Q. J. R. Meteorol. Soc. (200x), **00**, pp. 1–999 doi: 10.1256/qj.yy.n

On forecast errors in variational data assimilation using high resolution advection schemes of the Lin-Rood finite volume shallow water model

By S. Akella[1] * and I. M. Navon[1] †
[1]*Florida State University, USA.*

(Received ; revised)

SUMMARY

We used the the Lin-Rood finite volume shallow water model in the framework of 4-D Var data assimilation addressing first the hierarchical implementation of high resolution (van Leer and PPM) advection schemes in both forward and adjoint models. The results obtained show that using the various advection schemes consistently improves variational data assimilation (VDA) in the strong constraint form, which does not include model error, but the cost functional included efficient and physically meaningful construction of the background term, $\mathcal{J}_b$ using balance and diffusion equation based correlation operators. This was then followed by an in-depth study of various approaches to model the systematic component of model error in the framework of a weak constraint VDA. Three simple forms, decreasing, invariant, and exponentially increasing in time forms of evolution of model error were tested. The inclusion of model error provides a substantial reduction in forecasting errors, in particular the exponentially increasing form in conjunction with the piecewise parabolic high resolution advection scheme provided the best results. Results presented in this article can be used to formulate sophisticated model error forms.

KEYWORDS: 4D-VAR Inverse Problems Model Error Strong Constraint Data Assimilation Weak Constraint Data Assimilation

1. INTRODUCTION

Variational data assimilation (VDA) aims to find a model trajectory that best fits (in a least squared sense) the observational data over an assimilation time interval by adjustment of the initial conditions supplied for forward model integration (Le Dimet and Talagrand 1986; Navon et al. 1992). In the so-called *strong constraint* or *classical* version of VDA, it is assumed that the forecast model perfectly represents evolution of the actual atmosphere. The best fit model trajectory is obtained by adjusting only the initial conditions via minimization of a cost functional, subject to the model equations as strong constraint. However numerical weather prediction (NWP) models are imperfect, since they are discretized, dissipative and dispersion errors arise, and, moreover subgrid processes are not included. In addition, most of the physical processes and their interactions in the atmosphere are parametrized, also a complete mathematical modeling of the boundary conditions and forcing terms can never be achieved. Usually all of these modeling drawbacks are collectively addressed by the term, *model error* (ME). Following Dee (1995), we would like to distinguish between forecasting and model errors. ME is one of the causes of forecasting errors, another cause being erroneous specification of initial conditions used to produce the forecast.

Studies indicate that ME can severely impact forecast errors, see Boer (1984); Dalcher and Kalnay (1987); Bloom and Shubert (1990); Zupanski and Zupanski (2002). For early methods on estimating modeling errors in operational NWP models see Thiébaux and Morone (1990); Saha (1992). Thus giving up the assumption that the model is perfect, in the context of strong constraint VDA leads us to *weak constraint* formulation of VDA, which is the main theme of this paper; since we include time evolution of the variables, we could say weak constraint 4D-Var (time plus three space

* E-mail: sakella@math.fsu.edu
† Corresponding author: School of Computational Science & Department of Mathematics, Florida State University, Tallahassee, Fl-32306, USA. E-mail: navon@csit.fsu.edu.

dimensions). Instead we prefer to use the general term VDA, because have we used a two dimensional global shallow water model for presenting our results.

In sequential data assimilation using Kalman filtering theory, the inclusion of ME forms an integral part of the filter formulation, various filtering approaches which include ME have been considered by Chepurin et al. (2005); Dee and Todling (2000); Dee and Da Silva (1999); Dee et al. (1999); Dee and Da Silva (1998); Dee (1995); Zupanski (2005). When the number of observations is considerably smaller, the method of *representers* (Bennett 1992) provides a computationally efficient (in storage/ space requirements) formulation of VDA. The incorporation of ME in such a framework has been shown by Bennett et al. (1993, 1996, 1997); Uboldi and Kamachi (2000).

Model error is formally introduced as a *correction* to the time derivatives of model variables in the weak constraint formulation of VDA. Let the vector $\mathbf{x}(t)$ be used to represent the state of the atmosphere, then its evolution accounting for ME in the NWP model is written as,

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{M}[\mathbf{x}(t)] + \mathbf{T}[\eta(t)], \tag{1}$$

where $\mathcal{M}[.]$ denotes all the mathematical operations involved in the NWP model, η represents ME and $\mathbf{T}[.]$ is an operator that accounts for the fact that only certain components of the state vector have modeling errors (none-the-less, often $\mathbf{T}[.]$ is set to be equal to the unit matrix). ME usually varies both spatially and temporally, and has both systematic and stochastic components. Comparing the strong and weak constraint VDA, in the formulation of former, it is assumed that η has mean, $E[\eta(t)] = 0, \forall\, t$ and model error covariance matrix, $\mathbf{Q} = E[\eta(t)\, \eta^T(t')] = 0, \forall\, t \,\&\, t'$, where $E[.]$ is the mathematical expectation operator. It should be noted that if the mean and (co)variance of a random vector are prescribed to be equal to zero, then all realizations of that random vector are identically equal to zero, thus, $\eta \equiv 0$. Whereas in the weak constraint version of VDA, the mean and covariance of ME are to be specified. However exact statistical details of ME are difficult to obtain (Daley 1992a,b; Dee and Da Silva 1998; Zhu and Kamachi 2000) a fact which led researchers to suggest a variety of assumptions to approximate ME.

Early efforts to model the systematic component of ME were pioneered by Derber (1989). He suggested a simplified approach to model η to be equal to $\lambda(t)\, \phi$. The temporal part, $\lambda(t)$ is a specified function of time alone, and ϕ is a spatially dependent, control variable. Three different forms of λ were considered, namely, parabolic, delta function and constant in time. It was observed that the parabolic variation of λ provided results comparable to a constant in time λ. Using a similar approach (Wergen 1992; Zupanski 1993) it was shown that inclusion of ME allows significant reduction in forecast RMSE (see table (1) for a list of acronyms and their definitions).

