INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2012; **70**:1207–1220
Published online 7 Dec 2011 in Wiley Online Library (wileyonlinelibrary.com/journal/nmf). DOI: 10.1002/fld.2736

A penalized four-dimensional variational data assimilation method for reducing forecast error related to adaptive observations

M. J. Hossen[1], I. M. Navon[2,*,†] and F. Fang[3]

[1]*Research School of Earth Sciences, Australian National University, Bldg 61 Mills Road Acton ACT, 0200 Australia*
[2]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA*
[3]*The Applied Modelling and Computation Group (AMCG), Department of Earth Science and Engineering, South Kensington Campus, Imperial College London, SW7 2AZ, UK*

SUMMARY

Four-dimensional variational (4D-Var) data assimilation method is used to find the optimal initial conditions by minimizing a cost function in which background information and observations are provided as the input of the cost function. The optimized initial conditions based on background error covariance matrix and observations improve the forecast. The targeted observations determined by using methods such as adjoint sensitivity, observation sensitivity, or singular vectors may further improve the forecast. In this paper, we are proposing a new technique—consisting of a penalized 4D-Var data assimilation method that is able to reduce the forecast error significantly. This technique consists in penalizing the cost function by a forecast aspect defined over the verification domain at the verification time. The results obtained using the penalized 4D-Var method show that the initial condition is optimally estimated, thus resulting in a better forecast by significantly reducing the forecast error over the verification domain at verification time. Copyright © 2011 John Wiley & Sons, Ltd.

Received 9 October 2011; Accepted 16 November 2011

KEY WORDS: variational method; shallow water; optimization; finite volume; inverse; forecast error

1. INTRODUCTION

Numerical weather prediction is based on the integration of a dynamic system of partial differential equations modeling the behavior of the atmosphere. Therefore, discrete initial conditions describing the state of the atmosphere have to be provided prior to the integration because they, along with the model equations and boundary conditions, control the evolution of the solution trajectory in space and time. To find the best estimate for the initial condition, we use four-dimensional variational (4D-Var) data assimilation (DA) techniques [1–4]. In this method, the initial condition is optimized by minimizing the cost function defined as the combination of deviations of the desired analysis from a forecast and observations weighted by the inverse of the corresponding forecast and observation error covariance matrices.

Four-dimensional variational DA method uses a flow-dependent background error covariance for estimating the atmospheric state and assimilates indirect observational data (such as satellite radiance without transforming them) into analysis variables. The computational expense of the variational assimilation can be reduced by using the adjoint of the numerical model to calculate

*Correspondence to: I. M. Navon, Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA.
†E-mail: inavon@fsu.edu

all of the components of the gradient of the cost function with respect to the initial conditions in one integration of the forward model followed by integration of the corresponding adjoint model. The adjoint model arises from the theory of optimization and optimal control of partial differential equations [5, 6]. Its theoretical aspects were presented by LeDimet and Talagrand [1], Talagrand and Courtier [4], and LeDimet *et al.* [7].

Results from 4D-Var experiments with a large scale numerical model were published in the early 1990s [8–10]. Thepaut *et al.* [11] demonstrated the ability of 4D-Var method to generate flow-dependent and baroclinic structure functions in meteorological analysis.

The forecast impact of targeting is determined by the distribution and types of routine and targeted observations, the quality of the background or the first guess, and the ability of the DA procedure to combine information from the background and observations. To deploy targeted observations, we need to define a target area. Typically, an objective procedure (often based on adjoint or ensemble techniques) is used a day or more in advance to identify a target region for the spatial observations.

It can also be determined on the basis of high probability for a large or a fast-growing initial condition error based on searches for the directions where errors in the state vector at the targeting time will propagate most at the verification time on the verification domain.

The goal of the adaptive observations is to add targeted observations inside the sensitive regions to improve the initial conditions so that the forecast error can be reduced significantly. Adjoint-based observation sensitivity techniques could be used to identify the adaptive observation space and time locations that are valuable for the assimilation procedure to conduct optimal data thinning and to design the cost-effective field experiments for collecting adaptive observations. Langland [12] showed that a small number of additional observational resources can be deployed to improve a specific forecast aspect. The design of cost-effective observation targeting strategies relies on the ability to a priori identification of optimal sites for collecting data of large impact on reducing forecast errors. LeDimet *et al.* [7] presented the theoretical formulation of the sensitivity analysis in variational DA in the context of optimal control. Daescu and Navon [13] proposed a new adjoint sensitivity approach where the interaction between adaptive observations and routine observations was studied. The singular vector (SV) approach provides a possibility of searching for directions in phase space where the errors in the initial condition will increase rapidly. The specification of the initial and final norms plays a crucial role. In the European Center for Medium-Range Weather Forecasts operational Ensemble Prediction System, SVs are computed with the so-called total energy norm at initial and final times. It can be shown that among simple norms, the total energy norm provides SVs that agree best with analysis error statistics [14]. Barkmeijer *et al.* [15, 16] have shown that the Hessian of the cost function in a variational DA scheme can be used to compute SVs that incorporate an estimate of the full analysis error covariance at initial time and total energy norm at final time. This type of SV is called Hessian SV. Ehrendorfer and Tribbia [17] stated that such an approach to determine SVs provides an efficient way to describe the forecast error covariance matrix when only a limited number of linear integrations are possible. Although finding the Hessian matrix explicitly involves a computationally intensive effort, we can calculate the Hessian vector product by using the second-order adjoint [18]. This also requires an efficient generalized eigenvalue problem solver to compute SVs. For advanced work on this topic, see Hodinez and Daescu [19].

In this paper, we are proposing a new cost function, which can be minimized to find an optimal estimate of the initial condition. This initial condition reduces the forecast error significantly over the verification domain at the verification time. The new cost function is obtained by penalizing the cost function with a term defined as being proportional to the square of the distance between analysis and both background and observation, with the forecast aspect defined over verification domain at verification time.

The structure of the paper is as follows. In Section 2, we present derivation of the equations and the algorithm of penalized 4D-Var, whereas in Section 3, we describe the twin numerical experiments implementing the algorithm. Numerical results are presented and discussed in Section 4 along with the pseudo-algorithm of the penalized 4D-Var approach. Section 5 is dedicated to the summary and conclusions.

