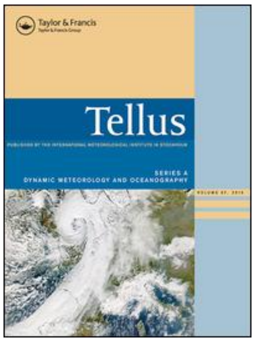


Tellus A: Dynamic Meteorology and Oceanography

ISSN: (Print) 1600-0870 (Online) Journal homepage: https://www.tandfonline.com/loi/zela20

A hybrid approach to estimating error covariances in variational data assimilation

Haiyan Cheng, Mohamed Jardak, Mihai Alexe & Adrian Sandu

To cite this article: Haiyan Cheng, Mohamed Jardak, Mihai Alexe & Adrian Sandu (2010) A hybrid approach to estimating error covariances in variational data assimilation, Tellus A: Dynamic Meteorology and Oceanography, 62:3, 288-297

To link to this article: https://doi.org/10.1111/j.1600-0870.2009.00442.x

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

Tellus (2010), 62A, 288–297

TELLUS

A hybrid approach to estimating error covariances in variational data assimilation

By HAIYAN CHENG, MOHAMED JARDAK, MIHAI ALEXE and ADRIAN SANDU*,
Computational Science Laboratory, Department of Computer Science, Virginia Polytechnic Institute and State University, 2202 Kraft Drive, Blacksburg, VA 24060, USA

(Manuscript received 2 April 2009; in final form 2 March 2010)

ABSTRACT

Data assimilation (DA) involves the combination of observational data with the underlying dynamical principles governing the system under observation. In this work we combine the advantages of the two prominent DA systems: the 4D-Var and the ensemble methods. The hybrid method described in this paper consists of identifying the subspace spanned by the major 4D-Var error reduction directions. These directions are then removed from the background covariance through a Galerkin-type projection, and are replaced by estimates of the analysis error obtained through a low-rank Hessian inverse approximation. The updated error covariance in one window can be used as the background covariance for the next window thus better capturing the 'error of the day'. The numerical results for a non-linear model demonstrate how the hybrid method leads to a good estimate of the true error covariance, and improves the 4D-Var analysis results.

1. Introduction

Data assimilation (DA) is a relatively novel and versatile multidisciplinary methodology. Given a consistent numerical model for a complex time-dependent system, and a set of observations of the model state taken at various time moments, DA allows us to combine these two sources of information to obtain a state approximation, which is, in a well defined statistical sense, the 'best' estimate of the true model state.

DA comprises three types of methods: interpolation, variational, and sequential methods (Daley, 1991; Malanotte-Rizzoli, 1996; Cohn, 1997; Ghil et al., 1997; Courtier et al., 1998; van Leeuwen, 2009). In the first method, the measurements are interpolated from the points of observation onto the grid points. The interpolation can be weighted by the statistics of the observations. While simple to implement, this approach is not justified by any physical arguments. The variational and sequential data assimilation methods fit into the framework of estimation theory. Sequential DA features the Kalman filter (Kalman, 1960; Jazwinski, 1970), the ensemble Kalman filter (EnKF) (Evensen, 1994, 2003, 2007; Burgers et al., 1998), as well as several types of particle filters (Doucet et al., 2000, 2001; Arulampalam et al., 2002; Berliner and Wikle, 2007). The main drawback of sequential data assimilation comes from the fact that the covariance matrices of the state variables have huge dimensions for operational models.

The variational methods were introduced by Sasaki (1958). These methods consider the equations governing the flow as constraints, and the problem is closed by using a variational principle, for example, the minimization of the discrepancy between the model and the observations. This problem fits into the framework of optimal control theory (Lions, 1971), as proposed by Le Dimet (1980), Le Dimet and Talagrand (1986) and Courtier and Talagrand (1987). The major drawback of the variational methods is in assuming the background error covariances being static, nearly homogeneous and isotropic. This has been pointed out in (Parrish and Derber, 1992; Cohn et al., 1998; Courtier et al., 1998; Lorenc, 2003). The ensemble methods provide an alternative to variational data assimilation. The EnKF estimates the background error covariances from an ensemble of short-term forecasts (Anderson, 2001; Houtekamer and Mitchell, 2001). Although sequential methods and variational methods are implemented separately, some connections between them have long been known. In fact, for a linear model with Gaussian errors, with the same input data (initial background state, background covariance, distribution of observations, and observation covariance), both the 4D-Var and the Kalman filter yield the same analysis at the end of the assimilation window (Talagrand and Courtier, 1987; Thépaut and Courtier, 1991).

*Corresponding author.
e-mail: sandu@cs.vt.edu
DOI: 10.1111/j.1600-0870.2010.00442.x

This equivalence between sequential and variational methods provides a possibility of combining the two approaches, leading to powerful hybrid DA algorithms.

There have been several recent research efforts in this direction. The 3D-Var/EnKF of Hamill and Snyder (2000) is a hybrid scheme blending the 3D-Var and the ensemble Kalman filter. It utilizes the ensemble framework to propagate the estimation statistics of the model. However, it does not fully exploit the smoothing characteristics of 4D-Var. The Ensemble Kalman Smoother (EnKS) of Evensen and van Leeuwen (2000) recomputes a new analysis for all recent measurements upon the receipt of each new observations. This approach is computationally intractable for real world problems. Zhang et al. (2009) devised the Ensemble 4D-Var (E4DVAR). E4DVAR is essentially constructed as a feedback loop between 4D-Var and the EnKF method. The ensemble mean in EnKF is periodically replaced by the 4D-Var analysis, and the covariance matrix built from the EnKF ensemble provides the background information in 4D-Var. Cessna et al. (2008) proposed the Ensemble Variational Estimation (EnVE) method. EnVE is a coupled hybrid of EnKF/EnKS and 4D-Var. It leverages the non-linear statistical propagation properties of the EnKF/EnKS to initialize and properly define a consistent variational iteration similar to 4D-Var.

