RMetS
Royal Meteorological Society

A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation

Guocan Wu,[a] Xiaogu Zheng,[a*] Liqun Wang,[b] Shupeng Zhang,[a] Xiao Liang[a] and Yong Li[c]

[a] *College of Global Change and Earth System Science, Beijing Normal University, Beijing, China*
[b] *Department of Statistics, University of Manitoba, Winnipeg, Canada*
[c] *School of mathematical sciences, Beijing Normal University, Beijing, China*

*Correspondence to: X. Zheng, College of Global Change and Earth System Science, Beijing Normal University, #19 Xinjie Kouwai St, Beijing 100875, China. E-mail: x.zheng@bnu.edu.cn

Correct estimation of the forecast and observational error covariance matrices is crucial for the accuracy of a data assimilation algorithm. In this article we propose a new structure for the forecast error covariance matrix to account for limited ensemble size and model error. An adaptive procedure combined with a second-order least squares method is applied to estimate the inflated forecast and adjusted observational error covariance matrices. The proposed estimation methods and new structure for the forecast error covariance matrix are tested on the well-known Lorenz-96 model, which is associated with spatially correlated observational systems. Our experiments show that the new structure for the forecast error covariance matrix and the adaptive estimation procedure lead to improvement of the assimilation results. Copyright © 2012 Royal Meteorological Society

Key Words: data assimilation; ensemble Kalman filter; error covariance inflation; second-order least squares estimation; adaptive estimation

Received 20 December 2011; Revised 2 June 2012; Accepted 15 June 2012; Published online in Wiley Online Library 13 August 2012

Citation: Wu G, Zheng X, Wang L, Zhang S, Liang X, Li Y. 2013. A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. *Q. J. R. Meteorol. Soc.* **139**: 795–804. DOI:10.1002/qj.2000

1. Introduction

Data assimilation is a procedure for producing an optimal combination of model outputs and observations. The combined result should be closer to the true state than either the model forecast or the observation are. However, the quality of data assimilation depends crucially on the estimation accuracy of the forecast and observational error covariance matrices. If these matrices are estimated appropriately, then the analysis states can be generated by minimizing an objective function, which is technically straightforward and can be accomplished using existing engineering solutions (Reichle, 2008).

The ensemble Kalman filter (EnKF) is a popular sequential data assimilation approach, which has been widely studied and applied since its introduction by Evensen (1994a, 1994b). In EnKF, the forecast error covariance matrix is estimated as the sampling covariance matrix of the ensemble forecast states, which is usually underestimated due to the limited ensemble size and model error. This may eventually lead to the divergence of the EnKF assimilation scheme (e.g. Anderson and Anderson, 1999; Constantinescu *et al.*, 2007).

One of the forecast error covariance matrix inflation techniques is additive inflation, in which a noise is added to the ensemble forecast states that samples the probability distribution of model error (Hamill and Whitaker, 2005). Another widely used forecast error covariance matrix inflation technique is multiplicative inflation, that is, to multiply the matrix by an appropriate factor.

In early studies of multiplicative inflation, researchers determined the inflation factor by repeated experimentation and chose a value according to their prior knowledge. Hence such experimental tuning is rather empirical and subjective. Wang and Bishop (2003) proposed an on-line estimation method for the inflation factor of the forecast error covariance matrix in a model with a linear observational operator. Building on that work, Li *et al.*(2009) further developed the algorithm. All these methods are based on the first moment estimation of the squared observation-minus-forecast residual, which was first introduced by Dee (1995). Anderson (2007, 2009) used a Bayesian approach to covariance matrix inflation for the spatially independent observational errors, and Miyoshi (2011) further simplified Anderson's inflation approach by making a number of additional simplifying assumptions.

In practice, the observational error covariance matrix may also need to be adjusted (Liang *et al.*, 2011). Zheng (2009) and Liang *et al.*(2011) proposed an approach to simultaneously optimize the inflation factor of the forecast error covariance matrix and the adjustment factor of the observational error covariance matrix. Their approach is based on the optimization of the likelihood function of the observation-minus-forecast residual, an idea proposed by Dee and colleagues (Dee and Da Silva, 1999; Dee *et al.*, 1999). However, the likelihood function of the observation-minus-forecast residual is nonlinear and involves the computationally expensive determinant and inverse of the residual's covariance matrix. In this article, the second-order least squares (SLS; Wang and Leblanc, 2008) statistic of the squared observation-minus-forecast residual is introduced as the objective function instead. The main advantage of the SLS objective function is that it is a quadratic function of the factors, and therefore the closed forms of the estimators of the inflation factors can be obtained. Compared with the method proposed by Liang *et al.*(2011), the computational cost is greatly reduced.

Another innovation of this article is to propose a new structure for the forecast error covariance matrix that is different from the sampling covariance matrix of the ensemble forecast states used in the conventional EnKF. In the ideal situation, an ensemble forecast state is assumed to be a random vector with the true state as its population mean. Hence it is more appropriate to define the ensemble forecast error by the ensemble forecast states minus true state rather than by the perturbed forecast states minus their ensemble mean (Evensen, 2003). This is because in a model with large error and limited ensemble size, the ensemble mean of the forecast states can be very different from the true state. Therefore, the sampling covariance matrix of the ensemble forecast states can be very different from the true forecast error covariance matrix. As a result, the estimated analysis state can be substantially inaccurate. However, in reality the true state is unknown, and the analysis state is a better estimate of the true state than the forecast state. Therefore, in this article we propose to use the information feedback from the analysis state to update the forecast error covariance matrix. In fact, our proposed forecast error covariance matrix is a combination of multiplicative and additive inflation. Bai and Li (2011) also used the feedback from the analysis state to improve assimilation but in a different way.

This article consists of four sections. Section 2 proposes an EnKF scheme with a new structure for the forecast error covariance matrix and its adaptive estimation procedure based on the second-order least squares method. Section 3 presents the assimilation results on the Lorenz model with a correlated observational system. Conclusions and discussion are provided in section 4.