2. DERIVATION OF THE EQUATIONS AND PENALIZED FOUR-DIMENSIONAL VARIATIONAL ALGORITHM

In 4D-Var DA, an initial condition is sought such that the forecast best fits the observations within an assimilation window $[t_0, t_f]$. 4D-Var DA provides an optimal estimate $x_0^a \in \mathbb{R}^n$ to the initial condition of a nonlinear forecast model by minimizing the cost function defined as

$$\begin{aligned}\mathcal{J}(x_0) &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\ &\quad + \frac{1}{2}\sum_{i=0}^{N}(y_i - H_i x_i)^T \mathbf{R}_i^{-1}(y_i - H_i x_i) \\ x_0^a &= \arg\min \mathcal{J},\end{aligned} \tag{1}$$

where $x_0 = x(t_0)$ denotes the initial state at the initial time t_0, x_b is a prior (background) estimate to the initial state, $y_i \in \mathbb{R}^{k_i}$, $i = 0, 1, 2, \ldots, N$ is the set of observations available at time $t_i \in [t_0, t_f]$ and $x_i = \mathcal{M}_{0,i}(x_0)$ is the nonlinear model forecast state at time t_i, and $H_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ is the observation operator that maps the state space into the observation space at time t_i. $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}_i$ is the observational error covariance matrix at time t_i. We assume that background and observation errors are uncorrelated with each other. In our case, we take the error covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$ to be diagonal. The control variable or the variable to be optimized is the initial state of the model x_0. The model $\mathcal{M}$ is assumed to be perfect by imposing the model equations as the strong constraint.

To minimize the cost function in Equation (1) with respect to x_0, we need to calculate the gradient of the cost function with respect to the control variable, that is, $\nabla_{x_0}\mathcal{J}$. The adjoint method provides an efficient approach to calculate the gradient of the cost function with respect to control variables. The gradient of the cost functional (1) is

$$\begin{aligned}\nabla_{x_0}\mathcal{J} &= \mathbf{B}^{-1}(x - x_0) - \sum_{i=0}^{N}\mathbf{M}_{0,i}^T\mathbf{R}_i^{-1}(y_i - H_i x_i) \\ &= \mathbf{B}^{-1}(x - x_0) - \mathbf{R}_0^{-1}(y_0 - H_0 x_0) \\ &\quad - \mathbf{M}_{0,t_a}^T\mathbf{R}_{t_a}^{-1}\left(y_{t_a} - H_{t_a}x_{t_a}\right) \\ &\quad - \mathbf{M}_{0,t_r}^T\mathbf{R}_{t_r}^{-1}\left(y_{t_r} - H_{t_r}x_{t_r}\right),\end{aligned} \tag{2}$$

where the linear operator $\mathbf{M}_{0,i} = \mathbf{M}(t_0, t_i)$ is called tangent linear model and its transpose $\mathbf{M}_{0,i}^T$ is the adjoint model. Routine observations are available at $t = t_r$, t_r being the time for routine observations, whereas adaptive observations are available at $t = t_a$, t_a for adaptive observations.

The background error covariance is estimated by using the well-known National Meteorological Center (NMC) method [20]. In this process, the background errors are assumed to be well approximated by an averaged forecast difference (e.g., month-long series of 24-h to 12-h forecasts valid at the same time) statistics:

$$\begin{aligned}B &= \overline{\epsilon_b^T \epsilon_b} = \overline{\left(x_b - x_0^t\right)^T \left(x_b - x_0^t\right)} \\ &\approx \overline{(x_{t_0+24} - x_{t_0+12})^T \left(x_{t_0+24} - x_{t_0+12}\right)},\end{aligned} \tag{3}$$

where x_0^t is the true atmospheric state and x_b is the background error. The bar denotes an average over time and/or space.

2.1. *The adjoint sensitivity approach*

The first approach to identify the adaptive observations locations is the adjoint sensitivity method. In practice, it is of interest to assess the observation impact on the forecast measure $\mathcal{J}_v$ over the verification domain at the verification time t_v. The verification domain, denoted by $\mathcal{D}_v$, is the domain

DOI: 10.1002/fld

where the forecast error is significant. The functional $\mathcal{J}_v$ is defined as a scalar measure of the forecast error over $\mathcal{D}_v$

$$\mathcal{J}_v = \frac{1}{2}\left(x_v^f - x_v^t\right)^T P^T E P \left(x_v^f - x_v^t\right), \tag{4}$$

where x_v^f is the model forecast at the verification time initialized from x_0^a and x_v^t is the verification state at t_v initialized from x_0^t that serves as a proxy to the true atmospheric state. P is a projection operator on $\mathcal{D}_v$ satisfying $P^*P = P^2 = P$, and E is a diagonal matrix of the total energy norm [13].

For the adaptive observations locations to be selected, the gradient of cost functional $\mathcal{J}_v$ defined in Equation (4) is used. The gradient of the function (4) at t_i is defined as

$$\nabla_{x_i}\mathcal{J}_v = \mathbf{M}_{i,v}^T P^T E P \left(x_v^f - x_v^t\right), \tag{5}$$

where $x_i = x(t_i)$.

2.1.1. Location of adaptive observations by adjoint sensitivity. We use the gradient of the function defined in Equation (5) to evaluate the sensitivity of the forecast error with respect to the model state at each targeting instant t_i. A large sensitivity value indicates that small variations in the model state x_i will have a significant impact on the forecast at the verification time. The adjoint sensitivity field with respect to the total energy metric is defined as

$$F_v(\lambda,\theta) = ||\nabla_{x_i}\mathcal{J}_v||_E, \tag{6}$$

where E is the total energy metric, and a weighted norm is defined as

$$||x||_E = \frac{1}{2}(u^2+v^2) + \frac{h^2}{h_0}, \tag{7}$$

where u and v are the zonal and meridional wind components, respectively, h is the geopotential height of the atmosphere, and h_0 is the mean geopotential height of the reference data at the initial time. The adaptive observations at the target instant t_i are deployed at the first n_i locations (λ,θ), where $F_v(\lambda,\theta)$ attains largest values. Here, θ and λ mean latitude and longitude coordinates of a point, respectively.