This research work proposes a novel hybrid method to generate covariance estimates at the end of the assimilation window. Ideally, at the end of the assimilation window, the error covariance matrix shall represent the true covariance statistics of the forecast (posterior) state error. The state errors are the vectors in a n-dimensional space that can be viewed as the sum of the components in the subspace spanned by the 4D-Var and the components in the orthogonal subspace. The variational approach with 4D-Var reduces the former components, while the latter parts are untouched. Our method is the first approach to explicitly consider the error components removed by the numerical optimization method in 4D-Var. Since 4D-Var is an iterative process, in practice, the errors in the subspace spanned by the 4D-Var can not be canceled completely. We use a low-rank Hessian inverse approximation to add back the error components that are eliminated by the hybrid projection.

This approach can be thought of as a Galerkin-like projection of the background covariance onto the orthogonal complement of the space spanned by the leading 4D-Var error reduction directions. The Hessian inverse information is added through a Karhunen-Loève expansion of the leading eigenpairs of the Hessian inverse matrix. The summation of these two parts leads to a hybrid ensemble, thus a hybrid covariance. The new background covariance estimate can be constructed as a convex linear combination of the static background covariance and the hybrid covariance. This update is carried out at the end of every assimilation window to provide a new flow-dependent background covariance for the next assimilation window.

The numerical experiments in Section 4 show that the hybrid covariance estimate is superior to the static background covariance, especially when many windows are used. Moreover, the hybrid covariance update leads to a significant improvement in the quality of the model forecast.

Similar to our present work which consists on updating the background error covariance at the end of each assimilation window and based on the fixed-lag smoother proposed by Cohn et al. (1994), Li and Navon (2001) put forward the cycling 4D-Var. They separated the observations into a sequence of batches and performed a standard 4D-Var procedure separately for each batch of observations. The solution at the end of each one batch is used as a the background for the subsequent batch. Our proposed method differs from the cycling 4D-Var of Li and Navon, as we do not act at all on the observations. It is also worthy of note that the only common point between our present work and the reduced-order approach to observation sensitivity of Daescu and Navon (2008) resides in the control space spanned by the descent directions.

Qiu et al. (2007) proposed the explicit 4D-Var implementation (E-4DVAR) as an alternative to the regular 4D-Var (I-4DVar). They used the singular value decomposition (SVD) to construct an orthogonal basis in which the control variables in the cost function appear explicitly. As a consequence the analysis does not require the use of an adjoint integration. An ensemble-based 4D-Var approach was proposed in (Qiu et al., 2007). This work is very similar to the previous one. This time the ensemble members are sampled from a time series produced by a single integration of the model over a long time period. In the same vein, Tian et al. (2008) proposed the POD-E4DVAR. They merged both Monte Carlo (MC) method and proper orthogonal decomposition (POD) into the 4D-Var method. A SVD-based method has been employed to generate a POD basis that captures the most possible kinetic energy. The solution for the analysis problem is then approximately expressed by a truncated expansion. With this, the observation operator in the cost function is replaced by its tangent linear counterpart and consequently the calculation of the model adjoint is no longer needed. Although our method is an ensemble-based 4D-Var method that uses the SVD method, it differs from the aforementioned work. Our present work is geared towards generating covariance estimates at the end of the assimilation window to capture the 'error of the day'. This has been achieved by first identifying the subspace spanned by the major 4D-Var error reduction directions and then by removing them from the background covariance matrix. Hence, the drawback of the variational methods consisting of assuming the background error covariances not time-evolving is alleviated. In our work, the SVD method has been used to construct the set of the most dominant singular vectors generated from the difference of the 4D-Var optimization directions.

The rest of this paper is structured as follows. Section 2 discusses the mathematical foundation of the 4D-Var method. In Section 3, we present a step by step description of our hybrid method. The hybrid DA scheme is then tested on the Lorenz 96 model (Lorenz, 1996) in Section 4. The summary and conclusions are presented in Section 5, along with further research directions.

2. Mathematical background on 4D-Var

Consider a time-dependent model given by the initial value problem

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mathbb{M}(t, x)\,, \quad t_0 \le t \le t_F, \qquad x(t_0) = x_0. \tag{1}$$

Let $\mathcal{M}$ be a discrete solution operator for the model (1). The discrete approximation to the exact model state at time $t = t_i$ is $x_i \in \mathbb{R}^n$

$$x_i = \mathcal{M}_{t_0 \to t_i}(x_0). \tag{2}$$

In the following, we will denote by t_0 and t_F the start time and the end time of a 4D-Var assimilation window, respectively.

Given a background model state $x_0^b \in \mathbb{R}^n$ at t_0, the associated background error covariance matrix $B_{t_0} \in \mathbb{R}^{n\times n}$, a set of noisy observations $y_i \in \mathbb{R}^m$ with the observation covariance $R_i \in \mathbb{R}^{m\times m}$ at selected times t_i, the *discrete* 4D-Var data assimilation problem consists of determining the maximum likelihood initial model state $x^a(t_0)$ as the minimizer of the following cost functional

$$\mathcal{J}(x_0) = \mathcal{J}^{\mathrm{b}} + \mathcal{J}^{\mathrm{o}} = \left(x_0 - x_0^b\right)^T B_{t_0}^{-1}\left(x_0 - x_0^b\right) + \sum_{i=1}^{N_{\mathrm{obs}}} (\mathcal{H}_i x_i - y_i)^T R_i^{-1} (\mathcal{H}_i x_i - y_i)\,.$$

Here $\mathcal{J}^{\mathrm{b}}$ represents the departure from the background value x_0^b, while $\mathcal{J}^{\mathrm{o}}$ measures the misfit between the model prediction and the observations. The $\mathcal{H}_i$ are observation operators that map the state variables $x_i \approx x(t_i)$ to the observation space.

