-1}$ and $\mathbf{P} = (\mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$, where $\mathbf{Q}$ is the model error covariance matrix, $\mathbf{H}$ is the observation operator and $\mathbf{R}$ is the observation error covariance matrix, and in which α_i is a free parameter. Using this expression in the expression for the weights, the weight of particle i becomes a nonlinear function of parameter α_i, for each particle index i. We now set a target weight as the lowest maximal weight of all particles, so

$$w_{\text{target}} = \min_i \{\max_{x_i}(w_i)\} \tag{14}$$

and fix the αs such that the weight of each particle is equal to this target weight by solving

$$w_i(\alpha_i) = w_{\text{target}}. \tag{15}$$

As mentioned, full details are given in Zhu *et al.* (2016).

2.3.2 | Convergence results

The proof of convergence is rather lengthy and is provided in Appendix 4. The numerical experiments in the next section confirm this convergence on a nonlinear 1000-dimensional problem.

3 | EXPERIMENTS

In this section we will verify our scheme in a moderately high-dimensional setting of the Lorenz 1996 model. The Lorenz 96 model (Lorenz, 1995) is a dynamical nonlinear model given by:

$$\frac{\mathrm{d}\mathbf{x}_j}{\mathrm{d}t} = -\mathbf{x}_{j-2}\mathbf{x}_{j-1} + \mathbf{x}_{j-1}\mathbf{x}_{j+1} - \mathbf{x}_j + F, \tag{16}$$

where x_j is the state variable of the model at position j and F is a forcing constant, which is typically chosen as 8 for chaotic behaviour. The dimension of Lorenz 96 model is chosen as 1000 for all experiments discussed here. We use a time step of $\Delta t = 0.05$ with a fourth Runge–Kutta scheme for the deterministic part and an Euler–Maruyama scheme for the stochastic part of the model.

The detailed description of the IEWPF scheme can be found in Zhu *et al.* (2016). For the IEWPF scheme we choose the initial background-error covariance matrix $\mathbf{B}$ as a tridiagonal matrix with main diagonal value 1 and sub-/super-diagonal value 0.25. The true model error covariance matrix $\mathbf{Q}_t$ is also chosen as a tridiagonal matrix with main diagonal value 0.2 and both sub- and super-diagonal values are 0.05 for all the experiments in this section. The observation error covariance denoted as $\mathbf{R}$ is chosen as a diagonal matrix, with variances varying with experiment. The initial ensemble member perturbations are generated by random sampling from a Gaussian

with covariance matrix the background-error covariance and with zero mean.

To illustrate the method, we observe every grid point at every time step for the model evolution process for all the experiments in this paper. This approach was taken because it is unclear how results from extensive experimentation with different observation strategies will carry over to other models.

The typical total evolving time of the modelling scheme is set to be 250 time steps, with a few longer experiments to study convergence. The temporal averaging time is set to 150 time steps and the second iterative part of the estimation scheme is set to 100 time steps for all the experiments except for those described in section 3.4.

3.1 | The importance of spatial smoothing

The spatial smoothing strategy is described in section 2.2, and the localization radius is chosen to be five grid points, with a ten-grid-point moving average along the diagonal.

We choose $\mathbf{R} = 0.005$ and the initial first guess $\mathbf{Q}0 = 0.25\mathbf{B}$. The number of ensemble members is set to be 40.

Figure 1 shows the 1000 main diagonal entries of the estimated model error covariance with and without smoothing after the 250 time steps, which are arranged in a 40×25 matrix in sequential order.

It shows that smoothing strongly reduces the variance of the noisy main diagonal entries, leading to a more accurate covariance than without smoothing. Part of the full estimated model error covariance matrix with and without smoothing is given in Figure 2.

3.2 | Sensitivity to ensemble size

Since the number of the ensemble members will always be much smaller than the rank of the model error covariance, we need to use the localization and smoothing described above. We investigate the impact of the ensemble size while incorporating this localization and smoothing. The ensemble size is increased from 40 to 200 and 400. The main diagonal value of observation error covariance is again set to 0.005. The initial first guess of model error covariance is again $\mathbf{Q}0 = 0.25\mathbf{B}$.

Figure 3 depicts the time evolution of the estimated main and sub-diagonals of the estimated $\mathbf{Q}$ matrix. What strikes the eye is the large value at initial times, much larger than the initial estimate of $\mathbf{Q}$, which is 0.25 for the main diagonal. The reason is simply that the DA needs some time to adjust, and the values we see here are actually close to those of the initial background covariance matrix $\mathbf{B}$. After a few time steps those values are forgotten and we settle quickly at the correct value, which is 0.2. This transient is visible in all experiments, both for the diagonal and the off-diagonal elements.

The true values are 0.2 and 0.05. The figure shows that the ensemble size has little impact on the estimation methods. Therefore a small ensemble size of 40 is used in the rest of
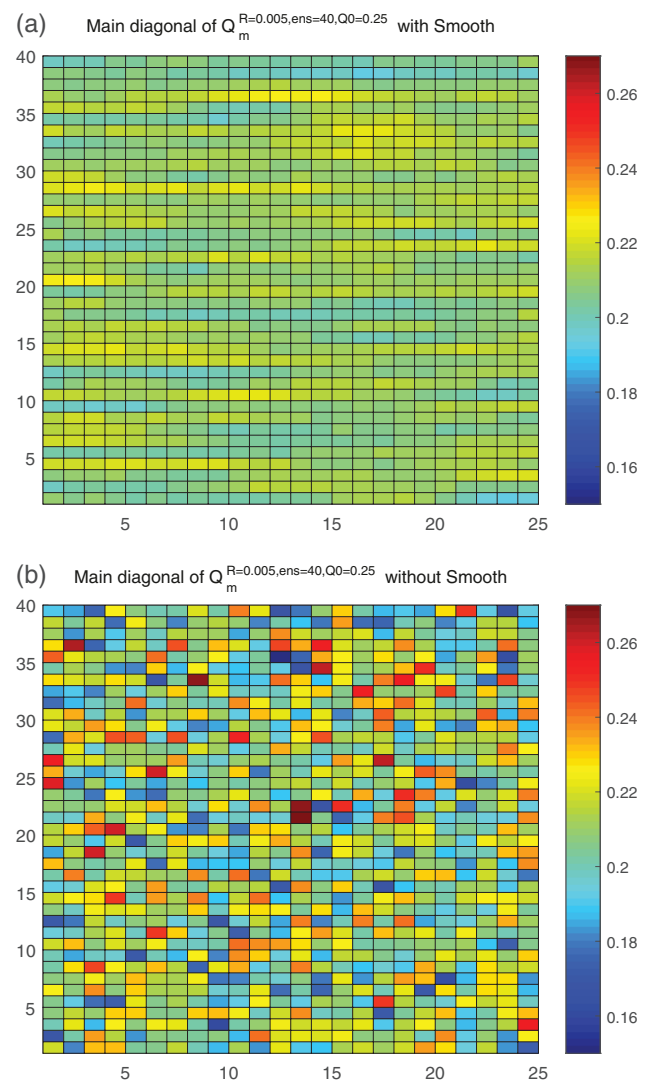
FIGURE 1 The main diagonal entries of estimated model error covariance $\mathbf{Q}_m$ with $\mathbf{R} = 0.005$, 40 ensemble members and $\mathbf{Q}0 = 0.25$ (a) with and (b) without smoothing [Colour figure can be viewed at wileyonlinelibrary.com]