2.2. Penalized four-dimensional variational method

The fundamental idea of the penalty method is to replace a constrained optimization problem by a series of unconstrained problems whose solutions ideally converge to the solution of the original constrained problem. The general form of constrained minimization is

$$\min_x \mathcal{J}(x)$$
$$\text{subject to } c(x) \leqslant 0,$$

where x is an n-dimensional vector and $c(x)$ is an m-dimensional vector. The unconstrained problems are formed on the basis of two methods: (i) quadratic penalty method; and (ii) augmented Lagrangian method. The outlines of the methods are given in the following text.

2.2.1. Quadratic penalty method. Instead of solving the constrained optimization problem, we can solve an unconstrained minimization problem by defining the quadratic penalty function,

$$Q(x;r) = \mathcal{J}(x) + \frac{1}{2}r|c(x)|^2$$

for any scalar $r > 0$ known as penalty parameter. We seek the approximate minimizer x_k of the function $Q(x;r)$ as $r_k \to \infty$ when $k \to \infty$.

DOI: 10.1002/fld

In this work, we penalize the cost functional $\mathcal{J}(x)$ defined in Equation (1) by adding a penalty term

$$r\mathcal{J}_v \tag{8}$$

to the cost function to reduce the forecast error over the verification domain at the verification time. $\mathcal{J}_v$ is the forecast aspect defined in Equation (4). In this work, we employ the penalty method in a weak sense that we try to find the minimizer by reducing the forecast aspect $\mathcal{J}_v$ until it reaches a prescribed small value δ instead of attaining a perfect steady state where the forecast error is absolutely zero, that is, $\mathcal{J}_v = 0$. That is, we are looking for the optimal initial condition so that the cost function is minimized subject to the constraint that the forecast error is very small. In this case, the penalty parameter is sufficiently large but does not tend to infinity, which is equivalent to imposing an inequality constraint of the form

$$\mathcal{J}_v \leqslant \delta \tag{9}$$

With this, we choose the inequality constraint

$$c(x) = \sqrt{\mathcal{J}_v} - \epsilon, \tag{10}$$

where $\epsilon = \sqrt{\delta}$.

The modified cost function is

$$Q(x;r) = \mathcal{J}(x) + \frac{1}{2} r |c(x)|^2. \tag{11}$$

That is,

$$\begin{aligned} Q &= \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\ &\quad + \frac{1}{2}\sum_{i=0}^{N}(y_i - H_i x_i)^T \mathbf{R}_i^{-1}(y_i - H_i x_i) \\ &\quad + \frac{r}{2}\left(\sqrt{\mathcal{J}_v} - \epsilon\right)^2 \\ x_0^p &= \arg\min Q, \end{aligned} \tag{12}$$

where $x_v^f = \mathcal{M}_{0,v}(x_0)$. The minimizer of the cost function (11) is obtained by using an unconstrained minimization routine that requires the gradient of the cost function (11). The gradient of the penalized cost function is obtained by using the following formula:

$$\nabla_{x_0} Q = \nabla_{x_0}\mathcal{J} + \frac{r}{2}\left(1 - \frac{\epsilon}{\sqrt{\mathcal{J}_v}}\right)\nabla_{x_0}\mathcal{J}_v. \tag{13}$$

The algorithm of quadratic penalty method (because it uses a sequence of infeasible points and feasibility is obtained only at the optimum) can be summarized as follows:

1. Start with an initial point x_0 and an initial value of parameter $r_0 > 0$. Set $k = 0$.
2. Minimize $Q(x;r)$ with x_k by using an unconstrained minimization method and obtain x_k^*.
3. Test whether x_k^* is a solution of the problem, that is, satisfying the constraints $c(x) \leqslant 0$ within some prescribed accuracy criteria. If this is true, the process is terminated; otherwise, r_k is updated on the basis of the value of $\mathcal{J}_v$ (details in Algorithm 2).
4. Set $k = k + 1$, use as a new starting point $x_k = x_k^*$, and go to step 2.

The method depends for its success on sequentially increasing the penalty parameter r to high values. The approximate minimizer becomes increasingly more accurate as r gets higher.

DOI: 10.1002/fld

2.2.2. Augmented Lagrangian method. We again consider the same constrained minimization problem

$$\min_x \mathcal{J}(x) \tag{14}$$

$$\text{subject to } \; c(x) \leqslant 0, \tag{15}$$

where x is an n-dimensional vector and $c(x)$ is an m-dimensional vector. We can define the augmented Lagrangian as

$$L(x;\lambda,r) = \mathcal{J}(x) + \frac{r}{2}\sum_{i=1}^{m}\left(c_i(x) + \frac{\lambda_i}{r}\right)^2$$

for any scalar $r > 0$ and $\lambda = \lambda_i, i = 1,\ldots,m,\ \lambda \in R^m$. We seek the approximate minimizer x_k of the function $L(x;\lambda,r)$ as $r_k \to \infty$ when $k \to \infty$. In this paper, we minimize the cost functional $\mathcal{J}(x)$ defined in Equation (1) with respect to x subject to $c(x) \leqslant 0$ where $c(x) = \sqrt{\mathcal{J}_v} - \epsilon$ defined as Equation (10).

The augmented Lagrangian function for unconstrained minimization problem is

$$L(x;\lambda,r) = \mathcal{J}(x) + \frac{r}{2}\left(c(x) + \frac{\lambda}{r}\right)^2$$

$$L(x;\lambda,r) = \mathcal{J}(x) + \frac{r}{2}\left(\sqrt{\mathcal{J}_v} - \epsilon + \frac{\lambda}{r}\right)^2, \tag{16}$$

and the gradient of this function is

$$\nabla L(x;\lambda,r) = \nabla\mathcal{J}(x) + \frac{r}{2}\left(1 - \frac{\epsilon - \frac{\lambda}{r}}{\sqrt{\mathcal{J}_v}}\right)\nabla_x \mathcal{J}_v. \tag{17}$$

The algorithm of augmented Lagrangian method proceeds as follows:

1. Start with an initial point x_0; an initial value of parameter $r_0 > 0$ and initial multiplier λ_0 can be selected on the basis of either prior knowledge or start with a zero. Set $k = 0$.
2. Minimize $L(x;\lambda,r)$ with x_k by using an unconstrained minimization method and obtain x_k^*.
3. Test whether x_k^* is a solution of the problem, that is, satisfying the constraints within some prescribed accuracy criteria. If this is true, terminate the process; otherwise, r_k is updated on the basis of the value of $\mathcal{J}_v$ (details in Algorithm 3), whereas λ_k is updated by $\lambda_{k+1} = \lambda_k + r_k * c(x_k)$.
4. Set $k = k + 1$, use as a new starting point $x_k = x_k^*$, and go to step 2.