Current 4D-Var assimilation studies use the same background covariance matrices for each assimilation window. Ideally however, this background information should be updated periodically to capture the system dynamics. Using 4D-Var with an improved background covariance can be expected to lead to a higher quality forecast with a larger range of validity.

Knowledge of the forecast (a posteriori) error statistics at the endpoints of the 4D-Var window is very valuable. A good estimation of the posterior covariance at the end of one assimilation window is supposed to be used as the background covariance for the next assimilation window, allowing one to account for the errors of the day, and ultimately improve the quality of the forecast.

3. The hybrid method

4D-Var experiments usually use a static prescribed background error covariance at the beginning of each assimilation window. This matrix is built using statistical information inferred from the background state, the available observations, and the model dynamics. The drawback of this approach is that an accurate analysis is only guaranteed within each window. The ideal DA method should account for the 'errors of the day', by updating the background error covariance matrix between assimilation windows.

The objective is to estimate the analysis error covariances, so that the updated error statistics can be used in the background covariance of the next assimilation window. Note that 4D-Var by itself does not yield any covariance information. Second order adjoint models (Wang et al., 1992; Le Dimet et al., 2002) can be used to obtain eigenpair information for the Hessian $\nabla^2\mathcal{J}$. The inverse Hessian is known to approximate the *a posteriori* error covariance matrix at t_0 (Gejadze et al., 2008). Thus, we can expect that the eigenvectors corresponding to the smallest Hessian eigenvalues give some indication of the directions of maximum error growth in the dynamic model (1).

The hybrid DA method makes use of two ensembles. The 'background ensemble' is generated using the error statistics given by B_{t_0}, and each of its members is propagated (in parallel) to t_F. The 4D-Var analysis is started from the mean of the background ensemble $x^b(t_0)$. The vectors in the subspace spanned by the 4D-Var process correspond to the directions of maximum error growth that are visited during the first optimization iterations. A 'hybrid ensemble' is generated by projecting the background ensemble perturbations out of the space spanned by the first several 4D-Var increments. The hybrid covariance matrix is then assembled from these hybrid ensemble members. This can be done both at t_0, and at t_F, with no additional computational cost, since the 4D-Var iterates are available at both endpoints of the assimilation window.

In the following we give a detailed description of our algorithm.

3.1. The background ensemble

(1) At $t = t_0$.

Suppose we are given the background state $x^b(t_0) \in \mathbb{R}^n$, the background covariance matrix $B_0 \in \mathbb{R}^{n\times n}$, a set of observations $y_j \in \mathbb{R}^m$ taken at times t_j, for $j = 1, \ldots, N_{\mathrm{obs}}$, and the observation covariances $R_j \in \mathbb{R}^{m\times m}$.

(i) Generate a set of N_e normally distributed perturbation vectors with zero mean, and covariance B_{t_0}:

$$\Delta x_i^b \in \mathcal{N}(0, B_{t_0}), \quad i = 1, \ldots N_e. \tag{3}$$

(ii) Construct a background ensemble of size N_e centered around the mean $x^b(t_0)$:

$$x_i^b(t_0) = x^b(t_0) + \Delta x_i^b, \quad i = 1, \ldots, N_e. \tag{4}$$

(2) At $t = t_F$.

(i) The background ensemble members at t_0 are propagated through the model up to t_F. We obtain

$$\left\{x_1^b(t_F),\ x_2^b(t_F),\ \ldots,\ x_{N_e}^b(t_F)\right\}, \tag{5}$$

where

$$x_i^b(t_F) = \mathcal{M}_{t_0 \to t_F}\left(x_i^b(t_0)\right), \quad i = 1, \ldots, N_e. \tag{6}$$

(ii) Compute the mean $x^b(t_F)$, and the background ensemble perturbations

$$\Delta x_i^b(t_F) = x_i^b(t_F) - x^b(t_F) \tag{7}$$

to prepare for the projection.

3.2. The hybrid ensemble

(1) At $t = t_0$.
Starting from $x^b(t_0)$, the numerical optimization routine used by 4D-Var generates a sequence of iterates $\{x^{(j)}\}_{j=1\ldots k}$. Here k is a truncation index, in that we retain only the first k iterates, with $k \ll n$. This strategy is consistent with the observation that the reduction in the cost function is fastest during the initial iterations of the optimizer. This phenomenon has been consistently observed in practice (Navon et al., 1992; Li et al., 1993; Zou et al., 1993; Sandu and Zhang, 2008). Note that all of the iterates $\{x^{(j)}\}_{j=1\ldots k}$ are calculated by the 4D-Var process at both endpoints of the assimilation window. Let $x_0^{(j)}$ be the jth 4D-Var iterate at t_0 (for all j), and

$$x_F^{(j)} = \mathcal{M}_{t_0 \to t_F}\left(x_0^{(j)}\right). \tag{8}$$

(2) At $t = t_F$.

(i) Let $x_F^{(j)}$ be the jth iterate propagated to t_F, and denote by $\mathcal{S}_{t_F}$ the matrix whose columns are the normalized 4D-Var increments

$$\mathcal{S}_{t_F} = \left\{ \frac{x_F^{(j)} - x_F^{(j-1)}}{\left\| x_F^{(j)} - x_F^{(j-1)} \right\|} \right\}_{j=1,\ldots,k}, \tag{9}$$

with $x_F^0 \equiv x^b(t_F)$.

(ii) Using the singular value decomposition (SVD) of $\mathcal{S}_{t_F}$:

$$\mathcal{S}_{t_F} = U \Sigma V^T, \tag{10}$$

we prune the error direction space by retaining only the first l right singular vectors $u_1, \ldots u_l$ that correspond to the largest l singular values $\sigma_1, \ldots, \sigma_l$, where

$$\frac{\sigma_j}{\sigma_1} < \tau < 1, \quad l < j \le k. \tag{11}$$

Here τ is a problem-dependent, user-controlled threshold. The SVD is used for two reasons. First, one should not remove too many of the 4D-Var directions, otherwise the subspace on which we define our hybrid error statistics will be too small to lead to a good posterior covariance estimate. Second, the SVD is a numerically robust orthogonalization procedure.