2. Methodology

2.1. *EnKF with SLS inflation*

Using the notations of Ide *et al.*(1997), a nonlinear discrete-time forecast and linear observational system is written as

$$\mathbf{x}_i^{\rm t} = M_{i-1}(\mathbf{x}_{i-1}^{\rm a}) + \eta_i \tag{1}$$

and

$$\mathbf{y}_i^{\rm o} = \mathbf{H}_i\mathbf{x}_i^{\rm t} + \varepsilon_i, \tag{2}$$

where i is the time index; $\mathbf{x}_i^{\rm t} = \{x_i^{\rm t}(1), x_i^{\rm t}(2), \ldots, x_i^{\rm t}(n)\}^{\rm T}$ is the n-dimensional true state vector at time step i; $\mathbf{x}_{i-1}^{\rm a} = \{x_{i-1}^{\rm a}(1), x_{i-1}^{\rm a}(2), \ldots, x_{i-1}^{\rm a}(n)\}^{\rm T}$ is the n-dimensional analysis state vector which is an estimate of $\mathbf{x}_{i-1}^{\rm t}, M_{i-1}$ is a nonlinear forecast operator such as a weather forecast model; $\mathbf{y}_i^{\rm o}$ is an observational vector with dimension p_i; $\mathbf{H}_i$ is an observational matrix of dimension $p_i \times n$ that maps model states to the observational space, $\boldsymbol{\eta}_i$ and $\boldsymbol{\varepsilon}_i$ are the forecast error vector and the observational error vector respectively, which are assumed to be statistically independent of each other, time-uncorrelated, and have mean zero and covariance matrices $\mathbf{P}_i$ and $\mathbf{R}_i$ respectively. The goal of the EnKF assimilation is to find a series of analysis states $\mathbf{x}_i^{\rm a}$ that are sufficiently close to the corresponding true states $\mathbf{x}_i^{\rm t}$, using the information provided by M_i and $\mathbf{y}_i^{\rm o}$.

It is well-known that any EnKF assimilation scheme should include a forecast error inflation scheme. Otherwise, the EnKF may diverge (Anderson and Anderson, 1999). In this article, a procedure for estimating the multiplicative inflation factor of $\mathbf{P}_i$ and adjustment factor of $\mathbf{R}_i$ is proposed based on the SLS principle (Wang and Leblanc, 2008). The basic filter algorithm in this article uses perturbed observations (Burgers *et al.*, 1998) without localization (Houtekamer and Mitchell, 2001). The estimation steps of this algorithm equipped with SLS inflation are as follows.

(1) Calculate the perturbed forecast states

$$\mathbf{x}_{i,j}^{\rm f} = M_{i-1}(\mathbf{x}_{i-1,j}^{\rm a}), \tag{3}$$

where $\mathbf{x}_{i-1,j}^{\rm a}$ is the perturbed analysis state derived from the previous time step ($1 \leq j \leq m$, with m the number of ensemble members).

(2) Estimate the inflated forecast and adjusted observational error covariance matrices.
Define the forecast state $\mathbf{x}_i^{\rm f}$ to be the ensemble mean of $\mathbf{x}_{i,j}^{\rm f}$ and suppose that the initial forecast error covariance matrix is

$$\hat{\mathbf{P}}_i = \frac{1}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\rm f} - \mathbf{x}_i^{\rm f}) \cdot (\mathbf{x}_{i,j}^{\rm f} - \mathbf{x}_i^{\rm f})^{\rm T}, \tag{4}$$

and the initial observational error covariance matrix is $\mathbf{R}_i$. Then, the adjusted forms of forecast and observational error covariance matrices are $\lambda_i\hat{\mathbf{P}}_i$ and $\mu_i\mathbf{R}_i$ respectively.

There are several approaches for estimating the inflation factor λ_i and adjustment factor μ_i. For example, Wang and Bishop (2003), Li *et al.*(2009)

and Miyoshi (2011) use the first order least square of the squared observation-minus-forecast residual $\mathbf{d}_i \equiv \mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{f}}$ to estimate λ_i; Liang *et al.*(2011) maximize the likelihood of $\mathbf{d}_i$ to estimate λ_i and μ_i. In this article, we propose to use the SLS approach for estimating λ_i and μ_i. That is, λ_i and μ_i are estimated by minimizing the objective function

$$L_i(\lambda,\mu) = \mathrm{Tr}[(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mu\mathbf{R}_i)(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mu\mathbf{R}_i)^{\mathrm{T}}]. \quad (5)$$

This leads to

$$\hat{\lambda}_i = \frac{\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2} \quad (6)$$

and

$$\hat{\mu}_i = \frac{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2}. \quad (7)$$

(See Appendix A for detailed derivation.) Similar to Wang and Bishop (2003) and Li *et al.*(2009), this procedure does not use the Bayesian approach (Anderson 2007, 2009; Miyoshi 2011).

(3) Compute the perturbed analysis states

$$\mathbf{x}_{i,j}^{\mathrm{a}} = \mathbf{x}_{i,j}^{\mathrm{f}} + \hat{\lambda}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}(\mathbf{H}_i\hat{\lambda}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} + \hat{\mu}_i\mathbf{R}_i)^{-1}(\mathbf{y}_i^{\mathrm{o}} + \boldsymbol{\varepsilon}'_{i,j} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{f}}), \quad (8)$$

where $\boldsymbol{\varepsilon}'_{i,j}$ is a normal random variable with mean zero and covariance matrix $\hat{\mu}_i\mathbf{R}_i$ (Burgers *et al.*, 1998). Here $(\mathbf{H}_i\hat{\lambda}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} + \hat{\mu}_i\mathbf{R}_i)^{-1}$ can be effectively calculated using the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996; Tippett *et al.*, 2003; Liang *et al.*, 2011). Further, the analysis state $\mathbf{x}_i^{\mathrm{a}}$ is estimated as the ensemble mean of $\mathbf{x}_{i,j}^{\mathrm{a}}$. Finally, set $i = i+1$ and return to step (1) for the assimilation at next time step.

2.2. *EnKF with SLS inflation and new structure for forecast error covariance matrix*

By Eqs (1) and (3), the ensemble forecast error is $\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{t}}.\mathbf{x}_i^{\mathrm{f}}$ is an estimate of $\mathbf{x}_i^{\mathrm{t}}$ without knowing observations. The ensemble forecast error is initially estimated as $\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}}$, which is used to construct the forecast error covariance matrix in section 2.1. However, due to limited ensemble size and model error, $\mathbf{x}_i^{\mathrm{f}}$ can be biased. Therefore, $\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}}$ can be a biased estimate of $\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{t}}$.

