Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2009), 61A, 601–609
Printed in Singapore. All rights reserved

TELLUS

Using the ensemble Kalman filter to estimate multiplicative model parameters

By XIAOSONG YANG[1,2]* and TIMOTHY DELSOLE[1,2], [1]*Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Rd., Suite 302, Calverton, MD, USA;* [2]*George Mason University, Fairfax, VA, USA*

(Manuscript received 28 July 2008; in final form 25 May 2009)

ABSTRACT

This paper proposes a simple approach to estimating multiplicative model parameters using the ensemble square root filter. The basic idea, following previous studies, is to augment the state vector by the model parameters. While some success with this approach has been reported if the model parameters enter as additive terms in the tendency equations, this approach is problematic if the model parameters are multiplied by the state variables. The reason for this difficulty is that multiplicative parameters change the dynamical properties of the model, and in particular can cause the model to become dynamically unstable. This paper shows that model instability can be avoided if the usual persistence model for parameter update is replaced by a temporally smoothed version of the update model. In addition, the augmentation approach can be interpreted as two simultaneously decoupled ensemble Kalman filters for the model state and the parameter state, respectively. Implementation of the parameter estimation does not require changing the data assimilation algorithm—it just has to be supplemented by a parameter estimation step that is similar to the state estimation step. Covariance localization is found to be necessary not only for the model state, but also for augmented model parameters, if they are spatially dependent. The new formulation is illustrated with the Lorenz-96 model and shown to be capable of estimating additive and multiplicative model parameters, as well as the state, under relatively challenging conditions (e.g. using 20 observations to estimate 120 unknown variables).

1. Introduction

The ensemble Kalman filter (EnKF), pioneered by Evensen (1994), has gained popularity for geophysical data assimilation due to its simple implementation without the need for an adjoint model. However, the EnKF effectively assumes a perfect model and hence does not account for deficiencies in the model, for example, model bias and uncertain model parameters. A commonly used method for accounting for uncertain model parameters in Kalman filtering is the ‘state space augmentation method’ (Anderson, 2001; Baek et al., 2006; Gillijns and De Moor, 2007), which involves augmenting the state vector with some model parameters and then constructing a Kalman filter (KF) for the augmented model. In the state augmented EnKF, correlations between the model state and the model parameters are estimated from a finite ensemble (Baek et al., 2006; Zupanski and Zupanski, 2006). The state space augmentation method has been successfully used to estimate model parameters that enter additively in the tendency equations (Baek et al., 2006). However, augmentation increases the computational burden because it increases the dimension of the error covariance matrix. This computational load can be reduced in the local ensemble Kalman filter (LEKF), since LEKF would solve the augmented EnKF equations in a series of local regions with small state dimension using advanced parallel computing techniques (Baek et al., 2006).

An alternative approach to dealing with model deficiencies is the two-stage KF proposed by Friedland (1969), where the estimation of the state and the bias in the innovation increments are separated. Dee and da Silva (1998) derived a useful, suboptimal variation of the two-stage filter and applied it to estimate forecast bias from the innovation increments in atmospheric data assimilation (Dee and Todling, 2000). An ensemble-based two-stage KF was developed by Gillijns and De Moor (2007). It is assumed in the two-stage filter that the uncertainties in the forecast model state and the bias are uncorrelated. Since this assumption is not generally satisfied, the two-stage filter is suboptimal.

Although the state augmentation method has been applied successfully to estimate model parameters that enter additively in the tendency equations (Baek et al., 2006), it can be problematic if the parameters are multiplied by the state variables in the tendency equations. The reason for this difficulty is that multiplicative parameters change the dynamical properties of the model. As a result, the ensemble of model parameters

*Corresponding author.
e-mail: xyang@cola.iges.org
DOI: 10.1111/j.1600-0870.2009.00407.x

specified by the filter may lead one realization of the model to be dynamically unstable, in which case the filter 'blows up', that is, the forecast and analysis become unbounded. We propose a method for avoiding filter blow up. In most applications of the state augmentation approach, the model parameters are evolved forward in time using a persistence model (Dee and da Silva, 1998; Baek et al., 2006; Zupanski and Zupanski, 2006). In this paper, we introduce a temporally smoothed version of the persistence model. This smoothing model damps the rapid temporal variation of the model parameters, leading to smooth temporal variation. It will be shown that this temporal smoothing can prevent model blow-up. We further suggest applying covariance localization to the parameters, when the parameters are spatially dependent.

The paper is organized as follows. In Section 2, a sequential square root EnKF for state and parameter estimation is derived. The numerical models and experimental setup are introduced in Section 3, and experimental results are presented in Section 4. The paper ends with conclusions.

2. Augmented ESRF for state and parameter estimation

Let $\mathbf{x}^f$ be the m-dimensional vector specifying the forecast state, and let $\mathbf{y}$ be the p-dimensional observation vector derived from

$$\mathbf{y} = \mathbf{H}_x \mathbf{x}^f + \varepsilon, \tag{1}$$

where $\mathbf{H}_x$ is the $p \times m$ operator mapping the model state to the observation space (here, assumed linear, though it need not be), ε is the observational noise, assumed to be Gaussian with $\mathbf{0}$ mean and covariance matrix $\mathbf{R}$. Suppose the model depends on q parameters that we want to estimate. In the augmentation method, the estimated state vector includes the model state vector plus the uncertain model parameters. Let $\mathbf{b}^f$ be a q-dimensional vector containing the uncertain model parameters. Then, define the augmented state vector $\mathbf{x}^{*f}$ with dimension $m + q$ as

$$\mathbf{x}^{*f} = \begin{bmatrix} \mathbf{x}^f \\ \mathbf{b}^f \end{bmatrix}. \tag{2}$$

Since model parameters usually are not observed, the observation operator for the augmented model is

$$\mathbf{H}^* = [\,\mathbf{H}_x \quad \mathbf{0}\,], \tag{3}$$

where $\mathbf{0}$ is a $p \times q$ zero matrix. Note that $\mathbf{H}^*\mathbf{x}^{*f} = \mathbf{H}\mathbf{x}^f$, so (1) is still the observation equation for the augmented filter.