this paper. We do note, however, that this result does depend on the amount of localization and smoothing used in the estimate of $\mathbf{Q}$.

3.3 | Sensitivity to initial model error covariance Q0

In the following experiments we study the sensitivity of the estimation scheme to different initial covariance estimates $\mathbf{Q}0$. Specifically, we test the behaviour of the scheme for $\mathbf{Q}0 = 0.5\mathbf{B}$, $0.25\mathbf{B}$ and $0.1\mathbf{B}$. The observation error variance is set to be 0.005 and the ensemble size is 40.

Figure 5 illustrates that the convergence rate of the mean diagonal and super- and sub-diagonal values is not sensitive to the value of $\mathbf{Q}0$, at least not in the range tested here.

3.4 | Sensitivity to observation error covariance R

In this section two studies will be performed. First we will test the sensitivity of the method to different values of the
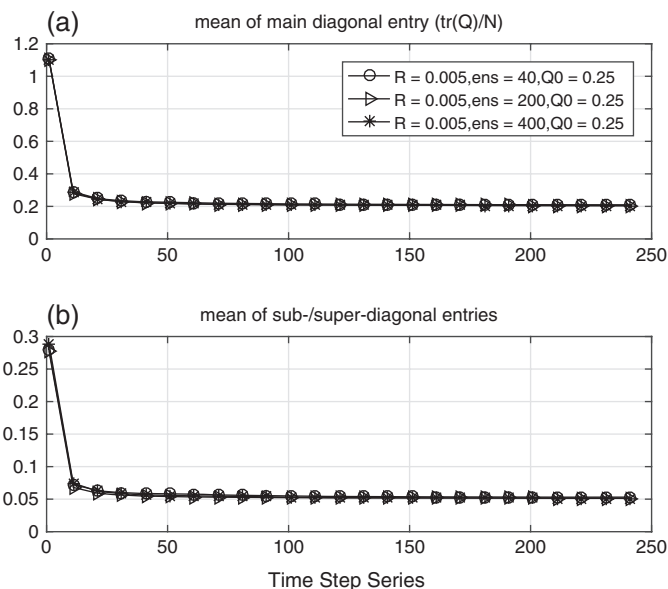
(a)

(b)
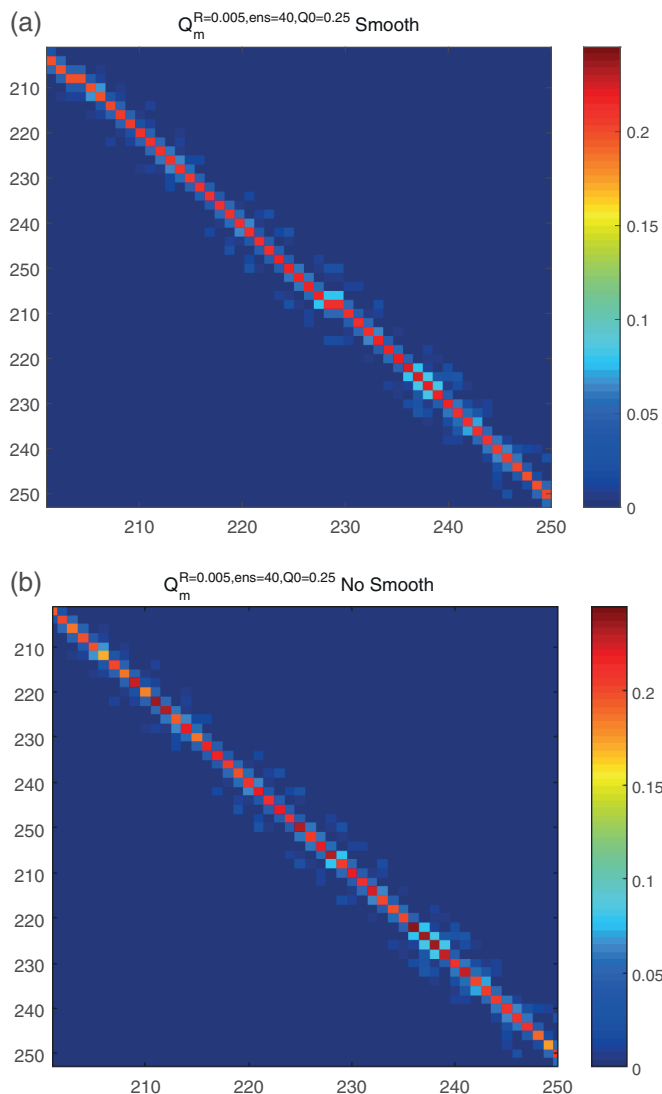
FIGURE 2 Enlarged part of the estimated model error covariance (a) with and (b) without smoothing [Colour figure can be viewed at wileyonlinelibrary.com]