3. DESCRIPTION OF TWIN NUMERICAL EXPERIMENTS

3.1. Experimental setup

The numerical experiments were performed in the twin experiment framework using a finite volume global two-dimensional shallow water (SW) equations model adapted as in [21] that has been widely used as an essential tool for testing promising numerical methods for solving geophysical science problems. The SW equations, a first prototype of the partial differential equations, describe the horizontal dynamics of the atmosphere. The SW equations in spherical coordinates are written as

DOI: 10.1002/fld

$$\frac{\partial h}{\partial t} + \frac{1}{a\cos\theta}\left[\frac{\partial}{\partial\lambda}(hu) + \frac{\partial}{\partial\theta}(hv\cos\theta)\right] = 0 \tag{18}$$

$$\frac{\partial u}{\partial t} + \frac{1}{a\cos\theta}\left[u\frac{\partial u}{\partial\lambda} + v\cos\theta\frac{\partial u}{\partial\theta}\right] - \left(f + \frac{u}{a}\tan\theta\right)v + \frac{g}{a\cos\theta}\frac{\partial h}{\partial\lambda} = 0 \tag{19}$$

$$\frac{\partial v}{\partial t} + \frac{1}{a\cos\theta}\left[u\frac{\partial v}{\partial\lambda} + v\cos\theta\frac{\partial v}{\partial\theta}\right] + \left(f + \frac{u}{a}\tan\theta\right)u + \frac{g}{a}\frac{\partial h}{\partial\theta} = 0, \tag{20}$$

where $f = 2\Omega\sin\theta$ is the Coriolis parameter, Ω is the angular speed of the rotation of the earth, h is the height of the homogeneous atmosphere, u and v are the zonal and meridional wind components, respectively, θ and λ are the latitudinal and longitudinal directions, respectively, a is the radius of the earth, and g is the gravitational constant.

We consider a spatial discretization on a 72×37 grid ($5° \times 5°$ resolution). As a result of this, the dimension of the discrete state vector $x = (h, u, v)$ is $7992 = 3\times72\times37$. For numerical stability, we choose the integration time step, $\Delta t = 900s$. For our numerical experiment, we consider the 500 mb European Center for Medium-Range Weather Forecasts ERA-40 data valid for March 15, 2002 00 h as a true (reference) atmospheric state x_0^t. The model states at the initial time and after 30-h integration are displayed in Figures 1 and 2, respectively. The background field x_b is obtained from a 6-h integration of SW model initialized at $t_0 - 6$ h with x_0^t. Observational data for the DA procedure is generated from the SW model trajectory initialized with x_0^t and corrupted with the random errors from a normal distribution $N(0, \sigma^2)$. We choose the standard deviation $\sigma_h = 5$ for the height and $\sigma_u = \sigma_v = 0.5$ for the velocities. The background error covariance matrix is calculated by using the NMC method as described previously. We assumed that the background and observation errors are uncorrelated. Therefore, the error covariance matrices are diagonal.

Four-dimensional variational DA is carried out in the assimilation window $[t_0, t_0 + 6\text{ h}]$. The routine observation for our experiment is available at t_0 and $t_0 + 6$ h only on a coarse $10° \times 10°$ mesh grid, and the total number of observation locations are 648. So, the observation operator is thus a $3 \times 648 \times 2664$ matrix with entries of 0 and 1 only. At the verification time $t_v = t_0 + 30$ h, the forecast error is calculated by using reference state, $x_v^t = \mathcal{M}_{0,v}(x_0^t)$, and the forecast from

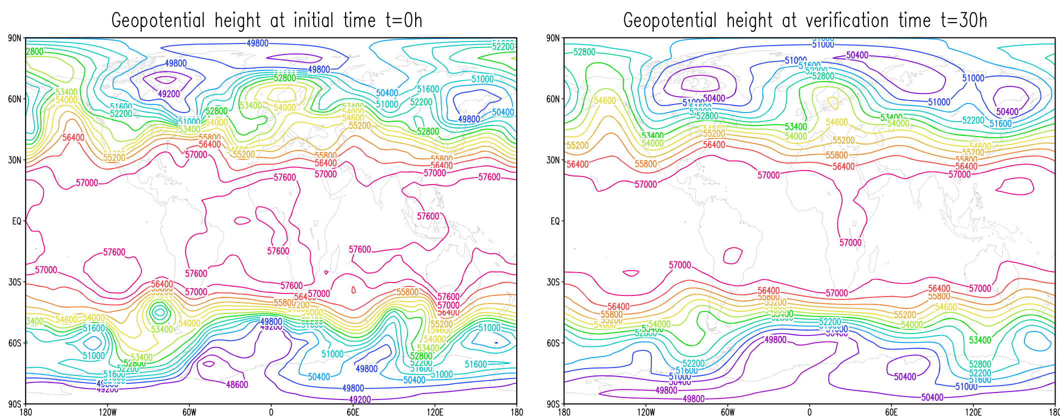

Figure 1. Graph of the geopotential height at the initial time $t = 0$ h (left) and at verification time $t = 30$ h (right).