Let

$$U_{t_F} = \left\{u_F^1\ u_F^2\ \ldots\ u_F^l\right\}. \tag{12}$$

(iii) The orthogonal projector onto the orthogonal complement of Range(U_{t_F}) is defined as

$$\mathcal{P}_{t_F} = I - U_{t_F} U_{t_F}^T. \tag{13}$$

Using $\mathcal{P}_{t_F}$, project the ensemble perturbations Δx_i^b (7) out of the range of U_{t_F}, to obtain the projected ensemble perturbations Δx_i^p:

$$\Delta x_i^p(t_F) = \mathcal{P}_{t_F} \Delta x_i^b(t_F). \tag{14}$$

These projected perturbations do not contain any components in the space Range(U_{t_F}), by the definition of $\mathcal{P}_{t_F}$.

3.3. Hybrid enhancement

As discussed previously, the 4D-Var data assimilation process implicitly builds a subspace S_{t_F} where the error reduction is greatest. Since in practice the number of optimization iterations is much smaller than the number of state variables, the dimension of this optimization subspace is small compared to the dimension of the entire state space. As constructed, the projected background ensemble perturbations Δx_i^p (14) has no error components in the optimization subspace. However, the resulting analysis error is not orthogonal to this subspace, but may contain small posterior error components that lie in the optimization subspace. While these errors are small at t_0, they might grow considerably during subsequent system evolution.

Therefore, we seek to approximate the analysis error components in the optimization subspace, and to account for them in the hybrid ensemble. We refer to this procedure as an enhancement of the hybrid ensemble. The basis of the proposed enhancement is the observation that the L-BFGS algorithm implicitly produces an approximation of the inverse Hessian using only information from the optimization subspace. This inverse Hessian approximates the posterior error covariance reduced to the optimization subspace (Gejadze et al., 2008).

The detailed enhancement procedure is now discussed.

(i) At $t = t_0$:
The analysis error covariance can be approximated by the inverse of the Hessian matrix H^{-1} (Thacker, 1989; Gejadze et al., 2008). The inverse Hessian approximation is constructed from the optimization iterates and corresponding gradients [for further details, the reader is kindly directed to Algorithms 7.4 and 7.5 in (Nocedal and Wright, 2006), and references therein]. The gradients can be computed through finite differences, or through the application of the adjoint of (1).

We perform an eigendecomposition of the approximate inverse Hessian to obtain its leading d ($d \ll n$) eigenpairs (λ_j, w_j), $j = 1, \ldots, d$. For large-scale models, these eigenpairs can be calculated with a matrix-free eigenvalue solver such as (P)ARPACK

(Lehoucq et al., 1998) or Anasazi (Sala et al., 2004). The following approximation is valid by the Eckart-Young theorem (Golub and Van Loan, 1996)

$$H^{-1} \approx \sum_{j=1}^{d} \lambda_j w_j w_j^T. \tag{15}$$

We sample the posterior error components in the optimization subspace using perturbation terms obtained from the Hessian inverse, and referred to as 'Hessian-based perturbations'. Based on the Karhunen-Loève decomposition (Loève, 1955), and making use of (15), the Hessian-based perturbation $\Delta x_i^{\mathrm{Hess}}$ at $t = t_0$ can be represented as:

$$\Delta x_i^{\mathrm{Hess}}(t_0) = \sum_{j=1}^{d} \xi_j^i \sqrt{\lambda_j} w_j, \quad \text{where } \xi_j^i \in \mathcal{N}(0, 1). \tag{16}$$

(ii) At $t = t_F$:

The Hessian-based perturbations are added to the analysis and propagated from t_0 to t_F using the non-linear model

$$\begin{aligned} x_i^{\mathrm{Hess}}(t_0) &= x^a(t_0) + \Delta x_i^{\mathrm{Hess}}(t_0) \\ x_i^{\mathrm{Hess}}(t_F) &= \mathcal{M}_{t_0 \to t_F}\, x_i^{\mathrm{Hess}}(t_0) \\ \Delta x_i^{\mathrm{Hess}}(t_F) &= x_i^{\mathrm{Hess}}(t_F) - x^a(t_F). \end{aligned} \tag{17}$$

3.4. Covariance update

The hybrid ensemble perturbations are obtained by adding the projection ensemble perturbations (14) and the Hessian-based perturbations (17)

$$\Delta x_i^h(t_F) = \Delta x_i^p(t_F) + \Delta x_i^{\mathrm{Hess}}(t_F). \tag{18}$$

To summarize, the background error components in the optimization subspace are removed from the ensemble by the orthogonal projection. However, we can expect that the most important error components remaining after the 4D-Var process will be re-introduced in the hybrid ensemble through the Hessian enhancement.

The hybrid ensemble covariance matrix is obtained from the hybrid ensemble

$$B_{t_F}^{\mathrm{ens}} = \frac{1}{\sqrt{N_e - 1}} \sum_{i=1}^{N_e} \left(\Delta x_i^h\right) \cdot \left(\Delta x_i^h\right)^T. \tag{19}$$

Due to the small size of the ensemble this covariance matrix can suffer from spurious long-distance correlations. This issue can be addressed by applying a localization procedure, as is typically done in ensemble based data assimilation.

The analysis covariance matrix can then be updated through a convex combination of the static background covariance B_0 and the hybrid ensemble covariance $B_{t_F}^{\mathrm{ens}}$ as

$$B_{t_F}^h = \alpha \cdot B_0 + (1 - \alpha) \cdot B_{t_F}^{\mathrm{ens}}. \tag{20}$$

The blending factor $0 < \alpha \leq 1$ ensures that the analysis covariance matrix is non-singular, and therefore it can be used as background in the next 4D-Var assimilation window.