The standard KF equation for the analysed ensemble mean is

$$\bar{\mathbf{x}}^{*a} = \bar{\mathbf{x}}^{*f} + \mathbf{K}(\mathbf{y} - \mathbf{H}^*\bar{\mathbf{x}}^{*f}), \tag{4}$$

where the Kalman gain matrix $\mathbf{K}$ is defined as

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{*T} (\mathbf{H}^* \mathbf{P}^f \mathbf{H}^{*T} + \mathbf{R})^{-1}, \tag{5}$$

$\mathbf{P}^f$ is the background-error covariance matrix for $\mathbf{x}^*$, the superscript a refers to the assimilated state, the overbar denotes the ensemble mean, and superscript T denotes the matrix transpose. Following Whitaker and Hamill (2002), we apply the ensemble square root filter to estimate the augmented state, and update the ensemble mean and ensemble perturbation separately without perturbed observations. Let n be the ensemble size. Let the difference between the jth ensemble forecast member and the corresponding ensemble mean be collected into matrices as follows

$$\mathbf{X}^{f'} = \left[\,\mathbf{x}_1^f - \bar{\mathbf{x}}^f, \quad \mathbf{x}_2^f - \bar{\mathbf{x}}^f, \quad \cdots, \quad \mathbf{x}_n^f - \bar{\mathbf{x}}^f\,\right], \tag{6}$$

$$\mathbf{B}^{f'} = \left[\,\mathbf{b}_1^f - \bar{\mathbf{b}}^f, \quad \mathbf{b}_2^f - \bar{\mathbf{b}}^f, \quad \cdots, \quad \mathbf{b}_n^f - \bar{\mathbf{b}}^f\,\right]. \tag{7}$$

An $(m + q) \times n$ dimensional ensemble perturbation matrix of the augmented vector can be defined as

$$\mathbf{X}^{*f'} = \begin{bmatrix} \mathbf{X}^{f'} \\ \mathbf{B}^{f'} \end{bmatrix}. \tag{8}$$

In the EnKF, $\mathbf{P}^f$ is approximated by sample estimates of covariance based on the ensemble. Thus

$$\mathbf{P}^f = \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{P}_{bx}^T \\ \mathbf{P}_{bx} & \mathbf{P}_{bb} \end{bmatrix}, \tag{9}$$

where $\mathbf{P}_{xx} = \mathbf{X}^{f'}\mathbf{X}^{f'T}/(n-1)$ is the forecast ensemble covariance of $\mathbf{x}$, $\mathbf{P}_{bx} = \mathbf{B}^{f'}\mathbf{X}^{f'T}/(n-1)$ is the forecast ensemble cross covariance between $\mathbf{b}$ and $\mathbf{x}$, and $\mathbf{P}_{bb} = \mathbf{B}^{f'}\mathbf{B}^{f'T}/(n-1)$ is the forecast ensemble covariance of $\mathbf{b}$.

Substituting (2) and (3) into (4) and (5), and invoking (9), the full KF eqs (4) and (5) can be written equivalently as

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}_x\left(\mathbf{y} - \mathbf{H}_x\bar{\mathbf{x}}^f\right), \tag{10}$$

$$\bar{\mathbf{b}}^a = \bar{\mathbf{b}}^f + \mathbf{K}_b\left(\mathbf{y} - \mathbf{H}_x\bar{\mathbf{x}}^f\right), \tag{11}$$

where the Kalman gain matrices for the model state and the augmented parameters are defined as

$$\mathbf{K}_x = \mathbf{P}_{xx}\mathbf{H}_x^T\left(\mathbf{H}_x\mathbf{P}_{xx}\mathbf{H}_x^T + \mathbf{R}\right)^{-1}, \tag{12}$$

and

$$\mathbf{K}_b = \mathbf{P}_{bx}\mathbf{H}_x^T\left(\mathbf{H}_x\mathbf{P}_{xx}\mathbf{H}_x^T + \mathbf{R}\right)^{-1}. \tag{13}$$

It is important to recognize that the state update eq. (10) is exactly the same as the original KF update eq. (4). Therefore, in the augmented KF, the state can be updated using precisely the same algorithm as the non-augmented filter. Furthermore, the parameter update eq. (11) has the same structural form as the state update, implying that the parameter can be updated using an algorithm that is structurally similar to the non-augmented filter.

If the observations are independent, then, following Whitaker and Hamill (2002), the observations can be assimilated one at a time in sequence as

$$\mathbf{x}_j^{\prime a} = \mathbf{x}_j^{\prime f} - \alpha\mathbf{K}_x\mathbf{H}_x\mathbf{x}_j^{\prime f}, \tag{14}$$

$$\mathbf{b}_j^{\prime a} = \mathbf{b}_j^{\prime f} - \alpha \mathbf{K}_b \mathbf{H}_x \mathbf{x}_j^{\prime f}, \tag{15}$$

where

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}_x \mathbf{P}_{xx} \mathbf{H}_x^T + \mathbf{R}}}\right)^{-1}, \tag{16}$$

and $j = 1 \ldots n$ denotes the jth ensemble member. For a single observation, $\mathbf{H}_x \mathbf{P}_{xx} \mathbf{H}_x^T$ and $\mathbf{R}$ are reduced to scalars.

Equations (10)–(16) define the decoupled augmented filter for estimating uncertain parameters in the EnKF. Note that the covariance matrix $\mathbf{P}_{bb}$ of the parameter state does not appear in the update equations. Although $\mathbf{P}_{bb}$ does not influence the assimilation of data, it does quantify the uncertainty in the estimated value of $\mathbf{b}$.

