A local ensemble Kalman filter for atmospheric data assimilation

By EDWARD OTT[1], BRIAN R. HUNT[2], ISTVAN SZUNYOGH[3]*, ALEKSEY V. ZIMIN[4],
ERIC J. KOSTELICH[5], MATTEO CORAZZA[6], EUGENIA KALNAY[6], D. J. PATIL[2] and
JAMES A. YORKE[2],   [1]*Institute for Research in Electronics and Applied Physics, University of Maryland,
College Park, MD, USA;* [2]*Institute for Physical Science and Technology and Department of Mathematics, University
of Maryland, College Park, MD, USA;* [3]*Institute for Physical Science and Technology and Department of Meteorology,
University of Maryland, College Park, MD, USA;* [4]*Institute for Physical Science and Technology, University of
Maryland, College Park, MD, USA;* [5]*Department of Mathematics and Statistics, Arizona State University, AZ, USA;*
[6]*Department of Meteorology, University of Maryland, College Park, MD, USA*

ABSTRACT

In this paper, we introduce a new, local formulation of the ensemble Kalman filter approach for atmospheric data assimilation. Our scheme is based on the hypothesis that, when the Earth's surface is divided up into local regions of moderate size, vectors of the forecast uncertainties in such regions tend to lie in a subspace of much lower dimension than that of the full atmospheric state vector of such a region. Ensemble Kalman filters, in general, take the analysis resulting from the data assimilation to lie in the same subspace as the expected forecast error. Under our hypothesis the dimension of the subspace corresponding to local regions is low. This is used in our scheme to allow operations only on relatively low-dimensional matrices. The data assimilation analysis is performed locally in a manner allowing massively parallel computation to be exploited. The local analyses are then used to construct global states for advancement to the next forecast time. One advantage, which may take on more importance as ever-increasing amounts of remotely-sensed satellite data become available, is the favorable scaling of the computational cost of our method with increasing data size, as compared to other methods that assimilate data sequentially. The method, its potential advantages, properties, and implementation requirements are illustrated by numerical experiments on the Lorenz-96 model. It is found that accurate analysis can be achieved at a cost which is very modest compared to that of a full global ensemble Kalman filter.

1. Introduction

The accuracy of a data assimilation scheme strongly depends on the accuracy of the assumed background error statistics. Because the true background errors are not known, the implementation of a data assimilation system requires the development of statistical models that provide an estimate of the background error covariance matrix (e.g. Lorenc, 1986; Daley, 1991; Kalnay, 2002).

In the case of linear dynamics, the mathematically consistent technique to define a background error covariance matrix is the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961), which utilizes the dynamical equations to evolve the most probable state and the error covariance matrix in time. In the case of linear systems with unbiased normally distributed errors, the Kalman filter provides estimates of the system state that are optimal in the

mean square sense. The method has also been adapted to non-linear systems, but, in this case, optimality no longer applies.

Although the Kalman filter approach has been successfully implemented for a wide range of applications and has long been considered for atmospheric data assimilation (e.g. Ghil et al., 1981), the computational cost involved does not allow for an operational implementation for weather forecasting in the foreseeable future (see Daley, 1991, for details). The recognition of this computational constraint led to the development of reduced-state estimates of the background error covariance matrix. Essentially, there are two ways to obtain a reduced rank estimate of the background error covariance matrix. Schemes based on linearizing the model dynamics reduce the rank of the background error covariance matrix by projecting the model state on a basis that is much lower-dimensional than the full model space (e.g. Fukumori and Malanotte-Rizzoli, 1995; Cane et al., 1996; Pham et al., 1998; Tangborn, 2004). [In a more recent paper, Fukumori (2002) also showed that this approach can be extended to obtain

*Corresponding author.
e-mail: szunyogh@ipst.umd.edu

a high-rank estimate of the background error covariance matrix by partitioning the full Kalman filter into a series of reduced rank problems.]

Another approach toward reducing the cost of the Kalman filter is to use a relatively small ensemble of forecasts to estimate the forecast error covariance (e.g. Evensen, 1994; Houtekamer and Mitchell, 1998, 2001; Anderson, 2001; Bishop et al., 2001; Hamill et al., 2001; Whitaker and Hamill, 2002; Keppenne and Rienecker, 2002). In ensemble-based data assimilation schemes, the ensemble of (background) forecasts is generated by evolving an ensemble of initial conditions distributed according to the result of the previous analysis. This approach has the advantage of not requiring the labor-consuming development and maintenance of a linearized forecast model.

The main difference between the existing ensemble-based schemes is in the generation of the analysis ensemble. One family of schemes is based on perturbed observations (Evensen and van Leeuwen, 1996; Houtekamer and Mitchell, 1998, 2001; Hamill and Snyder, 2000; Hamill et al., 2001; Keppenne and Rienecker, 2002). In this approach, the analysis ensemble is obtained by assimilating a different set of observations to each member of the background ensemble. The different sets of observations are created by adding random noise to the real observations, where the random noise component is generated according to the observational error covariance matrix. Such a system has been developed at the Canadian Meteorological Service and was recently reported to break even with the operational 3D-Var scheme (Peter Houtekamer et al., private communication).

The other family of schemes, the Kalman square-root filters, uses a different approach to reduce the size of the ensemble. These techniques perform the analysis only once, to obtain both the mean analysis and the analysis error covariance matrix. Then the analysis ensemble perturbations (to the mean analysis) are generated by transforming the background ensemble perturbations to a set of vectors that can be used to represent the analysis error covariance matrix. Thus, the analysis is confined to the subspace of the ensemble. Because there is an infinite set of analysis perturbations that can be used to represent the analysis error covariance matrix, many different schemes can be derived following this approach (Tippett et al., 2002). Existing examples of the square root filter approach are the ensemble transform Kalman filter (Bishop et al., 2001), the ensemble adjustment filter (Anderson, 2001), and the ensemble square-root filter (Whitaker and Hamill, 2002).

A key practical aspect of most atmospheric ensemble Kalman filters is 'covariance localization' which restricts the use of the ensemble-based covariance information to small subsets of variables defined at model grid points within local regions much smaller than the surface of the Earth. The local approach makes possible a high-dimensional estimate of the global background error covariance statistics based on a small ensemble. In the sequential data assimilation schemes, proposed in the earlier papers, observations are treated one at a time. The localization is done by updating the analysis at all grid points within a predefined correlation length from each observation. The scheme we propose (Ott et al., 2002) is a Kalman square-root filter, but not a sequential scheme. The basic idea is that we perform the analysis at each grid point simultaneously using the state variables and all observations in the local region centered at that point. The possibility of a similar computational model was mentioned in Anderson (2001), while Kalnay and Toth (private communication) used this kind of local method and a single bred vector to increment the state along the direction determined by the difference between the bred vector and the background. In a sense, our technique is related to previous work that attempted to construct a simplified Kalman filter by explicitly taking into account the dominant unstable directions of the state space (Fisher, 1998).

In the following sections we describe and test our scheme, which we call a local ensemble Kalman filter (LEKF). Section 2 is an outline of the algorithm, while in Section 3 we define our local regions and explain how the dimension of the local state vector can be further reduced. Specifically, we discuss the possibility of approximating the covariance as being restricted to a subspace whose dimension may be substantially lower than the number of ensemble members. We speculate that this may be computationally advantageous and, at the same time, may not noticeably degrade the accuracy of the assimilation. In Section 4 we explain the analysis scheme for the local regions. These local analyses can be processed in parallel, involve relatively small matrices, and treat all data simultaneously. These features suggest the potential for the LEKF method in rapidly and efficiently assimilating large amounts of data (e.g. as will become available from future satellite observing systems). In Section 5, the local analyses are pieced together to obtain the ensemble of analysis fields. In particular, we discuss a novel consideration for achieving a smooth physical ensemble of analysis fields by minimization of the distance between the analysis and background ensembles (see also the Appendix). In Section 6 we illustrate our data assimilation scheme by numerical experiments applying it to a toy spatio-temporally chaotic model system introduced by Lorenz (1996) and Lorenz and Emanuel (1998). Among other things, these numerical experiments confirm and quantify the possible advantage of localization in allowing reduced ensemble size, test the efficiency of a new variance inflation technique (Section 4.4), and show LEKF accuracy similar to that of a full Kalman filter, but at much lower computational cost.

2. Outline of the algorithm

An outline of the algorithm is as follows.

(1) Advance the analysis ensemble of global atmospheric states to the next analysis time, thus obtaining a new background ensemble of global atmospheric states.

(2) Associate a local region with each grid point and, for each local region and each member of the background ensemble, form

Fig 1. Illustration of the local ensemble Kalman filter scheme as given by the seven steps listed in Section 2. The symbols in the figure are as follows: $\mathbf{x}^{a(i)}(\mathbf{r}, t)$, the analysis ensemble fields as a function of position $\mathbf{r}$ on the globe at time t; $\mathbf{x}^{b(i)}_{mn}$, ensemble of background atmospheric states in local region mn; $\mathbb{S}_{mn}$, the local low dimensional subspace in region mn; $\delta\hat{\mathbf{x}}^{b(i)}_{mn}$, perturbations of the background ensemble members from the most probable background state; $\hat{\bar{\mathbf{x}}}^{a}_{mn}(t)$, the mean analysis state in $\mathbb{S}_{mn}$; $\hat{\mathbf{P}}^{a}_{mn}(t)$, the analysis error covariance matrix in $\mathbb{S}_{mn}$.
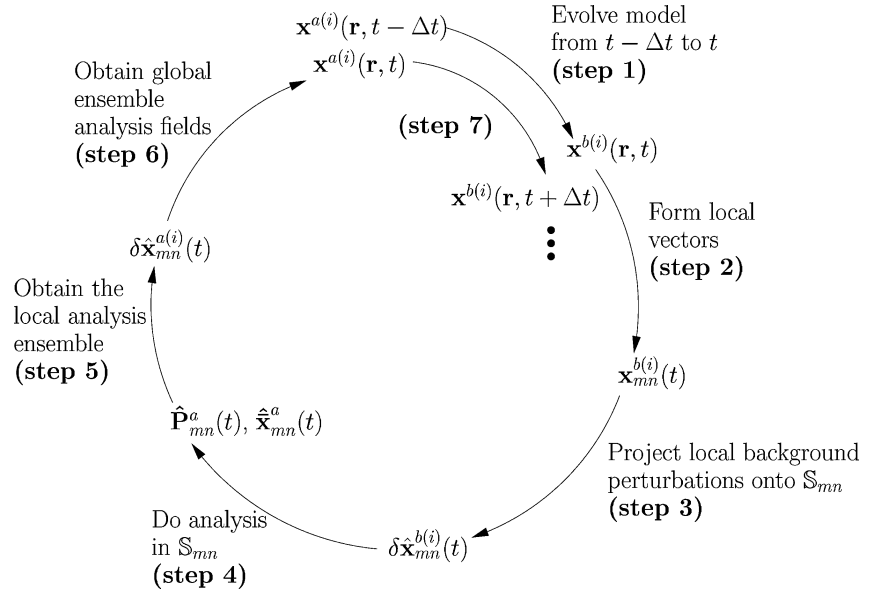
vectors of the atmospheric state information in that local region (Section 3).

(3) For each local ensemble member vector obtained in step 2, obtain its perturbation from the mean over the background ensemble, and project these perturbations on to the low-dimensional subspace that best represents the ensemble in that region (Section 3).