In this article, we propose to use observations for improving the estimation of the ensemble forecast error. The idea is as follows: after analysis state $\mathbf{x}_i^{\mathrm{a}}$ is derived, it is generally a better estimate of $\mathbf{x}_i^{\mathrm{t}}$ than the forecast state $\mathbf{x}_i^{\mathrm{f}}$. So $\mathbf{x}_i^{\mathrm{f}}$ in Eq. (4) is substituted by $\mathbf{x}_i^{\mathrm{a}}$ for updating the forecast error covariance matrix. This procedure is repeated iteratively until the corresponding objective function (Eq. (5)) converges. For the computational details, step (2) in section 2.1 is modified to the following adaptive procedure.

(2a) Use step (2) in section 2.1 to inflate the initial forecast error covariance matrix to ${}_0\hat{\lambda}_{i0}\hat{\mathbf{P}}_i$ and adjust initial observational error covariance matrix to ${}_0\hat{\mu}_i\mathbf{R}_i$. Then use step (3) in section 2.1 to estimate the initial analysis state ${}_0\mathbf{x}_i^{\mathrm{a}}$ and set $k = 1$.

(2b) Update the forecast error covariance matrix as

$${}_k\hat{\mathbf{P}}_i = \frac{1}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})\cdot(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}}. \quad (9)$$

Then, adjust the forecast and observational error covariance matrices to ${}_k\hat{\lambda}_i\,{}_k\hat{\mathbf{P}}_i$ and ${}_k\hat{\mu}_i\mathbf{R}_i$, where

$${}_k\hat{\lambda}_i = \frac{\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2}, \quad (10)$$

and

$${}_k\hat{\mu}_i = \frac{\mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2}, \quad (11)$$

are estimated by minimizing the objective function

$${}_kL_i(\lambda,\mu) = \mathrm{Tr}[(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mu\mathbf{R}_i)(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mu\mathbf{R}_i)^{\mathrm{T}}]. \quad (12)$$

If ${}_k\mathrm{L}_i({}_k\hat{\lambda}_i, {}_k\hat{\mu}_i) < {}_{k-1}\mathrm{L}_i({}_{k-1}\hat{\lambda}_i, {}_{k-1}\hat{\mu}_i) - \delta$, where δ is a predetermined threshold to control the convergence of Eq. (12), then estimate the k-th updated analysis state as

$${}_k\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{f}} + {}_k\hat{\lambda}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}(\mathbf{H}_{ik}\hat{\lambda}_{ik}\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} + {}_k\hat{\mu}_{ik}\mathbf{R}_i)^{-1}(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{f}}), \quad (13)$$

set $k = k+1$ and return to Eq. (9); otherwise, accept ${}_{k-1}\hat{\lambda}_{ik-1}\hat{\mathbf{P}}_i$ and ${}_{k-1}\hat{\mu}_i\mathbf{R}_i$ as the estimated forecast and observational error covariance matrices at the ith time step, and go to step (3) in section 2.1.

A flowchart of our proposed assimilation scheme is shown in Figure 1. Moreover, our proposed forecast error covariance matrix (Eq. (9)) can be expressed as

$${}_k\lambda_{ik}\hat{\mathbf{P}}_i = \frac{{}_k\lambda_i}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}} + \frac{{}_k\lambda_i m}{m-1}(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}}, \quad (14)$$

which is a multiplicatively inflated sampling error covariance matrix plus an additive inflation matrix (see Appendix B for the proof).

2.3. *Notes*

2.3.1. *Correctly specified observational error covariance matrix*

If the observational error covariance matrix $\mathbf{R}_i$ is correctly known, then its adjustment is no longer required. In this

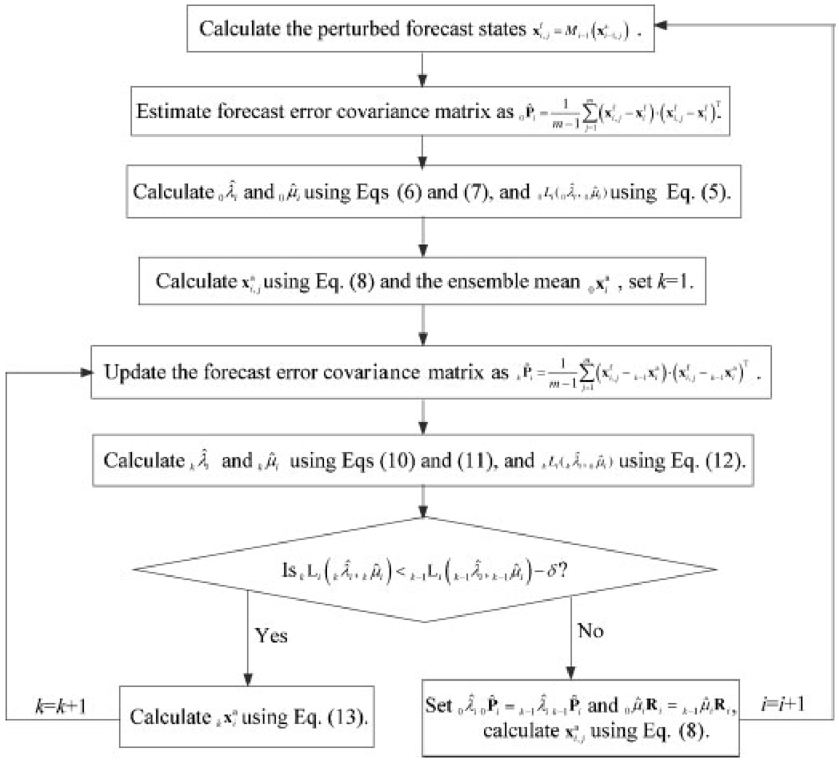

Figure 1. Flowchart of our proposed assimilation scheme.

case, the inflation factor ${}_k\hat{\lambda}_i$ can be estimated by minimizing the following objective function

$$L_i(\lambda) = \mathrm{Tr}[(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_i\,{}_k\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mathbf{R}_i)(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \lambda\mathbf{H}_i\,{}_k\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}} - \mathbf{R}_i)^{\mathrm{T}}]. \quad (15)$$

This leads to a simpler estimate

$${}_k\hat{\lambda}_i = \frac{\mathrm{Tr}[\mathbf{H}_i\,{}_k\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \mathbf{R}_i)]}{\mathrm{Tr}[\mathbf{H}_i\,{}_k\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\,{}_k\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}]}. \quad (16)$$

2.3.2. *Smoothing observational adjustment factor*

If the observational error covariance matrix $\mathbf{R}_i$ is assumed to be time invariant, then the adjustment factor μ_i should also be time invariant. In this case we propose to estimate μ_i as

$$\hat{\mu}_i = \frac{1}{K}(\tilde{\mu}_k + \sum_{k=i-K}^{i-1}\hat{\mu}_k), \quad (17)$$

where K is the number of previous assimilation time steps that involve smoothing adjustment factor of $\mathbf{R}_i$, $\tilde{\mu}_i$ is estimated using Eq. (11) and $\hat{\mu}_k (k = i-K, \cdots, i-1)$ are previous estimates using the smoothing procedure. A larger K does not increase the computational burden, but if K is too small, the smoothing effect may be weakened. In this article, K is selected as 10. As we demonstrate later, the estimated $\hat{\mu}_i$ using Eq. (11) has relatively large error and formula (17) gives a much better estimate.