The model state and parameter values are coupled through the forecast model. Let the mapping from the initial state to the final state by the non-linear forecast model be represented as

$$\mathbf{x}_{k+1,j}^{f} = \mathbf{M}\left(\mathbf{x}_{k,j}^{a}, \mathbf{b}_{k,j}^{a}\right), \tag{17}$$

where the operator $\mathbf{M}$ maps the state vector at time instant k to time instant $k + 1$, and j denotes the jth ensemble member. The state augmentation approach also requires an evolution model for the parameter state. A persistence model, commonly used in previous studies (Baek et al., 2006; Zupanski and Zupanski, 2006), is expressed as

$$\mathbf{b}_{k+1,j}^{f} = \mathbf{b}_{k,j}^{a}. \tag{18}$$

We find that in practice a persistence model for multiplicative model parameters leads to model blow-up for small ensemble sizes. To mitigate model blow-up, we introduce a temporally smoothed version of the persistence model

$$\mathbf{b}_{k+1,j}^{f} = (1-\beta)\mathbf{b}_{k,j}^{a} + \beta \mathbf{b}_{k,j}^{f}, \tag{19}$$

where β is a smoothing coefficient between 0 and 1. By introducing this smoothing model, rapid temporal variation of the model parameters is damped, leading to smooth temporal variation. It will be shown in Section 4 that this temporal smoothing can prevent model blow-up. The smoothing coefficient β is set to 0.8 for most experiments; the sensitivity of β will be discussed in Section 4.

In order to avoid filter divergence of the EnKF due to the finite ensemble size and model non-linearities, we apply covariance inflation (Anderson and Anderson, 1999) and localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001). We further suggest applying localization to the parameters, provided the parameters are spatially dependent and hence a distance between parameter and observation can be defined. The value of the covariance inflation factor is chosen using the adaptive covariance inflation algorithm proposed by Anderson (2007). The localization applied here is the fifth order polynomial function of Gaspari and Cohn (1999) with half-width c, and c is a tuning parameter. If the distance between the observation and the state variable is greater than $2c$, then the localization function is zero, which implies that the observation has no impact on the state variable; otherwise, it approximates a Gaussian. A correlation matrix $\mathbf{S}$ is computed by the fifth order function, and the localization can be made by a Schur product (an element-by-element multiplication)

$$\hat{\mathbf{P}}_{xx} = \mathbf{S}_x \circ \mathbf{P}_{xx}, \tag{20}$$

$$\hat{\mathbf{P}}_{bx} = \mathbf{S}_b \circ \mathbf{P}_{bx}. \tag{21}$$

$\mathbf{P}_{xx}$ and $\mathbf{P}_{bx}$ in eqs (12) and (13) will be replaced by $\hat{\mathbf{P}}_{xx}$ and $\hat{\mathbf{P}}_{bx}$ if localization is applied. The optimal value of c was tuned based on the rms errors. If the parameter $\mathbf{b}$ has no spatial dependence, then no localization can be applied, that is, the correlation matrix $\mathbf{S}_b$ is a matrix of all 1.

3. Experimental setup

The model used here is a modified version of the Lorenz-96 model (Lorenz and Emanuel, 1998). The original Lorenz-96 model has governing equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = (x_{i+1} - x_{i-2})\,x_{i-1} - x_i + f_0, \tag{22}$$

where $i = 1 \ldots N$ with cyclic indices. The specific model we use is

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = (x_{i+1} - x_{i-2})\,x_{i-1} - \frac{x_i}{1+d_i} + f_0 + f_i, \tag{23}$$

where the dissipation parameters d_i and forcing parameters f_i vary with i. Here, N is 40 and f_0 is 8.0. With small values of $\mathbf{d} = [d_1, d_2, \ldots, d_{40}]$ and $\mathbf{f} = [f_1, f_2, \ldots, f_{40}]$, the system of (23) should behave similar to (22). This setup is similar to the experiment setup by Baek et al. (2006), but instead of assuming a sine function of the bias, we randomly choose model parameters according to

$$\mathbf{f}_i \approx N(0, 1), i = 1 \ldots N, \tag{24}$$

$$\mathbf{d}_i \approx N(0.25, 0.2), i = 1 \ldots N, \tag{25}$$

where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. Since $\mathbf{d_i}$ is identified as a dissipation parameter, we reject values of $\mathbf{d}_i$ less than or equal to -1 (although these rare events never occurred in our simulations). While we restrict the true values of $\mathbf{d}$ while defining the model, we do not restrict the estimated values of $\mathbf{d}$ during data assimilation. After selecting the parameters $\mathbf{f}$ and $\mathbf{d}$ in this way, these parameters are held fixed throughout all model integrations.

The effect of $\mathbf{f}_i$ and $\mathbf{d}_i$ on the stability of the model is an important consideration. Following Lorenz (1963), an energy equation for (22) can be derived by multiplying each side by x_i and summing over all i, and making use of the periodic boundary conditions. Performing this operation on (23) yields

$$\frac{\mathrm{d}E}{\mathrm{d}t} \leq -\frac{2E}{1+d_{\max}} + \sum_{i=1}^{N}(f_0 + f_i)x_i, \tag{26}$$

where $d_{\max}$ is the largest value of $\mathbf{d}_i$, $i = 1, \ldots N$, and $E = \frac{1}{2}\sum_{i=1}^{N} x_i^2$. Note that the distance from the origin in state space is $\sqrt{2E}$. Since the first term on the right-hand side of (28) grows with the square of x_i whereas the second grows only linearly, it follows that all solutions tend to contract towards the origin if they are sufficiently far from the origin. Thus, this model is stable for all values of $f_0 + f_i$ (provided $1 + d_i > 0$ for all i, as assumed in the model). However, the forecast model can become unstable if $d_i \leq -1$ for some i, which can occur randomly in some realizations of the EnKF. For this reason, we argue that estimation of multiplicative model parameters is generally more difficult than additive model parameters, since the former parameters affect model stability whereas the latter do not. For complex geophysical models, the exact stability boundary is unknown and difficult to identify. This fact motivates the simpler approach of using the temporally smoothed version of the persistence model (19), which is found to prevent parameter values from reaching the stability boundary.