(4) Perform the data assimilation in each of the local low-dimensional subspaces, obtaining the analysis mean and covariance in each local region (Section 4).

(5) From the local analysis mean and covariance, obtain a suitable local analysis ensemble of local atmospheric states (Sections 5.1 and 5.2).

(6) Use the local analyses, obtained in step 5, to form a new global analysis ensemble (this is where the square root filter comes in; Section 5.3).

(7) Go back to step 1.

These steps are summarized, along with a key of important symbols that we use, in Fig. 1 and its caption.

3. Local vectors and their covariance

A model state of the atmosphere is given by a vector field $\mathbf{x}(\mathbf{r}, t)$ where $\mathbf{r}$ is two-dimensional and runs over discrete values $\mathbf{r}_{mn}$ (the grid in the physical space used in the numerical computations). Typically, the two components of $\mathbf{r}$ are the geographical longitude and latitude, and $\mathbf{x}$ at a fixed $\mathbf{r}$ is a vector of all relevant physical state variables of the model (e.g. wind velocity components, temperature, surface pressure, humidity, etc., at all height levels included in the model). Let u denote the dimensionality of $\mathbf{x}(\mathbf{r}, t)$ (at fixed $\mathbf{r}$); e.g. when five independent state variables are defined at 28 vertical levels, $u = 140$.

Data assimilation schemes generally treat $\mathbf{x}(\mathbf{r}, t)$ as a random variable characterized by a probability distribution. The characterization of $\mathbf{x}$ is updated over time in two ways: (i) it is evolved according to the model dynamics; (ii) it is modified periodically to take into account recent atmospheric observations.

We perform our analysis locally in model space. In this section we introduce our local coordinate system and the approximations we make to the local probability distribution of $\mathbf{x}(\mathbf{r}, t)$. Because all the analysis operations take place at a fixed time t, we will suppress the t dependence of all vectors and matrices introduced henceforth.

Motivated by the work of Patil et al. (2001) we introduce at each point 'local vectors' $\mathbf{x}_{mn}$ of the information $\mathbf{x}(\mathbf{r}_{m+m',n+n'})$ for $-l \leq m', n' \leq l$. That is, $\mathbf{x}_{mn}$ specifies the model atmospheric state within a $(2l + 1)$ by a $(2l + 1)$ patch of grid points centered at $\mathbf{r}_{mn}$. The particular shape of the local region was chosen to keep the notations as simple as possible, but different (e.g. circular) shape regions and localization in the vertical direction can also be considered. The dimensionality of $\mathbf{x}_{mn}$ is $(2l + 1)^2 u$. We now consider local vectors obtained from the model as forecasts, using initial conditions distributed according to the result of the previous analysis, and we denote these by $\mathbf{x}^{b}_{mn}$ (where the superscript 'b' denotes 'background'). Let $F_{mn}(\mathbf{x}^{b}_{mn})$ be our approximation to the probability density function for $\mathbf{x}^{b}_{mn}$ at the current analysis time t. A fundamental assumption is that this probability distribution can be usefully approximated as Gaussian

$$F_{mn}\left(\mathbf{x}^{b}_{mn}\right) \sim \exp\left[-\frac{1}{2}\left(\mathbf{x}^{b}_{mn} - \bar{\mathbf{x}}^{b}_{mn}\right)^{T}\left(\mathbf{P}^{b}_{mn}\right)^{-1}\left(\mathbf{x}^{b}_{mn} - \bar{\mathbf{x}}^{b}_{mn}\right)\right],$$

(1)

where $\mathbf{P}^{b}_{mn}$ and $\bar{\mathbf{x}}^{b}_{mn}$ are the local background error covariance matrix and most probable state associated with $F_{mn}(\mathbf{x}^{b}_{mn})$. We

emphasize that the Gaussian form for the background probability distribution, $F_{mn}(\mathbf{x}^b_{mn})$, is rigorously justifiable only for a linear system, but not for a non-linear system such as the atmosphere.

As explained subsequently, the rank of the $(2l + 1)^2u \times (2l + 1)^2u$ covariance matrix $\mathbf{P}^b_{mn}$ for our approximate probability distribution function F_{mn} is much less than $(2l + 1)^2u$. Let

$$k = \mathrm{rank}\left(\mathbf{P}^b_{mn}\right). \tag{2}$$

Thus, $\mathbf{P}^b_{mn}$ has a $(2l + 1)^2u - k$ dimensional null space $\bar{\mathbb{S}}_{mn}$ and the inverse $(\mathbf{P}^b_{mn})^{-1}$ is defined for the component of the vectors $(\mathbf{x}^b_{mn} - \bar{\mathbf{x}}^b_{mn})$ lying in the k-dimensional subspace $\mathbb{S}_{mn}$ orthogonal to $\bar{\mathbb{S}}_{mn}$

In the data assimilation procedure we describe in this paper, the background error covariance matrix $\mathbf{P}^b_{mn}$ and the most probable background state $\bar{\mathbf{x}}^b_{mn}$ are derived from a $k' + 1$ member ensemble of global state field vectors $\{\mathbf{x}^{b(i)}(\mathbf{r})\}$, $i = 1, 2, \ldots, k' + 1$; $k' \geq k \geq 1$. The most probable state is given by

$$\bar{\mathbf{x}}^b_{mn} = (k' + 1)^{-1} \sum_{i=1}^{k'+1} \mathbf{x}^{b(i)}_{mn}, \tag{3}$$

where $\mathbf{x}^{b(i)}_{mn}$ is the local vector for the patch centered at $\mathbf{r}_{mn}$ and the ensemble member $\mathbf{x}^{b(i)}(\mathbf{r})$. To obtain the local background error covariance matrix $\mathbf{P}^b_{mn}$ that we use in our analysis, we first consider a matrix $\mathbf{P}^{b'}_{mn}$ given by

$$\mathbf{P}^{b'}_{mn} = \mathbf{X}^b_{mn}\mathbf{X}^{bT}_{mn}, \tag{4}$$

where the superscript 'T' denotes transpose, and

$$\mathbf{X}^b_{mn} = (k')^{-1/2}\left[\delta\mathbf{x}^{b(1)}_{mn}\middle|\delta\mathbf{x}^{b(2)}_{mn}\middle|\cdots\middle|\delta\mathbf{x}^{b(k'+1)}_{mn}\right], \tag{5}$$

$$\delta\mathbf{x}^{b(i)}_{mn} = \mathbf{x}^{b(i)}_{mn} - \bar{\mathbf{x}}^b_{mn}. \tag{6}$$

We assume that forecast uncertainties tend to lie in a low-dimensional subset of the $(2l + 1)^2u$ dimensional local vector space. [Preliminary results with an implementation of our data assimilation scheme on the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) model support this view.] Thus, we anticipate that we can approximate the background error covariance matrix by one of much lower rank than $(2l + 1)^2u$, and this motivates our assumption that an ensemble of size of $k' + 1$, where $k' + 1$ is substantially less than $(2l + 1)^2u$, will be sufficient to yield a good approximate representation of the background covariance matrix. Typically, $\mathbf{P}^{b'}_{mn}$ has rank k', i.e. it has k' positive eigenvalues. Let the eigenvalues of the matrix $\mathbf{P}^{b'}_{mn}$ be denoted by $\lambda^{(j)}_{mn}$, where the labeling convention for the index j is

$$\lambda^{(1)}_{mn} \geq \lambda^{(2)}_{mn} \geq \cdots \geq \lambda^{(k)}_{mn} \geq \cdots \geq \lambda^{(k')}_{mn}. \tag{7}$$

Because $\mathbf{P}^b_{mn}$ is a symmetric matrix, it has k' orthonormal eigenvectors $\{\mathbf{u}^{(j)}_{mn}\}$ corresponding to the k' eigenvalues (7). Thus

$$\mathbf{P}^{b'}_{mn} = \sum_{j=1}^{k'} \lambda^{(j)}_{mn}\mathbf{u}^{(j)}_{mn}\left[\mathbf{u}^{(j)}_{mn}\right]^T. \tag{8}$$

Because the size of the ensemble is envisioned to be much less than the dimension of $\mathbf{x}^b_{mn}$, $(k' + 1) \ll (2l + 1)^2u$, the computation of the eigenvalues and eigenvectors of $\mathbf{P}^b_{mn}$ is most effectively done in the basis of the ensemble vectors. That is, we consider the eigenvalue problem for the $(k' + 1) \times (k' + 1)$ matrix $\mathbf{X}^{bT}_{mn}\mathbf{X}^b_{mn}$, whose non-zero eigenvalues are those of $\mathbf{P}^{b'}_{mn}$ (7) and whose corresponding eigenvectors left-multiplied by $\mathbf{X}^b_{mn}$ are the k' eigenvectors $\mathbf{u}^{(j)}_{mn}$ of $\mathbf{P}^{b'}_{mn}$. We approximate $\mathbf{P}^{b'}_{mn}$ by truncating the sum at $k \leq k'$

$$\mathbf{P}^b_{mn} = \sum_{j=1}^{k} \lambda^{(j)}_{mn}\mathbf{u}^{(j)}_{mn}\left[\mathbf{u}^{(j)}_{mn}\right]^T. \tag{9}$$

The basic justification for the approximation of the covariance by $\mathbf{P}^b_{mn}$ is our supposition that, for reasonably small values of k, the total error variance in all other directions is much less than the variance:

$$\sum_{j=1}^{k} \lambda^{(j)}_{mn}. \tag{10}$$

In order to meaningfully compare eigenvalues, eq. (7), the different components of $\delta\mathbf{x}^{b(i)}_{mn}$ (e.g. wind and temperature) should be properly scaled to ensure that, if the variance (10) approximates the full variance, then the first k eigendirections, $\{\mathbf{u}^{(j)}_{mn}\}$, $j = 1, 2, \ldots, k$, explain the important uncertainties in the background, $\bar{\mathbf{x}}^b_{mn}$. For instance, the weights for the different variables can be chosen so that the Euclidean norm of the transformed vectors is equal to their total energy norm derived in Talagrand (1981). In what follows, we assume that the vector components are already properly scaled. (If $k = k'$, the comparison of eigenvalues is not used and thus such a consistent scaling of the variables is not necessary.) We also note that similar truncation procedures have been discussed by Lermusiaux and Robinson (1999) and Heemink et al. (2001) in a general setting.