FIGURE 3 Time evolution of (a) the mean of the main diagonal entry and (b) the mean of the sub-/super-diagonal entries using three different ensemble sizes
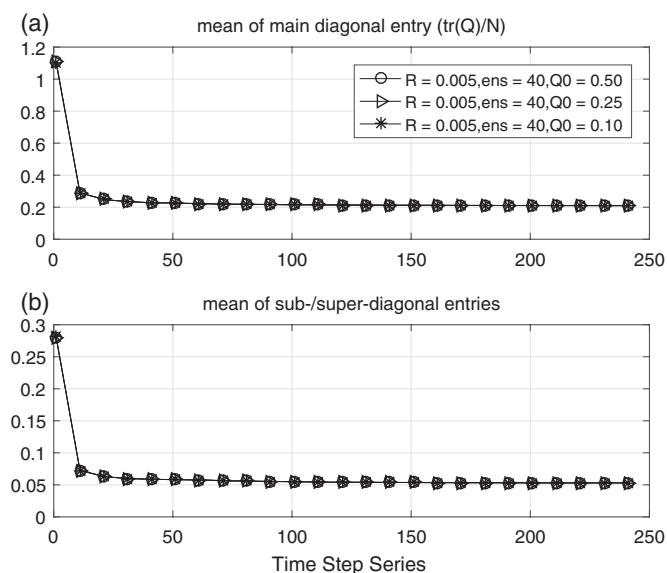

FIGURE 4 As Figure 3, but for three different values of $\mathbf{Q}0$

observation error covariance matrix $\mathbf{R}$, followed by experiments in which we assume wrong values for $\mathbf{R}$ in the assimilation scheme.

We first test the behaviour of the scheme for diagonal values of the observation error covariance set to 0.005, 0.05, 0.1 and 0.2. Forty ensemble members are used in this experiment and the initial first guess is set to $\mathbf{Q}0 = 0.25\mathbf{B}$.

All values for $\mathbf{R}$ showed convergence of $\mathbf{Q}$ to the true value as above, except the largest value of $\mathbf{R} = 0.2\,\mathbf{I}$. We found that when $\mathbf{R} = 0.2\,\mathbf{I}$ the estimation scheme is struggling to find good values for $\mathbf{Q}$ because the estimated matrix is not positive definite. Instead of fixing that directly, we used the practical approach which can easily be applied to larger systems too. Instead of using $\mathbf{Q} = \mathbf{C} - \mathbf{R} - 2\mathbf{V}$, we use $\mathbf{Q} \sim \mathbf{C} - \mathbf{R} - 1.5\mathbf{V}$ in our initial 150 estimates, and use $\mathbf{Q} = \mathbf{C} - \mathbf{R} - 2\mathbf{V}$ in the resulting 100 time steps.

Figures 5 and 6 show that accurate estimates for $\mathbf{Q}$ can be obtained, but the convergence is to a too high value of $\mathbf{Q}$ in the first 150 time steps. Using the correct equation for $\mathbf{Q}$ in the rest of the time steps gives a slow but definite convergence to the true value. The slow convergence is thought to be due to two processes. Firstly, new estimates for $\mathbf{Q}$ contribute a small amount to the $\mathbf{Q}$ used in the DA scheme because the new estimate is averaged with all the old estimates of $\mathbf{Q}$. Secondly, a larger value of elements in $\mathbf{R}$ will lead to a larger ensemble spread, so a larger amplitude of spurious fluctuations, so slower convergence to the actual value. With reference to Appendix 4, the slow convergence suggests that the matrices $\mathbf{A}^n$ and $\mathbf{B}^n$, which contain these ensemble averages, are larger here then for smaller $\mathbf{R}$.

We now test the performance of the scheme when $\mathbf{R}$ is not known very accurately. This is the more realistic case. Our tests allow for a 50% error in $\mathbf{R}$, so we test the scheme for
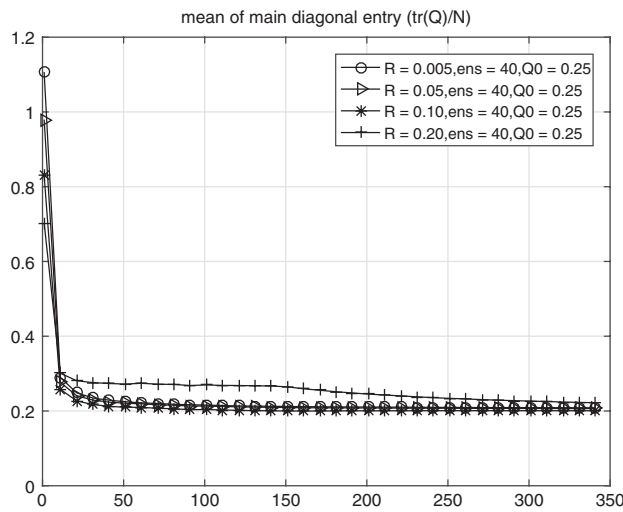
FIGURE 5 Time evolution over 350 time steps of the main diagonal entries of estimated model error covariance with four different values for **R**
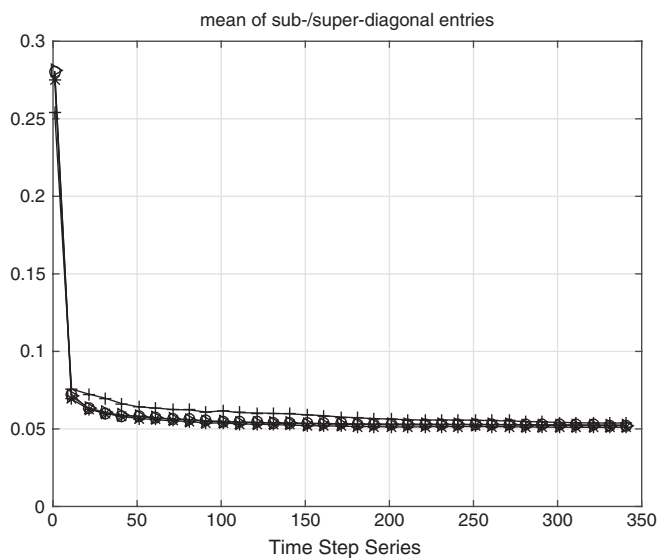
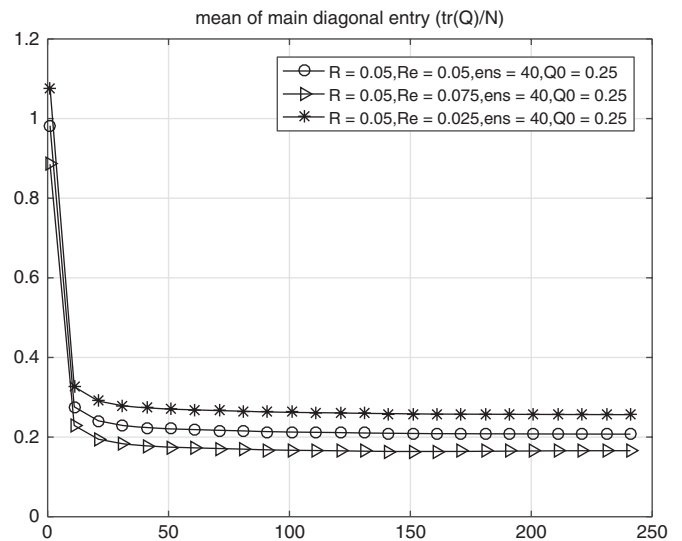**FIGURE 6** As Figure 5, but for the super- and sub-diagonal entries of estimated model error covariance