2.3.3. *Validation statistics*

In any toy model, the 'true' state $\mathbf{x}_i^{\mathrm{t}}$ is known by experimental design. In this case, we can use the root-mean-square error (RMSE) of the analysis state to evaluate the accuracy of the assimilation results. The RMSE at the ith step is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\|\mathbf{x}_i^{\mathrm{a}} - \mathbf{x}_i^{\mathrm{t}}\|^2}, \quad (18)$$

where $\|\cdot\|$ represents the Euclidean norm and n is the dimension of the state vector. A smaller RMSE indicates a better performance of the assimilation scheme.

3. Experiment on the Lorenz-96 model

In this section we apply our proposed data assimilation schemes to a nonlinear dynamical system with properties relevant to realistic forecast problems: the Lorenz-96 model (Lorenz, 1996) with model error and a linear observational system. We evaluate the performances of the assimilation schemes in section 2 through the following experiments.

3.1. *Description of dynamic and observational systems*

The Lorenz-96 model (Lorenz, 1996) is a strongly nonlinear dynamical system with quadratic nonlinearity, governed by the equation

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = (X_{k+1} - X_{k-2})X_{k-1} - X_k + F, \quad (19)$$

where $k = 1, 2, \cdots, K$ ($K = 40$, hence there are 40 variables). For Eq. (19) to be well defined for all values of k, we define $X_{-1} = X_{K-1}, X_0 = X_K, X_{K+1} = X_1$. The dynamics of Eq. (19) are 'atmosphere-like' in that the three terms on the right-hand side consist of a nonlinear advection-like term, a damping term and an external forcing term respectively. These terms can be thought of as some atmospheric quantity (e.g. zonal wind speed) distributed on a latitude circle.

In our assimilation schemes, we set $F = 8$, so that the leading Lyapunov exponent implies an error-doubling time of about eight time steps, and the fractal dimension of the attractor is 27.1 (Lorenz and Emanuel, 1998). The initial condition is chosen to be $X_k = F$ when $k \neq 20$ and $\mathbf{X}_{20} = 1.001F$. We solve Eq. (19) using a fourth-order Runge-Kutta time integration scheme (Butcher, 2003) with a time step of 0.05 non-dimensional unit to derive the true state. This is roughly equivalent to 6 h in real time, assuming that the characteristic time-scale of the dissipation in the atmosphere is 5 days (Lorenz, 1996).

In this study, we assume the synthetic observations are generated at every model grid point by adding random noises that are multivariate normally distributed with mean zero and covariance matrix $\mathbf{R}_i$ to the true states. The leading diagonal elements of $\mathbf{R}_i$ are $\sigma_o^2 = 1$ and the off-diagonal elements at site pair (j, k) are

$$\mathbf{R}_i(j,k) = \sigma_o^2 \times 0.5^{\min\{|j-k|, 40-|j-k|\}} \quad (20)$$

By considering spatially correlated observational errors, the scheme may potentially be applied for assimilating remote sensing observations and radiances data.

We added model errors in the Lorenz-96 model because it is inevitable in real dynamic systems. Thus, we chose different values of F in our assimilation schemes, while retaining $F = 8$ when generating the 'true' state. We simulate observations every four time steps for 100 000 steps to ensure robust results (Sakov and Oke, 2008; Oke *et al.*, 2009). The ensemble size is selected as 30. The predetermined threshold δ to control the convergence of Eq. (12) is set to be 1, because the values of objective functions are in the order of 10^5. In most cases of the following experiment, the objective functions converge after 3–4 iterations, and the estimated analysis states also converge.

3.2. Comparison of assimilation schemes

In section 2.1 we outlined the EnKF assimilation scheme with SLS error covariance matrix inflation. In section 2.2, we summarized the EnKF assimilation scheme with the SLS error covariance matrix inflation and the new structure for the forecast error covariance matrix. In this section, we assess the impacts of these estimation methods on EnKF data assimilation schemes using the Lorenz-96 model.

The Lorenz-96 model is a forced dissipative model with a parameter F that controls the strength of the forcing (Eq. (19)). The model behaves quite differently with different values of F and produces chaotic systems with integer values of F larger than 3. As such, we used a set of values of F to simulate a wide range of model errors. In all cases, the true states were generated by a model with $F = 8$. These observations were then assimilated into models with $F = 4, 5, \ldots, 12$.

3.2.1. Correctly specified observational error covariance matrix

Suppose the observational error covariance matrix $\mathbf{R}_i$ is correctly specified, we first take inflation adjustment on $\hat{\mathbf{P}}_i$ in each assimilation cycle and estimate the inflation factor λ_i by the method described in section 2.1. Then we conduct the adaptive assimilation scheme with the new structure for the forecast error covariance matrix proposed in section 2.2.

Figure 2 shows the time-mean analysis RMSE of the two assimilation schemes averaged over 100 000 time steps, as a function of F. Overall, the analysis RMSE of the two assimilation schemes gradually grows when the model error increases. When F is around the true value 8, the two assimilation schemes have almost indistinguishable values of the analysis RMSE. However, when F becomes increasingly distant from 8, the analysis RMSE of the assimilation scheme with the new structure for the forecast error covariance matrix becomes progressively smaller than that of the assimilation scheme with the forecast error covariance matrix inflation only.