For the purpose of subsequent computation, we consider the coordinate system for the k-dimensional space $\mathbb{S}_{mn}$ determined by the basis vectors $\{\mathbf{u}^{(j)}_{mn}\}$. We call this the 'internal coordinate system' for $\mathbb{S}_{mn}$. To change between the internal coordinates and those of the local space, we introduce the $(2l + 1)^2u \times k$ matrix,

$$\mathbf{Q}_{mn} = \left\{\mathbf{u}^{(1)}_{mn}\middle|\mathbf{u}^{(2)}_{mn}\middle|\cdots\middle|\mathbf{u}^{(k)}_{mn}\right\}. \tag{11}$$

We denote the projection of vectors into $\mathbb{S}_{mn}$ and the restriction of matrices to $\mathbb{S}_{mn}$ by a superscript circumflex (hat). Thus, for a $(2l + 1)^2u$ dimensional column vector $\mathbf{w}$, the vector $\hat{\mathbf{w}}$ is a k-dimensional column vector given by

$$\hat{\mathbf{w}} = \mathbf{Q}^T_{mn}\mathbf{w}. \tag{12}$$

Note that this operation consists of both projecting $\mathbf{w}$ into $\mathbb{S}_{mn}$ and changing to the internal coordinate system. Similarly, for a $(2l + 1)^2u \times (2l + 1)^2u$ matrix $\mathbf{M}$, the matrix $\hat{\mathbf{M}}$ is $k \times k$ and given by

$$\hat{\mathbf{M}} = \mathbf{Q}^T_{mn}\mathbf{M}\mathbf{Q}_{mn}. \tag{13}$$

To go back to the original $(2l + 1)^2 u$ dimensional local vector space, note that $\mathbf{Q}_{mn}^{\mathrm{T}} \mathbf{Q}_{mn} = \mathbf{I}$ while $\mathbf{Q}_{mn}\mathbf{Q}_{mn}^{\mathrm{T}}$ represents projection on $\mathbb{S}_{mn}$, i.e. it has null space $\bar{\mathbb{S}}_{mn}$ and acts as the identity on $\mathbb{S}_{mn}$. We may write $\mathbf{w}$ as

$$\mathbf{w} = \mathbf{w}^{(\|)} + \mathbf{w}^{(\perp)}, \tag{14}$$

$$\mathbf{w}^{(\|)} = \Lambda_{mn}^{(\|)}\mathbf{w} = \mathbf{Q}_{mn}\hat{\mathbf{w}}, \qquad \mathbf{w}^{(\perp)} = \Lambda_{mn}^{(\perp)}\mathbf{w}, \tag{15}$$

where $\mathbf{w}^{(\|)}$ and $\mathbf{w}^{(\perp)}$ denote the components of $\mathbf{w}$ in $\mathbb{S}_{mn}$ and $\bar{\mathbb{S}}_{mn}$, respectively, and the projection operators $\Lambda_{mn}^{(\|)}$ and $\Lambda_{mn}^{(\perp)}$ are given by

$$\Lambda_{mn}^{(\|)} = \mathbf{Q}_{mn}\mathbf{Q}_{mn}^{\mathrm{T}}, \qquad \Lambda_{mn}^{(\perp)} = \mathbf{I} - \mathbf{Q}_{mn}\mathbf{Q}_{mn}^{\mathrm{T}}. \tag{16}$$

In addition, if $\mathbf{M}$ is symmetric with null space $\bar{\mathbb{S}}_{mn}$

$$\mathbf{M} = \mathbf{Q}_{mn}\hat{\mathbf{M}}\mathbf{Q}_{mn}^{\mathrm{T}}. \tag{17}$$

Note that $\hat{\mathbf{P}}_{mn}^{\mathrm{b}}$ is diagonal,

$$\hat{\mathbf{P}}_{mn}^{\mathrm{b}} = \mathrm{diag}\left[\lambda_{mn}^{(1)}, \lambda_{mn}^{(2)}, \ldots, \lambda_{mn}^{(k)}\right], \tag{18}$$

and thus it is trivial to invert.

4. Data assimilation

With Section 3 as background, we now consider the assimilation of observational data to obtain a new specification of the probability distribution of the local vector. For simplicity, we assume that all observations collected for the current analysis were taken at the same time t. Let $\mathbf{y}_{mn}^{\mathrm{o}}$ be the vector of current observations within the local region, and assume that the errors in these observations are unbiased, are uncorrelated with the background, and are normally distributed with covariance matrix $\mathbf{R}_{mn}$. An ideal (i.e. noiseless) measurement is a function of the true atmospheric state. Considering measurements within the local region (m, n), we denote this function $\mathcal{H}_{mn}(\cdot)$. That is, if the true local state is $\mathbf{x}_{mn}^{\mathrm{a}}$, then the error in the observation is $\mathbf{y}_{mn}^{\mathrm{o}} - \mathcal{H}_{mn}(\mathbf{x}_{mn}^{\mathrm{a}})$. Assuming that the true state is near the mean background state $\bar{\mathbf{x}}_{mn}^{\mathrm{b}}$, we approximate $\mathcal{H}_{mn}(\mathbf{x}_{mn}^{\mathrm{a}})$ by linearizing about $\bar{\mathbf{x}}_{mn}^{\mathrm{b}}$

$$\mathcal{H}_{mn}\left(\mathbf{x}_{mn}^{\mathrm{a}}\right) \approx \mathcal{H}_{mn}\left(\bar{\mathbf{x}}_{mn}^{\mathrm{b}}\right) + \mathbf{H}_{mn}\Delta\mathbf{x}_{mn}^{\mathrm{a}}, \tag{19}$$

where

$$\Delta\mathbf{x}_{mn}^{\mathrm{a}} = \mathbf{x}_{mn}^{\mathrm{a}} - \bar{\mathbf{x}}_{mn}^{\mathrm{b}}, \tag{20}$$

and the matrix $\mathbf{H}_{mn}$ is the Jacobian matrix of partial derivatives of $\mathcal{H}_{mn}$ evaluated at $\bar{\mathbf{x}}_{mn}^{\mathrm{b}}$. [If there are s scalar observations in the local $(2l + 1) \times (2l + 1)$ region at analysis time t, then $\bar{\mathbf{y}}_{mn}^{\mathrm{o}}$ is s-dimensional and the rectangular matrix $\mathbf{H}_{mn}$ is $s \times (2l + 1)^2 u$.] Then, because we have assumed the background (pre-analysis) state $\mathbf{x}_{mn}^{\mathrm{b}}$ to be normally distributed, it will follow below that $\mathbf{x}_{mn}^{\mathrm{a}}$ is also normally distributed. Its distribution is determined by the most probable state $\bar{\mathbf{x}}_{mn}^{\mathrm{a}}$ and the associated covariance matrix $\mathbf{P}_{mn}^{\mathrm{a}}$. The data assimilation step determines $\bar{\mathbf{x}}_{mn}^{\mathrm{a}}$ (the local analysis) and $\mathbf{P}_{mn}^{\mathrm{a}}$ (the local analysis covariance matrix).

Because our approximate background covariance matrix $\mathbf{P}_{mn}^{\mathrm{b}}$ has null space $\bar{\mathbb{S}}_{mn}$, we consider the analysis increment component $\Delta\mathbf{x}_{mn}^{\mathrm{a}(\|)} = \Lambda_{mn}^{(\|)}(\mathbf{x}_{mn}^{\mathrm{a}} - \bar{\mathbf{x}}_{mn}^{\mathrm{b}})$ within the k-dimensional subspace $\mathbb{S}_{mn}$, and perform the data assimilation in $\mathbb{S}_{mn}$. Thus, the data assimilation is performed by minimizing the quadratic form

$$\begin{aligned} J\left(\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}\right) &= \left(\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}\right)^{\mathrm{T}}\left(\hat{\mathbf{P}}_{mn}^{\mathrm{b}}\right)^{-1}\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}} \\ &\quad + \left[\hat{\mathbf{H}}_{mn}\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}} + \mathcal{H}_{mn}\left(\bar{\mathbf{x}}_{mn}^{\mathrm{b}}\right) - \mathbf{y}_{mn}^{\mathrm{o}}\right]^{\mathrm{T}}\mathbf{R}_{mn}^{-1} \\ &\quad \times \left[\hat{\mathbf{H}}_{mn}\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}} + \mathcal{H}_{mn}\left(\bar{\mathbf{x}}_{mn}^{\mathrm{b}}\right) - \mathbf{y}_{mn}^{\mathrm{o}}\right]. \end{aligned} \tag{21}$$

Here $\hat{\mathbf{H}}_{mn} = \mathbf{H}_{mn}\mathbf{Q}_{mn}$ maps $\mathbb{S}_{mn}$ to the observation space, using the internal coordinate system for $\mathbb{S}_{mn}$ introduced in the previous section, so that $\Delta\mathbf{x}_{mn}^{\mathrm{a}(\|)} = \mathbf{Q}_{mn}\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}$. The most probable value of $\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}$

$$\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}} = \hat{\mathbf{P}}_{mn}^{\mathrm{a}}\hat{\mathbf{H}}_{mn}^{\mathrm{T}}\mathbf{R}_{mn}^{-1}\left[\mathbf{y}_{mn}^{\mathrm{o}} - \mathcal{H}_{mn}\left(\bar{\mathbf{x}}_{mn}^{\mathrm{b}}\right)\right], \tag{22}$$

is the minimizer of $J\left(\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}\right)$, where the analysis covariance matrix $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ is the inverse of the matrix of second derivatives (Hessian) of $J\left(\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}\right)$ with respect to $\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}}$:

$$\begin{aligned} \hat{\mathbf{P}}_{mn}^{\mathrm{a}} &= \left[\left(\hat{\mathbf{P}}_{mn}^{\mathrm{b}}\right)^{-1} + \hat{\mathbf{H}}_{mn}^{\mathrm{T}}\mathbf{R}_{mn}^{-1}\hat{\mathbf{H}}_{mn}\right]^{-1} \\ &= \hat{\mathbf{P}}_{mn}^{\mathrm{b}}\left[\mathbf{I} + \hat{\mathbf{H}}_{mn}^{\mathrm{T}}\mathbf{R}_{mn}^{-1}\hat{\mathbf{H}}_{mn}\hat{\mathbf{P}}_{mn}^{\mathrm{b}}\right]^{-1}. \end{aligned} \tag{23}$$

Finally, going back to the local space representation, we have

$$\bar{\mathbf{x}}_{mn}^{\mathrm{a}} = \mathbf{Q}_{mn}\Delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}} + \bar{\mathbf{x}}_{mn}^{\mathrm{b}}. \tag{24}$$

5. Updating the ensemble

We now wish to use the analysis information, $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ and $\bar{\mathbf{x}}_{mn}^{\mathrm{a}}$, to obtain an ensemble of global analysis fields $\{\mathbf{x}^{\mathrm{a}(i)}(\mathbf{r}, t)\}$; $i = 1, 2, \ldots, k' + 1$. Once these fields are determined, they can be used as initial conditions for the atmospheric model. Integrating these global fields forward in time to the next analysis time $t + \Delta t$, we obtain the background ensemble $\{\mathbf{x}^{\mathrm{b}(i)}(\mathbf{r}, t + \Delta t)\}$. This completes the loop, and, if the procedure is stable, it can be repeated for as long as desired.