Solutions of (23) are obtained by integrating the model forward with the time interval 0.05, and a fourth-order Runge–Kutta numerical method is applied at each model time step. The truth is one single integration of the model (23) with the prescribed **d** and **f** for 6000 time steps after a spin-up period of 30 000 time steps. The observational data set was constructed by adding Gaussian white noise with zero mean and unit variance to the truth. This implies that the observation covariance matrix **R** is the identity matrix. Observations were taken at every other grid point, so the number of observations equals 20.

Initial conditions for the ensemble assimilation are randomly chosen from the truth. The augmented EnKF is used to assimilate 6000 observation times, but results are reported only for the last 3000 times. In all experiments the root mean square error (RMSE) of the filter is computed as the root mean square of the difference between the analysis and the truth over the 40 grid points for the model state and from model time steps 3000–6000. We emphasize that RMSE measures the error in the model state, not the error in the parameter estimates.

4. Numerical results

4.1. Additive parameter estimation

In our first experiment, **d** is fixed and equal to the true values, but **f** is considered to be unknown and therefore augmented in the EnKF. The true value of **f** is prescribed as in (24). In the decoupled augmented EnKF solutions (10)–(15), **b** represents the model parameter **f**. The initial ensemble mean of **f** was **0** and initial ensemble perturbations were randomly drawn from zero mean white noise with variance of 0.1. The localization half-width c is 10 relative to the model domain size 40 for both **x** and **b**. Here, we have 80 unknowns (**x** and **f**) to be estimated, based on 20 observations.

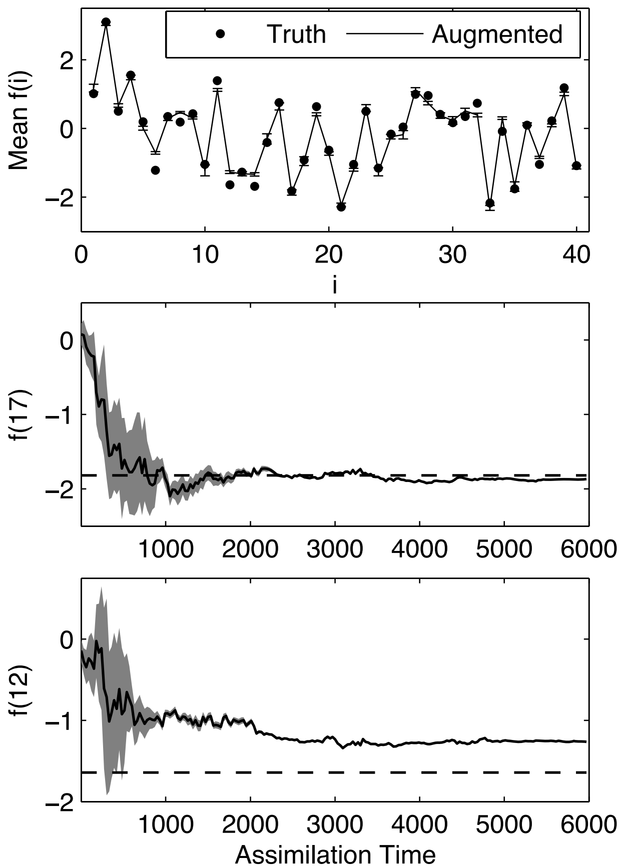

Fig 1. The time average (solid) and ± 2 *SD* (error bars) of the augmented additive parameter **f** over the model time step from 3000 to 6000 and the true values (closed circle) at each grid point (top panel). The augmented parameter **f** (solid) and **f** ± 2 *SE* (shaded) derived from EnKF at the model grid points 17 (middle panel) and 12 (bottom panel) as a function of assimilation time, and the true value (dashed) for comparison. The ensemble size is 20.

In the top panel of Fig. 1, we plot the true values of **f** as dots, the time mean as solid line, and ± 2 *SD* of the estimated value of **f** over the model time step from 3000 to 6000 as error bars. The ensemble size here is 20. We see that the estimated values of **f** agree well with the true values, for example, the error bars are small relative to the mean values, indicating convergence to the true values. The EnKF gives not only the most probable value of the model parameters, but also a measure of their uncertainties. To illustrate this capability, we show in Fig. 1 the time series of the estimated values of **f** at two selected points, as well as their associated confidence interval, as estimated from the EnKF as plus/minus twice the square roots of the diagonal elements of P_{bb}. Note that $\sqrt{\mathbf{P_{bb}}}$ measures the ensemble spread of the parameters. In the first case, at grid point 19, the assimilated value of the forcing parameter converges to the true value and the confidence interval brackets the true value. Thus, the EnKF behaves appropriately by converging to the true value at long times. In the second case, at grid point 12, the assimilated value does not converge to the true value and, moreover, the corresponding confidence interval does not bracket the true value consistently. We find that some estimates do not converge to the true values even with 40 observations (figures not shown). In general, the number of grid points that converge to the true value (to within the confidence interval) increases with increasing number of observations and ensemble members. A plausible reason for why some estimates converge to the correct values whereas others do not is the following. The EnKF determines estimates by minimizing the mean square error. As the model parameters approach their true values, the mean square error becomes less sensitive to the model parameters. Eventually, the parameters become sufficiently close to their true values that further (small) changes in the parameters result in statistically indistinguishable differences in the mean square error. Accordingly, the EnKF effectively reduces the associated uncertainty estimates because parameter variations within the regions of uncertainty yield no statistically significant improvement in RMSE.