FIGURE 7 Time evolution of the main diagonal entries of estimated model error covariance with inaccurate **R**

FIGURE 8 Time evolution of the sub-/super-diagonal entries of estimated model error covariance with inaccurate **R**

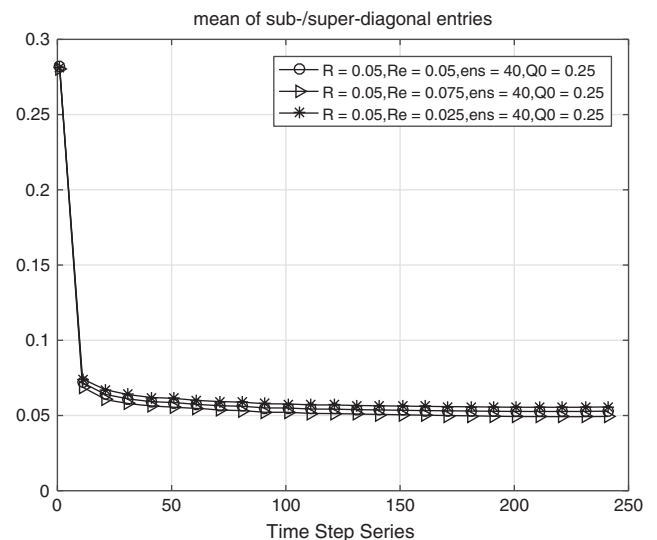diagonal values of **R** in the range 0.025, 0.05, and 0.075, with a true value of **R** = 0.05.

Figure 7 shows that the diagonal elements for **Q** are well estimated when the true **R** is used, but systematically too high and too low when the diagonal values of **R** are chosen too low and too high, respectively. This is a direct consequence of the way we calculate **Q**. On the other hand, as shown in Figure 8, the super- and sub-diagonal elements in **Q** are affected to a much lesser extent, showing that only those elements in **Q** which are directly affected by **R** suffer most from an inaccurate **R** estimate.

4 | CONCLUSIONS AND DISCUSSION

A new iterative scheme of estimating model error covariance in observation space, $\mathbf{HQH}^{\mathrm{T}}$, using particle filters is presented, exploring the innovation vectors of all particles.

A simple temporal and spatial smoothing strategy is applied in the new scheme to speed up and stabilize the estimating process, increasing the rank of the estimated matrix and adding smoothness where we expect that to be appropriate. A convergence proof for the new method has also been developed.

The new scheme is easy to implement with the implicit equal-weights particle filter. It was tested using numerical experiments on the 1000-dimensional Lorenz (1996) model. The ensemble size has little impact on the estimating procedure, related to the temporal and spatial smoothing and localization applied. The resulting model error covariance matrices quickly converge to the correct matrix. These results are robust to varying initial guesses for **Q**, and varying sizes of observation error. Large values for the observation errors lead

to slower convergence of the method as the ensemble spread will be larger, so the sampling errors are larger.

We also tested the method in the situation where $\mathbf{R}$ is not known accurately. The results are as reported for the Desroziers method when trying to estimate $\mathbf{HBH}^T$, i.e. too low $\mathbf{R}$ values lead to too large $\mathbf{HBH}^T$ values and vice versa (Desroziers *et al.*, 2005; 2009). Our numerical experiments indicate that only those elements of $\mathbf{Q}$ in which $\mathbf{R}$ is inaccurate are affected, but the generality of this statement is not tested, and is difficult to test exhaustively using only the simple model used here.

A weakness of the method is that we can only estimate $\mathbf{HQH}^T$, so $\mathbf{Q}$ in observation space. Indeed, in the experiments we assumed that the whole state space is observed directly, which is not very realistic. For a full $\mathbf{Q}$ one will have to rely on spatial interpolation, with all problems related to that. Furthermore, $\mathbf{H}$ is assumed to be linear, or its linearization has to be accurate. However, apart from spatial smoothness we can explore the connection to the physics. Since $\mathbf{Q}$ is related to model physics, one could use that part of $\mathbf{Q}$ estimated in an observed area to infer $\mathbf{Q}$ in unobserved areas with similar physics. This is a clear advantage of estimating $\mathbf{Q}$ over $\mathbf{B}$: $\mathbf{Q}$ is a matrix related to the model physics, while $\mathbf{B}$ is a matrix related to the accuracy of the state, which will vary over space and time, not directly related to the physics (Bannister, 2008a; 2008b).

We assumed that the $\mathbf{HQH}^T$ matrix can be stored, which will not be true in realistic high-dimensional applications. However, $\mathbf{Q}$, and hence $\mathbf{HQH}^T$ will be a narrow matrix, as it represents the influence of model errors which are largest at the model grid scale, with small horizontal length-scales. So we only have to store the (block) diagonal and a few super- and sub- (block) diagonals, which should not be too problematic. This statement is not true for non-local observations, such as several satellite observations that are line integrals. In that case $\mathbf{Q}$ is still sparse, but off-diagonal elements can be large, and localization has to be done with care. Ideally one would like to store $\mathbf{Q}$ in operator form, but it is unclear at this moment how to estimate these operations directly. That would mean that one would need to generate $\mathbf{Q}^{1/2}\mathbf{x}$ for arbitrary $\mathbf{x}$; this needs further research, which could be quite exciting in its own right.