For dynamically evolving systems such as discrete NWP models, ME is expected to depend on the model state and should be evolving in time (Griffith and Nichols 1996, 2000). Various simple forms of evolution of ME in time were considered by Griffith and Nichols (2000), henceforth referred to as GN00. At any time step, t_k, the evolution of ME is assumed to be given by the following equation,

$$\eta_k = T_k(\mathbf{e}_k) + \mathbf{q}_k, \tag{2}$$

where $\mathbf{e}_k$ represents time-varying systematic components of ME, T_k describes the distribution of systematic errors in the NWP model equations, and $\mathbf{q}_k$ (stochastic component) is an unbiased, serially correlated, normally distributed random vector, with

known covariance. The evolution of $\mathbf{e}_k$, is in-turn modeled by assuming that it depends on the state vector, $\mathbf{x}_k$,

$$\mathbf{e}_{k+1} = g_k(\mathbf{x}_k, \mathbf{e}_k).$$

GN00 suggested three forms for the evolution of the above systematic component of ME,

1. constant in time: $\mathbf{e}_{k+1} = \mathbf{e}_k$, $T_k \equiv I$. It is inferred that this form is suitable for modeling errors in source terms and boundary conditions.
2. Evolving in time: $\mathbf{e}_{k+1} = \mathcal{F}_k\, \mathbf{e}_k$, $T_k \equiv I$, where $\mathcal{F}_k$ is a linear model, which is appropriate for representing discretization errors.
3. Spectral form: $\mathbf{e}_{k+1} = \mathbf{e}_k$, T_k is a block diagonal matrix, with diagonal entries given by $I, I \cdot \sin(\frac{\kappa}{N\tau}), I \cdot \cos(\frac{\kappa}{N\tau})$, where τ is a constant time scale.

It is to be noted that the control of ME as well as the model initial conditions in weak constraint VDA doubles the size of the optimization problem (compared to strong constraint VDA), in addition if the stochastic component is included in the ME formulation, then one would have to save every random realization at each model time step, which amounts to tripling the size of the optimization problem. The computational results in GN00 were provided by neglecting $\mathbf{q}_k$, the stochastic component of ME and using the constant and evolving forms of the systematic component, see GN00 for additional details. Similar approaches for modeling the systematic component of ME was considered by Martin et al. (2002) and reduction of ME control vector size by projecting it on to the subspace of eigenvectors corresponding to the leading eigenvalues of the adjoint-tangent linear operators was illustrated by Vidard et al. (2000).

The above described approach (of GN00) provides the systematic component of ME at any discrete time step, t_k, in other words, the evolution of ME has been considered as a *discrete process*. Vidard et al. (2004) (from now onwards referred to as VPLD04) considered a continuous in time form for the evolution of ME. This approach is consistent with the fact that model equations are first written as differential equations and then descretized in space and time. If the initial ME, $\eta(t_0) = \eta_0$, then VPLD04 modeled the evolution of ME as,

$$\frac{d\eta}{dt} = \Phi[\eta(t), \mathbf{x}(t)] + \mathbf{q}(t), \tag{3}$$

where $\mathbf{q}(t)$ is the stochastic component of ME. Once again, neglecting the stochastic component reduces the size of the control vector, as in the case considered by VPLD04; they also assumed that $\Phi[\eta(t), \mathbf{x}(t)] = \eta(t)$. This implies that the evolution of ME term is modeled by the following simple exponential growth equation,

$$\frac{d\eta}{dt} = \eta(t).$$

Such a *deterministic* approach to model the evolution of ME significantly simplifies the weak constraint VDA, since only the initial ME (η_0) is to be obtained via solution of the optimization problem (see VPLD04 for additional details).

Daley (1992a) suggested that model error is correlated in time and used a Markov process to model its evolution in a simple Kalman filtering (KF) framework. The Markovian assumption is based on the observation that as the numerical model is integrated in time, errors show a trend of serial correlation in both time and space.The most important property of a Markov process is that the state at any time in future is

dependent only on its present value, but not on its value in the past. Considering the model error as a Markov process, at any two successive time steps, t_k and t_{k+1},

$$\eta_{k+1} = \mu \mathbf{G}_k[\eta_k] + (1-\mu)\mathbf{q}_k, \tag{4}$$

where μ is a scalar, such that $0 \leqslant \mu < 1$, $\mathbf{q}_k$ is the random component of ME and $\mathbf{G}_k[.]$ is a linear operator (see Daley, 1992a for discussion on the implementation of two different forms of this operator, again in a KF setting). Using $\mathbf{G}_k \equiv I$, in the above form of ME, D. Zupanski (1997) and Zupanski et al. (2005) (henceforth referred to as ZZ05) provided results obtained using the NCEP' s regional weather prediction system in weak constraint VDA framework. They assumed the initial value of the model error, η_0 to be equal to zero and a coarse time scale for the evolution of $\mathbf{q}_k$, such that at-most four of such vectors are present in a 12 hour time window of data assimilation (to limit the size of control vector in minimization). In addition, all of those $\mathbf{q}_k$' s were obtained via minimization of the weak constraint VDA cost functional, thereby implicitly assuming that each $\mathbf{q}_k$ is a *m*ean deterministic forcing term. The model error covariance matrix was derived in a novel way, please see ZZ05 and references therein for details.

Often discontinuities (on the scale of the model grid resolution) in solutions to NWP models arise due to sharp fronts formed in low-pressure systems, hydraulic jumps, etc. High resolution advection schemes provide means to capture such discontinuous solutions accurately. A consistent formulation to improve the accuracy of the numerical high resolution advection scheme, ranging from first order to second and third order accuracy in space was provided by Akella and Navon (2005), hereafter referred to as AN05. In the same paper was provided a comparison of the impact of using different slope limited monotone upstream centered schemes for conservation laws (MUSCL) on data assimilation (in strong constraint form) in one space dimension using a viscous Burgers equation model and in two dimensions using a shallow water equations model. Another consistent method of decreasing the discretization errors (principally truncation errors) is to refine the model resolution, such an approach in VDA was suggested by Le Dimet and Shutyaev (2005), henceforth referred to as LDS05. However this approach is limited by the resolution of the observational system, indeed one of the conclusions of LDS05 was that the improvement in predictability is most sensitive to the observational errors. In the present article we extend the studies in AN05, by conducting various VDA experiments in a more practical setting than that presented in AN05. In particular the cost functional includes $\mathcal{J}_b$, the background cost functional, the formulation of which seriously impacts the performance of data assimilation system. The square root formulation using linear balance operator and diffusion operator has been used so that the inverse of the background error covariance matrix is not required to be specified (details are described in Appendix A).