To test the above hypothesis, we compared the performance of the augmented model with two other models, which we call the perfect model and the imperfect model. Here, the perfect model refers to the model that generates the observations, and the imperfect model refers to the model with $\mathbf{f} = 0$. The experiments of the perfect model and imperfect model define the upper and lower limits of the expected performance of the data assimilation. Note that the adaptive covariance inflation tends to be larger in the imperfect model case to account for model errors (Anderson, 2007). Figure 2 shows the RMSE as a function of ensemble size for the perfect model, the augmented model and the imperfect model. For ensemble size 20, which was used in the experiments whose results are displayed in Fig. 1, the RMSE of the augmented model is close to that of the perfect model. In general, the RMSE of the imperfect model is about

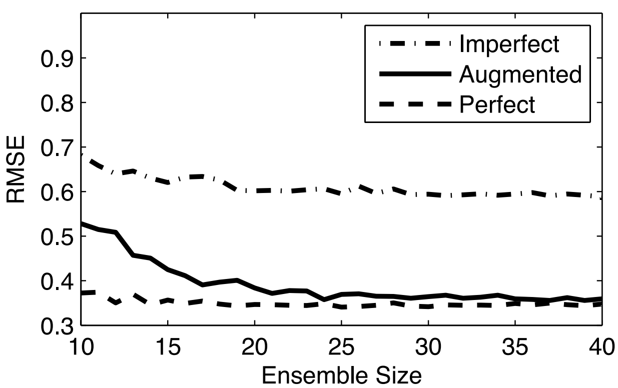

Fig 2. The root mean square error (RMSE) over the assimilation time steps from 3000 to 6000 of the analysed model states as a function of ensemble size for the perfect model (dashed), the additive parameter-augmented model (solid) and the imperfect model (dot–dashed).

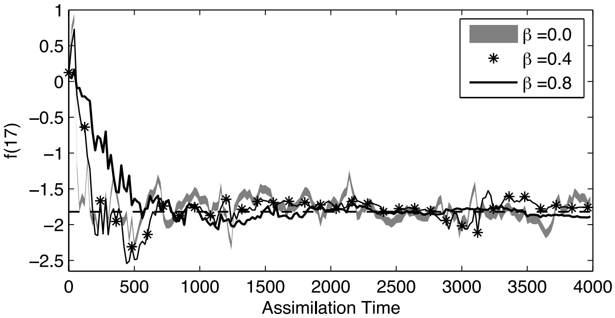

Fig 3. The augmented additive parameter at the model grid point 17 as a function of assimilation time for different smoothing coefficients: 0.0 (shaded), 0.4 (solid-star), and 0.8 (solid). The corresponding true value (dashed) is plotted for comparison. The ensemble size is 20.

50% larger than that of the perfect model. The RMSE for the augmented model converges close to that for the perfect model when the ensemble size is large. It is noteworthy that the augmented model performs better than the imperfect model even for small ensemble size.

In the above experiment, the smoothing coefficient β in (19) was set to 0.8. Sensitivity experiments with respect to the smoothing coefficient β were conducted. Figure 3 depicts the estimated additive parameter at grid point 17 as a function of assimilation time for different smoothing coefficients. The estimated values for the persistence model without temporal smoothing ($\beta = 0$) varies rapidly with time, especially in the first 200 time steps, and undergoes larger fluctuations about the mean value. As the smoothing coefficient increases, the temporal variations decrease substantially. The time series for $\beta = 0.8$ are much smoother than those for $\beta = 0.4$ or 0.0. The RMSE for different smoothing coefficients are listed in the second column of Table 1. The temporal smoothing slightly improves the RMSE, but there is no significant difference of RMSE among

Table 1. Analysis RMSE for different smoothing coefficients β

β	Additive parameter 20 ensembles	Multiplicative parameter 20 ensembles	Additive and multiplicative parameters 20 ensembles
0.9	0.3867	0.3728	0.4532
0.8	0.3751	0.3775	0.4197
0.7	0.3691	0.3754	0.4159
0.6	0.3738	0.3537	0.4224
0.5	0.3753	0.3814	–
0.4	0.3875	0.3591	–
0.3	0.3906	–	–
0.2	0.3950	–	–
0.0	0.3833	–	–

Note: '–' denotes filter divergence.

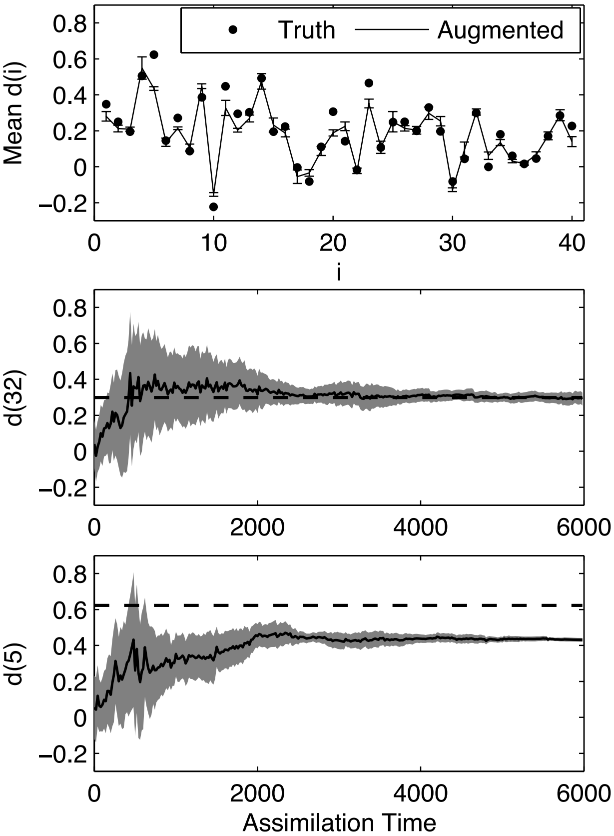

Fig 4. The time average (solid) and ± 2 *SD* (error bars) of the augmented multiplicative parameter $\mathbf{d}$ over the model time steps from 3000 to 6000 and the true model parameter (closed circle) at each grid point (top panel). The augmented parameter $\mathbf{d}$ (solid) and $\mathbf{d} \pm 2$ *SE* (shaded) derived by EnKF at the model grid points 32 (middle panel) and 5 (bottom panel) as a function of assimilation time (bottom panel), and the true value (dashed) for comparison. The ensemble size is 20.