The accuracy of the method for estimating $\mathbf{Q}$ is dependent on the accuracy which with we know $\mathbf{R}$. Unfortunately, when our estimate of $\mathbf{R}$ is poor, we will get less accurate estimates for $\mathbf{Q}$ without knowing this to be the case; the method cannot detect issues like this. The most straightforward test might be to study the statistics of predictions, which is a non-trivial exercise given the complex relation in a nonlinear model between errors in $\mathbf{Q}$ and errors in prediction statistics. In our experiments we only looked at diagonal $\mathbf{R}$, while it is well known that $\mathbf{R}$ tends to be non-diagonal, for instance due to representation errors (Stewart *et al.*, 2008; van Leeuwen, 2014; Hodyss and Nichols, 2015). Although not studied, the expectation is that inaccurate off-diagonal elements in $\mathbf{R}$

will lead to inaccurate non-diagonal elements in $\mathbf{Q}$. Another related issue is that we assumed that the observation operator $\mathbf{H}$ is known exactly, which is often not the case. To first order (Lorenc, 1986; van Leeuwen, 2014), these errors can be incorporated in $\mathbf{R}$, and the discussion above holds also for this case.

Although estimating $\mathbf{Q}$ is in its infancy, it is good to realise the importance of increasing the effort in this area. Estimated model errors give us a direct insight into which parts of the model are in error *at the level of the model equations*, allowing for direct model improvement. That would make DA into a tool much more powerful than is used at present. It is true that there are numerous issues with estimating $\mathbf{Q}$, and no doubt we will never be able to estimate it completely, but that is not an argument for not trying. The potential benefits are enormous.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, J.L. (2007) Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, 230(1), 99–111.

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

Bannister, R.N. (2008a) A review of forecast-error covariance statistics in atmospheric variational data assimilation. I: characteristics and measurements of forecast-error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 1951–1970.

Bannister, R.N. (2008b) A review of forecast-error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast-error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 1971–1996.

Buehner, M., Houtekamer, P., Charette, C., Mitchell, H.L. and He, B. (2010) Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: description and single-observation experiments. *Monthly Weather Review*, 138(5), 1550–1566.

Chorin, A. and Tu, X. (2009) Implicit sampling for particle filters. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 17 249–17 254.

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

Desroziers, G., Berre, L., Chabot, V. and Chapnik, B. (2009) A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137(10), 3420–3436.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99, 10143–10162. https://doi.org/10.1029/94JC00572.

Evensen, G. and van Leeuwen, P.J. (2000) An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128, 1852–1867.

Frei, M. and Künsch, H.R. (2013) Bridging the ensemble Kalman and particle filters. *Biometrika*, 100(4), 781–800.

Hodyss, D. and Nichols, N.K. (2015) The error of representation: basic understanding. *Tellus A*, 67, 24822. https://doi.org/10.3402/tellusa.v67.24822.

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

van Leeuwen, P.J. (2010) Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136, 1991–1999.

van Leeuwen, P.J. (2014) Representation errors and retrievals in linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1612–1623.

Lei, J. and Bickel, P. (2011) A moment matching ensemble filter for nonlinear non-gaussian data assimilation. *Monthly Weather Review*, 139(12), 3964–3973.

Liu, C., Xiao, Q. and Wang, B. (2008) An ensemble-based four-dimensional variational data assimilation scheme. PartI: Technical formulation and preliminary test. *Monthly Weather Review*, 136(9), 3363–3373.

Lorenc, A.C. (1986) Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112, 1177–1194.

Lorenz, E.N. (1995) Pre dictability: a problem partly solved. In: *Seminar on Predictability*, Vol. I, Reading, UK: ECMWF, pp. 1–18.

Poterjoy, J. (2016) A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144(1), 59–76.

Reich, S. (2013) A non-parametric ensemble transform method for Bayesian inference. *SIAM Journal of Scientific Computing*, 35(4), A2013–A2024.

Stewart, L.M., Dance, S. and Nichols, N. (2008) Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 56(8), 1521–1527.

Stewart, L.M., Dance, S.L. and Nichols, N.K. (2013) Dat a assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A*, 65, 19546. https://doi.org/10.3402/tellusa.v65i0.19546.

Todling, R. (2015a) A complementary note to lag-1 smoother approach to system-error estimation: The intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 141(692), 2917–2922.

Todling, R. (2015b) A lag-1 smoother approach to system-error estimation: sequential method. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1502–1513.

Weston, P.P., Bell, W. and Eyre, J.R. (2014) Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140, 2420–2429. https://doi.org/doi.org/10.1002/qj.2306.

Zhu, M., van Leeuwen, P.J. and Amezcua, J. (2016) Implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 142, 1904–1919. https://doi.org/10.1002/qj.2784.

APPENDIX A: ESTIMATING Q USING THE ANALYSIS ENSEMBLE

In this appendix we describe two other methods for estimating $\mathbf{HQH}^T$, and show that they will be less efficient than the method presented in the main text.

Estimating Q based on $(\mathbf{y}^n - \mathbf{Hx}_i^n)(\mathbf{y}^n - \mathbf{Hx}_i^n)^T$

This method uses the analysis instead of the forecast to form the covariance, as:

$$\mathbf{C} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{y}^n - \mathbf{Hx}_i^n)(\mathbf{y}^n - \mathbf{Hx}_i^n)^T. \quad (A1)$$

To evaluate this, we use, suppressing the time index n,

$$\mathbf{y} - \mathbf{Hx}_i = \mathbf{y} - \mathbf{Hx}_t + \mathbf{Hx}_t - \mathbf{Hx}_i$$
$$= \mathbf{y} - \mathbf{Hx}_t + \mathbf{Hx}_t - \mathbf{H}f_t + \mathbf{H}f_t - \mathbf{Hx}_i. \quad (A2)$$