For the Lorenz-96 model with large error ($F = 12$), the time-mean analysis RMSEs of the two assimilation schemes are listed in Table 1, as well as the time-mean values of the objective functions. The EnKF assimilation scheme is also included for comparison. These results show clearly that our two schemes have significantly smaller RMSEs than the EnKF assimilation scheme. Moreover, the assimilation scheme with the new structure of the forecast error covariance

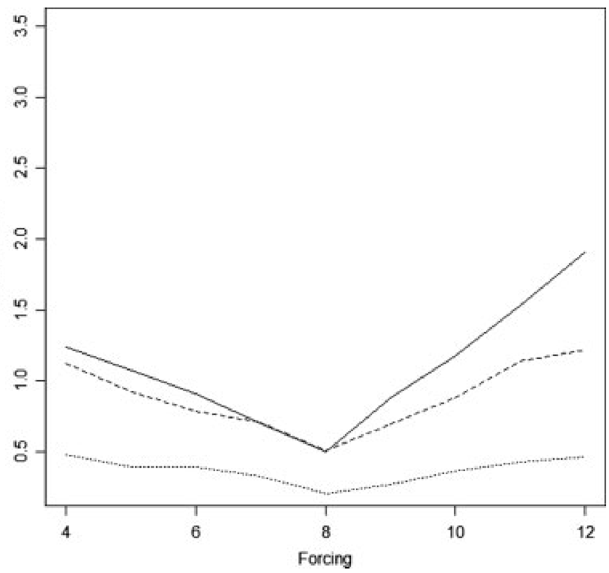

Figure 2. Time-mean values of the analysis RMSE as a function of forcing F when observational errors are spatially correlated and their covariance matrix is correctly specified, by using three EnKF schemes. (1) SLS only (solid line, described in section 2.1); (2) SLS and new structure (dashed line, described in section 2.2); and (3) SLS and true ensemble forecast error (dotted line, described in section 2.2).

Table 1. The time-mean analysis RMSE and the time-mean objective function values in four EnKF schemes for the Lorenz-96 model when observational errors are spatially correlated and their covariance matrix is correctly specified: (1) EnKF (non-inflation); (2) the SLS scheme in section 2.1 (SLS); (3) the SLS scheme in section 2.2 (SLS and new structure); (4) the SLS scheme in the discussion (SLS and true ensemble forecast error). The forcing term $F = 12$.

EnKF schemes	Time-mean RMSE	Time-mean L
Non-inflation	5.65	2298754
SLS	1.89	148468
SLS and new structure	1.22	38125
SLS and true ensemble forecast error	0.48	19652

matrix performs much better than the assimilation scheme with forecast error covariance matrix inflation only.

3.2.2. *Incorrectly specified observational error covariance matrix*

In this section, we suppose that the observational error covariance matrix is correct only up to a constant factor. We estimate this factor using different estimation methods and evaluate the corresponding assimilation results.

We set the observational error covariance matrix $\mathbf{R}_i$ as four times the true matrix and introduce another factor μ_i to adjust $\mathbf{R}_i$. We conduct the assimilation scheme in four cases: (i) inflate forecast and adjust observational error covariance matrices only (section 2.1); (ii) inflate forecast and adjust observational error covariance matrices and use the smoothing technique for μ_i (section 2.3.2) where the number K (the smoothing interval) is 10; (iii) inflate forecast and adjust observational error covariance matrices and use the new structure for the forecast error covariance matrix (section 2.2); (iv) inflate forecast and adjust observational error covariance matrices with smoothing μ_i and use the new structure for the forecast error covariance matrix. The model parameter F again takes values 4, 5, ..., 12 when assimilating observations, but $F = 8$ is used when generating the true states in all cases.

Figure 3 shows the time-mean analysis RMSE of the four cases averaged over 100 000 time steps, as a function of F. Overall, the analysis RMSE of the four cases gradually grows when the model error increases. However, the analysis RMSE in the cases using the new structure for the forecast error covariance matrix (cases 3 and 4) are smaller than those in the cases using the error covariance matrix inflation technique only (cases 1 and 2). Moreover, using the smoothing technique for estimating the adjustment factor μ_i (cases 2 and 4) can improve the assimilation result to some extent.

For the Lorenz-96 model with model parameter $F = 12$, the time-mean analysis RMSE of the four cases are listed in Table 2, along with the time-mean values of the objective functions. These results show clearly that when the observational error covariance matrix is specified incorrectly, the assimilation result is much better if the new structure for the forecast error covariance matrix is used (cases 3 and 4). Moreover, the assimilation result can be further improved by using the smoothing technique to estimate μ_i (cases 2 and 4).

The estimated $\hat{\mu}_i$ over 100 000 time steps in the two cases of using the new structure of the forecast error covariance matrix (cases 3 and 4) are plotted in Figure 4. It can be seen that the time-mean value of estimated $\hat{\mu}_i$ is 0.75 in case 3, but is 0.36 after using the smoothing technique (case 4).

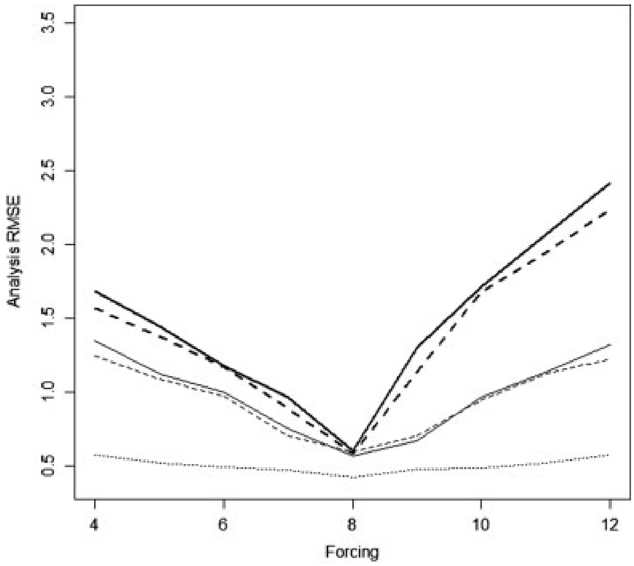

Figure 3. Time-mean values of the analysis RMSE as a function of forcing F when observational errors are spatially correlated and their covariance matrix is incorrectly specified, by using five EnKF schemes. (1) SLS only (thick solid line); (2) SLS with smoothing μ_i (thick dashed line); (3) SLS and new structure (thin solid line); (4) SLS with smoothing μ_i and new structure (thin dashed line); and (5) SLS and true ensemble forecast error (dotted line).