Our remaining task is to specify the ensemble of global analysis fields $\{\mathbf{x}^{\mathrm{a}(i)}(\mathbf{r}, t)\}$ from our analysis information, $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ and $\bar{\mathbf{x}}_{mn}^{\mathrm{a}}$. Denote $(k' + 1)$ local analysis vectors by

$$\mathbf{x}_{mn}^{\mathrm{a}(i)} = \bar{\mathbf{x}}_{mn}^{\mathrm{a}} + \delta\mathbf{x}_{mn}^{\mathrm{a}(i)}. \tag{25}$$

Using eqs. (14) and (15) we write

$$\delta\mathbf{x}_{mn}^{\mathrm{a}(i)} = \delta\mathbf{x}_{mn}^{\mathrm{a}(i)(\|)} + \delta\mathbf{x}_{mn}^{\mathrm{a}(i)(\perp)} = \mathbf{Q}_{mn}\delta\hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} + \delta\mathbf{x}_{mn}^{\mathrm{a}(i)(\perp)}. \tag{26}$$

In addition, we let

$$\delta\mathbf{x}_{mn}^{\mathrm{a}(i)(\perp)} = \delta\mathbf{x}_{mn}^{\mathrm{b}(i)(\perp)} = \Lambda_{mn}^{(\perp)}\delta\mathbf{x}_{mn}^{\mathrm{b}(i)}, \tag{27}$$

because our analysis uses the observations only to reduce the variance in the space $\mathbb{S}_{mn}$, leaving the variance in $\bar{\mathbb{S}}_{mn}$ unchanged. [We note, however, that by our construction of $\bar{\mathbb{S}}_{mn}$ in Section 3, the total variance in $\bar{\mathbb{S}}_{mn}$ is expected to be small

compared to that in $\mathbb{S}_{mn}$. Also, in the case $k = k'$ all members of the analysis perturbation ensemble will lie in $\mathbb{S}_{mn}$, so that projection on to $\mathbb{S}_{mn}$ is superfluous, and $\delta \mathbf{x}_{mn}^{\mathrm{a}(i)(\perp)}$ in eq. (26) and the term $\Lambda_{mn}^{(\perp)} \delta \mathbf{x}_{mn}^{\mathrm{b}(i)}$ in eq. (28) may be omitted.] Combining eqs. (25)–(27), we have

$$\mathbf{x}_{mn}^{\mathrm{a}(i)} = \bar{\mathbf{x}}_{mn}^{\mathrm{a}} + \mathbf{Q}_{mn} \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} + \Lambda_{mn}^{(\perp)} \delta \mathbf{x}_{mn}^{\mathrm{b}(i)}. \tag{28}$$

We require that

$$\sum_{i=1}^{k'+1} \delta \mathbf{x}_{mn}^{\mathrm{a}(i)} = \mathbf{0}, \tag{29}$$

which, by virtue of eq. (26), and eq. (27)

$$\sum_{i=1}^{k'+1} \delta \mathbf{x}_{mn}^{\mathrm{b}(i)} = \mathbf{0}, \tag{30}$$

is equivalent to

$$\sum_{i=1}^{k'+1} \delta \mathbf{x}_{mn}^{\mathrm{a}(i)(\|)} = \mathbf{Q}_{mn} \sum_{i=1}^{k'+1} \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} = \mathbf{0}. \tag{31}$$

Thus we require that

$$\sum_{i=1}^{k'+1} \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} = \mathbf{0}. \tag{32}$$

In addition, $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ is given by

$$\hat{\mathbf{P}}_{mn}^{\mathrm{a}} = k'^{-1} \sum_{i=1}^{k'+1} \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} \big[\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} \big]^{\mathrm{T}}. \tag{33}$$

We now turn to the task of determining the analysis perturbations $\{\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}\}$. Once these are known $\{\mathbf{x}_{mn}^{\mathrm{a}(i)}\}$ is determined from eq. (28).

5.1. Determining the ensemble of local analysis perturbations

There are many choices for $\{\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}\}$ that satisfy eqs. (32) and (33), and in this section we will describe possible methods for computing a set of solutions to these equations. [See also Tippett et al. (2002) for different approaches to this problem in the global setting.] There are two main criteria we have in mind in formulating these methods. First, the method for computing $\{\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}\}$ should be numerically stable and efficient. Second, because we wish to specify global fields that we think of as being similar to physical fields, we desire that these fields be slowly varying in m and n. That is, if $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ is slowly varying, we do not want to introduce any artificial rapid variations in the individual $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}$ through our method of constructing a solution of eqs. (32) and (33). For this purpose we regard the background vectors as physical states, and hence slowly varying in m and n. (This is reasonable because the background ensemble is obtained from evolution of the atmospheric model from time $t - \Delta t$ to time t.) In Section 5.2 we use the desired slow variation of the analysis in formulating our method of constructing $\{\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}\}$.

Thus we are motivated to express the analysis ensemble vectors $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}$ as formally linearly related to the background ensemble vectors. $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(2)}, \cdots, \delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(k'+1)}$,

$$\hat{\mathbf{X}}_{mn}^{\mathrm{a}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn}, \tag{34}$$

where the matrix $\mathbf{Y}_{mn}$ is $(k' + 1) \times (k' + 1)$, and

$$\hat{\mathbf{X}}_{mn}^{\mathrm{a,b}} = (k')^{-1/2} \big\{ \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a,b}(1)} \big| \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a,b}(2)} \big| \cdots \big| \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a,b}(k'+1)} \big\}. \tag{35}$$

In the atmospheric science literature a global equivalent of transformation (34) was first applied by Bishop et al. (2001). It was later pointed out by Tippett et al. (2002) that the method obtained by Bishop et al. is essentially one particular implementation of the more general class of square-root Kalman filters, known by engineers since the 1960s (Andrews, 1968; Bierman, 1977).

Using eq. (35) the analysis and the background covariance matrices can be expressed as

$$\hat{\mathbf{P}}_{mn}^{\mathrm{a}} = \hat{\mathbf{X}}_{mn}^{\mathrm{a}} \hat{\mathbf{X}}_{mn}^{\mathrm{aT}}, \qquad \hat{\mathbf{P}}_{mn}^{\mathrm{b}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} \hat{\mathbf{X}}_{mn}^{\mathrm{bT}}. \tag{36}$$

Considering eq. (34), we see that eq. (36) yields the following equation for $\mathbf{Y}_{mn}$

$$\hat{\mathbf{P}}_{mn}^{\mathrm{a}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn} \mathbf{Y}_{mn}^{\mathrm{T}} \hat{\mathbf{X}}_{mn}^{\mathrm{bT}}. \tag{37}$$

We note that eq. (32) can be written as

$$\hat{\mathbf{X}}_{mn}^{\mathrm{a}} \mathbf{v} = \mathbf{0}, \tag{38}$$

where $\mathbf{v}$ is a column vector of $(k' + 1)$ ones. Thus, in addition to eq. (37), we demand that $\mathbf{Y}_{mn}$ must also satisfy

$$\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn} \mathbf{v} = \mathbf{0}. \tag{39}$$

Equations (37) and (39) have infinitely many solutions for $\mathbf{Y}_{mn}$. See Tippett et al. (2002) for an extensive discussion.

5.2. 'Optimal' choices for $\mathbf{Y}_{mn}$

Because our assimilations are performed independently in each local region, the smoothness and physical realizability of our analysis fields are a concern. Certainly, in order for the results to vary slowly from one grid point to the next, it is important that we use an algorithm for computing a solution of eqs. (37) and (39) that depends continuously on $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ and $\hat{\mathbf{P}}_{mn}^{\mathrm{b}}$. Furthermore, because we think of the background ensemble members as reasonable physical fields, we adopt the idea that we will seek to choose the analysis perturbations $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}$ so as to minimize their difference with the background

$$\begin{aligned} \mathcal{F}_D \big[\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} \big] &= \sum_{i=1}^{k'+1} \big\| \delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} - \delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(i)} \big\|_D^2 \\ &= \sum_{i=1}^{k+1} \big[\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} - \delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(i)} \big]^{\mathrm{T}} \\ &\quad \times \mathbf{D}_{mn}^{-1} \big[\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)} - \delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(i)} \big], \end{aligned} \tag{40}$$

where the positive definite symmetric matrix $\mathbf{D}_{mn}$ specifies the metric. With this minimization condition imposed on eqs. (37) and (39) the solution for $\mathbf{Y}_{mn}$ is now unique, although it depends on the choice of the metric $\mathbf{D}_{mn}$. As shown in the Appendix, the two choices, $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ and $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{\mathrm{b}}$, both yield the same solution for $\mathbf{Y}_{mn}$. These choices appear to be favorable in that they provide a natural normalization of the distance by the appropriate uncertainty. Accordingly, we will henceforth use this solution for $\mathbf{Y}_{mn}$

$$\mathbf{Y}_{mn} = \left[\mathbf{I} + \mathbf{X}_{mn}^{\mathrm{bT}} \left(\hat{\mathbf{P}}_{mn}^{\mathrm{b}} \right)^{-1} \left(\hat{\mathbf{P}}_{mn}^{\mathrm{a}} - \hat{\mathbf{P}}_{mn}^{\mathrm{b}} \right) \left(\hat{\mathbf{P}}_{mn}^{\mathrm{b}} \right)^{-1} \mathbf{X}_{mn}^{\mathrm{b}} \right]^{1/2}, \quad (41)$$

where the square root indicated above is the positive symmetric square root. See the Appendix and Ott et al. (2002) for the solution of eqs. (37) and (39) subject to the condition that eq. (40) be minimum.

5.3. Construction of the global fields

Using eqs. (34) and (41) we now have the ensemble of analysis perturbations $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}$. Putting these in eq. (28) we now obtain $(k' + 1)$ local analyses $\mathbf{x}_{mn}^{\mathrm{a}(i)}$ at each point $\mathbf{r}_{mn}$, and it now remains to use these to construct an ensemble of global fields $\{\mathbf{x}^{\mathrm{a}(i)}(\mathbf{r})\}$ that can be propagated forward in time to the next analysis time. There are various ways of doing this. One method is to take the state of the global vector, $\mathbf{x}^{\mathrm{a}(i)}$, at the point $\mathbf{r}_{mn}$ directly from the state information at $\mathbf{r}_{mn}$ contained in the local vector, $\mathbf{x}_{mn}^{\mathrm{a}(i)}$. This approach uses only the analysis results at the center point of each local region to form the global analysis vectors. A variant of this method is used in Section 6.4.

5.4. Variance inflation

In past work on ensemble Kalman filters (Anderson and Anderson, 1999 ; Whitaker and Hamill, 2002) it was found that inflating the covariance by a constant multiplicative factor slightly larger than one, on each analysis step, leads to more stable and improved analyses. One rationale for doing this is to compensate for the effect of finite sample size, which can be shown to, on average, underestimate the covariance.

In our numerical experiments in Section 6 we will consider two methods of variance inflation. One method, which we refer to as 'regular variance inflation', multiplies all background perturbations $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{b}(i)}$ by a constant $(1 + \delta)$. This corresponds to multiplying $\hat{\mathbf{P}}_{mn}^{\mathrm{b}}$ by $(1 + \delta)^2$. This method has been previously used by Anderson and Anderson (1999) and Whitaker and Hamill (2002). In addition to this method, we introduce a second variance inflation method

$$\hat{\mathbf{P}}_{mn}^{\mathrm{a}} \rightarrow \hat{\mathbf{P}}_{mn}^{\mathrm{a}} + \frac{\varepsilon \Lambda}{k} \mathbf{I}_k, \quad (42)$$

where $\Lambda = \mathrm{Trace}\{\hat{\mathbf{P}}_{mn}^{\mathrm{a}}\}$. Our results of Section 5 indicate that eq. (42) may yield superior performance. We refer to this method as enhanced variance inflation.

The variance inflation (42) has the effect of enhancing the probability of error in directions that formally show only very small error probability (i.e. eigendirections corresponding to small eigenvalues of the covariance matrices). Note that, through eqs. (41) and (34), such a modification of $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ also modifies the ensemble perturbations $\delta \hat{\mathbf{x}}_{mn}^{\mathrm{a}(i)}$.