3.5. The computational cost of the hybrid DA method

The most expensive steps of the hybrid DA method are:

(i) the forward propagation of the background ensemble members (4) through the non-linear model, from t_0 to t_F,

(ii) the computation of 4D-Var iterates $x_0^{(j)}$ starting from $x^b(t_0)$ and the propagation to $x_F^{(j)}$ and

(iii) the generation of additional Hessian-based perturbations, and their propagation through the system.

Note that the first two phases are independent. They can be performed in parallel on a multiprocessor computer, thus substantially reducing the run time of the hybrid method. Once the final time background ensemble (5), and the optimization iterates at t_F have been computed, the hybrid projection method can be readily applied. The only additional cost is incurred when computing the SVD of $\mathcal{S}_{t_F}$.

The quasi-Newton approximation of the inverse Hessian H^{-1} is computed at almost no extra cost from the sequence of gradients used by L-BFGS. Additional computational costs for enhancing the hybrid ensemble are due to the eigendecomposition of H^{-1}, the Karhunen-Loève expansion, and the propagation of the new perturbations (17). In practice one can use only a small number of Hessian-based perturbations, in which case only a subset of the projected perturbations are enhanced via (18).

4. Numerical experiments

4.1. Experiment setting

We test the hybrid data assimilation strategy with the Lorenz 96 model, which is described by the following set of equations

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = -x_{j-1}(x_{j-2} - x_{j+1} - x_j) + F, \quad j = 1, \ldots 40, \tag{21}$$

with periodic boundary conditions. The forcing term is $F = 8.0$. The initial state before the model spin-up is chosen to be

$$x_i(t_{-10}) = 1 + 0.1 \,\mathrm{mod}(i, 5), \quad i = 1, \ldots, 40. \tag{22}$$

This state is integrated forward to t_0 to provide the reference trajectory $x^t(t_{-k})$, $k = 0, 1, \ldots$.

A static background covariance matrix B_{t_0} is constructed as follows. First define the covariance matrix $B_{t_{-3}}$ at t_{-3} using

$$(B_{t_{-3}})_{ij} = \sigma_i \cdot \sigma_j \cdot \exp\left(-\frac{d^2}{L^2}\right), \quad i,\ j = 1, \ldots, 40, \tag{23}$$

where the standard deviation for the state variable i is $\sigma_i = 0.03\, x_i^t(t_{-3})$, and the correlation distance is set to $L = 5$. We account for the periodic boundary conditions in that $d = \min\{|i - j|, n - |i - j|\}$ ($n = 40$). A background ensemble with 100 members is generated with mean $x^t(t_{-3})$ and covariance $B_{t_{-3}}$. This background ensemble is propagated forward in time to t_0. The ensemble covariance at t_0 is computed, and is localized using a correlation distance of $L = 5$ (Houtekamer and Mitchell 1998; Hamill and Snyder, 2000). Next, to prevent the covariance from having a large condition number, due the model dynamics collapsing on a smaller-dimensional manifold, we regularize the covariance eigenvalues, while keeping its eigenvectors the same. The regularization changes the smaller eigenvalues of the covariance matrix (such that the smallest eigenvalue becomes 10^{-4}), while the largest eigenvalue remains unchanged. This procedure yields a non-singular B_{t_0}, as required by the 4D-Var algorithm.

The background initial state is obtained from the reference solution, plus a random perturbation consistent with the background error statistics

$$x^b(t_0) = x^t(t_0) + B_{t_0}^{1/2}\,\xi, \quad \xi \in \mathcal{N}(0,1)^n. \tag{24}$$

The observation covariance is a diagonal matrix with $R_{ii} = 0.4$ and assumed to be constant in time for all the observations. The linear observation operator $\mathcal{H}$ captures only a subset of 30 model states, which includes every other state from the first 20 states plus all of the last 20 states. The length of one assimilation window is set to be one time unit. The observation data is generated from the reference ('true') solution with added Gaussian noise (whose standard deviation is equal to 3% of the solution value). There are $N_{\text{obs}} = 20$ (uniformly spaced) observation times inside each window. Our implementation of 4D-Var makes use of the L-BFGS routine of Zhu et al. (1997), the de facto 'gold standard' of gradient-based optimizers used in data assimilation studies. During the 4D-Var optimization the gradients and the solution increments are saved. The Hessian inverse information is obtained using a six-term L-BFGS approximation (Nocedal and Wright, 2006).

The simulation time spans seven consecutive time windows. In each window a 4D-Var data assimilation is carried out; the background value is the analysis solution at the end of the previous window, and the background covariance matrix is the hybrid covariance at the end of the previous window. A background ensemble is generated, consistent with this background mean and error covariance matrix.

In the projection step we perform the SVD of the directions that span the optimization space. We set $\tau = 0.1$ in (11), to retain the most important optimization directions. A low rank approximation of the Hessian inverse is obtained from the last six gradient and solution increments in the optimization, via the L-BFGS approximation formula (Nocedal and Wright, 2006; Gejadze et al., 2008). The hybrid ensemble covariance matrices are localized using a Gaussian de-correlation function with a correlation distance $L = 5$, see (23). A blending factor $\alpha = 0.2$ is used in (20) to obtain the hybrid covariance matrices.

In each window we also perform an ensemble of 100 4D-Var runs, each initialized with a different background state (consistent with the background error distribution at the beginning of the window), with a different set of observations (consistent with the observation statistics). The ensemble of 4D-Var solutions is propagated to the end of the window, and is used to compute the reference analysis covariance.