Following is the outline of present paper. In section 2 we focus on the issue of uncertainty in the specification of initial conditions only (recall that it contributes to forecast errors) with no ME term in VDA. We first show that changing the advection scheme used in discretization of the non-linear terms in the governing equations (which can be considered as altering the numerical model) leads to a decrease in forecasting error. Next we provide results obtained using various schemes by conducting VDA in strong constraint form, and once again an improvement in predictability is achieved by improving the numerical model used in VDA. Section 3 focuses on the issue of accounting for model error in VDA, via weak constraint formulation. We provide a detailed formulation of feasible forms of modeling the ME. Using three different forms of modeling the evolution of ME, an analysis of the obtained results is discussed. Finally

in section 4 we summarize our results emphasizing what is novel in our contribution and outline future research directions in this area.

2. IMPACT OF USING DIFFERENT NUMERICAL ADVECTION SCHEMES

(*a*) *Model forecasts*

In this article a global two-dimensional shallow water equations (SWE) model has been used for numerical experiments and analysis. The solutions of SWE exhibit some of the important properties of large scale atmospheric flow and the equations have certain important features (such as, horizontal dynamical aspects) in common with more complicated NWP models. Therefore derivation and testing (Williamson et al. 1992) of various algorithms for solving SWE has often been a first step towards developing new atmosphere and ocean GCMs. In spherical coordinates the vorticity divergence form of the SWE can be written as following,

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{V}h) = 0 \tag{5}$$

$$\frac{\partial u}{\partial t} = \Omega v - \frac{1}{a\ cos\,\theta}\frac{\partial}{\partial \lambda}(\kappa + \varphi) \tag{6}$$

$$\frac{\partial v}{\partial t} = -\Omega u - \frac{1}{a}\frac{\partial}{\partial \theta}(\kappa + \varphi) \tag{7}$$

where h represents the fluid height (above the surface height, h_s), $\mathbf{V} = (u, v)$, u and v represent the zonal and meridional wind velocity components respectively, θ and λ are the latitude and longitudinal directions respectively, ω is the angular speed of rotation of the earth, a is radius of the earth. The free surface potential is given by

$$\varphi = \varphi_s + g\ h,$$

$\varphi_s = gh_s$, $\kappa = \frac{1}{2}\mathbf{V} \cdot \mathbf{V}$ is the kinetic energy, and $\Omega = 2\omega sin\theta + \nabla\ \times\ \mathbf{V}$ is the absolute vorticity. Details on other forms of writing the SWE and their development can be found in Williamson et al. (1992) and Haltiner and Williams (1980).

The explicit flux-form semi-Lagrangian, finite volume shallow water equations model of Lin and Rood (1997), henceforth referred to as LR97, has been used for forward model integration. This model serves as the dynamical core in the community atmosphere model (CAM), version 3.0, and its operational version implemented at NCAR and NASA is known as finite volume-general circulation model (FV-GCM). A two grid combination based on C-grid and D-grids was used for advancing from time step t_n to $t_n + \Delta t$. In the first half of the time step, advective winds (time centered winds on the C-grid: (u^*, v^*)) are updated on the C-grid, and in the other half of the time step, the prognostic variables (h, u, v) are updated on the D-grid. We will follow the suggestion in LR97, and always use unconstrained van Leer scheme to advect winds on the C-grid (this strategy provides solutions whose accuracy is comparable to those obtained by using more CPU demanding advection schemes, for e.g., constrained van Leer and PPM schemes), except for the first order advection scheme, in which case, we will use first order scheme on both C- and D-grids. Therefore on the D-grid, we will be using the unconstrained, constrained van Leer, and the PPM schemes. Using the finite volume method, within each cell of the discrete grid, if we consider a piecewise linear approximation to the solution, whose slope is *limited* in a certain way depending on the values of the solution at the neighboring grid cells, one can consistently derive a family

of van Leer schemes. Alternatively, if we assume a piecewise parabolic approximation to the solution within each cell, then we obtain the PPM scheme. For further details on formulation of these schemes see Lin et al. (1994) and AN05. From now onwards we will use the following convention to refer to our **test cases**,

1. *first order advection scheme*: first order advection on both C- and D-grids,
2. *unconstrained van Leer scheme*: unconstrained van Leer on both C- and D-grids,
3. *constrained van Leer scheme*: constrained van Leer on D-grid and unconstrained van Leer scheme on C-grid,
4. *PPM scheme*: PPM scheme on D-grid and unconstrained van Leer scheme on C-grid.

The poles have been treated in a fashion similar to that in Suarez and Takacs (1995) using a polar Fourier filter. Further details of the model can be found in LR97 and references therein. Also a comparison of these schemes for the test cases proposed in Williamson et al. (1992) are provided in LR97.

Unless specified otherwise, here we consider a regular latitude-longitude discretization on the sphere, using a $2.5^o \times 2.5^o$ grid resolution, and a time step of $\Delta t = 450$ seconds. Reanalyzed data at 500 hPa pressure level obtained from the ERA-40, ECMWF 40-year reanalysis (ECMWF (2002)) system, valid for 00 UTC 2 February 2001 (henceforth denoted by T_{-06}) was used to specify the geopotential height field (winds fields were obtained using geostrophic assumption) as initial conditions for forward model integration. Using the above specified advection schemes, we integrated the model for 36 hours, saving forecasts at every 6 hour interval. In the forecast and adjoint models, to introduce systematic errors ω was set to 0.95 times the value used for generating observations, which was specified to be equal to $7.292 \times 10^{-5} s^{-1}$. In order to obtain the observations, a twin experiment framework is considered. To simulate real-life *noisy* observations, a 1% random perturbation in the initial conditions prescribed at T_{-06} was added and the PPM advection scheme was used on both C- and D- grids (though the PPM scheme is expensive to implement, it provides a very accurate forecast, see LR97, AN05). We integrated the model for 36 hours, once again saving the states after every 6 hours, as *observations*, see table (1) for nomenclature of the different time intervals. RMSE between model forecasts and observations are provided for first order, unconstrained van Leer, constrained van Leer and PPM schemes (test cases: 1- 4) in Fig. 1. The RMSE indicate a trend of decreasing errors, the first order scheme being the most erroneous whereas the PPM scheme being characterized by the least error when compared with the observational data (since the first order scheme performs poorly when compared to the other schemes, we will discontinue its usage in our further studies), and the unconstrained van Leer scheme exhibits larger errors than the constrained van Leer scheme. These results are consistent with previous results (see for instance AN05; LR97; Lin et al. (1994)), the larger implicit diffusive property of the van Leer schemes (the constrained van Leer scheme is better than the unconstrained van Leer scheme due to the monotonicity constraint applied in the former) when compared to the PPM scheme has been argued to be the reason for the above trend in errors. In the following subsection, we further analyze these schemes, particularly in the context of strong constraint VDA (thus dealing with the issue of erroneous specification of initial conditions only).