6. Numerical experiments

6.1. Lorenz-96 model

In this section we will test the skill of the proposed local ensemble Kalman filter scheme by carrying out observing system simulation experiments (OSSEs) on the Lorenz-96 (L96) model (Lorenz, 1996; Lorenz and Emanuel, 1998)

$$\frac{\mathrm{d}x(j, t)}{\mathrm{d}t} = [x(j + 1, t) - x(j - 2, t)]x(j - 1, t) - x(j, t) + F. \quad (43)$$

Here, $j = 1, \ldots, J$, where $x(-1) = x(J - 1)$, $x(0) = x(J)$, and $x(J + 1) = x(1)$. This model mimics the time evolution of an unspecified scalar meteorological quantity, x, at J equidistant grid points along a latitude circle. We solve eq. (43) with a fourth-order Runge–Kutta time integration scheme with a time-step of 0.05 non-dimensional unit [which may be thought of as nominally equivalent to 6 h in real-world time assuming that the characteristic time-scale of dissipation in the atmosphere is 5 d; see Lorenz (1996) for details]. We emphasize that this toy model, eq. (43), is very different from a full atmospheric model, and that it can, at best, only indicate possible trends and illustrate possible behaviors.

For our chosen forcing, $F = 8$, the steady-state solution, $x(j) = F$ for $j = 1, \ldots, J$, in eq. (43) is linearly unstable. This instability is associated with unstable dispersive waves characterized by westward (i.e. in the direction of decreasing j) phase velocities and eastward group velocities. Lorenz and Emanuel (1998) demonstrated by numerical experiments for $F = 8$ and $J = 40$ that the x field is dominated by a wavenumber 8 structure, and that the system is chaotic; it has 13 positive Lyapunov exponents, and its Lyapunov dimension (Kaplan and Yorke, 1979) is 27.1. It can be expected that, due to the eastward group velocities, growing uncertainties in the knowledge of the model state propagate eastward. A similar process can be observed in operational numerical weather forecasts, where dispersive short (longitudinal wavenumber 6–9) Rossby waves, generated by baroclinic instabilities, play a key role in the eastward propagation of uncertainties (e.g. Persson, 2000; Szunyogh et al., 2002; Zimin et al., 2003).

We carried out experiments with three different size systems ($J = i \times 40$, $i = 1, 2, 3$) and found that increasing the number of variables did not change the wavelength, i.e. the x fields were dominated by wavenumber $i \times 8$ structures.

6.2. Root mean square analysis error

The 40-variable version of the L96 model was also used by Whitaker and Hamill (2002) to validate their ensemble square root filter (EnSRF) approach. In designing our OSSEs we follow their approach of first generating the 'true state', $x^t(j, t)$, $j = 1, \ldots, J$, by a long (40 000 time-step) model integration; then first creating 'observations' of all model variables at each time-step by adding uncorrelated normally distributed random noise with unit variance to the 'true state' (i.e. $\mathbf{R}_m = \mathbf{I}$). (The rms random observational noise variance of 1.00 is to be compared with the value 3.61 of the time mean rms deviation of solutions, $x(j, t)$, of eq. (43) from their mean.) We found that our results were the same for Gaussian noise and for truncated Gaussian noise (we truncated at three standard deviations). The effect of reduced observational networks is studied by removing observations one by one, starting from the full network, at randomly selected locations. The reduced observational networks are fixed for all experiments. That is, the difference between a network with O observations and another with $O + 1$ observations is that there is a fixed location at which only the latter takes observations.

The observations are assimilated at each time-step, and the accuracy of the analysis is measured by the time mean of the rms error:

$$E = \left\{ \frac{1}{J} \sum_{j=1}^{J} [\bar{x}^a(j) - x^t(j)]^2 \right\}^{1/2}. \tag{44}$$

6.3. Reference data assimilation schemes

In order to the assess the skill of our data assimilation scheme in shadowing the true state, we considered three alternative schemes for comparison.

6.3.1. Full Kalman filter. For the sake of comparison with our local ensemble Kalman filter results, we first establish a standard that can be regarded as the best achievable ensemble Kalman filter result that could be obtained given that computer resources placed no constraint on computations of the analysis. (In contrast with operational weather prediction, for our simple J-variable Lorenz model, this is indeed the case.) For this purpose, we considered the state $\mathbf{x}(t) = [x(1, t), x(2, t), \ldots, x(J, t)]$ on the entire domain rather than on a local patch. Then, several ensemble Kalman filter runs were carried out with different numbers of ensemble members. In these integrations, full (k') rank estimates of the covariance matrices were considered and the ensemble perturbations were updated using eqs. (34) and (41).

We found that stable cycling of the full ensemble Kalman filter requires increasing variance inflation when the number of observations is reduced, even if several hundred ensemble members are used (e.g. the assimilation of 21 observations required 2% variance inflation). This suggests that variance inflation is
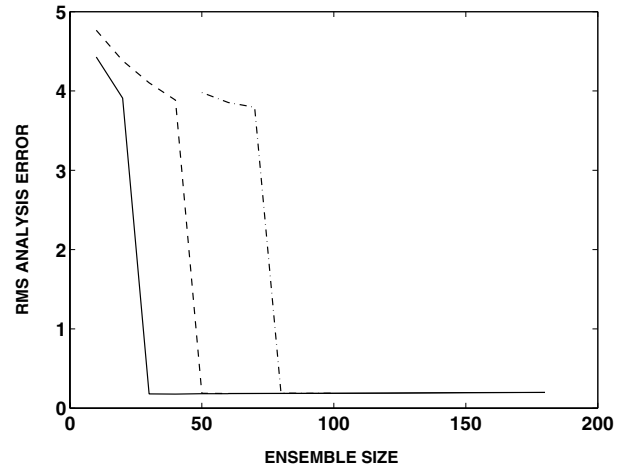

Fig 2. The rms error of the full Kalman filter as function of the number of ensemble members. The results are shown for $J = 40$ (solid line), $J = 80$ (dashed line) and $J = 120$ (dotted-dashed line).

needed, not to compensate for sampling errors due to limited sample size, but to correct for variance lost due to non-linear effects.

It can be seen that, by increasing the number of ensemble members, the time mean of E converges to ≈ 0.20 regardless of J (Fig. 2). The only difference between the different size systems (characterized by different values of J) is that more ensemble members are required to reach the minimum value for the larger systems. We refer to 0.2 as the 'optimal' error, and we regard it as a comparison standard for our local Kalman filter method. (However, we note that it may not be truly optimal because Kalman filters are rigorously optimal only for linear dynamics.)

6.3.2. Conventional method. We designed another comparison scheme that we call the conventional method, to obtain an estimate of the analysis error that can be expected from a procedure analogous to a 3D-Var scheme adapted to the L96 model. In this scheme, only the best estimate of the true state is sought (not an ensemble of analyses) using a constant estimate of the background error covariance matrix that does not change with time or position. This background error covariance matrix was determined by an iterative process. In the first step, the background error covariance matrices from the full Kalman filter were averaged over all locations and time-steps to obtain a first estimate. Then, a time series of the true background error vector $\mathbf{b} = \mathbf{x}^t - \bar{\mathbf{x}}^b$ was generated and used to obtain an estimate of the background error covariance matrix for the next iteration step. This step was repeated until the estimated background error covariance matrix converged, where the convergence was measured by the Frobenius matrix norm. We found that this procedure was always convergent when all variables were observed. The estimate obtained this way is not necessarily optimal in the sense of providing the smallest possible analysis error of any constant background error matrix, but it has the desirable feature that the

background error statistics are correctly estimated by the analysis scheme. This is a big advantage compared to the operational schemes, for which the estimate of the background error covariance matrix has to be computed by rather ad hoc techniques, because the true state, and therefore the true background error statistics, are not known. Thus, it might be assumed that our 'conventional method' provides an estimate of the analysis error that is of good accuracy as compared to analogous operational schemes.

For reduced observational networks ($O < J$), the background error covariance matrix was determined by starting the iteration from the background error covariance matrix for $O + 1$. It was found that, when more than a few observations (more than six for $J = 40$) were removed, our iterative determination of background error covariance matrices started to diverge after an initial phase of convergence. This probably occurs because the background error becomes inhomogeneous, due to the inhomogeneous observing network, and the average background error underestimates the error at the locations where the background error is larger than average. This leads to a further increase of the background error at some locations, resulting in an overall underestimation of the background error. This highlights an important limitation of the schemes based on a static estimate of the background error covariance matrix: the data assimilation scheme must overestimate the average background error in order to prevent the large local background errors from further growth. Keeping this in mind, we chose that member of our iteration scheme that provided the smallest analysis error.

6.3.3. Direct insertion. We now give a third standard designed to decide whether the data assimilation schemes provide any useful information compared to an inexpensive and simple scheme, not requiring matrix operations. This scheme, called direct insertion, updates the state estimate by replacing the background with the observations, where observations are available, and leaving the background unchanged, where there are no observations.

6.4. Implementation of the local ensemble Kalman filter

Because eq. (43) is a system with one spatial dimension (rather than the two spatial dimensions assumed in Sections 2–5), we here label local regions by a single spatial index (rather than the label *mn* used in Sections 2–5). Thus, by local region *m* we mean the points located at $j = (m - l), \ldots, m, \ldots, (m + l)$.

In our experiments, the local analysis covariance matrix is computed by eq. (23) and the local analysis mean is obtained by eq. (22). The analysis ensemble is updated by eqs. (34) and (41) and the variance of the analysis ensemble is increased by a factor of $1 + \varepsilon$ in each step using eq. (42). The final analysis ensemble at each point j is computed by averaging the values of $\mathbf{x}^{(i)}(j)$ that result from the analyses for the $2l' + 1$ local regions $j - l', \ldots, j, \ldots, j + l'$ where $l' < l$. We found that this averaging can give better results for small ensembles (for about $k < 20$) than simply taking $x^{(i)}(j)$ as the value at the mid-
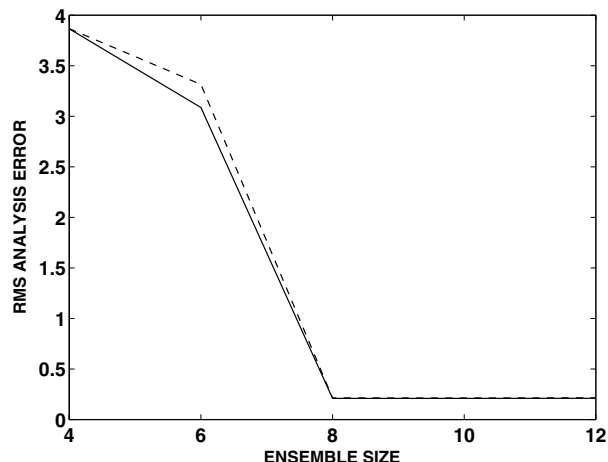

Fig 3. The rms error of the local ensemble Kalman filter as function of the number of ensemble members. The results are shown for $J = 40$ (solid line), $J = 80$ and $J = 120$ which coincide (dashed line).

dle of the local region (Section 5.3). Choosing $l = 6$, $k = k'$, $l' = 2$, and $\varepsilon = 0.12$, values that for $J = 40$ gave the lowest mean error (≈ 0.2), we have found that the mean error does not change with increasing J.

Figure 3 shows, that when the ensemble has at least eight members, the analysis error settles at the level (≈ 0.2) of the 'optimal' scheme, independent of J. This independence of our scheme (and presumably other schemes that use covariance localization) on system size is roughly consistent with the supposition of an effective correlation that is less than J. Thus, our method appears to be effective on large systems of this type. Moreover, the (nonparallelized) analysis computational time scales linearly with the number of local regions (i.e. with J). This favorable scaling is to be expected, because the analysis computation size in each local region is independent of J.