Now we use the expression for $\mathbf{x}_i$ from the IEWPF, given by:

$$\mathbf{x}_i = f_i + \mathbf{K}(\mathbf{y}^n - f_i) + \alpha_i^{1/2}\mathbf{P}^{1/2}\xi_i \quad (A3)$$

leading to

$$\mathbf{y} - \mathbf{Hx}_i = \mathbf{L}(\mathbf{y} - \mathbf{Hx}_t) + \mathbf{LH}(\mathbf{x}_t - f_t)$$
$$+ \mathbf{LH}(f_t - f_i) - \mathbf{HP}^{1/2}\alpha_i^{1/2}\xi_i, \quad (A4)$$

in which we defined $\mathbf{L}$ via

$$\mathbf{L} = (1 - \mathbf{HK})$$
$$= \left\{ 1 - \mathbf{HQH}^T(\mathbf{HQH}^T + \mathbf{R})^{-1} \right\}$$
$$= \mathbf{R}(\mathbf{HQH}^T + \mathbf{R})^{-1}. \quad (A5)$$

Using this expression in $\mathbf{C}$, and assuming that all cross-terms vanish, leads to

$$\mathbf{C} \approx \mathbf{L R L}^T + \mathbf{L H Q H}^T\mathbf{L}^T + 2\mathbf{L V L}^T + \mathbf{W}, \quad (A6)$$

in which

$$\mathbf{V} \approx \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{H}f_i - \bar{f})(\mathbf{H}f_i - \bar{f})^T \quad (A7)$$

and

$$\mathbf{W} = \mathbf{HP}^{1/2}\frac{1}{N-1} \sum_{i=1}^{N} \alpha_i \xi_i \xi_i^T \mathbf{P}^{1/2^T}\mathbf{H}^T. \quad (A8)$$

We thus find

$$\mathbf{L H Q H}^T\mathbf{L}^T \approx \mathbf{C} - \mathbf{L}(\mathbf{R} + 2\mathbf{V})\mathbf{L}^T - \mathbf{W}, \quad (A9)$$

which is similar to the expression for $\mathbf{HQH}^T$ from the main text, but contains an extra matrix $\mathbf{L}$, and $\mathbf{W}$. To solve this equation, one can rewrite it as

$$\mathbf{LHQH}^T\mathbf{L}^T = \mathbf{D} \quad (A10)$$

with

$$\mathbf{D} = \mathbf{C} - \mathbf{L}(\mathbf{R} + 2\mathbf{V})\mathbf{L}^T - \mathbf{W}. \quad (A11)$$

One way to solve this equation is to define $\mathbf{Z} = \mathbf{HQ H}^T\mathbf{L}^T$ leading to

$$\mathbf{LZ} = \mathbf{D}, \quad (A12)$$

and form an SVD of $\mathbf{L} = \mathbf{U\Lambda V}^T$ to find

$$\mathbf{Z} = \mathbf{V\Lambda}^{-1}\mathbf{U}^T\mathbf{D}. \quad (A13)$$

Now solve $\mathbf{Z} = \mathbf{HQH}^T\mathbf{L}^T$ using the same SVD of $\mathbf{L}$ as

$$\mathbf{HQH}^T = \mathbf{V\Lambda}^{-1}\mathbf{U}^T\mathbf{D} \, \mathbf{U\Lambda}^{-1}\mathbf{V}^T. \quad (A14)$$

This method is much more involved as it needs the SVD of $\mathbf{L}$, which can be prohibitively expensive when the dimension of the system is large. Other methods than the SVD could be explored, but that is beyond the scope of this paper.

Estimating Q via $\{\mathbf{y}^n - \mathbf{H}f(\mathbf{x}_i^{n-1})\}(\mathbf{y}^n - \mathbf{H}\mathbf{x}_i^n)^{\mathrm{T}}$

Finally, $\mathbf{Q}$ could be estimated exploring the cross-covariance via

$$\mathbf{C} = \frac{1}{N-1}\sum_{i=1}^{N}\{\mathbf{y}^n - \mathbf{H}f\left(\mathbf{x}_i^{n-1}\right)\}(\mathbf{y}^n - \mathbf{H}\mathbf{x}_i^n)^{\mathrm{T}}. \quad \text{(A15)}$$

To evaluate this, we again suppress the time index and use the expressions for $(\mathbf{y} - \mathbf{H}f_i)$ and $(\mathbf{y} - \mathbf{H}\mathbf{x}_i)$ to find for $\mathbf{C}$

$$\mathbf{C} \approx \mathbf{RL}^{\mathrm{T}} + \mathbf{HQH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} + 2\mathbf{V}\,\mathbf{L}^{\mathrm{T}}, \quad \text{(A16)}$$

in which

$$\mathbf{V} \approx \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{H}f_i - \mathbf{H}\bar{f})(\mathbf{H}f_i - \mathbf{H}\bar{f})^{\mathrm{T}} \quad \text{(A17)}$$

and

$$\mathbf{L} = (1 - \mathbf{HK}). \quad \text{(A18)}$$

Furthermore, we assumed that the α_i term cancels with the other terms on average.

We can rewrite this equation as

$$\mathbf{HQH}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}} = \mathbf{D}, \quad \text{(A19)}$$

with

$$\mathbf{D} = \mathbf{C} - (\mathbf{R} + 2\mathbf{V})\mathbf{L}^{\mathrm{T}}. \quad \text{(A20)}$$

Define the SVD of $\mathbf{L}$ as $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$ to find

$$\mathbf{HQH}^{\mathrm{T}} = \mathbf{D}\,\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathrm{T}}. \quad \text{(A21)}$$

This method is better conditioned than the method above, but still needs the SVD of $\mathbf{L}$.

APPENDIX B: CONVERGENCE OF THE ESTIMATION METHOD

First recall that the update equation for the IEWPF from section 2.3.1 is

$$\mathbf{x}_i^n = f(\mathbf{x}_i^{n-1}) + \mathbf{K}^n(\mathbf{y}^n - \mathbf{H}f_i^n) + (\mathbf{P}^n)^{1/2}(\alpha_i^n)^{1/2}\xi_i^n \quad \text{(B1)}$$

in which now

$$f_i^n = f(\mathbf{x}_i^{n-1}), \quad \text{(B2)}$$

$$\mathbf{K}^n = \mathbf{Q}^n\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}^n\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \quad \text{(B3)}$$

and

$$\mathbf{P}^n = \{(\mathbf{Q}^n)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\}^{-1}, \quad \text{(B4)}$$

and $\mathbf{Q}^n$ is our current estimate of $\mathbf{Q}$.