(*b*) *VDA experiments in strong constraint formalism*

Data assimilation schemes determine the analysed atmospheric state as an optimal combination of a-priori background information and observational information. Let $\mathbf{x}^t$

be the *true* state of the atmosphere, $\mathbf{x}^b$ be the background field and $\mathbf{y}^o$ denote the observations. Usually a short-range forecast provides $\mathbf{x}^b$. Then the error in the background field is equal to $\mathbf{e}^b = \mathbf{x}^b - \mathbf{x}^t$, and the error in the observations field is given by $\mathbf{e}^o = \mathbf{y}^o - \mathbf{H}(\mathbf{x}^t)$. $\mathbf{H}$ is an observation operator that maps model variables to observations, if all the model variables are observed and observations occur at every model grid point, then $\mathbf{H} \equiv I$. Denoting the mathematical expectation operator by $E[.]$, the background error and observation error covariances are given by $\mathbf{B} = E[(\mathbf{e}^b - E[\mathbf{e}^b])(\mathbf{e}^b - E[\mathbf{e}^b])^T]$ and $\mathbf{R} = E[(\mathbf{e}^o - E[\mathbf{e}^o])(\mathbf{e}^o - E[\mathbf{e}^o])^T]$, respectively. Error covariances measure the uncertainty involved with both of these data sources, hence they determine the quality of data assimilation. Due to lack of knowledge of the true state of the atmosphere, $\mathbf{x}^t$, we can only guess what $\mathbf{B}$ and $\mathbf{R}$ should be. Thus they are approximations of the *true* error covariances.

In strong constraint version of VDA, neglecting the ME, minimization of the following nonlinear quadratic cost functional, $\mathcal{J}_o$, accomplishes the goal of fitting model states ($\mathbf{x}(t_i)$) and observations ($\mathbf{y}^o(t_i)$) in an assimilation time interval, $[t_0, t_n]$. Often the so-called background cost, $\mathcal{J}_b$ is added to $\mathcal{J}_o$ to regularize the following cost functional. Its minimization with respect to the initial state, $\mathbf{x}(t_0)$ as a control variable (Kalnay 2003),

$$\begin{aligned} \mathcal{J}[\mathbf{x}(t_0)] = \mathcal{J}(\mathbf{x}_0) &= \underbrace{\frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^b]^T \, \mathbf{B}^{-1} \, [\mathbf{x}(t_0) - \mathbf{x}^b]}_{\mathcal{J}_b} \\ &+ \underbrace{\frac{1}{2} \sum_{i=0}^{n} [H(\mathbf{x}(t_i)) - \mathbf{y}^o(t_i)]^T \, \mathbf{R}^{-1} \, [H(\mathbf{x}(t_i)) - \mathbf{y}^o(t_i)]}_{\mathcal{J}_o}, \end{aligned} \quad (8)$$

subject to the following model equations as strong constraint,

$$\begin{aligned} \mathbf{x}(t_0) &= \mathbf{x}_0, \\ \frac{d\mathbf{x}(t)}{dt} &= \mathcal{M}[\mathbf{x}(t)], \end{aligned} \quad (9)$$

is achieved by using iterative minimization algorithms, such as quasi-Newton or truncated-Newton methods. These algorithms require availability of gradient of the cost functional with respect to the control variables, which is in-turn efficiently obtained by backward integration of the adjoint model (Lorenc 1986; Navon et al. 1992; Zou et al. 1993). Note that in the above model equations, we did not account for ME, i.e, $\eta(t) \equiv 0, \forall t$.

One of the principal causes of observational errors is instrumentation error, which is sequentially correlated in space and time. Accurate specification of the observation error statistics is very important in the implementation of data assimilation techniques. However in this study we deal with model errors and the observational error and background error covariances have been assumed to be invariant in time, for further details regarding the impact of observational errors on data assimilation, please see Daley (1992a, 1993). Further we assume that the observations are not biased and that the background state and observations are mutually uncorrelated.

The formulation of the $\mathcal{J}_b$ term is crucial to the performance of the data assimilation system. Considering a single observation, at a single grid point, the analysis increment

is proportional to a column of **B**. Hence background error covariance spreads out the information in the analysis from the observations and provides statistically consistent increments at the neighboring grid points and levels of the model. It also ensures that observations of one model variable produce dynamically consistent increments in the other model variables. Using background knowledge makes the VDA problem well-posed even when there are only a few observations, also it fills any data voids with *good quality* information (Navon et al. 2005). In addition the background state, $\mathbf{x}^b$ provides an initial guess for minimization of $\mathcal{J}$. Ideally the optimal design of background error covariance should take into account the average variances, autocorrelations and balance properties of the background errors, so that the covariances of short range forecast errors in data assimilation are adequately represented (Derber and Bouttier 1999). Dynamic or flow dependent formulation of **B** could improve analyses and subsequent forecasts (Rishøjgaard 1998), particularly if the observations are nonuniformly distributed. However most of the studies and in particular, operational implementations use a *static* background error covariance; since the focus of this paper is model error, we do not deal with these issues anymore and follow the approaches of Weaver and Courtier (2001); Derber and Bouttier (1999) to construct **B** as a multivariate and cross correlated operator, see Appendix A for further details (see also Gaspari and Cohn 1999 for an alternative formulation).

We used the same observations as used in the previous subsection to compare different model forecasts to conduct DA using different advection schemes for a time interval of 24 hours in a twin experiment framework. These observations were obtained by introducing random perturbations in the initial condition, which can be looked upon as introducing uncertainty in initial conditions and using a slightly different version of the model (namely, using PPM advection scheme on both C- and D-grids). The fact that observations were generated by usage of a different model than that used for DA, introduces a systematic model bias. We conducted three DA experiments, in each case, we used either unconstrained van Leer (test case 2), constrained van Leer (test case 3) or PPM advection scheme (test case 4) in both forward and adjoint modes (see AN05 for details on derivation of the adjoint model for high resolution advection schemes).