We note that the aforementioned scaling property of the local Kalman filter is in contrast to the behavior of the full Kalman filter, which, in order to achieve the 'optimal' precision, requires many more ensemble members as J is increased. This demonstrates the potential superiority of the local Kalman filter in terms of computational efficiency when applied to large systems. Furthermore, because the minimum error was independent of J, it suffices to use the smallest, 40-variable, system for further experimentation.

6.5. Comparison of the data assimilation schemes

The four data assimilation schemes (local ensemble Kalman filter, full Kalman filter, conventional method, and direct insertion) are compared for different numbers of observations (Fig. 4). The two Kalman filter schemes give almost identical error results, although the full Kalman filter has a very small advantage. The two Kalman filter schemes and the conventional data assimilation scheme are always more accurate than direct insertion, indicating that they are always able to retrieve non-trivial,
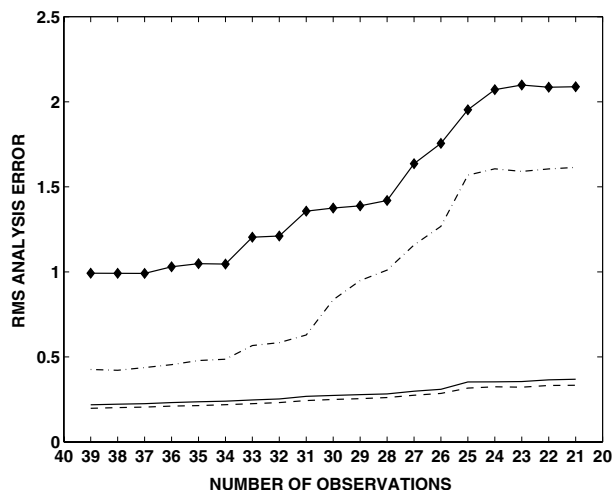
Fig 4. The rms error of the different analysis schemes as function of the number of observations. The results are shown for the full Kalman filter (4% variance inflation; dashed line), conventional scheme (dashed-dotted line), direct insertion (solid line with diamonds), and the local ensemble Kalman filter (3% variance inflation; solid line).

Table 1. Dependence of the time mean rms error on the box size $(2l + 1)$ and the rank (k) of the background covariance matrix. The coefficient of the enhanced variance inflation is $\epsilon = 0.012$

				k			
$2l + 1$	3	4	5	6	7	8	9
5	0.23	0.23	0.23				
7	0.22	0.21	0.22	0.22	0.22		
9	0.23	0.21	0.21	0.21	0.21	0.20	0.21
11	0.21	0.22	0.20	0.20	0.20	0.20	0.20
13	1.43	1.79	0.20	0.20	0.20	0.20	0.20
15	2.51	2.61	0.24	0.20	0.19	0.20	0.19

useful information about the true state. The two Kalman filter schemes, in addition, have a growing advantage over the conventional scheme as the number of observations is decreased. This shows that, as the observational network and the background error become more inhomogeneous, the adaptive nature of the background error covariance matrix in the Kalman filters leads to a growing advantage over the static scheme.

The above numerical experimentation results provide a guide for making good parameter choices in the case of the L96 model. In future applications to actual weather models, choices for the analysis parameters might similarly be determined by experimentation, but it would also be useful to obtain some guides for initial guesses of good parameter choices.

6.6. Sensitivity to the free parameters

The free parameters of our scheme are the dimensionality of the local regions (which is $2l + 1$), the rank of the covariance matrices (k), and the coefficient (ε) in the enhanced variance inflation algorithm. These parameters have been fixed so far. In what follows, the sensitivity of the data assimilation scheme to the tunable free parameters is investigated by numerical experiments (k' and l' are held fixed at $k' = 9$ and $l' = 2$). In these experiments, our 'true state' and observations are generated in the same way as in Whitaker and Hamill (2002) ($O = J$). Also, we use the same ensemble size as Whitaker and Hamill ($k' + 1 = 10$). Hence our analysis error results and theirs can be directly compared.

In the first experiment, the variance inflation coefficient is constant, $\varepsilon = 0.012$, while the dimension of the local vectors $(2l + 1)$ and the rank (k) of the background covariance matrix

are varied. The results are shown in Table 1. The scheme seems to be stable and accurate for a wide range of parameters. The optimal size local region consists of $2l + 1 = 9, 11, 13$ grid points, at which rank $k = 5, 6, 7, 8, 9$ estimates of the background covariance matrix provide similarly accurate analyses. Moreover, rank 3 and 4 estimates lead to surprisingly accurate analyses for the smaller size ($2l + 1 = 5, 7, 9$) local regions. This indicates that the background uncertainty in a local region at a given time ($\hat{\mathbf{P}}^b_m$) can be well approximated in a low-dimensional (k-dimensional) linear space. Conversely, the large error for $2l + 1 = 13, 14$ and $k = 2, 3$ can be explained on the basis that the state in these larger patch sizes is correspondingly higher-dimensional. Our premise, that the dimension of this space can be significantly lower than the number of ensemble members ($k' + 1$) needed to evolve the uncertainty, proved to be correct for the L96 model. (We note that the local dimensionality k is also much smaller than the 'global' Lyapunov dimension, 27.1, of the system.) We note that our best results are at least as good as the best results published in Whitaker and Hamill (2002) and attain the optimal value (0.20) from Section 6.3. The fact that the error remains small for $(2l + 1) = 5, 7, 9$ even when we use small k ($k = 3, 4$ as compared to our $k' = 9$) may be significant, particularly for implementation on a real weather model, because the computational cost of the relevant matrix operations scales as k^3.

In the second experiment, the dimension of the local regions is constant ($2l + 1 = 13$), while the rank and the variance inflation coefficient are varied. The results are shown in Table 2. The second experiment was then repeated by using the regular variance inflation of Anderson and Anderson (1999) and Whitaker and Hamill (2002). In the regular variance inflation, all background ensemble perturbations are multiplied by $r = 1 + \delta$, where δ is small, $1 \gg \delta > 0$. This inflation strategy increases the total variance in the background ensemble by a factor of $(1 + \Delta) = 1 + \delta^2 + 2\delta$. It can be seen from Table 3 that, to achieve near optimal results, values of Δ at least three times as large as those of ε are required. The main difference between the two inflation schemes is that the enhanced scheme inflates the dominant eigendirections of the background covariance matrix less aggressively, and the

Table 2. Dependence of the time mean rms error on the coefficient (ε) of the enhanced variance inflation scheme and the rank (k) of the background covariance matrix. The window size is $2l + 1 = 13$

ϵ	k				
	5	6	7	8	9
0.008	1.16	0.20	0.20	0.19	0.19
0.010	0.22	0.19	0.19	0.20	0.20
0.012	0.21	0.20	0.19	0.20	0.19
0.014	0.20	0.20	0.20	0.20	0.19
0.016	0.21	0.20	0.19	0.19	0.20
0.018	0.21	0.19	0.19	0.20	0.20
0.202	0.20	0.20	0.20	0.19	0.20

Table 3. Dependence of the rms analysis error on Δ in the regular variance inflation scheme and the rank (k) of the background error covariance matrix. The window size is 13

Δ	k				
	5	6	7	8	9
0.024	0.87	0.42	0.21	0.21	0.21
0.028	0.36	0.20	0.20	0.20	0.20
0.032	0.20	0.20	0.20	0.20	0.20
0.036	0.20	0.20	0.20	0.29	0.20
0.040	0.20	0.20	0.20	0.20	0.20
0.044	0.20	0.20	0.20	0.20	0.20
0.048	0.20	0.20	0.20	0.20	0.20

least dominant eigendirections more aggressively. The numerical results suggest that this feature of the scheme is beneficial, indicating that the ensemble-based estimate of the background error is more reliable in the more unstable directions than in the other directions. For further discussion, see Ott et al. (2002).

Finally, we re-emphasize that the significance of the results of all our numerical experiments on the toy model (43) is limited. Many important factors of real weather forecasting are not represented (e.g. model error), and very idealized conditions are assumed (e.g. known, normal, uncorrelated, unbiased, observation errors, and no 'subgrid scale' stochastic-like input to the evolution of the 'truth' state). On the other hand, it is also probably reasonable to assume that, if our assimilation procedure gave unfavorable results for our idealized toy model situation, then the scheme would also be unlikely to be effective in the real case. Thus, one can view the good results obtained with our assimilation scheme in these numerical experiments as necessary, but certainly not sufficient, for future successful performance in a real situation.

7. Summary and conclusions

In this paper, we have introduced a local method for assimilating atmospheric data to determine best-guess current atmospheric states. The main steps in our method are summarized in Section 2 and Fig. 1.

Numerical tests of our method using the Lorenz model, eq. (43), have been performed. These tests indicate that the method is potentially very effective in assimilating data. Other potential favorable features of our method are that only low-dimensional matrix operations are required, and that the analyses in each of the local regions are independent, suggesting the use of efficient parallel computation. These features should make possible fast data assimilation in operational settings. This is supported by preliminary work in which we have implemented our method on the T62, 28-level version of the NCEP GFS model. The assimilation of a total number of 1.5×10^6 observations (including wind, temperature, and surface pressure observations) at $k' = k = 39$ and $2l + 1 = 9$ takes about 6 min CPU time on 40 2.8-GHz Xeon processors. Experiments with the NCEP GFS also show that the strategy of assembling the global fields by using information only from the center point of the local regions provides well-balanced analysis fields. Regardless of the number of assimilated observation, the spatial and temporal smoothness of the analyzed fields was the same as that of the nature run. This indicates that the localization strategy we propose does not lead to spurious gravity wave excitation. A detailed account of these results will be published in a forthcoming paper.

Finally, we note that, although in this paper we assume that observations are synchronous with analyses, the extension to the non-synchronous case can be carried out by a technique discussed by Hunt et al. (2004). The latter paper shows that ensemble Kalman filters are, in general, relatively easily adaptable to accommodating non-synchronous observations.