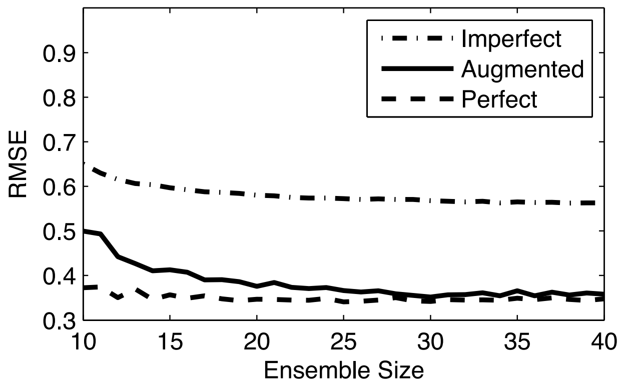

Fig 5. The root mean square error (RMSE) over the assimilation time steps from 3000 to 6000 of the analysed model states as a function of ensemble size for the perfect model (dashed), the multiplicative parameter-augmented model (solid) and the imperfect model (dot–dashed).

different smoothing coefficients in the estimation of additive model parameters.

4.2. Multiplicative parameter estimation

In the second experiment, $\mathbf{f}$ is fixed and equal to the true values in (24), but $\mathbf{d}$ is considered to be unknown and therefore augmented in the EnKF. The true value of $\mathbf{d}$ is prescribed as in (25). In the decoupled augmented EnKF solutions (10)–(15), $\mathbf{b}$ represents the model parameter $\mathbf{d}$. The initial ensemble mean of $\mathbf{d}$ was $\mathbf{0}$ and initial ensemble perturbations were randomly drawn from zero mean white noise with variance of 0.01. The localization half-width c is 10 relative to the model domain size 40 for both $\mathbf{x}$ and $\mathbf{d}$. The temporal smoothing coefficient in (19) is 0.8. In this experiment, we have 80 unknowns ($\mathbf{x}$ and $\mathbf{d}$) to be estimated based on 20 observations.

In our first experiment we used the persistence model (18) for the augmented parameters. Unfortunately, we found that the model became unstable and the filter estimates diverged to infinity (i.e. the filter 'blew up'). Consequently, we abandoned the persistence model (18) and adopted the temporally smoothed version of the persistence model (19). The time average and ± 2 *SD* of the augmented damping parameter $\mathbf{d}$ over the assimilation time steps from 3000 to 6000 are plotted in the top panel of Fig. 4. The corresponding true values are also shown for comparison. The ensemble size here is 20. The estimated mean values of the damping parameter agree fairly well with the true values. We also show time series of the assimilated parameter values and an associated confidence interval for two representative grid points. At grid point 32, the estimated value starts from the initial value 0, and then gradually converges close to the true value. The ensemble spread eventually decreases as the estimated value converges to the true value. Furthermore, the uncertainty in the estimated values, derived from the ensemble standard deviations, provides a realistic interval within which the true value can be found. At grid point 5, the estimated value starts from the initial value 0, gradually passes through the true value, and eventually converges to the 'wrong' value. It turns out, however, that the RMSE of the augmented model with the 'wrong' parameter values is close to the perfect model case, as shown in Fig. 5. Thus, as far as RMSE is concerned, the forecast model with a 'wrong' parameter performs almost as well as the perfect model. Figure 5 shows the RMSE as a function of ensemble size for the perfect model, the damping parameter-augmented model and the imperfect model. Here, the perfect model refers to the model which generates the observations, and the imperfect model is the model with $\mathbf{d} = \mathbf{0}$ in (25). The RMSE for the augmented model converges close to that for the perfect model for ensemble sizes around 25 or larger. It should be emphasized that the augmented model still performs better than the imperfect model for small ensemble sizes (less than 20).

The damping parameter **d** is a multiplicative term in the model, while the forcing parameter **f** is an additive term. In addition, the damping parameter is in the denominator, and may cause a singularity if the denominator is zero. As shown in Fig. 3, the augmented state changes rapidly if no temporal smoothing is applied. The rapid changes often cause the filter to be unstable, since the denominator term in the damping goes to zero. In this case, the temporal smoothing makes the filter stable. The third column of Table 1 lists the RMSE for different smoothing coefficients. The filter is unstable when the smoothing coefficient is less than 0.4. In practice, the optimal value of β cannot be determined in terms of RMSE since no truth is available. Even without the truth, we can tune the smoothing coefficient to make the filter stable. The RMSEs for different smoothing coefficients are marginally different once the filter is stable (see Table 1).

Note that the augmented EnKF of additive parameters is stable even without smoothing (see the second column of Table 1). We also tried experiments with increasing the mean value of **f** from 0 to 2, and found that the model is still stable even for a 10 member ensemble. Increasing observational error may cause ensemble collapse (ensemble covariances go to zero), thus leading to filter divergence. In this study, we mainly emphasize that the multiplicative parameters may change model stability and cause model blow-up (solutions go to infinity), for instance, the multiplicative damping parameter d in eq. (21) cannot be equal to or less than -1, while the additive forcing parameter does not cause model blow-up.