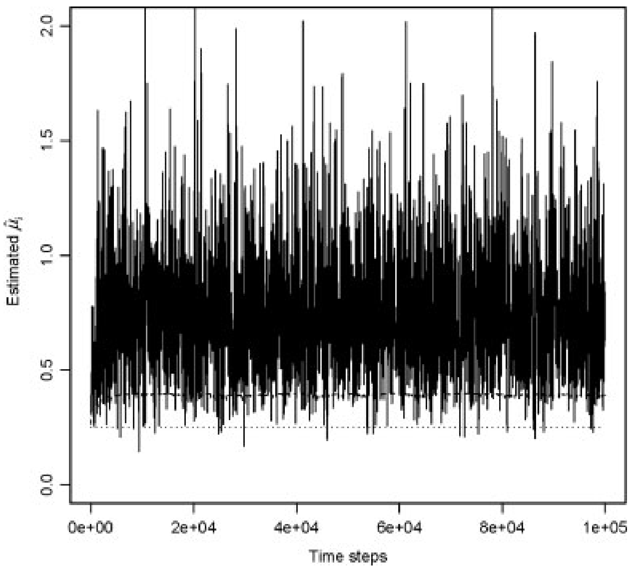

Figure 4. The times series of estimated $\hat{\mu}_i$ when observational error covariance matrix is incorrectly specified (solid line) and using smoothing technique (dashed line). The dotted line is the correct scale (0.25) of the observational error covariance matrix.

Table 2. The time-mean analysis RMSE and the time-mean objective function values in five EnKF schemes for Lorenz-96 model when observational errors are spatially correlated and their covariance matrix is incorrectly specified: (1) SLS; (2) SLS with smoothing μ_i; (3) SLS and new structure; (4) SLS with smoothing μ_i and new structure; (5) SLS and true ensemble forecast error. The forcing term $F = 12$.

EnKF schemes	Ensemble size 30		Ensemble size 20	
	Time-mean RMSE	Time-mean L	Time-mean RMSE	Time-mean L
SLS	2.43	1426541	3.51	1492685
SLS with smoothing μ_i	2.25	127643	2.86	207964
SLS and new structure	1.35	41326	1.45	95685
SLS with smoothing μ_i and new structure	1.22	37953	1.40	58466
SLS and true ensemble forecast error	0.58	21585	0.60	21355

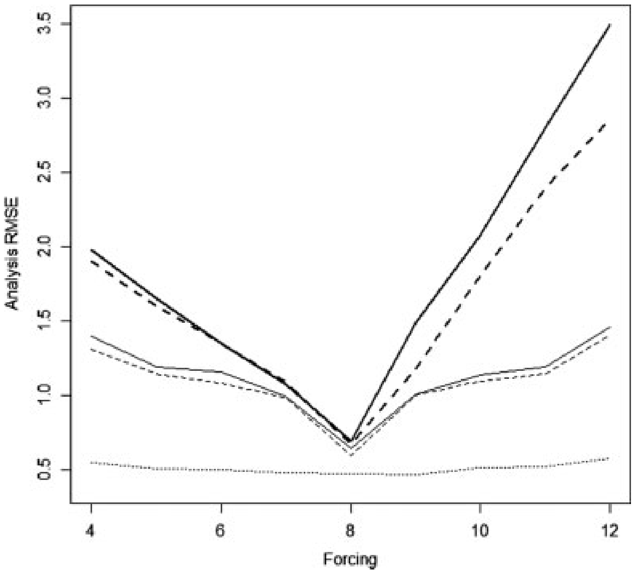

Figure 5. Similar to Figure 3, but ensemble size is 20.

The latter is closer to the reciprocal of the constant that we multiplied to the observational error covariance matrix (0.25).

To investigate the effect of ensemble size on the assimilation result, Figure 3 is reproduced with the ensemble size 20. The results are shown in Figure 5 as well as in Table 2. Generally speaking, Figures 5 is similar to Figure 3, but with larger analysis error. This indicates that the smaller ensemble size corresponds to the larger forecast error and the analysis error. We also repeated our analysis with the ensemble size 10. However in this case, both the inflation and new structure are not effective. This could be due to the ensemble size 10 being too small to generate robust covariance estimation.

4. Discussion and conclusions

It is well-known that accurately estimating the error covariance matrix is one of the most important steps in data assimilation. In the EnKF assimilation scheme, the forecast error covariance matrix is estimated as the sampling covariance matrix of the ensemble forecast states. Due to the limited ensemble size and model error, however, the forecast error covariance matrix is usually underestimated, which may lead to the divergence of the filter. So we multiplied the estimated forecast error covariance matrix by an inflation factor λ_i and proposed the SLS estimation for this factor.

In fact, the true forecast error should be represented as the ensemble forecast states minus the true state. However, since in real problems the true state is not available, we use the ensemble mean of the forecast states instead. Consequently the forecast error covariance matrix is represented as the sampling covariance matrix of the ensemble forecast states. If the model error is large, however, the ensemble mean of the forecast states may be far from the true state. In this case, the estimated forecast error covariance matrix will remain far from the truth, no matter which inflation technique is used.

To verify this point, a number of EnKF assimilation schemes with necessary error covariance matrix inflation are applied to the Lorenz-96 model, but with the forecast state $\mathbf{x}_i^{\mathrm{f}}$ in the forecast error covariance matrix (Eq. (4)) substituted by the true state $\mathbf{x}_i^{\mathrm{t}}$. The corresponding RMSEs are shown in Figures 2, 3 and 5 and Tables 1 and 2. All the figures and tables show that the analysis RMSE is significantly reduced.

However, since the true state $\mathbf{x}_i^{\mathrm{t}}$ is unknown, we use the analysis state $\mathbf{x}_i^{\mathrm{a}}$ to replace the forecast state $\mathbf{x}_i^{\mathrm{f}}$, because $\mathbf{x}_i^{\mathrm{a}}$ is closer to $\mathbf{x}_i^{\mathrm{t}}$ than $\mathbf{x}_i^{\mathrm{f}}$. To achieve this goal, a new structure for the forecast error covariance matrix and an adaptive procedure for estimating the new structure are proposed here to iteratively improve the estimation. As shown in section 3, the RMSEs of the corresponding analysis states are indeed smaller than those of the EnKF assimilation scheme with the error covariance matrix inflation only. For example, in the experiment in section 3.1, when the error covariance matrix inflation technique is applied, the RMSE is 1.89, which is much smaller than that for the original EnKF. When we further used the new structure of the forecast error covariance matrix in addition to the inflation, the RMSE is reduced to 1.22 (see Table 1).

In practice, the observational error covariance matrix is not always known correctly, and hence needs to be adjusted too. Another factor μ_i is introduced to adjust the observational error covariance matrix, which can be estimated simultaneously with λ_i by minimizing the second-order least squares function of the squared observation-minus-forecast residual. If the observational error covariance matrices are time invariant, a smoothing technique (Eq. (17)) can be used to improve the estimation. Our experiment in section 3.2 shows that the smoothing technique plays an important role in retrieving the correct scale of the observational error covariance matrix.