8. Acknowledgments

9. Appendix A: finding the analysis ensemble that minimizes the distance to the background ensemble

A.1. Minimizing solution in terms of $\mathbf{Z}_{mn}$

As will be shown subsequently, an equivalent alternative to the expression (34) for the analysis ensemble $\hat{\mathbf{X}}_{mn}^{a}$ is

$$\hat{\mathbf{X}}_{mn}^{a} = \mathbf{Z}_{mn}\hat{\mathbf{X}}_{mn}^{b}. \tag{A1}$$

Using eq. (A1) in eq. (36) we obtain

$$\hat{\mathbf{P}}_{mn}^{a} = \mathbf{Z}_{mn}\hat{\mathbf{P}}_{mn}^{b}\mathbf{Z}_{mn}^{T}. \tag{A2}$$

Considering eq. (A1), we see that eq. (38) is automatically satisfied by eq. (A1) because the background perturbations sum to zero, eq. (30). We seek the solution of eq. (A2) that minimizes eq. (40). Defining

$$\tilde{\mathbf{X}}_{mn}^{a,b} = (\mathbf{D}_{mn})^{-1/2}\,\hat{\mathbf{X}}_{mn}^{a,b}, \quad \text{and}$$
$$\delta\tilde{\mathbf{x}}_{mn}^{a,b} = (\mathbf{D}_{mn})^{-1/2}\,\delta\hat{\mathbf{x}}_{mn}^{a,b}, \tag{A3}$$

we obtain from eqs. (40) and (A3)

$$\mathcal{F}_D\big[\delta\hat{\mathbf{x}}_{mn}^{a(i)}\big] = \sum_{i=1}^{k'+1}\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)} - \delta\tilde{\mathbf{x}}_{mn}^{b(i)}\big]^T\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)} - \delta\tilde{\mathbf{x}}_{mn}^{b(i)}\big]. \tag{A4}$$

Introducing a $k \times k$ matrix $\mathbf{B}_{mn}$ of Lagrange multipliers, we form the following quantity

$$\mathcal{L} = \sum_{i=1}^{k'+1}\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)} - \delta\tilde{\mathbf{x}}_{mn}^{b(i)}\big]^T\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)} - \delta\tilde{\mathbf{x}}_{mn}^{b(i)}\big]$$
$$- \sum_{p,q=1}^{k}(\mathbf{B}_{mn})_{p,q}$$
$$\times \left\{(\tilde{\mathbf{P}}_{mn}^{a})_{p,q} - \frac{1}{k'}\sum_{i=1}^{k'+1}\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)}\big]_p\big[\delta\tilde{\mathbf{x}}_{mn}^{a(i)}\big]_q\right\}, \tag{A5}$$

where

$$\tilde{\mathbf{P}}_{mn}^{a} = (\mathbf{D}_{mn})^{-1/2}\,\hat{\mathbf{P}}_{mn}^{a}\,(\mathbf{D}_{mn})^{-1/2},$$
$$\tilde{\mathbf{P}}_{mn}^{b} = (\mathbf{D}_{mn})^{-1/2}\,\hat{\mathbf{P}}_{mn}^{b}\,(\mathbf{D}_{mn})^{-1/2}. \tag{A6}$$

We now minimize eq. (A5) with respect to $\delta\tilde{\mathbf{x}}_{mn}^{a(i)}$ and $\mathbf{B}_{mn}$. Forming the first and second derivatives of $\mathcal{L}$ with respect to $\delta\tilde{\mathbf{x}}_{mn}^{a(i)}$, we have

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\delta\tilde{\mathbf{x}}_{mn}^{a(i)}} = \tilde{\mathbf{Z}}_{mn}^{-1}\delta\tilde{\mathbf{x}}_{mn}^{a(i)} - \delta\tilde{\mathbf{x}}_{mn}^{b(i)}, \tag{A7}$$

$$\frac{1}{2}\frac{\partial^2\mathcal{L}}{\partial\delta\tilde{\mathbf{x}}_{mn}^{a(i)}\partial\delta\tilde{\mathbf{x}}_{mn}^{a(i)}} = \tilde{\mathbf{Z}}_{mn}^{-1}, \tag{A8}$$

where we have defined $\tilde{\mathbf{Z}}_{mn}^{-1}$ as

$$\tilde{\mathbf{Z}}_{mn}^{-1} = \mathbf{I} + \frac{1}{2k'}\big(\mathbf{B}_{mn} + \mathbf{B}_{mn}^T\big). \tag{A9}$$

Because $\mathcal{L}$ is stationary

$$\tilde{\mathbf{X}}_{mn}^{a} = \tilde{\mathbf{Z}}_{mn}\tilde{\mathbf{X}}_{mn}^{b}, \tag{A10}$$

and the derivative with respect to $\mathbf{B}_{mn}$ yields

$$\tilde{\mathbf{P}}_{mn}^{a} = \tilde{\mathbf{Z}}_{mn}\tilde{\mathbf{P}}_{mn}^{b}\tilde{\mathbf{Z}}_{mn}^{T} \tag{A11}$$

Equation (A11) for $\tilde{\mathbf{Z}}_{mn}$ is the same as eq. (A2) for $\mathbf{Z}_{mn}$, except that we have additional conditions that $\tilde{\mathbf{Z}}_{mn}$ must satisfy. Namely, by virtue of eq. (A9), we know that the solution of eq. (A11) that minimizes $\mathcal{F}_D[\delta\hat{\mathbf{x}}_{mn}^{a(i)}]$ is symmetric:

$$\tilde{\mathbf{Z}}_{mn} = \tilde{\mathbf{Z}}_{mn}^{T}. \tag{A12}$$

Furthermore, from eq. (A8) and the condition for a minimum, the matrix $\tilde{\mathbf{Z}}_{mn}$ must be positive definite. With these extra conditions,

the solution of eq. (A11) is unique. To see this, we pre- and post-multiply eq. (A11) by the positive definite symmetric square root of $\tilde{\mathbf{P}}_{mn}^{b}$ [which we denote by $(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}$] to obtain

$$(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}\,\tilde{\mathbf{P}}_{mn}^{a}\,(\tilde{\mathbf{P}}_{mn}^{b})^{1/2} = \big[(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}\,\tilde{\mathbf{Z}}_{mn}\,(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}\big]^2 \tag{A13}$$

where we have made use of eq. (A12) and the fact that $\tilde{\mathbf{P}}_{mn}^{b}$ is symmetric. All the matrices in eq. (A13) are positive definite symmetric, as are the left and right sides of the equation. Thus, taking the unique positive definite symmetric square root of eq. (A13), we obtain the solution for $\tilde{\mathbf{Z}}_{mn}$

$$\tilde{\mathbf{Z}}_{mn} = (\tilde{\mathbf{P}}_{mn}^{b})^{-1/2}\big[(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}\tilde{\mathbf{P}}_{mn}^{a}(\tilde{\mathbf{P}}_{mn}^{b})^{1/2}\big]^{1/2}(\tilde{\mathbf{P}}_{mn}^{b})^{-1/2}. \tag{A14}$$

From eqs. (A1), (A3) and (A10) we have

$$\mathbf{Z}_{mn} = (\mathbf{D}_{mn})^{1/2}\,\tilde{\mathbf{Z}}_{mn}\,(\mathbf{D}_{mn})^{-1/2}, \tag{A15}$$

which is the solution for $\mathbf{Z}_{mn}$ that minimizes the quadratic form (40), where in eq. (A15) $\tilde{\mathbf{Z}}_{mn}$ is defined by eqs. (A14) and (A6). Due to the fact that the metric $\mathbf{D}_{mn}$ is so far still unspecified, our result still gives an infinite family of solutions for $\mathbf{Z}_{mn}$. However, we feel that the specification of a suitable metric provides a reasonable basis for choosing a particular solution. In particular, $\mathbf{D}_{mn}$ equal to $\hat{\mathbf{P}}_{mn}^{b}$ or $\hat{\mathbf{P}}_{mn}^{a}$ appears to be favorable in that it yields a natural, intuitive normalization of the distance between $\delta\hat{\mathbf{x}}_{mn}^{a(i)}$ and $\delta\hat{\mathbf{x}}_{mn}^{b(i)}$ based on the uncertainty of these quantities. Furthermore, both $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{a}$ and $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{b}$ yield the same $\mathbf{Z}_{mn}$. To see this, we note that eq. (A2) can be put in the form

$$(\hat{\mathbf{P}}_{mn}^{a})^{1/2}\big[(\hat{\mathbf{P}}_{mn}^{a})^{-1/2}\mathbf{Z}_{mn}(\hat{\mathbf{P}}_{mn}^{a})^{1/2}\big]^{-1}(\hat{\mathbf{P}}_{mn}^{a})^{1/2}$$

$$= (\hat{\mathbf{P}}_{mn}^{b})^{1/2}\big[(\hat{\mathbf{P}}_{mn}^{b})^{-1/2}\mathbf{Z}_{mn}(\hat{\mathbf{P}}_{mn}^{b})^{1/2}\big]^T(\hat{\mathbf{P}}_{mn}^{b})^{1/2}. \tag{A16}$$

Thus, symmetry of $(\hat{\mathbf{P}}_{mn}^{a})^{-1/2}\mathbf{Z}_{mn}(\hat{\mathbf{P}}_{mn}^{a})^{1/2}$ (corresponding to the choice $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{a}$) implies symmetry of $(\hat{\mathbf{P}}_{mn}^{b})^{-1/2}\mathbf{Z}_{mn}(\hat{\mathbf{P}}_{mn}^{b})^{1/2}$ (i.e. $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{b}$). We thus conjecture that $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{a,b}$ will yield better performance than other choices.

A.2. Minimizing solution of eqs. (37) and (39) for $\mathbf{Y}_{mn}$

We now wish to translate what we have found above for $\mathbf{Z}_{mn}$ to the corresponding $\mathbf{Y}_{mn}$. Another way of solving for the analysis fields is to use the 'Potter method' (Bierman, 1977). To see how this solution is obtained, let

$$\mathbf{A}_{mn} = \mathbf{Y}_{mn}\mathbf{Y}_{mn}^{T} \tag{A17}$$

so that eq. (37) becomes

$$\hat{\mathbf{P}}_{mn}^{a} = \hat{\mathbf{X}}_{mn}^{b}\mathbf{A}_{mn}\hat{\mathbf{X}}_{mn}^{bT}. \tag{A18}$$

A solution for $\mathbf{A}_{mn}$ consistent with eqs. (A17) and (A18) is

$$\mathbf{A}_{mn} = \mathbf{I} + \hat{\mathbf{X}}_{mn}^{bT}(\hat{\mathbf{P}}_{mn}^{b})^{-1}\big[\hat{\mathbf{P}}_{mn}^{a} - \hat{\mathbf{P}}_{mn}^{b}\big](\hat{\mathbf{P}}_{mn}^{b})^{-1}\hat{\mathbf{X}}_{mn}^{b}. \tag{A19}$$

This solution for $\mathbf{A}_{mn}$ is symmetric and can also be shown to be positive definite. Equation (A19) yields $\mathbf{A}_{mn} = \mathbf{I}$ if $\hat{\mathbf{P}}_{mn}^{\mathrm{a}} = \hat{\mathbf{P}}_{mn}^{\mathrm{b}}$, and satisfaction of eq. (A18) by eq. (A19) can be verified by direct substitution and making use of $\hat{\mathbf{P}}_{mn}^{\mathrm{b}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} \hat{\mathbf{X}}_{mn}^{\mathrm{bT}}$. Thus, we have as a possible solution

$$\mathbf{Y}_{mn} = (\mathbf{A}_{mn})^{1/2}. \tag{A20}$$

We now show that eqs. (A19) and (A20) also satisfy eq. (39). By eq. (30) we have $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{v} = \mathbf{0}$ and from eq. (A19) we have $\mathbf{A}_{mn} \mathbf{v} = \mathbf{v}$; i.e. $\mathbf{v}$ is an eigenvector of $\mathbf{A}_{mn}$ with eigenvalue one. Because the positive square root is employed in eq. (A20), $\mathbf{v}$ is also an eigenvector of $\mathbf{Y}_{mn}$ with eigenvalue one. Hence $\mathbf{X}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn} \mathbf{v} = \mathbf{X}_{mn}^{\mathrm{b}} \mathbf{v}$, which is identically zero by eq. (30), thus satisfying eq. (39). We now ask whether each solution $\mathbf{Z}_{mn}$ of eq. (A2) has a corresponding $\mathbf{Y}_{mn}$ such that $\mathbf{Z}_{mn} \hat{\mathbf{X}}_{mn}^{\mathrm{b}}$ and $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn}$ yield the same result for $\hat{\mathbf{X}}_{mn}^{\mathrm{a}}$. To see that such a $\mathbf{Y}_{mn}$ does indeed exist, we note that the matrix $\hat{\mathbf{X}}_{mn}^{\mathrm{b}}$ (which consists of k rows and $k' + 1$ columns) has a (non-unique) right inverse $(\hat{\mathbf{X}}_{mn}^{\mathrm{b}})^{-1}$ such that $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} (\hat{\mathbf{X}}_{mn}^{\mathrm{b}})^{-1} = \mathbf{I}_k$, where