In our estimate of $\mathbf{C}$ and $\mathbf{V}$, we need to calculate f_i^n which can be written as:

$$f_i^n = f(\mathbf{x}_i^{n-1})$$
$$= f\{f_i^{n-1} + \mathbf{K}^{n-1}(\mathbf{y}^{n-1} - \mathbf{H}f_i^{n-1})$$
$$+ \alpha_i^{n-1\,1/2}\mathbf{P}^{n-1\,1/2}\xi_i^{n-1}\}. \quad \text{(B5)}$$

For ease of notation, we drop the time index from now on, noting that $\mathbf{Q} = \mathbf{Q}^{n-1}$, $\mathbf{P} = \mathbf{P}^{n-1}$, $\mathbf{K} = \mathbf{K}^{n-1}$, $\alpha = \alpha^{n-1}$, and

$\xi_i = \xi_i^{n-1}$. Let us now write our estimate $\mathbf{Q}$ in terms of the true covariance $\mathbf{Q}_{\mathrm{t}}$ and a systematic perturbation, as

$$\mathbf{Q} = \mathbf{Q}_{\mathrm{t}} + \epsilon, \quad \text{(B6)}$$

in which we assume that the matrix ϵ is much smaller in magnitude than $\mathbf{Q}_{\mathrm{t}}$. Also note that $\epsilon = \epsilon^{n-1}$, so it is a time-dependent matrix. Then we can write, to first order in ϵ,

$$\mathbf{K} = (\mathbf{Q}_t + \epsilon)\mathbf{H}^{\mathrm{T}}\{\mathbf{H}(\mathbf{Q}_t + \epsilon)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\}^{-1}$$
$$\approx \mathbf{K}_t + \epsilon\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}_t\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$
$$- \mathbf{K}_t\mathbf{H}\epsilon\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}_t\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$
$$= \mathbf{K}_t + (1 - \mathbf{K}_t\mathbf{H})\epsilon\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}_t\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}. \quad \text{(B7)}$$

For matrix $\mathbf{P}$ we find:

$$\mathbf{P} = \{(\mathbf{Q}_t + \epsilon)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\}^{-1}$$
$$= \mathbf{Q} - \mathbf{QH}^{\mathrm{T}}(\mathbf{HQH}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HQ}$$
$$= \mathbf{Q}_t + \epsilon - (\mathbf{Q}_t + \epsilon)\mathbf{H}^{\mathrm{T}}\{\mathbf{H}(\mathbf{Q}_t + \epsilon)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\}^{-1}(\mathbf{Q}_t + \epsilon)$$
$$\approx \mathbf{Q}_t - \mathbf{Q}_t\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}_t\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{HQ}_t + (1 - \mathbf{K}_t\mathbf{H})\epsilon$$
$$- (1 - \mathbf{K}_t\mathbf{H})\epsilon\mathbf{H}^{\mathrm{T}}\mathbf{K}_t^{\mathrm{T}}$$
$$= \mathbf{P}_t + (1 - \mathbf{K}_t\mathbf{H})\epsilon(1 - \mathbf{K}_t\mathbf{H})^{\mathrm{T}}$$
$$= \mathbf{P}_t + \mathbf{P}_t\mathbf{Q}_t^{-1}\epsilon\mathbf{Q}_t^{-1}\mathbf{P}_t. \quad \text{(B8)}$$

We need the square root of $\mathbf{P}^n$ in the direction of ϵ. A first-order Taylor series expansion for matrix functions reads, for μ a small scalar,

$$g(X + \mu Y) \approx g(X) + \mu\left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu=0} g(X + \mu Y). \quad \text{(B9)}$$

So we write $\epsilon = \mu\mathbf{E}$, in which μ is a small scalar, and use $g(X) = X^{1/2}$ in the above equation to find:

$$\mathbf{P}^{1/2} = \left(\mathbf{P}_t + \mu\mathbf{P}_t\mathbf{Q}_t^{-1}\mathbf{E}\mathbf{Q}_t^{-1}\mathbf{P}_t\right)^{1/2}$$
$$\approx \mathbf{P}_t^{1/2} + \mu\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{\mu=0}\left(\mathbf{P}_t + \mu\mathbf{P}_t\mathbf{Q}_t^{-1}\mathbf{E}\mathbf{Q}_t^{-1}\mathbf{P}_t\right)^{1/2}$$
$$= \mathbf{P}_t^{1/2} + \frac{1}{2}\mu\mathbf{P}_t^{-1/2}\mathbf{P}_t\mathbf{Q}_t^{-1}\mathbf{E}\mathbf{Q}_t^{-1}\mathbf{P}_t$$
$$= \mathbf{P}_t^{1/2} + \frac{1}{2}\mathbf{P}_t^{1/2}\mathbf{Q}_t^{-1}\epsilon\,\mathbf{Q}_t^{-1}\mathbf{P}_t. \quad \text{(B10)}$$

$\mathbf{Q}$ also appears in α_i, but that dependence is rather complicated as the argument of an exponential that is part of a complicated argument of the Lambert W function. Instead we will argue below that we do not need to explore this further.

Using these first-order approximations, we find for $f(\mathbf{x}_i^n)$, to first order in ϵ,

$$f_i^n = f(\mathbf{x}_i^{n-1})$$
$$= f\left\{f_i^{n-1} + \mathbf{K}_t(\mathbf{y}^{n-1} - \mathbf{H}f_i^{n-1}) + \alpha_{i_t}^{n-1\,1/2}\mathbf{P}_t^{1/2}\xi_i^{n-1}\right\}$$
$$+ F_i^n(1 - \mathbf{K}_t\mathbf{H})\epsilon\mathbf{H}^{\mathrm{T}}(\mathbf{HQ}_t\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y}^{n-1} - \mathbf{H}f_i^{n-1})$$
$$- \frac{1}{2}F_i^n\alpha_{i_t}^{1/2}\mathbf{P}_t^{1/2}\mathbf{Q}_t^{-1}\epsilon\,\mathbf{Q}_t^{-1}\mathbf{P}_t\xi_i$$
$$+ F_i^n\left(\frac{\partial\alpha_{i_t}^{1/2}}{\partial\mathbf{Q}_t}\right)^{\mathrm{T}}\epsilon\,\mathbf{P}_t^{1/2}\xi_i, \quad \text{(B11)}$$

in which $\mathbf{F}_i^n$ is the linearized model at state $\mathbf{x}_i^{n-1}$.