The second-order least squares function of the squared observation-minus-forecast residual seems to be a good objective function to quantify the goodness of fit of the error covariance matrix. The SLS can be used to estimate the factors for adjusting both the forecast and observational error covariance matrices, while the first order method can estimate only the inflation factor of the forecast error covariance matrix. The SLS can also provide a stopping criterion for the iteration in the adaptive estimation procedure when the new structure of the forecast error covariance matrix is used. This is important for preventing the proposed forecast error covariance matrix to depart from the truth in the iteration. In most cases in this study, the minimization algorithms converge after 3–4 iterations and the objective function decreases sharply. On the other hand, the improved forecast error covariance matrix indeed leads to the improvement of analysis state. In fact, as shown in Tables 1 and 2, a small objective function value always corresponds to a small RMSE of the analysis state.

We also investigated the difference between our proposed scheme and the assimilation scheme proposed by Liang *et al.*(2011). Generally speaking, the RMSE of the analysis state derived using the MLE inflation scheme proposed by Liang *et al.*(2011) is slightly smaller than that derived using the SLS inflation scheme only, but is larger than that derived using the SLS inflation with the new structure for forecast error covariance matrix. For example, for the Lorenz-96 model with forcing $F = 12$, the RMSE is 1.69 for MLE inflation (table 3 in Liang *et al.*, 2011), 1.89 for SLS inflation only and 1.22 for SLS inflation and new structure (Table 1). Whether this is a general rule or not is unclear, and is subject to further investigation. However, in the MLE inflation scheme, the objective function is nonlinear and especially involves the determinant of the observation-minus-forecast residual's covariance matrix, which is computationally expensive. The objective function in our proposed scheme is quadratic, so its minimizer is analytic and can be calculated easily.

Similar to other EnKF assimilation schemes with single parameter inflation, this study also assumes the inflation factor to be constant in space. Apparently this is not the case in many practical applications, especially when the observations are distributed unevenly. Persistently applying the same inflation values that are reasonably large to address problems in densely observed areas to all state variables can systematically overinflate the ensemble variances in sparsely observed areas (Hamill and Whitaker, 2005; Anderson, 2009). Even if the adaptive procedure for estimating the error covariance matrix is applied, the problem may still exist to some extent. In the two case studies conducted in this article, the observational systems are distributed relatively evenly.

In our future study we will investigate how to modify the adaptive procedure to suit the system with unevenly distributed observations. We also plan to apply our methodology to error covariance localization and to validate the proposed methodologies using more sophisticated dynamic and observational systems.

Acknowledgements

This work was supported by National High-tech R&D Program of China (Grant No. 2009AA122104), National Program on Key Basic Research Project of China (Grant Nos 2010CB951604), the National Natural Science foundation of China General Program (Grant Nos 40975062), and the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors gratefully acknowledge the anonymous reviewers for their constructive and relevant comments, which helped greatly in improving the quality of this manuscript. The authors are also grateful to the editors for their hard work and suggestions on this manuscript.

Appendix A

The forecast error covariance matrix $\hat{\mathbf{P}}_i$ is inflated to $\lambda\hat{\mathbf{P}}_i$. The estimation of the inflation factor λ is based on the observation-minus-forecast residual

$$\mathbf{d}_i = \mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{f}} = (\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}) + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}) \tag{A1}$$

The covariance matrix of the random vector $\mathbf{d}_i$ can be expressed as a second-order regression equation (Wang and Leblanc, 2008):

$$E[((\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}) + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))((\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}) + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] = \mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} + \Xi \tag{A2}$$

where E is the expectation operator and Ξ is the error matrix. The left-hand side of (A2) can be decomposed as

$$\begin{aligned} &E[((\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}) + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))((\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}) + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] \\ &= E[(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})^{\mathrm{T}}] \\ &\quad + E[(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] \\ &\quad + E[(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] \\ &\quad + E[(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})^{\mathrm{T}}] \end{aligned} \tag{A3}$$

Since the forecast and observational errors are statistically independent, we have

$$\begin{aligned} &E[(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})^{\mathrm{T}}] \\ &\quad = \mathbf{H}_i E[(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})^{\mathrm{T}}] = \mathbf{0} \end{aligned} \tag{A4}$$

and

$$\begin{aligned} &E[(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] \\ &\quad = E[(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}}]\mathbf{H}_i^{\mathrm{T}} = \mathbf{0}. \end{aligned} \tag{A5}$$

From Eq. (2), $\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}}$ is the observational error at the ith time step, and hence

$$E[(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})(\mathbf{y}_i^{\mathrm{o}} - \mathbf{H}_i\mathbf{x}_i^{\mathrm{t}})^{\mathrm{T}}] = \mathbf{R}_i. \tag{A6}$$

Further, since the forecast state $\mathbf{x}_{i,j}^{\mathrm{f}}$ is treated as a random vector with the true state $\mathbf{x}_i^{\mathrm{t}}$ as its population mean,

$$\begin{aligned} &E[(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))(\mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}}))^{\mathrm{T}}] \\ &= \mathbf{H}_i E[(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_i^{\mathrm{t}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}}]\mathbf{H}_i^{\mathrm{T}} \\ &\approx \mathbf{H}_i \frac{\lambda}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}} \\ &= \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}. \end{aligned} \tag{A7}$$

Substituting Eqs (A3)–(A7) into Eq (A2), we have

$$\mathbf{R}_i + \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{T} \approx \mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} + \Xi. \tag{A8}$$

It follows that the second-order moment statistic of error Ξ can be expressed as

$$\begin{aligned} \mathrm{Tr}[\Xi\Xi^{\mathrm{T}}] &\approx \mathrm{Tr}[(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \mathbf{R}_i - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}) \\ &\quad (\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \mathbf{R}_i - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})^{\mathrm{T}}] \equiv L_i(\lambda). \end{aligned} \tag{A9}$$

Therefore, λ can be estimated by minimizing objective function $L_i(\lambda)$. Since $L_i(\lambda)$ is a quadratic function of λ with positive quadratic coefficients, the inflation factor can be easily expressed as

$$\hat{\lambda}_i = \frac{\mathrm{Tr}[\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \mathbf{R}_i)]}{\mathrm{Tr}[\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}]}. \tag{A10}$$