$$\begin{aligned} (\hat{\mathbf{X}}_{mn}^{\mathrm{b}})^{-1} &= \hat{\mathbf{X}}_{mn}^{\mathrm{bT}} (\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \hat{\mathbf{X}}_{mn}^{\mathrm{bT}})^{-1} + \mathbf{E}_{mn} \\ &= \hat{\mathbf{X}}_{mn}^{\mathrm{bT}} (\hat{\mathbf{P}}_{mn}^{\mathrm{b}})^{-1} + \mathbf{E}_{mn}, \end{aligned} \tag{A21}$$

and $\mathbf{E}_{mn}$ is any $k \times (k' + 1)$ matrix for which $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{E}_{mn} = \mathbf{0}_{mn}$. Thus, from $\hat{\mathbf{X}}_{mn}^{\mathrm{a}} = \mathbf{Z}_{mn} \hat{\mathbf{X}}_{mn}^{\mathrm{b}}$, we have

$$\hat{\mathbf{X}}_{mn}^{\mathrm{a}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} (\hat{\mathbf{X}}_{mn}^{\mathrm{b}})^{-1} \mathbf{Z}_{mn} \hat{\mathbf{X}}_{mn}^{\mathrm{b}}. \tag{A22}$$

From the definition of $\mathbf{Y}_{mn}$, namely $\hat{\mathbf{X}}_{mn}^{\mathrm{a}} = \hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn}$, we see that eqs. (A21) and (A22) yield

$$\mathbf{Y}_{mn} = \hat{\mathbf{X}}_{mn}^{\mathrm{bT}} (\hat{\mathbf{P}}_{mn}^{\mathrm{b}})^{-1} \mathbf{Z}_{mn} \hat{\mathbf{X}}_{mn}^{\mathrm{b}} + \mathbf{G}_{mn}, \tag{A23}$$

where $\mathbf{G}_{mn}$ is any $(k' + 1) \times (k' + 1)$ matrix satisfying $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{G}_{mn} = \mathbf{0}$. Because we desire that $\mathbf{Y}_{mn} = \mathbf{I}_{k'+1}$, when $\mathbf{Z}_{mn} = \mathbf{I}_k$, a possible choice for $\mathbf{G}_{mn}$ is

$$\mathbf{G}_{mn} = \mathbf{I}_{k'+1} - \hat{\mathbf{X}}_{mn}^{\mathrm{bT}} (\hat{\mathbf{P}}_{mn}^{\mathrm{b}})^{-1} \hat{\mathbf{X}}_{mn}^{\mathrm{b}}. \tag{A24}$$

[We note that $\mathbf{G}_{mn}$ given by eq. (A24) is a projection operator, $(\mathbf{G}_{mn})^p = \mathbf{G}_{mn}$ for any integer exponent p.] Thus, from eqs. (A23) and (A24), a $\mathbf{Y}_{mn}$ corresponding to any solution of eq. (A2) for $\mathbf{Z}_{mn}$ is

$$\mathbf{Y}_{mn} = \hat{\mathbf{X}}_{mn}^{\mathrm{bT}} (\hat{\mathbf{P}}_{mn}^{\mathrm{b}})^{-1} (\mathbf{Z}_{mn} - \mathbf{I}_k) \hat{\mathbf{X}}_{mn}^{\mathrm{b}} + \mathbf{I}_{k'+1}. \tag{A25}$$

Using eqs. (A25) and (A2) it can be verified that $\mathbf{Y}_{mn} \mathbf{Y}_{mn}^{\mathrm{T}} = \mathbf{A}_{mn}$ with $\mathbf{A}_{mn}$ given by eq. (A19). Thus, $\mathbf{Y}_{mn} \mathbf{Y}_{mn}^{\mathrm{T}}$ is the same $(k' + 1) \times (k' + 1)$ matrix for all solutions $\mathbf{Z}_{mn}$ given by eq. (A15).

The general solution of $\mathbf{Y}_{mn} \mathbf{Y}_{mn}^{\mathrm{T}} = \mathbf{A}_{mn}$ is

$$\mathbf{Y}_{mn} = (\mathbf{A}_{mn})^{1/2} \mathbf{O}_{mn}, \tag{A26}$$

where $\mathbf{O}_{mn}$ is an arbitrary orthogonal matrix. However, to ensure that eq. (39) is satisfied we also require that $\mathbf{O}_{mn} \mathbf{v} = \pm \mathbf{v}$ (where $\mathbf{v}$ is a column vector of $(k' + 1)$ ones); i.e. that $\mathbf{v}$ is an eigenvector of $\mathbf{O}_{mn}$ with eigenvalue ± 1. For example, $\mathbf{O}_{mn}$ can be any rotation about $\mathbf{v}$. Thus, there is still a large family of allowed orthogonal

matrices $\mathbf{O}_{mn}$. (Note that $\mathbf{O}_{mn}$ can depend on $\hat{\mathbf{P}}_{mn}^{\mathrm{a}}$ and $\hat{\mathbf{P}}_{mn}^{\mathrm{b}}$, and $\mathbf{O}_{mn}$ must be $\mathbf{I}$ whenever $\hat{\mathbf{P}}_{mn}^{\mathrm{a}} = \hat{\mathbf{P}}_{mn}^{\mathrm{b}}$.)

Note from eq. (A15) that $(\hat{\mathbf{P}}_{mn}^{\mathrm{b}})^{-1} \mathbf{Z}_{mn}$ is symmetric for $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{\mathrm{b}}$. Thus, the resulting $\mathbf{Y}_{mn}$ from eq. (A25) is symmetric and must therefore coincide with eq. (A20). Thus, we have the key result of this appendix: $\mathbf{Z}_{mn} \hat{\mathbf{X}}_{mn}^{\mathrm{b}}$ with $\mathbf{Z}_{mn}$ from eq. (A15) with $\mathbf{D}_{mn} = \hat{\mathbf{P}}_{mn}^{\mathrm{a},b}$ and $\hat{\mathbf{X}}_{mn}^{\mathrm{b}} \mathbf{Y}_{mn}$ with $\mathbf{Y}_{mn}$ given by eqs. (A19) and (A20) both yield the same result for $\hat{\mathbf{X}}_{mn}^{\mathrm{a}}$.

References

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

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

Andrews, A. 1968. A square root formulation of the Kalman covariance equations. *IAAA J.* **6**, 1165–1168.

Bierman, G. J. 1977. *Factorization Methods for Discrete Sequential Estimation*. Academic Press, San Diego, pp. 241.

Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **129**, 420–436.

Cane, M. A., Kaplan, A., Miller, R. N., Tang, B., Hackert, E. C. et al. 1996. Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.* **101**, 22599–22617.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, New York.

Evensen, G. 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99C5**, 10143–10162.

Evensen, G. and van Leeuwen, P. J. 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman Filter with a quasi-geostrophic model. *Mon. Wea. Rev.* **124**, 85–96.

Fisher, M. 1998. *Development of a simplified Kalman Filter. ECMWF Research Department Tech. Memo. 260*. European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, Berkshire, United Kingdom.

Fukumori, I. 2002. A partitioned Kalman filter and smoother. *Mon. Wea. Rev.* **130**, 1370–1383.

Fukumori, I. and Malanotte-Rizzoli, P. 1995. An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *J. Geophys. Res.* **100**, 6777–6793.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K. and Isaacson, E. 1981. Applications of estimation theory to numerical weather prediction. In: *Dynamic Meteorology: Data Assimilation Methods* (eds. A. L. Bengtsson, M. Ghil and E. Kallen). Springer-Verlag, New York 139–224.

Hamill, T. M. and Snyder, C. 2000. A hybrid ensemble Kalman Filter-3D variational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.

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

Heemink, A. W., Verlaan, M. and Segers, A. J. 2001. Variance Reduced Ensemble Kalman Filtering. *Mon. Wea. Rev.* **129**, 1718–1728.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman Filter technique. *Mon. Wea. Rev.* **126**, 796–811.

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

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E. and Patil, D. J. et al. 2004. Four-dimensional ensemble Kalman filtering. *Tellus* **56A**, in print.

Jones, R. 1965. An experiment in non-linear prediction. *J. Appl. Meteorol.* **4**, 701–705.

Kalman, R. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.* **82**, 35–45.

Kalman, R. and Bucy, R. 1961. New results in linear filtering and prediction theory. *Trans. ASME, Ser. D, J. Basic Eng.* **83**, 95–108.

Kalnay, E., 2002. *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, Cambridge.

Kaplan, J. L. and Yorke, J. A. 1979. Chaotic behavior of multidimensional difference equations. In: *Functional Differential Equations and Approximation of Fixed Points. Lecture Notes in Mathematics*, Volume 730 (eds.H.-O. Peitgen and H.-O. Walter). Springer-Verlag, Berlin, 204–227.

Keppenne, C. and Rienecker, H. 2002. Initial testing of a massively parallel ensemble Kalman Filter with the Poseidon Isopycnal Ocean General Circulation Model. *Mon. Wea. Rev.* **130**, 2951–2965.

Lermusiaux, P. F. J. and Robinson, A. B., 1999. Data assimilation via error subspace statistical estimation. Part I: theory and schemes. *Mon. Wea. Rev.* **127**, 1385–1407.

Lorenc, A. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.

Lorenz, E. N., 1996. Predictability: a problem partly solved. In: *Proc. Seminar on Predictability,* Volume 1. European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, Berkshire, United Kingdom.

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.

Ott, E., Hunt, B. R., Szunyogh, I., Corazza, M. Kalnay, E. et al. 2002. Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering. Preprint (physics/0203058).

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878–5881.

Persson, A. 2000. Synoptic-dynamic diagnosis of medium range weather forecast system. In: *Proceedings of the Seminars on Diagnosis of Models and Data Assimilation Systems*. European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, Berkshire, United Kingdom.

Pham, D. T., Verron, J. and Roubaud, M. C. 1998. A singular evolutive extended Kalman filter data assimilation in oceanography. *J. Mar. Syst.* **16**, 323–340.

Szunyogh, I., Toth, Z., Zimin, A. V., Majumdar, S. J. and Persson, A. 2002. Propagation of the effect of targeted observations: the 2000 Winter Storm Reconnaissance Program. *Mon. Wea. Rev.,* **130**, 1144–1165.

Talagrand, O. 1981. A study of the dynamics of four-dimensional data assimilation. *Tellus* **33**, 43–60.

Tangborn, A. 2004. Wavelet approximation of error covariance propagation in data assimilation. *Tellus* **56A**, 16–28.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2002. Ensemble square-root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Whitaker, J. S. and Hamill, T. H. 2002. Ensemble Data Assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.

Zimin, A. V., Szunyogh, I., Patil, D. J., Hunt, B. R. and Ott, E. 2002. Extracting envelopes of Rossby wave packets. *Mon. Wea. Rev.* **131**, 1011–1017.