4.2. Analysis errors

We first study the effect of using the hybrid covariance in the assimilation on the root mean square error (RMSE) of the analysis solution. The RMSE for the time window $[t_i,\ t_{i+1}]$ is defined as follows:

$$RMSE_i = \sqrt{\frac{1}{N}\sum_{j=0}^{N}\frac{\|x^a(t_i + j\Delta t) - x^t(t_i + j\Delta t)\|_2^2}{\|x^t(t_i + j\Delta t)\|_2^2}}, \tag{25}$$

where x^t is the 'true' solution and x^a is the analysis solution in the current time window (the solution initialized at t_i with the optimal initial condition obtained by 4D-Var, and propagated throughout the window). The comparisons are made at $N = 100$ equidistant time points inside each window $[t_i,\ t_{i+1} = t_i + 1]$, $i = 0, \ldots, 6$, with $\Delta t = 1/N$.

We run three different 4D-Var data assimilation experiments. The first one uses a fixed background covariance equal to B_{t_0} in all windows (the 'static B' case); this corresponds to a traditional 4D-Var setting. The second experiment uses a hybrid background error covariance obtained using only projection (the 'P' case), and the third experiment uses an enhanced hybrid covariance matrix using both projection and Hessian-based perturbations (the 'P+H' case). In both hybrid covariance tests a localization of ensemble covariances is performed with a correlation length $L = 5$, and the blending factor is $\alpha = 0.2$.

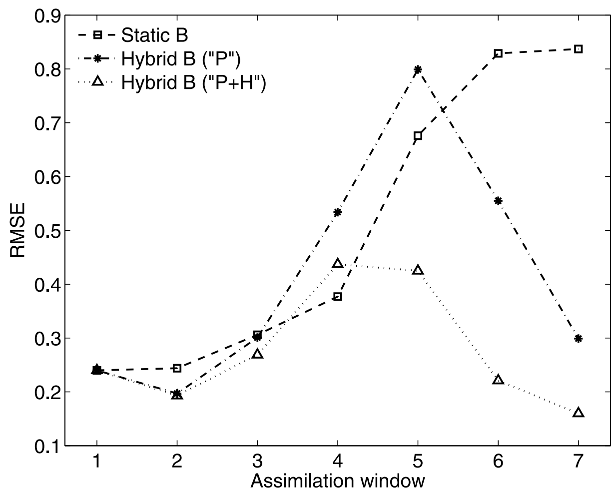
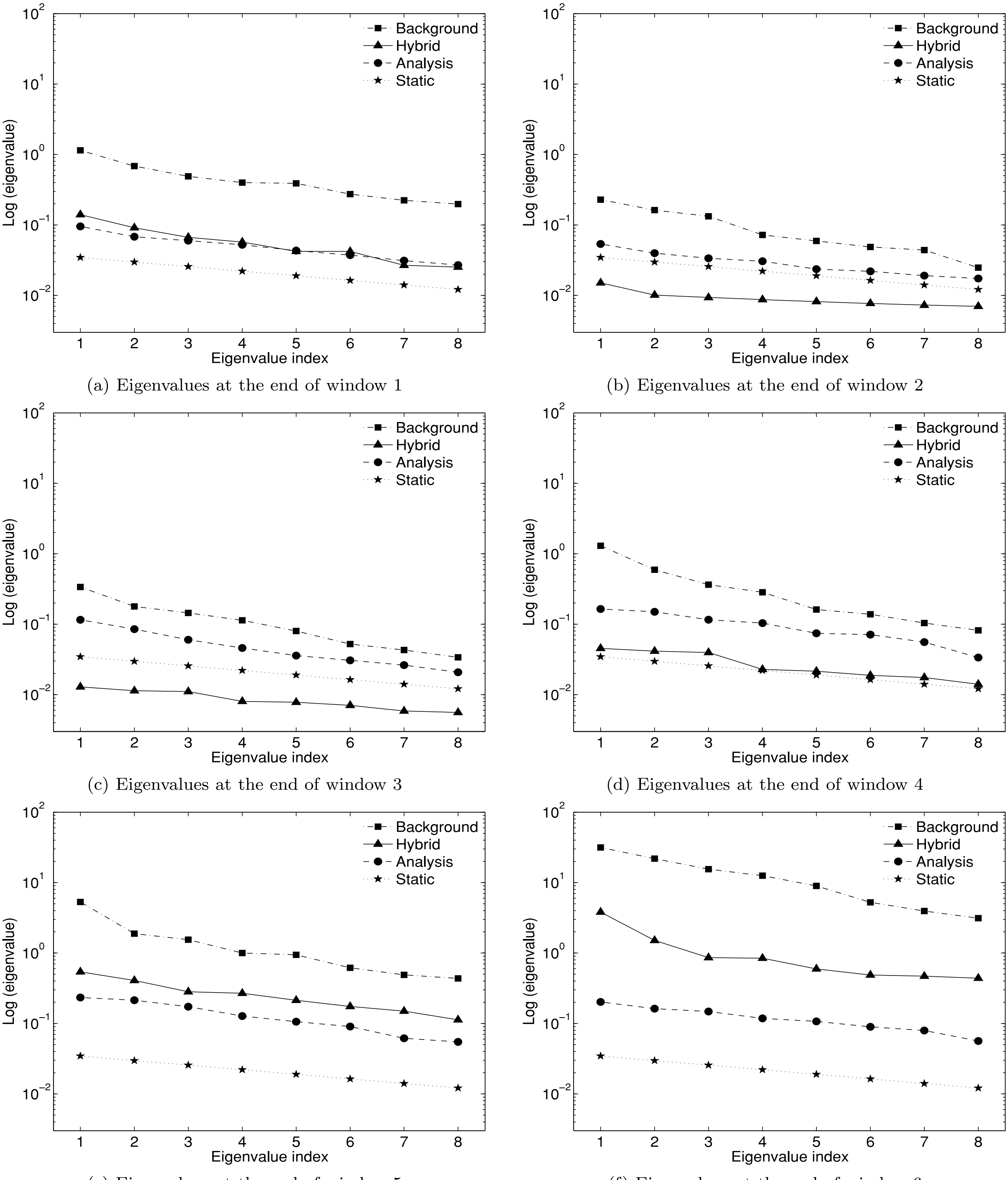


Fig. 1. Root-mean-square error (RMSE) comparison for seven assimilation windows, using different background covariance matrices (static and hybrid covariances with localization length $L = 5$, and blending factor $\alpha = 0.2$; P is projection only, P+H is projection with Hessian enhancement).