4.3. *Additive and multiplicative parameter estimation*

In this experiment, we assume the parameters **f** and **d** are unknown and thus augment them in the EnKF. The true values of **f** and **d** are prescribed in (24) and (25) respectively, and then held fixed throughout all integrations. In this case, we have a difficult condition for data assimilation: 20 observations and 120 unknowns. Following the previous two sections, the initial ensemble mean of **f** was **0** and initial ensemble perturbations were randomly drawn from zero mean white noise with variance of 0.1; the initial ensemble mean of **d** was **0** and initial ensemble perturbations were randomly drawn from zero mean white noise with variance of 0.01. The localization half-width c is 10 relative to the model domain size 40 for **x**, **f** and **d**. The temporal smoothing coefficient is 0.8 for both **f** and **d**.

The time average and ± 2 *SD* of the estimated additive and multiplicative parameters over the model time steps from 3000 to 6000 are shown in Fig. 6. The ensemble size here is 20. The aug-

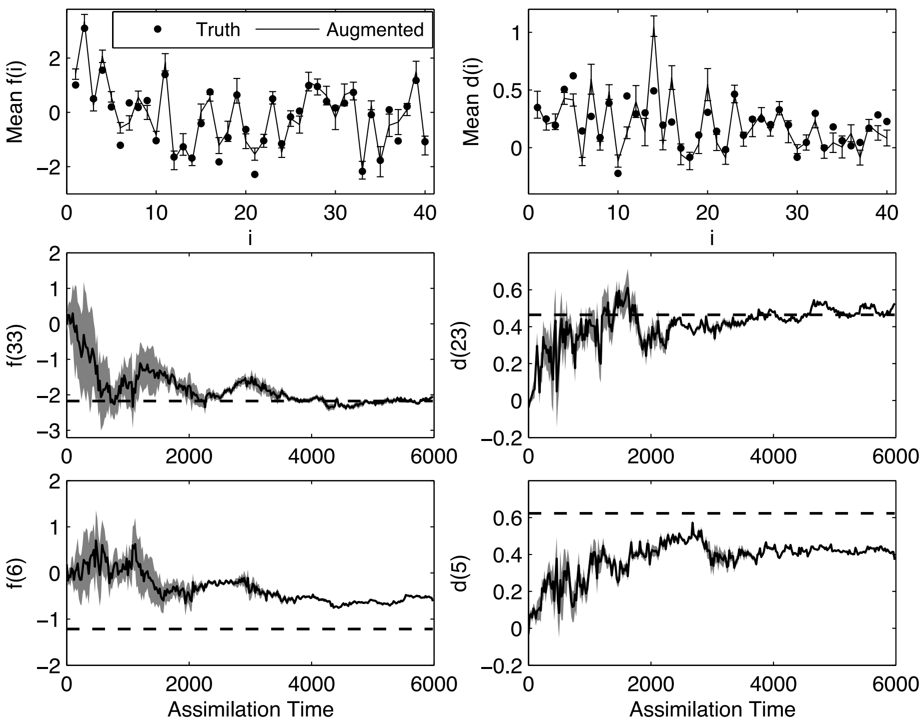

Fig 6. The time average (solid) and ± 2 *SD* (error bars) of the estimated additive parameter **f** (top, left-hand panel) and multiplicative parameter **d** (top, right-hand side) over the model time step from 3000 to 6000 and the corresponding true values (closed circle) at each grid point for the additive and multiplicative parameter-augmented model. The estimated additive parameter **f** (solid) and **f** ± 2 *SE* (shaded) derived by EnKF at the model grid points 33 (middle, left-hand panel) and 6 (bottom, left-hand panel), and the estimated multiplicative parameter **d** (solid) and **d** ± 2 *SE* (shaded) derived by EnKF at the model grid points 23 (middle, right-hand panel) and 5 (bottom, right-hand panel), as a function of assimilation time, and the true value (dashed) for comparison. The ensemble size is 20.

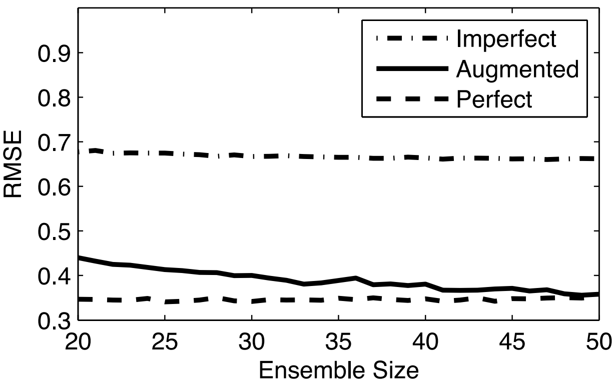

Fig 7. The root mean square error (RMSE) over the assimilation time steps from 3000 to 6000 of the analysed model states as a function of ensemble size for the perfect model (dashed), the additive and multiplicative parameter-augmented model (solid) and the imperfect model (dot–dashed).

mented EnKF is able to capture the spatially varying additive and multiplicative parameters, and the estimated mean values agree well with the true values. For the grid points where the augmented EnKF overestimates or underestimates the true values, the estimated values are still relatively close to the truth. Note that more estimates of the augmented parameters converged to the 'wrong' values, compared to the case of only augmenting additive or multiplicative parameter, since 40 more unknowns are augmented in this experiment but for the same ensemble size. As the ensemble size increases, the 'wrong' estimates are found to be improved (figures not shown). In this experiment, the temporal smoothing coefficient is 0.8. The sensitivity experiments for the smoothing coefficients are summarized in the fourth column of Table 1. The temporal smoothing has to be applied to avoid the model blow-up due to the singularity problem associated with the damping parameter. For ensemble size 20, the model became unstable for smoothing coefficient less than 0.6.

Figure 7 shows the RMSE as a function of ensemble size for the perfect model, the additive and multiplicative parameter-augmented model and the imperfect model. Here, the perfect model refers to the model which generates the observations, and the imperfect model is the model with $\mathbf{f} = \mathbf{0}$ and $\mathbf{d} = \mathbf{0}$. There are 40 more unknowns in this experiments than those in the additive or multiplicative parameter-augmented experiments, thus the compensation for more ensemble size is required. It was found that the filter blowup may occur if the ensemble size is less than 20. The RMSE for the augmented model converges close to that for the perfect model as the ensemble size increases up to 50. Even for ensemble size 20, the augmented EnKF can substantially reduce the RMSE compared to the EnKF with imperfect model. These results are quite encouraging, indicating that the augmented EnKF is capable of estimating and correcting the additive and multiplicative model parameters.