The background state, $\mathbf{x}^b$, which was the first guess for DA was obtained by a 6 hour forward integration of the reanalyzed data, at time, T_{-06}; see table (1) and Fig. 2 for naming and illustration of the different data sets used in DA experiment. Five observational data sets (at times, $T_{00}, T_{+06}, T_{+12}, T_{+18}, T_{+24}$, such that every 6 hours we have an observation) within a 24 hour interval are assimilated using the observations to model space operator, $\mathbf{H} = I$, i.e, the observations occur at every grid point of the model resolution. The observation error covariance matrix has been taken to be a block diagonal matrix, $R = [10^4 I, 100I, 100I]$, such that observational errors at every grid point are only autocorrelated and stationary in time. We used an unconstrained limited memory quasi-Newton (L-BFGS) minimization algorithm (Liu and Nocedal 1989; Nash and Nocedal 1991) for minimization of the cost functional given in Eq. (8). The following termination criteria was used to conduct DA experiments,

$$\|(\nabla \mathcal{J})_k\| \leq EPS \cdot MAX(1, \|\mathbf{x}_k(t_0)\|), \tag{10}$$

where $\|\cdot\|$ is the L_2 norm, $(\nabla \mathcal{J})_k$ is the gradient and $\mathbf{x}_k(t_0)$ is the optimal initial state vector at the $k^{th.}$ minimization iteration, and EPS was set to 5×10^{-5}.

During the minimization process, due to the regularization property of the minimization algorithm, the differences on larger scales are fit in the first few iterations, yielding the largest decrease in the cost functional, thereafter minimization proceeds to fit the smaller discrepancies, or small decreases in the value of the functional. In general

the observations occurring in the middle of the DA time window (in this case, at times, T_{+12} and T_{+18}) are best fit. We used the same minimization termination criteria given by Eq. (10) for all three DA experiments, and it took 114 function and gradient evaluations for the unconstrained van Leer, 78 for the constrained van Leer and the least, 63 for the PPM. A comparison of the RMSE in geopotential height field before and after DA is provided in Fig. 3 for the different advection schemes. Comparison of the differences in geopotential height field between model forecast using PPM advection scheme and observations at T_{+30}, forecast verification time, before and after data assimilation is provided in fig. 4 (similar results were obtained for the wind fields, as well as, using the unconstrained and constrained van Leer advection schemes). In all three cases we achieved more than 50% reduction in the RMSE by DA. Also the optimized initial condition is able to provide a better forecast at T_{+30} (30 hour forward integration) in all the cases. Clearly the PPM scheme is much superior when compared to the constrained and unconstrained van Leer schemes using lesser function and gradient computations in achieving the same reduction in forecast RMSE. We have demonstrated that it is possible to decrease the component of forecasting error associated with the mis-specification of initial conditions only by consistently improving the numerical advection scheme used for discretizing the nonlinear advection terms in the model, holding everything else fixed (resolution of the model and observational system as well). In the next section we will study the impact of introducing various forms of model error in VDA via weak constraint formulation.

3. WEAK CONSTRAINT VDA

Now we turn our attention to one other cause of forecasting error, namely the model error. In the strong constraint VDA, the model equations are assumed to be *perfect*, therefore modeling errors (whose causes have been described earlier in the introduction) are not taken into account. The weak constraint VDA provides a framework for incorporation of ME in the model equations, via explicit introduction of an extra term, $\eta(t)$, (as in Eq. 1),

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{M}[\mathbf{x}(t)] + \mathbf{T}[\eta(t)].$$

The operator $\mathbf{T}$ maps the space of the ME to the space of the model state, $\mathbf{x}$. If one has a-priori knowledge that the numerical model has some drawbacks, for e.g., modeling of the atmosphere in certain regions of the globe, (say, one of the poles) then the operator, $\mathbf{T}$ should be specified in such a way that only those model grid points (at the poles) have modeling errors, and the rest of the model states do not have any ME. In the literature (for instance, see GN00 and VPLD04), it has been often assumed that the model state at every grid point has an associated error, which implies that $\mathbf{T}$ is identically equal to the unit matrix, I, and, the dimension of η is equal to that of the model state, $\mathbf{x}$; in the present article we assume $\mathbf{T} = I$.

Past research work by Dee and Da Silva (1998) indicated that ME has contributions that are both systematic (or, deterministic) and random in nature. Following Derber (1989) and VPLD04 in the spirit of *variational continuous assimilation*, we will model the evolution of ME as a continuous process, using the following initial value problem (IVP), which is a continuous-in-time differential equation, (Eq. 3),

$$\frac{d\eta}{dt} = \Phi[\eta(t), \mathbf{x}(t)] + \mathbf{q}(t).$$

In this article, we are concerned only with the systematic part of ME, hence, the stochastic component, $\mathbf{q}(t)$, is neglected; the sequential filtering approaches based on KF, such as the ensemble KF provide an implicit framework for the inclusion of the stochastic terms (see the next section for a simple stochastic differential equation to model the evolution of the stochastic component in VDA). Therefore the above differential equation simplifies to,

$$\frac{d\eta}{dt} = \Phi[\eta(t), \mathbf{x}(t)]. \tag{11}$$

For closure of the above IVP, we need to specify the initial value of $\eta(t_0) = \eta_0$, and the nature of the mapping, $\Phi[.]$; both of them being very important parameters. First we describe the methodology used to calculate the initial value of ME and then address the issue of different approaches for modeling the evolution of ME, using different forms of $\Phi[.]$. To obtain η_0 the following weak constraint VDA cost functional ($\mathcal{J}$) is minimized (note that it is similar to the cost functional in Eq. (8), but includes an extra term, $\mathcal{J}_\eta$),

$$\begin{aligned}
\mathcal{J}[\mathbf{x}(t_0), \eta(t_0)] &= \mathcal{J}(\mathbf{x}_0, \eta_0) \\
&= \underbrace{\frac{1}{2}[\mathbf{x}_0 - \mathbf{x}^b]^T \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b]}_{\mathcal{J}_b} \\
&+ \underbrace{\frac{1}{2}\sum_{i=0}^{n}[H(\mathbf{x}(t_i)) - \mathbf{y}^o(t_i)]^T \mathbf{R}^{-1} [H(\mathbf{x}(t_i)) - \mathbf{y}^o(t_i)]}_{\mathcal{J}_o}, \\
&+ \underbrace{\frac{1}{2}[\eta_0 - \eta^b]^T \mathbf{Q}^{-1} [\eta_0 - \eta^b]}_{\mathcal{J}_\eta},
\end{aligned} \tag{12}$$

where $\mathbf{Q}$ is the model error covariance matrix (explained in Appendix B). Just as the background state, $\mathbf{x}^b$ was used as an initial guess for $\mathbf{x}_0$, to minimize the strong constraint VDA cost functional, we use η^b as an initial guess for η_0 to minimize the above $\mathcal{J}_\eta$ in the weak constraint VDA. Hence the above cost functional, $\mathcal{J}(\mathbf{x}_0, \eta_0)$ is minimized subject to the following equations as constraints,

$$\left.\begin{aligned}
&\mathbf{x}(t_0) = \mathbf{x}_0; \eta(t_0) = \eta_0, \\
&\frac{d\mathbf{x}(t)}{dt} = \mathcal{M}[\mathbf{x}(t)] + \eta(t); \frac{d\eta}{dt} = \Phi[\eta(t), \mathbf{x}(t)].
\end{aligned}\right\} \tag{13}$$