(a) Eigenvalues at the end of window 1

(b) Eigenvalues at the end of window 2

(c) Eigenvalues at the end of window 3

(d) Eigenvalues at the end of window 4

(e) Eigenvalues at the end of window 5

(f) Eigenvalues at the end of window 6

Fig. 2. The leading eigenvalues of the background, hybrid, analysis, and static error covariance matrices, for six data assimilation windows. Here $\alpha = 0.2$, and the spatial correlation $L = 5$. The hybrid covariance was generated using both the hybrid ensemble and Hessian information ('P+H').

The solution RMSE is shown in Fig. 1 for each of the seven windows. As can be seen from the data, using the static background covariance, the RMSE error increases from window 1 to window 7. We observe an error oscillation between windows for both the hybrid covariance test scenarios. In general, the solutions obtained with hybrid covariances have smaller errors, which shows that the hybrid covariance improves the static background covariance by incorporating the 'error of the day'. Furthermore, the enhanced covariance using both the projection and the Hessian-based perturbations leads to a smaller error then that of the projection alone. Since the projection directions, as well as the Hessian inverse approximation are dynamically generated at the beginning of each assimilation window, the system dynamics is naturally captured and reflected in the hybrid covariance matrix.

4.3. *Covariance approximations*

We next assess the effect of the proposed hybrid technique on capturing the posterior error covariances. For this we plot eight leading eigenvalues of different covariance matrices at the end of each of the first six assimilation windows. The results are shown in Fig. 2. At the beginning of each window a 100-member background ensemble is initialized with the mean given by the background state and perturbations consistent with a Gaussian background error distribution. this ensemble is propagated to the end of the window, and is used to compute the background covariance whose eigenvalues are reported in Fig. 2. The hybrid covariance is computed from the enhanced hybrid ensemble using both the projection and the Hessian inverse information. The reference analysis covariance is computed from a 100-ensemble of 4D-Var runs, each initialized with a member of the background ensemble, with different perturbations in the observations. The static covariance is the fixed background covariance B_{t_0}. The results in Fig. 2 indicate that the hybrid covariance is a good approximation to the analysis covariance. For later windows (5 and 6), the fixed background covariance eigenvalues slightly underestimates the reference analysis covariance eigenvalues, while the hybrid covariance eigenvalues tends to overestimate them to some extent.

5. Conclusions

Recently, considerable research efforts have been put into the development of variational data assimilation procedures that update the background covariance matrices such that the specific flow dynamics (hence, the 'errors of the day') are taken into account. This paper proposes a hybrid data assimilation method that yields an estimate of the a posteriori error covariance matrix. This hybrid covariance is built by removing the most significant error directions explored by 4D-Var (during the first few optimization iterations) from a set of background ensemble perturbations, and adding back estimates of the posterior error in the same subspace. The projection operator is easy to construct at the end of an assimilation window because the basis vectors are provided 'for free' by the 4D-Var process. Estimates of the posterior error are obtained using a low rank approximation of the Hessian inverse built implicitly by the L-BFGS algorithm. The projected background perturbations plus the posterior (Hessian-based) perturbations lead to a hybrid ensemble, which allows to construct an estimate of the posterior error covariance. Both the 4D-Var and the background ensemble runs can be performed in parallel on a multiprocessor machine. This significantly lowers the computational cost of the hybrid method, making it feasible for large-scale models. The additional cost of the method comes from the background ensemble projection at the window end points (for the hybrid projection strategy), and from the propagation of the Hessian-based ensemble (also embarassingly parallel), in case of the Hessian-enhanced hybrid strategy.

We demonstrate the hybrid data assimilation approach on the Lorenz-96 system. The use of hybrid covariances results in an analysis trajectory whose errors are smaller than that of the analysis obtained when reusing a constant background covariance. The hybrid projection method seeks to correctly account for the 'errors of the day', and improves the quality of the model forecast. This work lays the foundation for future research into hybridizing the 4D-Var and EnKF methods.

6. Acknowledgments

This research has been partially supported by NSF grants CCF-0515170, CCF-0635194, DMS-0915047, and by the NASA ROSES-2005 AIST project led by Dr. Meemong Lee.

References

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

Arulampalam, M. S., Maskell, S., Gordon, N. J. and Clapp, T. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **150**(2), 174–188.

Berliner, M. and Wikle, C. 2007. A Bayesian tutorial for data assimilation. *Physica D* **30**, 1–16.

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.

Cessna, J., Colburn, C. and Bewley, T. 2008. EnVE: a new estimation algorithm for weather forecasting and flow control. *Proceedings from 4th AIAA Flow Control Conference*, Seattle, WA, USA.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **75**, 257–288.

Cohn, S. E., da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. 1998. Assessing the effects of data selection with the DAO physical space statistical analysis system. *Mon. Wea. Rev.* **126**, 2913–2926.

Cohn, S. E., Sivakumaran, N. S. and Todling, R. 1994. A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.* **122**, 2838–2867.

Courtier, P., Anderson, E., Heckley, W., Pailleux, J., Vasiljevic, D. and co-authors. 1998. The ECMWF implementation of three dimensional variational assimilation (3D-Var) I: formulation. *Quart. J. Roy. Meteorol. Soc.* **124**, 1783–1807.

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint equations. Part 2: numerical results. *Quart. J. Roy. Meteorol. Soc.* **113**, 1329–1347.