DOI: 10.1002/fld

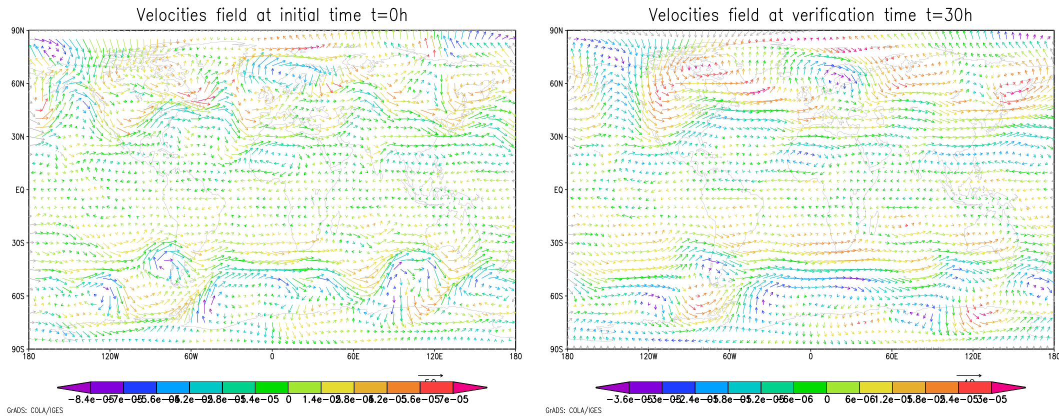

Figure 2. Graph of the velocities at the initial time $t = 0$ h (left) and at verification time $t = 30$ h (right).

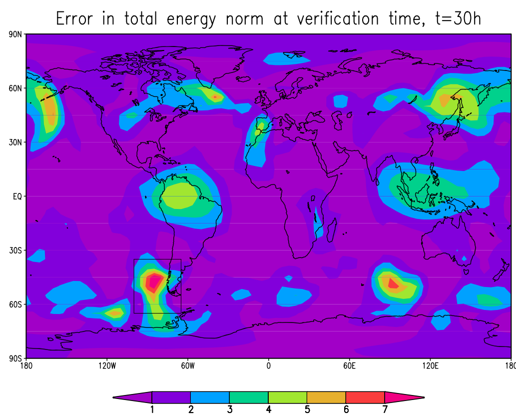

Figure 3. Graph of the forecast error at the verification time with the background estimate.

background $x_v^f = \mathcal{M}_{0,v}(x_b)$. The forecast error displayed in Figure 3 is calculated by using $||\mathcal{M}_{0,v}(x_b) - \mathcal{M}_{0,v}(x_0^t)||_E$ at t_v, where the background estimate is used as the initial conditions. We have found that the forecast error is large over the domain $\mathcal{D}_v = [65°\text{S}, 35°\text{S}] \times [100°\text{W}, 65°\text{W}]$, which is considered as the verification domain for our experiment.

4. RESULTS

In our twin experiment, we first minimize the cost function without adding the penalty term to the cost function. The minimization process using the Limited Memory Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton (L-BFGS Q-N) method terminates successfully after 16 iterations and 21 function evaluations. The cost function is decreased by 10 orders of magnitude, whereas the norm of the gradient is decreased by four orders of magnitude. We use the resulting optimal initial condition to compute the forecast error $||\mathcal{M}_{0,v}(x_0^a) - \mathcal{M}_{0,v}(x_0^t)||_E$ at t_v. The result is shown in Figure 4.

We then estimate the optimal initial condition by taking some adaptive observations with the routine observations by using the adjoint sensitivity method. The result shows that adaptive observation added to the routine observations based on adjoint sensitivity improves the forecast only slightly. To compute the adjoint sensitivity, we used Algorithm 1. The adjoint sensitivity and adaptive observation locations are displayed in Figures 5(a)–(f). The forecast error over $\mathcal{D}_v$ at t_v obtained by using the initial condition x_0^a optimized by assimilating routine and adaptive observations is displayed in Figure 6. The algorithm to compute the adjoint sensitivity is provided as follows:

DOI: 10.1002/fld

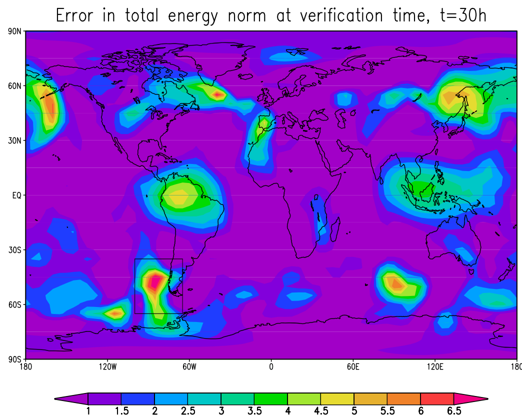

Figure 4. Graph of the forecast error $t = 30$ h from the optimal analysis obtained by using only routine observations.

Algorithm 1

1. Calculate model solution x_v^t at t_v with initial condition (true atmospheric state) x_0^t by

$$x_v^t = \mathcal{M}_{0,v}(x_0^t) \tag{21}$$

2. Obtain optimal initial condition x_0^a by minimizing the cost functional $\mathcal{J}$ defined in (1) with only routine observations. Calculate model forecast

$$x_v^f = \mathcal{M}_{0,v}(x_0^a) \tag{22}$$

3. Compute $\nabla_{x_v}\mathcal{J}_v = P^T E P\left(x_v^f - x_v^t\right)$ and use it as initial condition for adjoint model.
4. Integrate adjoint model backward from t_v to t_i : $\nabla_{x_i}\mathcal{J}_v = \mathbf{M}_{i,v}^T \nabla_{x_v}\mathcal{J}_v$

We then carry out several sequential minimizations of the penalized cost function defined in Equations (11) and (16) aiming at reducing the forecast error to certain minimum level. The minimum forecast error can be attained when a large value of the penalty parameter r is employed. The value of the penalty parameter and Lagrange multiplier λ are adaptively increased on the basis of the value of the cost function $\mathcal{J}_v$ over the verification domain. In our experiment, the initial values for both the penalty parameter and multiplier are $r_0 = 1$ and $\lambda_0 = 0$, respectively. The values are sequentially increased on the basis of the values of the $\mathcal{J}_v(k)$ for each call of the unconstrained minimization routine. We have also set the upper bound for $\mathcal{J}_v$ as $\delta = 10^{-4}$. Therefore, $\epsilon = \sqrt{\delta} = 10^{-2}$.