Introducing the following augmented Lagrangian functional, the above constrained minimization problem becomes an unconstrained problem,

$$\begin{aligned}
\mathcal{L}(\mathbf{x}, \eta, \mathbf{x}^*, \eta^*) &= \mathcal{J}(\mathbf{x}_0, \eta_0) \\
&+ \int_{t_0}^{t_n} \langle\, \mathbf{x}^*,\, \{\frac{d\mathbf{x}(t)}{dt} - \mathcal{M}[\mathbf{x}(t)] - \eta(t)\}\rangle\, dt \\
&+ \int_{t_0}^{t_n} \langle\, \eta^*,\, \{\frac{d\eta}{dt} - \Phi[\eta(t), \mathbf{x}(t)]\}\rangle\, dt,
\end{aligned} \tag{14}$$

where $\mathbf{x}^*, \eta^*$ are the Lagrange multiplier vectors corresponding to $\mathbf{x}, \eta$ respectively, in other words, $\mathbf{x}^*$ is the adjoint state corresponding to $\mathbf{x}$ and η^* is the adjoint state of η, and $\langle \cdot, \cdot \rangle$ denotes Euclidean inner product.

Using calculus of variations, the extrema of $\mathcal{L}$ are the solutions of the Euler-Lagrange equations (the extrema of $\mathcal{L}$ are the same as the extrema of $\mathcal{J}(\mathbf{x}_0, \eta_0)$). Using the first order optimality criteria, at the extrema of the Lagrangian, $\mathcal{L}$, the following equations are satisfied,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0, \frac{\partial \mathcal{L}}{\partial \eta} = 0, \tag{15a}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}^*} = 0, \frac{\partial \mathcal{L}}{\partial \eta^*} = 0. \tag{15b}$$

Equations (15b) yield the equations describing the evolution of model state and ME,

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{M}[\mathbf{x}(t)] - \eta(t), \frac{d\eta}{dt} = \Phi[\eta(t), \mathbf{x}(t)].$$

respectively. Equations (15a) yield the following adjoint equations which describe the evolution of the adjoint variables $\mathbf{x}^*, \eta^*$,

$$\mathbf{x}^*(t_n) = 0, \quad \eta^*(t_n) = 0, \tag{16a}$$

$$-\frac{d\mathbf{x}^*(t)}{dt} = [\frac{\partial \mathcal{M}}{\partial \mathbf{x}}]^T \mathbf{x}^* + [\frac{\partial \Phi}{\partial \mathbf{x}}]^T \eta^* + \delta(t - t_i) \sum_{i=0}^{n} [\frac{\partial H}{\partial \mathbf{x}}]^T \mathbf{R}^{-1} [H(\mathbf{x}(t_i)) - \mathbf{y}^o(t_i)], \tag{16b}$$

$$-\frac{d\eta^*(t)}{dt} = [\frac{\partial \Phi}{\partial \eta}]^T \eta^* + \mathbf{x}^*. \tag{16c}$$

Note that the evolution of $\mathbf{x}^*$ and η^* is coupled via the $\Phi[.]$ operator. Also the gradient of the cost functional, $\mathcal{J}(\mathbf{x}_0, \eta_0)$, with respect to the model state, $\mathbf{x}_0$ and ME state, η_0 is given by,

$$\nabla_{\mathbf{x}_0} \mathcal{J} = \nabla_{\mathbf{x}_0} \mathcal{J}_b + \nabla_{\mathbf{x}_0} \mathcal{J}_o = \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}^b] + \mathbf{x}^*(t_0), \tag{17a}$$

$$\nabla_{\eta_0} \mathcal{J} = \nabla_{\eta_0} \mathcal{J}_\eta + \nabla_{\eta_0} \mathcal{J}_o = \mathbf{Q}^{-1}[\eta_0 - \eta^b] + \eta^*(t_0), \tag{17b}$$

as usual, backward integration of the adjoint models (16b) and (16c) from time $t_n \rightarrow t_0$, provides us the values of initial adjoint states: $\mathbf{x}^*(t_0)$ and $\eta^*(t_0)$. Therefore the gradient in weak constraint VDA is given by $(\nabla_{\mathbf{x}_0} \mathcal{J}, \nabla_{\eta_0} \mathcal{J})^T$. Comparing this to the gradient in strong constraint VDA, which was only $\nabla_{\mathbf{x}_0} \mathcal{J}$, the size of the optimization problem is doubled. Recall that $\mathbf{T}$ was set equal to I, hence the size of the initial ME control vector, $\eta^*(t_0)$ is equal to size of the initial model state control vector, $\mathbf{x}_0$.

In strong constraint version of VDA we used a square-root formulation for the background error covariance matrix, $\mathbf{B}$, and transformed the space in which minimization was performed, such that there was no need for calculating $\mathbf{B}^{-1}$. In Appendix B we provide a similar treatment which involves both $\mathbf{B}$ and the ME covariance matrix, $\mathbf{Q}$, thereby circumventing the need to specify $\mathbf{B}^{-1}$ and $\mathbf{Q}^{-1}$.

Now we address possible approaches to model the evolution of ME, using different forms of the mapping, $\Phi[\eta(t), \mathbf{x}(t)]$, which maps the space of state variables, $\mathbf{x}$ and the

space of ME, η, on-to η only. As noted above, this mapping couples the evolution of the adjoint variables corresponding to the model states and ME variables, and it also increases the complexity involved in the backward integration of the adjoint models, Eq. (16). To the best of our knowledge, the issue of model errors in solutions of inverse problems using high resolution advection schemes has not been addressed as yet, hence to begin with, in this article we assume that $\frac{\partial \Phi}{\partial \mathbf{x}} = 0$, i.e, $\Phi[.]$ maps ME on-to itself. This assumption significantly simplifies the adjoint model equations, since the evolution of $\mathbf{x}^*$ is unchanged, and we can concentrate only the evolution of ME and its corresponding adjoint state.