We now develop $\mathbf{C}$ to first order in ϵ, leading to

$$\mathbf{C}^n = \mathbf{C}_t^n + \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}^n - \mathbf{H}f_i^n)(\mathbf{y}^{n-1} - \mathbf{H}f_i^{n-1})^T$$
$$\times(\mathbf{HQ}_t\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\epsilon(1 - \mathbf{K}_t\mathbf{H})^T\mathbf{F}^T\mathbf{H}^T + (..)^T$$
$$-\frac{1}{2}\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}^n - \mathbf{H}f_i^n)\xi_i^T\alpha_{i_t}^{1/2}\mathbf{P}_t\mathbf{Q}_t^{-1}\epsilon\mathbf{Q}_t^{-1}\mathbf{P}_t^{1/2}\mathbf{F}^T\mathbf{H}^T$$
$$+ (..)^T$$
$$+\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}^n - \mathbf{H}f_i^n)\xi_i^T\mathbf{P}_t^{1/2}\epsilon\frac{\partial\alpha_{i_t}}{\partial\mathbf{Q}_t}\mathbf{F}^T\mathbf{H}^T + (..)^T, \quad (B12)$$

where $(..)^T$ denotes the transpose of the previous term in the equation. First we note that, since ξ_i is independent from all other factors and has mean zero when averaged over the ensemble, so all terms that contain ξ_i will be small, and even zero for an infinite ensemble size. We cannot use that argument for the second term and its transpose because the $(\mathbf{y}^n - \mathbf{H}f_i^n)$, $(\mathbf{y}^{n-1} - \mathbf{H}f_i^{n-1})$ and $\mathbf{F}_i^n$ have non-zero ensemble means. In fact, using the same development as in Eq. (2), we find that the ensemble mean for these terms is $\epsilon^n + \beta_t^n$, $\epsilon^{n-1} + \beta_t^{n-1}$, and $\overline{\mathbf{F}_i^n}^T = \mathbf{F}^T$ respectively, for large ensemble sizes. However, these terms are dependent over time, so when we average over time in the first part of the scheme used to estimate $\mathbf{Q}$, these terms become small also, and this is further amplified by smoothing over space.

For the sample matrix $\mathbf{V}^n$ we find

$$\mathbf{V}^n = \mathbf{V}_t^n + \frac{1}{N-1}\sum_{i}(\mathbf{H}\overline{f^n} - \mathbf{H}f_i^n)(\mathbf{H}\overline{f^{n-1}} - \mathbf{H}f_i^{n-1})^T$$
$$\times(\mathbf{HQ}_t\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\epsilon(1 - \mathbf{K}_t\mathbf{H})^T\mathbf{F}^T\mathbf{H}^T + (..)^T$$
$$-\frac{1}{2}\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{H}\overline{f^n} - \mathbf{H}f_i^n)\xi_i^T\alpha_{i_t}^{1/2}$$
$$\times\mathbf{P}_t\mathbf{Q}_t^{-1}\epsilon\mathbf{Q}_t^{-1}\mathbf{P}_t^{1/2}\mathbf{F}^T\mathbf{H}^T + (..)^T$$
$$+\frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{H}\overline{f^n} - \mathbf{H}f_i^n)\xi_i^T\mathbf{P}_t^{1/2}\epsilon\frac{\partial\alpha_{i_t}}{\partial\mathbf{Q}_t}\mathbf{F}^T\mathbf{H}^T + (..)^T, \quad (B13)$$

and a similar convergence argument applies as for the RHS terms of $\mathbf{C}^n$.

In the second part of the iterative scheme, we use the estimated $\mathbf{Q}$ in the DA. The update scheme is

$$\mathbf{Q}^n = \mathbf{C}^n - \mathbf{R} - 2\mathbf{V}^n$$
$$= \mathbf{C}_t - \mathbf{R} - 2\mathbf{V}_t + \mathbf{A}^n\epsilon^{n-1}(\mathbf{B}^n)^T + \mathbf{B}^n\epsilon^{n-1}(\mathbf{A}^n)^T$$
$$= \mathbf{Q}_t + \mathbf{A}^n\epsilon^{n-1}(\mathbf{B}^n)^T + \mathbf{B}^n\epsilon^{n-1}(\mathbf{A}^n)^T, \quad (B14)$$

in which $\mathbf{A}^n$ and $\mathbf{B}^n$ are small and depend only weakly on ϵ^{n-1}, as argued above. Furthermore, we can write

$$\mathbf{Q}^n = \mathbf{Q}_t + \epsilon^n. \quad (B15)$$

Combining these two equations, we find

$$\mathbf{Q}^n = \mathbf{Q}_t + \mathbf{A}^n\epsilon^{n-1}\mathbf{B}^{nT} + \mathbf{B}^n\epsilon^{n-1}\mathbf{A}^{nT} = \mathbf{Q}_t + \epsilon^n. \quad (B16)$$

We thus find for ϵ^n the recursion

$$\epsilon^n = \mathbf{A}^n\epsilon^{n-1}\mathbf{B}^{nT} + \mathbf{B}^n\epsilon^{n-1}\mathbf{A}^{nT}, \quad (B17)$$

and we can write

$$|\epsilon^n| \leq 2|\mathbf{A}^n|\,|\epsilon^{n-1}|\,|\mathbf{B}^n|. \quad (B18)$$

Since $|\mathbf{A}^n|$ and $|\mathbf{B}^n|$ are small as argued above, this iteration will make ϵ^n smaller and smaller with increasing n, so that $\mathbf{Q}^n$ converges to $\mathbf{Q}_t$. The numerical experiments confirm this convergence on a nonlinear 1000-dimensional problem.