The initial ensemble variance we chose is 0.1 for **f** and 0.01 for **d** for all experiments in this study. We have tried some experiments with different initial ensemble variance (1, 0.1 and 0.01 for **f**, 0.1, 0.01 and 0.001 for **d**), and it turns out that the ensemble spread converged to the same level after several hundred time steps. Thus, the filter divergence is not sensitive to the initial ensemble variance of the parameters. This behaviour is probably due to the adaptive inflation we used. The inflation factors at early stages tends to be smaller if the initial ensemble variance is larger, and vice versa, hence the adaptive inflation automatically accounts for the initial ensemble variance.

5. Conclusions

This paper proposed a simple approach to estimating multiplicative model parameters using the ensemble square root filter. Specifically, we propose augmenting the state vector with the uncertain parameters, as commonly done in data assimilation, but updating the parameter estimates using a temporally smoothed update equation rather than simple persistence. The temporal smoothing was found to damp the rapid time variation of the uncertain parameters and thereby prevent model blow-up. Model blow-up is caused when multiplicative parameters drift into a parameter regime characterized by unbounded solutions. The temporal smoothing depends on a parameter β that controls the degree of smoothing, with large values corresponding to strong smoothing and hence longer convergence times for the estimated parameters. We expect larger values of β to be needed for decreasing ensemble size, increasing number of model parameters, or decreasing number of observations. The choice of β requires balancing the convergence rate to the true values against the degree of parameter variability that can be allowed while maintaining bounded solutions. We found that the RMSE of the analysis was insensitive to the choice of β, so β can be tuned as small as possible to maintain a stable filter.

In addition to the above method, we propose applying covariance localization to the model parameters, assuming the parameters are spatially dependent. This localization procedure was found to be essential when the number of uncertain parameters was large. The augmented EnKF can be viewed as two filters, one for the state that is exactly the same as in the non-augmented case, and one for the parameter that has the same structural form as the state form. As a result, implementation of the decoupled augmented EnKF does not require changing the data assimilation algorithm—it just has to be supplemented by a parameter estimation step that is similar to the state estimation step. Thus, any of the specialized techniques for solving the EnKF, for example, the advanced sequential parallel algorithm (Anderson and Collins, 2007; Whitaker et al., 2008), can be applied to solve the augmented EnKF. Presumably, this decoupling has been recognized by previous authors, but does not appear to have been expressed in the mathematical forms (10)–(13) previously, even though the derivation is simple and straightforward. Note that the decoupling is not restricted to the ensemble square root filter by Whitaker and Hamill (2002), and it can be implemented by other standard ensemble square root filters (Anderson, 2001; Bishop et al., 2001; Tippett et al., 2003).

Additive and/or multiplicative model parameter estimation experiments were performed to validate the proposed decoupled augmented method using the Lorenz-96 model. The designed experiments posed difficult conditions for data assimilation, which included observations available only at every other grid point and model parameter values randomly distributed in space. The time average of the estimated additive and multiplicative parameters agreed quite well with the corresponding true values, even when both additive and multiplicative parameters were present in the forecast model. As the ensemble size increases, the performance of the augmented EnKF converged to that of the EnKF with the perfect model. The results demonstrate that the augmented EnKF is able to properly estimate the additive parameters as well as multiplicative parameters. The EnKF not only gives accurate estimates of parameters, but also provides reasonable estimates of the uncertainty in the estimates.

In this study, the prediction model was in the same class as the true model, in the sense that for some choice of parameter values the prediction model is exactly the same as the model that generated the observations. In reality, the prediction model is not in the same class as the true model, and hence the performance of the methodology may not be as good as suggested here. Another problem is that atmospheric and oceanic models have hundreds or even thousands of parameters, so some selection of parameters needs to be made to render the estimation problem computationally feasible. Finally, this study considered estimation only for deterministic parameters, whereas a more appropriate approach might be to allow parameters to vary randomly. A method for estimating stochastic parameters is proposed in a forthcoming paper (DelSole and Yang, 2009).

6. Acknowledgments

This research is supported by NOAA grant NA06OAR4310001. We thank two anonymous reviewers for their critical comments that led to substantial improvements in the manuscript.

References

Anderson, J. L. 2001. An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

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

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

Anderson, J. L. and Collins, N. 2007. Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.* **24**, 1452–1463.

Baek, S.-J., Hunt, B. R., Kalney, E., Ott, E. and Szunyogh, I. 2006. Local ensemble Kalman filtering in the presence of model bias. *Tellus* **58A**, 293–306.

Bishop, C. H., Etherton, B. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

Dee, D. P. and da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Quart. J. R. Met. Soc.* **124**, 269–297.

Dee, D. P. and Todling, R. 2000. Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev.* **128**, 3268–3282.

DelSole, T. and Yang, X. 2009. Stochastic parameter estimation with the ensemble Kalman filter. *Physica D*, submitted.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 1043–1062.

Friedland, B. 1969. Treatment of bias in resursive filtering. *IEEE Trans. Autom. Control* **14**, 359–367.

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. R. Met. Soc.* **125**, 723–757.

Gillijns, S. and De Moor, B. 2007. Model error estimation in ensemble data assimilation. *Nonlin. Process. Geophys.* **14**, 59–71.

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci* **20**, 130–141.

Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci* **55**, 399–414.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square-root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Whitaker, J. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. and Toth, Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.* **136**, 463–482.

Zupanski, D. and Zupanski, M. 2006. Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.* **134**, 1337–1354.