The strong constraint VDA can significantly reduce the component of forecast errors due to inaccurate specification of model initial conditions. Therefore through weak constraint VDA, we aspire to further reduce the forecasting errors by reduction of errors such as those arising from discretization. We have used a range of schemes which have different dissipative and dispersive errors, the unconstrained van Leer being most dissipative among all three of the advection schemes and the PPM scheme, which is well known to be least dissipative and dispersive. Hence we expect that efficient modeling of the evolution of ME should provide further improvement of results obtained using the unconstrained van Leer scheme, for example. Note that other causes of model error such as those in limited area models due to mis-specification of boundary conditions can be tackled by following the approach used in ZZ05. Based on our experience about the evolution of forecast errors, we can say that they exhibit a trend of anywhere between linear to exponential growth. Hence we desire to model $\eta(t)$ to be an increasing function of time. Since the ME evolution is given by the following equation,

$$\eta(t_0) = \eta_0, \frac{d\eta}{dt} = \Phi[\eta(t)],$$

the rate of increase of $\eta(t)$ in time is given by the particular form of $\Phi[\eta]$. If $\Phi[\eta] < 0, \forall \eta$ then the ME decreases in time, if $\Phi[\eta] = 0, \forall \eta$ then ME is constant in time, and if $\Phi[\eta] > 0, \forall \eta$ then ME increases in time. We considered all these possibilities and investigated the following three forms of Φ,

1. Decreasing ME, $\Phi[\eta] = -\beta\eta$,
2. Constant ME, $\Phi[\eta] = 0$,
3. Exponentially increasing ME, $\Phi[\eta] = \gamma\eta$,

where β and γ are constants, and in our numerical results we specified $\beta = 0.2/\Delta t$, and $\gamma = 0.01/\Delta t$. We used the same termination criteria as in Eq.(10),

$$\|(\nabla \mathcal{J})_k\| \leq EPS \cdot MAX[1, (\|\mathbf{x}_k(t_0)\| + \|\eta_k(t_0)\|)],$$

where the gradient now includes the model error information, $\nabla \mathcal{J} = (\nabla_{\mathbf{x}_0} \mathcal{J}, \nabla_{\eta_0} \mathcal{J})^T$, and $\eta_k(t_0)$ is the optimal initial state vector at the $k^{th.}$ minimization iteration. Note that since the size of the gradient vector has doubled, the size of the Hessian of the cost functional is increased by four times. Due to the monotonicity criteria, increasing the size of the Hessian increases its condition number, which implies that a larger number of minimization iterations would have to be performed to achieve the same termination criteria (the same value of EPS was used in both strong and weak VDA for comparison sake). In fig. 5 we provide a comparison of the RMSE in geopotential height field after data assimilation using above forms of the ME for the unconstrained, constrained van Leer and PPM schemes (though not shown, the RMSE in wind fields was significantly reduced using different forms of ME). Clearly inclusion of ME

provides further reduction in forecast errors. As we said earlier, the model errors usually increase in time, therefore as expected the results obtained using the decreasing form of ME are inferior to those obtained with the other two forms. This conclusion is further substantiated by the fact that since the PPM scheme is least dissipative (compared to the van Leer schemes) the decreasing form of ME yields the same results as given by the strong constraint VDA. Though the constant form of ME is very simple, the results obtained using the van Leer schemes indicate that it yields results comparable to the increasing form. Noteworthy is the fact that in the case of PPM scheme, the increasing form is the best. To indicate that the inclusion of ME positively impacts DA via efficient preconditioning and weak constraint VDA, we provide a plot of $\mathcal{J}_o$ versus the number of cost functional evaluations using the constant ME form and the unconstrained van Leer scheme in fig. 6. Note the markedly improved fit of model states and observations using even the simplest form of constant ME. Figures 7(a)- 9(a) show the differences in the geopotential height field between model forecast and observation at forecast verification time, T_{+30} for various forms of model error using the PPM advection scheme (though not shown, the van Leer schemes yielded similar results). Note the improvement in forecast using the ME, when compared to the strong constraint VDA (i.e, with no ME), fig. 4. In figures 7(b)- 9(b) we show isolines of different initial model error state corresponding to the geopotential height field for the different forms of ME and PPM advection scheme. The decreasing form of ME is very dispersed when compared to the localized nature of the other two forms. As expected the obtained initial ME state for the constant and increasing forms with the PPM scheme is even more localized than that obtained with the van Leer schemes, indicating the lack of dissipative effect with the PPM scheme.

4. SUMMARY AND CONCLUSIONS

In this work we have shown that various high resolution advection schemes such as unconstrained, constrained van Leer and PPM provide different forecasting errors in a twin experiment framework. Two principal causes for forecast errors being the erroneous specification of model initial conditions and modeling errors. Various strong constraint VDA (which does not include model error) experimental results obtained using different advection schemes indicate that a mere change in the advection scheme alone provides more than 50% reduction in forecast RMSE. Next we studied in depth the nature of modeling errors and suggested a decreasing, constant and increasing in time forms of ME. Implementation of these forms in a weak constraint VDA framework yielded a further reduction in forecast errors. If it is a-priori known (perhaps through forecasts or numerical analysis using 1-D cases) that dissipative schemes were used for weak VDA, then even the simplest forms of ME, such as a constant in time form provides significantly better results. For highly accurate advection schemes such as the PPM scheme, the increasing form of ME is the best (when tested in the framework of a twin experiment). As such, based on the results obtained from our preliminary investigation of impact of various forms of ME in the context of weak VDA, the increasing form of model error is a good candidate for further research on this topic. To sum up, three different forms of ME using high resolution advection schemes in the presence of non-linear advection terms were studied in both strong and weak constraint VDA framework.

A discussion of related topics of future research is provided in what follows. For the sake of simplicity in the implementation of weak VDA, we selected forms of ME which were independent of the model state, i.e, $\Phi[\eta, \mathbf{x}] = \Phi[\eta]$. However at the expense

of extra computational work and developmental challenges, one could consider the tangent linear model (TLM) to be Φ, such a case was studied in KF framework by Dee and Da Silva (1998). Recall that we used the first order optimality criteria to derive the adjoint equations which provided us the means to obtain the gradient of the cost functional with respect to the control variables. But if the TLM is used for modeling the evolution of ME, i.e, $\Phi = \frac{\partial \mathcal{M}}{\partial \mathbf{x}}$ then one also needs to specify the action of $[\frac{\partial \Phi}{\partial \mathbf{x}}]^T$ and $[\frac{\partial \Phi}{\partial \eta}]^T$ on η^* (the adjoint state corresponding to model error variables) which amounts to using second order information, see Le Dimet et al. (2002). The complexity of such a second order adjoint model certainly depends on the complexity of the NWP model as well as the equation used for the evolution of ME. The second order optimality criteria provide the necessary and sufficient conditions for extrema of the cost functional, whereas first order criteria are only necessary but not sufficient. Also availability of the Hessian (of the cost functional) information, via Hessian/vector product obtained from the second order adjoint model, speeds up the minimization process, since implementation of the Newton methods is now possible.