Similarly, if the amplitude of the observational error covariance matrix is not correct, we can adjust $\mathbf{R}_i$ to $\mu_i\mathbf{R}_i$ as well (Li *et al.*, 2009; Liang *et al.*, 2011). Then the objective function becomes

$$\begin{aligned} L_i(\lambda, \mu) &= \mathrm{Tr}[(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} - \mu\mathbf{R}_i - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}} \\ &\quad - \mu\mathbf{R}_i - \lambda\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})^{\mathrm{T}}]. \end{aligned} \tag{A11}$$

As a bivariate function of λ and μ, the first partial derivative with respect to the two parameters respectively are

$$\lambda \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}) + \mu \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i) - \mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}), \quad (A12)$$

and

$$\lambda \mathrm{Tr}(\mathbf{H}_i, \hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i) + \mu \mathrm{Tr}(\mathbf{R}_i^{\mathrm{T}}\mathbf{R}_i) - \mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i). \quad (A13)$$

Setting Eqs (A12) and (A13) to zero and solving them lead to

$$\hat{\lambda}_i = \frac{\mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i)\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2} = \frac{\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_i\mathbf{P}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2} \quad (A14)$$

and

$$\hat{\mu}_i = \frac{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i) - \mathrm{Tr}(\mathbf{d}_i\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2} = \frac{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i\mathbf{d}_i) - \mathrm{Tr}(\mathbf{d}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{d}_i)\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)}{\mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}})\mathrm{Tr}(\mathbf{R}_i^2) - \mathrm{Tr}(\mathbf{H}_i\hat{\mathbf{P}}_i\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i)^2}. \quad (A15)$$

Appendix B

In fact,

$$\begin{aligned}
&\frac{{}_k\lambda_i}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \\
&= \frac{{}_k\lambda_i}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}} + \mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}}) (\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}} + \mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \\
&= \frac{{}_k\lambda_i}{m-1}\left\{\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}} + \sum_{j=1}^{m}(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \right. \\
&\quad \left. + \sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} + \sum_{j=1}^{m}(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}}\right\}.
\end{aligned} \quad (B1)$$

Since $\mathbf{x}_i^{\mathrm{f}}$ is the ensemble mean forecast, we have

$$\begin{aligned}
&\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \\
&= \left[\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})\right](\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \\
&= \left[\sum_{j=1}^{m}\mathbf{x}_{i,j}^{\mathrm{f}} - m\frac{1}{m}\sum_{j=1}^{m}\mathbf{x}_{i,j}^{\mathrm{f}}\right](\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_{i\ i}^{\mathrm{a}})^{\mathrm{T}} \\
&= \mathbf{0}
\end{aligned} \quad (B2)$$

and similarly

$$\sum_{j=1}^{m}(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}} = \mathbf{0}. \quad (B3)$$

That is, the last two terms of Eq. (B1) vanish. Therefore our proposed forecast error covariance matrix can be expressed as

$$\begin{aligned}
&\frac{{}_k\lambda_i}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_{i,j}^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}} \\
&= \frac{{}_k\lambda_i}{m-1}\left\{\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}} + m(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}}\right\} \\
&= \frac{{}_k\lambda_i}{m-1}\sum_{j=1}^{m}(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})(\mathbf{x}_{i,j}^{\mathrm{f}} - \mathbf{x}_i^{\mathrm{f}})^{\mathrm{T}} + \frac{{}_k\lambda_i m}{m-1}(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})(\mathbf{x}_i^{\mathrm{f}} - {}_{k-1}\mathbf{x}_i^{\mathrm{a}})^{\mathrm{T}}.
\end{aligned} \quad (B4)$$

References

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**: 2741–2758.

Anderson JL. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus* **59A**: 210–224.

Anderson JL. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus* **61A**: 72–83.

Bai YL, Li X. 2011. Evolutionary algorithm-based error parameterization methods for data assimilation. *Mon. Weather Rev.* **139**: 2668–2685.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**: 1719–1724.

Butcher JC. 2003. *Numerical Methods for Ordinary Differential Equations.* John Wiley & Sons: Chichester; 425 pp.

Constantinescu M, Sandu A, Chai T, Carmichael GR. 2007. Ensemble-based chemical data assimilation. I: General approach. *Q. J. R. Meteorol. Soc.* **133**: 1229–1243.

Dee DP. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.* **123**: 1128–1145.

Dee DP, Da Silva AM. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Weather Rev.* **127**: 1822–1834.

Dee DP, Gaspari G, Redder C, Rukhovets L, Da Silva AM. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Weather Rev.* **127**: 1835–1849.

Evensen G. 1994a. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162.

Evensen G. 1994b. Inverse methods and data assimilation in nonlinear ocean models. *Physica D* **77**: 108–129.
Evensen G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**: 343–367.
Golub GH, Van Loan CF. 1996. *Matrix Computations.* The Johns Hopkins University Press: Baltimore, MD: 694 pp.
Hamill TM, Whitaker JS. 2005. Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Weather Rev.* **133**: 3132–3147.
Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.
Ide K, Courtier P, Michael G, Lorenc AC. 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn* **75**: 181–189.
Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflatioin and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 523–533.
Liang X, Zheng XG, Zhang SP, Wu GC, Dai YJ, Li Y. 2011. Maximum likelihood estimation of inflation factors on forecast error covariance matrix for ensemble Kalman filter assimilation. *Q. J. R. Meteorol. Soc.* **138**: 263–273.
Lorenz EN. 1996. Predictability – a problem partly solved. Paper presented at *Seminar on Predictability*, ECMWF: Reading, UK.
Lorenz EN, Emanuel KA. 1998. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.* **55**: 399–414.
Miyoshi T. 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Weather Rev.* **139**: 1519–1535.
Oke PR, Sakov P, Schulz E. 2009. A comparison of shelf observation platforms for assimilation in an eddy-resolving ocean model. *Dyn. Atmos. Oceans* **48**: 121–142.
Reichle RH. 2008. Data assimilation methods in the earth sciences. *Adv. Water Resour.* **31**: 1411–1418.
Sakov P, Oke PR. 2008. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus* **60A**: 361–371.
Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Notes and correspondence: ensemble square root filters. *Mon. Weather Rev.* **131**: 1485–1490.
Wang L, Leblanc A. 2008. Second-order nonlinear least squares estimation. *Ann. Inst. Stat. Math.* **60**: 883–900.
Wang X, Bishop CH. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**: 1140–1158.
Zheng XG. 2009. An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Adv. Atmos. Sci.* **26**: 154–160.