Algorithms 2 and 3 are used to find the optimal minimizer x_0^p of the penalized cost function defined by using quadratic penalty and augmented Lagrangian method. We have found that the minimization routine performed well for a smaller value of the penalty parameter ($r < 10^5$). For the large value of r, that is, $r \geqslant 10^5$, the minimization fails to converge. However, in our experiment, we have found that the forecast error was reduced significantly, which means that the initial condition is estimated optimally by adding the penalty term to the cost function. The forecast error computed with the optimal initial condition x_0^p is displayed in Figure 7.

Algorithm 2 (Quadratic penalty)

1. Initialization: $r_0 = 1$, $\beta_0 = 6$
2. Calculation of $\mathcal{J}_v(0)$ with the starting point x_0.
3. Do loop $k = 1, 2, \ldots$
4. Forward integration of the forecast model

DOI: 10.1002/fld

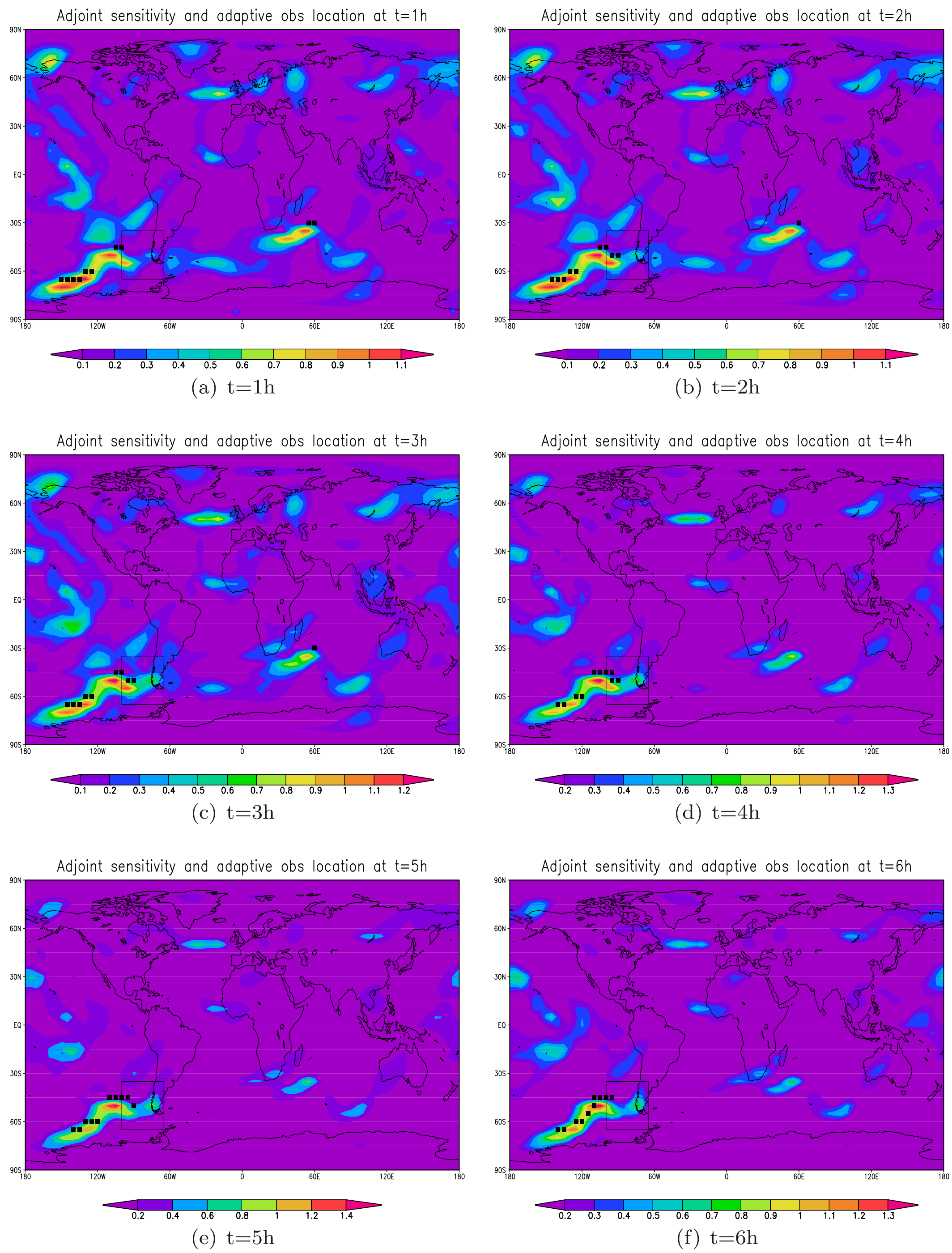

Figure 5. Time evolution of the sensitivity field corresponding to adjoint sensitivity method. The location of the adaptive observation at target instant t_i over assimilation window is marked with '□'.

5. Calculation of $\mathcal{J}_v(k)$ by using x_k^*
6. Calculation of $\beta_k = \frac{\mathcal{J}_v(0)}{\mathcal{J}_v(k)}$
7. If $\beta_k > 1$ then $r_{k+1} = \beta_k r_k$ else $r_{k+1} = 6r_k$
8. Find new initial condition x_k^* by an unconstrained minimization routine implementing a limited memory quasi-Newton technique
9. End do loop

DOI: 10.1002/fld

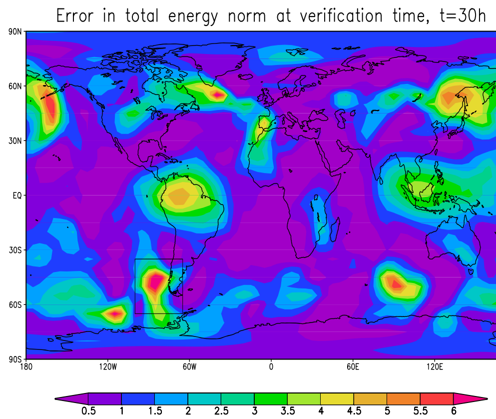

Figure 6. Graph of the forecast error $t = 30$ h from the optimal analysis obtained by using routine plus adaptive observations.