Daescu, D. and Navon, I. M. 2008. Reduced-order observation sensitivity in 4d-var data assimilation, *Reduced-order Observation Sensitivity in 4D-Var Data Assimilation*, 12th Conference on IOAS-AOLS.

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

Doucet, A., de Freitas, N. and Gordon, N. 2001. *Sequential Monte Carlo Methods in Practice* 1st Edition, Springer, New York, NY, USA.

Doucet, A., Godsill, S. and Andrieu, C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**(30), 197–208.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5 10), 108–129.

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Evensen, G. 2007. *Data Assimilation: The Ensemble Kalman Filter.* Springer, Berlin.

Evensen, G. and van Leeuwen, P. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.* **128**, 1852–1867.

Gejadze, I. Y., Le Dimet, F. X. and Shutyaev, V. 2008. On analysis error covariances in variational data assimilation. *SIAM J. Scient. Comput.* **30**(4), 1847–1874, http://dx.doi.org/10.1137/07068744X.

Ghil, M., Ide, K., Bennett, A. F., Courtier, P., Kimoto, M. and co-authors. 1997. Data assimilation in meteorology and oceanography: theory and practice. *J. Meteorol. Soc. Jpn.* **75**, 257–288.

Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations* 3rd Edition. Johns Hopkins University Press, Baltimore, MD, USA.

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

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. *Am. Meteorol. Soc.* **129**(1), 123–137.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory.* Academic Press, New York.

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

Lehoucq, R. B., Sorensen, D. C. and Yang, C. 1998. *ARPACK User's Guide: Solution of Large-Scale Eigenvalue Problems With Implicityly Restorted Arnoldi Methods*, SIAM, Philadelphia, PA, USA.

Le Dimet, F. X. 1980. Une étude générale d'analyse objective variationnelle des champs météorologiques, Technical Report LAMP 28, Universite de Clermond II, BP 45, 63170, Aubiére, France.

Le Dimet, F. X., Navon, I. M. and Daescu, D. 2002. Second order information in data assimilation. *Mon. Wea. Rev.* **130**(3), 629–648.

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations. *Tellus* **38A**, 97–110.

Li, Y., Navon, I. M., Courtier, P. and Gauthier, P. 1993. Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equation model and its adjoint. *Mon. Wea. Rev.* **121**(6), 1759–1769.

Li, Z. and Navon, I. M. 2001. Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother. *Quart. J. Roy. Meteorol. Soc.* **127**(572), 661–684.

Lions, J. L. 1971. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer Verlag, New York, NY, USA.

Loève, M. M. 1955. *Probability Theory*. Van Nostrand, Princeton, NJ.

Lorenc, A. C. 2003. The potential of the Ensemble Kalman Filter for NWP-A comparison with the 4D-VAR. *Quart. J. Roy. Meteorol. Soc.* **129**, 3183–3203.

Lorenz, E. N. 1996. Predictability: a problem partly solved. In: *Seminar on Predictability, ECMWF*, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading, UK.

Malanotte-Rizzoli, P. 1996. *Modern Approaches to Data Assimilation in Ocean Modeling Malanotte-Rizzoli*, P. Elsevier, Amsterdam.

Navon, I. M., Zou, X., Derber, J. and Sela, J. 1992. Variational data assimilation with an adiabatic version of the NMC Spectral Model. *Mon. Wea. Rev.* **120**(7), 1433–1446.

Nocedal, J. and Wright, S. J. 2006. *Numerical Optimization* 2nd Edition. Springer, New York, NY, USA.

Parrish, D. F. and Derber, J. C. 1992. The national meteorological center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Qiu, C., Shao, A., Xu, Q. and Wei, L. 2007. Fitting model fields to observations by using singular value decomposition: an ensemble-based 4dvar approach. *J. Geophys. Res.* **112**(D11105).

Qiu, C., Zhang, L. and Shao, A. 2007. An explicit four-dimensional variational data assimilation method. *Sci. Chin. Ser. D: Earth Sci.* **50**(8), 1232–1240.

Sala, M., Heroux, M. A. and Day, D. M. 2004. Trilinos Tutorial. Technical Report SAND2004-2189, Sandia National Laboratories, Albuquerque, NM, USA.

Sandu, A. and Zhang, L. 2008. Discrete second order adjoints in atmospheric chemical transport modeling. *J. Comput. Phys.* **227**(12), 5949–5983.

Sasaki, Y. K. 1958. An objective analysis based on the variational method. *J. Meteorol. Soc. Jpn.* **II**(36), 77–88.

Talagrand, O. and Courtier, P. 1987. Variational assimilation of meteorological observations with the adjoint equations Part I: theory. *Quart. J. Roy. Meteorol. Soc.* **113**, 1311–1328.

Thacker, W. C. 1989. The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.* **94**(C5), 6177–6196.

Thépaut, J. N. and Courtier, P. 1991. Four-dimensional data assimilation using the adjoint of multi-level primitive equation model. *Quart. J.Roy. Meteorol. Soc.* **117**, 1225–1254.

Tian, X., Xie, Z. and Dai, A. 2008. An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.* **113**(D21124), doi:10.1029/2008JD010358.

van Leeuwen, P. J. 2009. Particle filtering in geophysical systems. *Mon. Wea. Rev.* **137**, 4089–4114.

Wang, Z., Navon, I. M., Le Dimet, F. X. and Zou, X. 1992. The second order adjoint analysis: theory and applications. *Meteorol. Atmos. Phys.* **50**(1-3), 3–20.

Zhang, F., Zhang, M. and Hansen, J. A. 2009. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.* **26**(1), 1–8.

Zhu, C., Byrd, R. H., Lu, P. and Nocedal, J. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Software* **23**(4), 550–560.

Zou, X., Navon, I. M. and Sela, J. 1993. Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus* **45A**, 370–387.