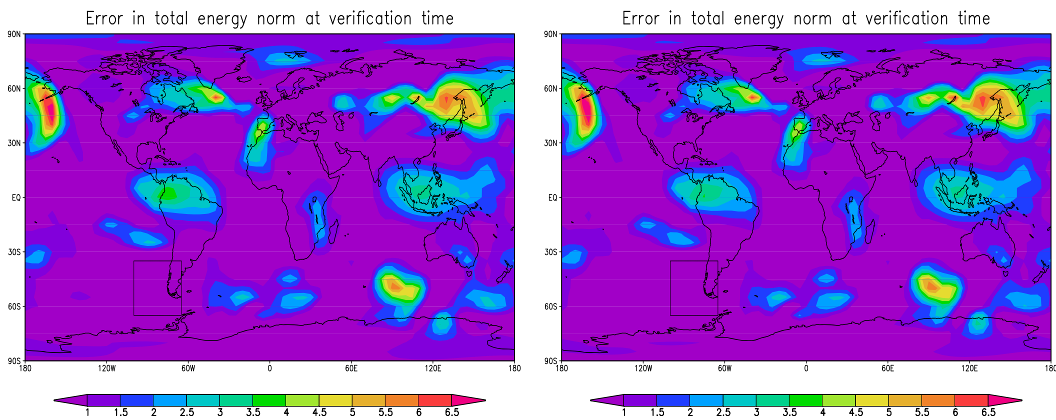

Figure 7. Graph of the forecast error $t = 30$ h with the optimal analysis x_0^p obtained by using quadratic penalty (left) and augmented Lagrangian (right) methods. In the figure, we see that forecast error over the verification domain is very small (less than 10^{-3}) if the optimal analysis x_0^p is used as initial condition for the forecast model.

Algorithm 3 (Augmented Lagrangian method)

1. Initialization: $r_0 = 1$, $\beta_0 = 6$, $\lambda_0 = 0$
2. Calculation of $\mathcal{J}_v(0)$ with the starting point x_0.
3. Do loop $k = 1, 2, \ldots$
4. Forward integration of the forecast model
5. Calculation of $\mathcal{J}_v(k)$ by using x_k^*
6. Calculation of $\beta_k = \frac{\mathcal{J}_v(0)}{\mathcal{J}_v(k)}$
7. If $\beta_k > 1$ then $r_{k+1} = \beta_k r_k$ else $r_{k+1} = 6r_k$
8. λ_k is updated by $\lambda_{k+1} = \lambda_k + r_k * (\sqrt{\mathcal{J}_v} - \epsilon)$.
9. Find new initial condition x_k^* by an unconstrained minimization routine implementing a limited memory quasi-Newton technique
10. End do loop

It is well known that the performance of minimization routine is very sensitive to the large values of penalty parameter. The reason is that the condition number of the Hessian matrix of the cost with

DOI: 10.1002/fld

respect to the control variables evaluated at the minimum increases as r is getting larger. Moreover, if the initial value of penalty parameter r is too large, it is very difficult to find the minima for any robust unconstrained minimization routine because of the slow convergence induced by the increasingly larger condition number of the Hessian of the penalized cost function. For this reason, we solved the problem of the quadratic penalty function as well as augmented Lagrangian function sequentially by using the unconstrained minimization routine M1QN3 [22] equivalent to L-BFGS routine [23] while moderately increasing values of the penalty parameter. We have found that both the methods perform very well if the penalty parameter is chosen by using a cost function $\mathcal{J}_v$ that is decreasing slowly and, consequently, the value of penalty parameter is increasing slowly. In our experiment, each successive x_k^* is used as the new starting point for solving an unconstrained minimization problem with the next increased value of the penalty parameter until an acceptable convergence criterion is attained.

5. SUMMARY AND CONCLUSIONS

In numerical weather prediction, we can reduce the forecast error by optimizing the initial condition. To obtain the optimal initial condition, we need to minimize the cost function defined by Equation (1), which depends on background information and observations. Studies show that only a few adaptive observations included along with the existing routine observations can improve the weather forecast. Several targeting methods have already been developed. In this paper, we use the newly proposed method, that is, the penalized 4D-Var DA and compare it with the adjoint sensitivity method. The approach proposed in this paper is found to be able to estimate the initial condition optimally by minimizing the penalized cost function defined in Equations (11) and (16). The evolution of the cost function and penalized cost function as well as their corresponding gradients are displayed in the Figure 8. We also display the evolution of the forecast error and the gradient of the cost function $\mathcal{J}_v$ obtained by using the aforementioned outlined different approaches in Figure 9. We have found that the forecast error is reduced significantly by using the new approach employing either the quadratic penalty method or the augmented Lagrangian method. We have found that the augmented Lagrangian method performs slightly better than the quadratic penalty method as expected from optimization theory. From the results obtained, we conclude that the proposed penalized 4D-Var approach (for both cases) performs better than the adaptive observations method. But the limitation of the proposed method is to the need to know the true model, x_0^t. However, in many cases, we can find a proxy that constitutes an approximation to the true model. Therefore, the penalized 4D-Var method enables us to obtain an optimal initial condition that may provide a better forecast than the aforementioned methods, for instance, adjoint sensitivity method without requiring us to provide additional observations to the existing routine network observations.

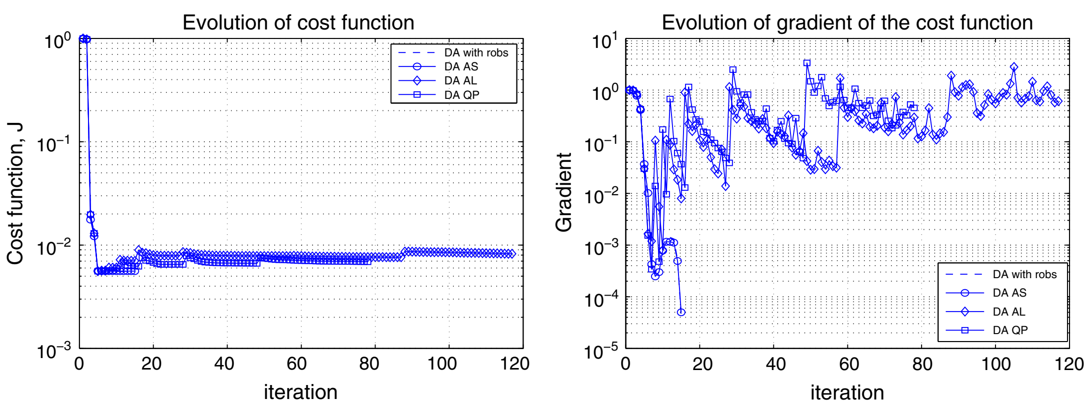

Figure 8. The normalized value of the cost function (left) and gradient of the cost function (right).

DOI: 10.1002/fld

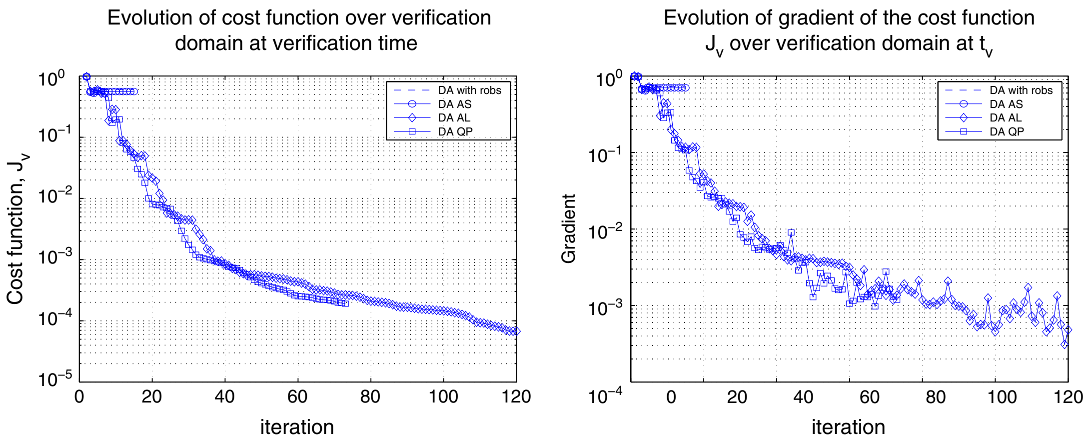

Figure 9. The normalized value of the cost function (left) and gradient of the cost function (right).

The proposed method is computationally more expensive in CPU time as several computational steps need to be carried out. We have found that the proposed method requires 15.045E + 02 s, whereas the adjoint method requires only 3.012E + 02 s. That is, the proposed method is five times slower than the adjoint method. The CPU time depends on the value of δ, that is, the expectation of the accuracy. For the proposed method, we use a larger value for the precision of stopping criterion (10^{-2}), whereas for adjoint sensitivity method, we use a smaller precision value (10^{-5}); otherwise, the minimization routine for the proposed method would require a large number of minimization iterations to attain the optimal value for each step.

ACKNOWLEDGEMENTS

This research was funded by the National Aeronautics and Space Administration (NASA) (grant no. NNG06GC67G). The first author would like to thank Professor A. A. Z. Ahmad for his critical reading and for making suggestions to correct the manuscript.

REFERENCES

1. LeDimet FX, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* 1986; **38 A**:97–110.
2. Derber J. Variational four-dimensional analysis using quasi-geostrophic constraints. *Monthly Weather Review* 1987; **115**:998–1008.
3. Lewis JM, Derber JC. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus* 1985; **37 A**:309–322.
4. Talagrand O, Courtier P. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: theory. *Quarterly Journal of the Royal Meteorological Society* 1987; **113**:1311–1328.
5. Lions JL. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer–Verlag: Berlin Heidelberg New York, 1971.
6. Glowinski R. *Numerical Methods for Nonlinear Variational Problems*. Springer–Verlag: New York, 1984.
7. LeDimet FX, Ngodock HE, Luong B, Verron J. Sensitivity analysis in data assimilation. *Journal of the Meteorological Society of Japan* 1997; **75**:245–255.
8. Thepaut JN, Courtier P. Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. *Quarterly Journal of the Royal Meteorological Society* 1991; **117**:1225–1254.
9. Navon IM, Zou X, Derber J, Sela J. Variational data assimilation with an adiabatic version of the NMC spectral model. *Monthly Weather Review* 1992; **120**:1433–1446.
10. Zupanski M. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Monthly Weather Review* 1993; **121**:2396–2408.
11. Thepaut JN, Hoffman RN, Courtier P. Interactions of dynamics and observations in a four-dimensional variational assimilation. *Monthly Weather Review* 1993; **121**:3393–3414.
12. Langland RH. Issues in targeted observing. *Quarterly Journal of the Royal Meteorological Society* 2005; **131**:3409–3425.

13. Daescu DN, Navon IM. Adaptive observations in the context of 4D-Var data assimilation. *Meteorology and Atmospheric Physics* 2004; **85**:205–226.
14. Palmer TN, Gelaro R, Barkmeijer J, Buizza R. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences* 1998; **55**:633–653.
15. Barkmeijer J, Van Gijzen M, Bouttier F. Singular vectors and estimates of the analysis-error covariance metric. *Quarterly Journal of the Royal Meteorological Society* 1998; **124**:1695–1713.
16. Barkmeijer J, Buizza R, Palmer TN. 3D-Var Hessian singular vectors and their potential use in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* 1999; **125**:2333–2351.
17. Ehrendorfer M, Tribbia JJ. Optimal prediction of forecast error covariances through singular vectors. *Journal of the Atmospheric Sciences* 1997; **54**:286–313.
18. LeDimet FX, Navon IM, Daescu DN. Second order information in data assimilation. *Monthly Weather Review* 2002; **130**(3):629–648.
19. Godinez HC, Daescu DN. Observation targeting with a second-order adjoint method for increased predictability. *Computational Geosciences* 2011; **15**:477–488.
20. Parrish DF, Derber JD. The national meteorological center spectral statistical interpolation analysis system. *Monthly Weather Review* 1992; **120**:1747–1763.
21. Akella S, Navon IM. Different approaches to model error formulation in 4D-Var: a study with high resolution advection schemes. *Tellus A* 2009; **61 A**:112–128.
22. Gilbert JCh, Lemarećhal C. Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical Programming* 1989; **45**:407–435.
23. Liu DC, Nocedal J. On the limited memory BFGS method for large-scale minimization. *Mathematical Programming* 1989; **45**:503–528.

DOI: 10.1002/fld