Another topic of further research addressing improvement of the models used for modeling the evolution of ME can be derived by using the additive property in both weak and strong constraint versions of VDA (see Li and Navon 2001). We can separate the observations into a few subsets and perform VDA for each subset (see Jarvinnen et al. 1996). Same can be done with the model error, provided errors are uncorrelated in time, allowing model error to adjust within a smaller time window. Hence, this could be beneficial for better estimation of model errors. This is referred to as the property of consistent optimality by Li and Navon (2001).

TABLE 1. LIST OF ACRONYMS

Acronym	Definition
ECMWF	European Center for Medium-Range Weather Forecasts
hPa	Hectopascals
RMSE	Root-mean-squared error
UTC	Universal time coordinate
PPM	Piecewise parabolic method
T_{-06}	Data set from ERA-40 reanalysis project valid for 00 UTC 2 February 2001
$T00$	Data set obtained by 6 hour integration of T_{-06}
T_{+06}	12 hour integration of T_{-06}
T_{+12}	18 hour integration
T_{+18}	24 hour integration
T_{+24}	30 hour integration
T_{+30}	36 hour integration

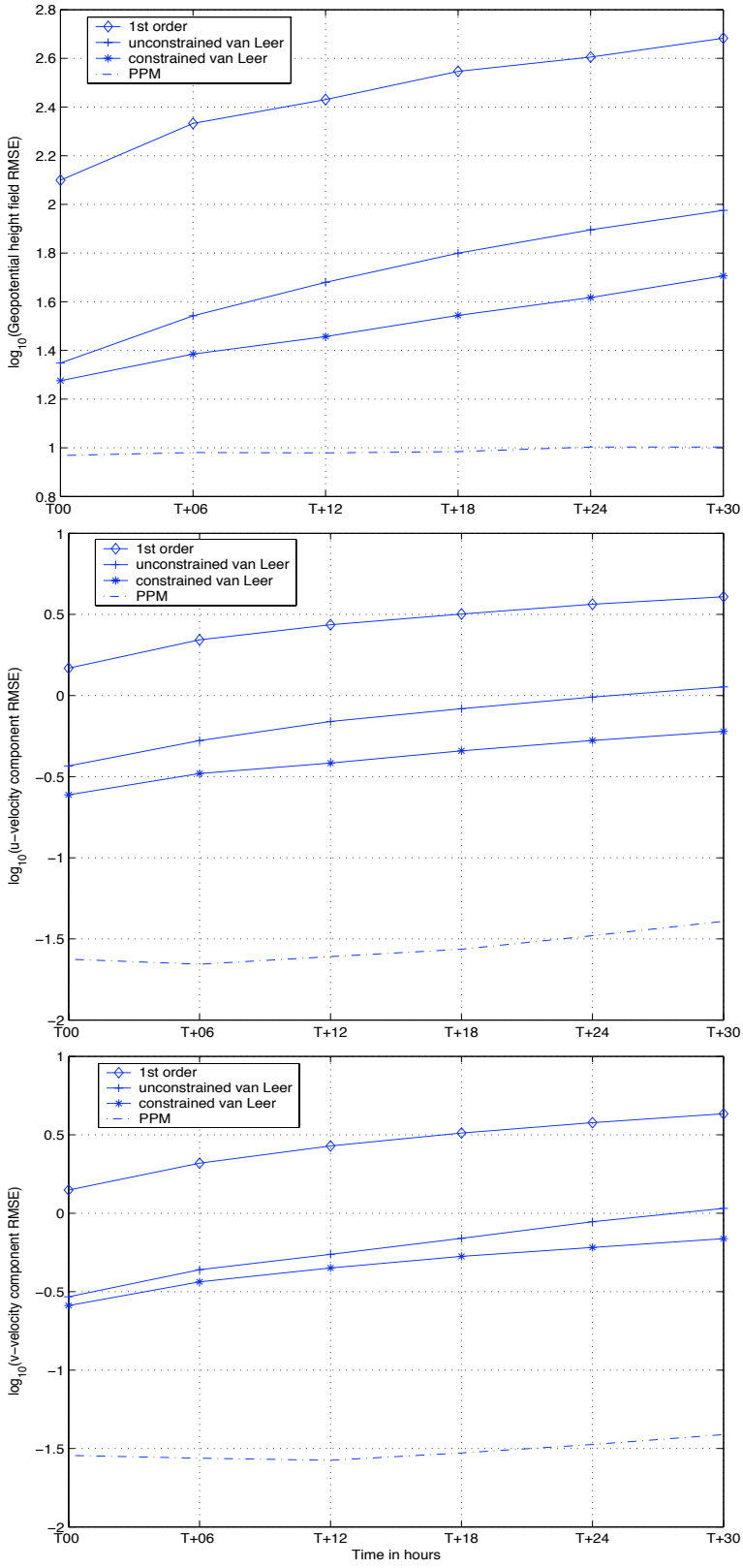

Figure 1. RMSE in the geopotential height and wind fields for different advection schemes.

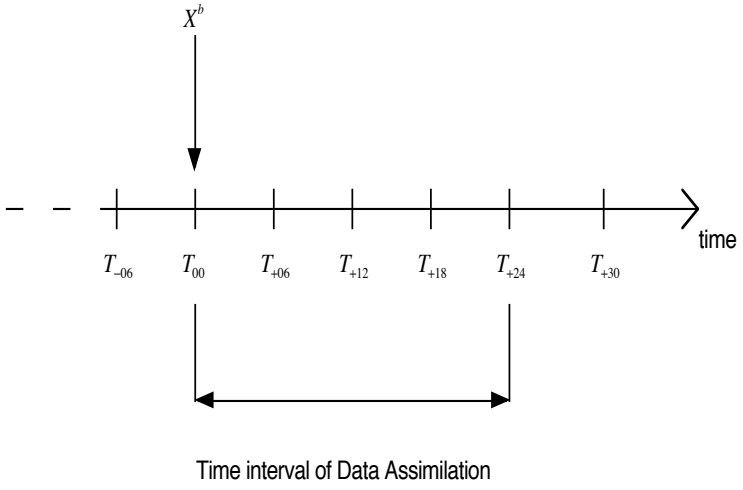

Figure 2. Illustration of data assimilation time window

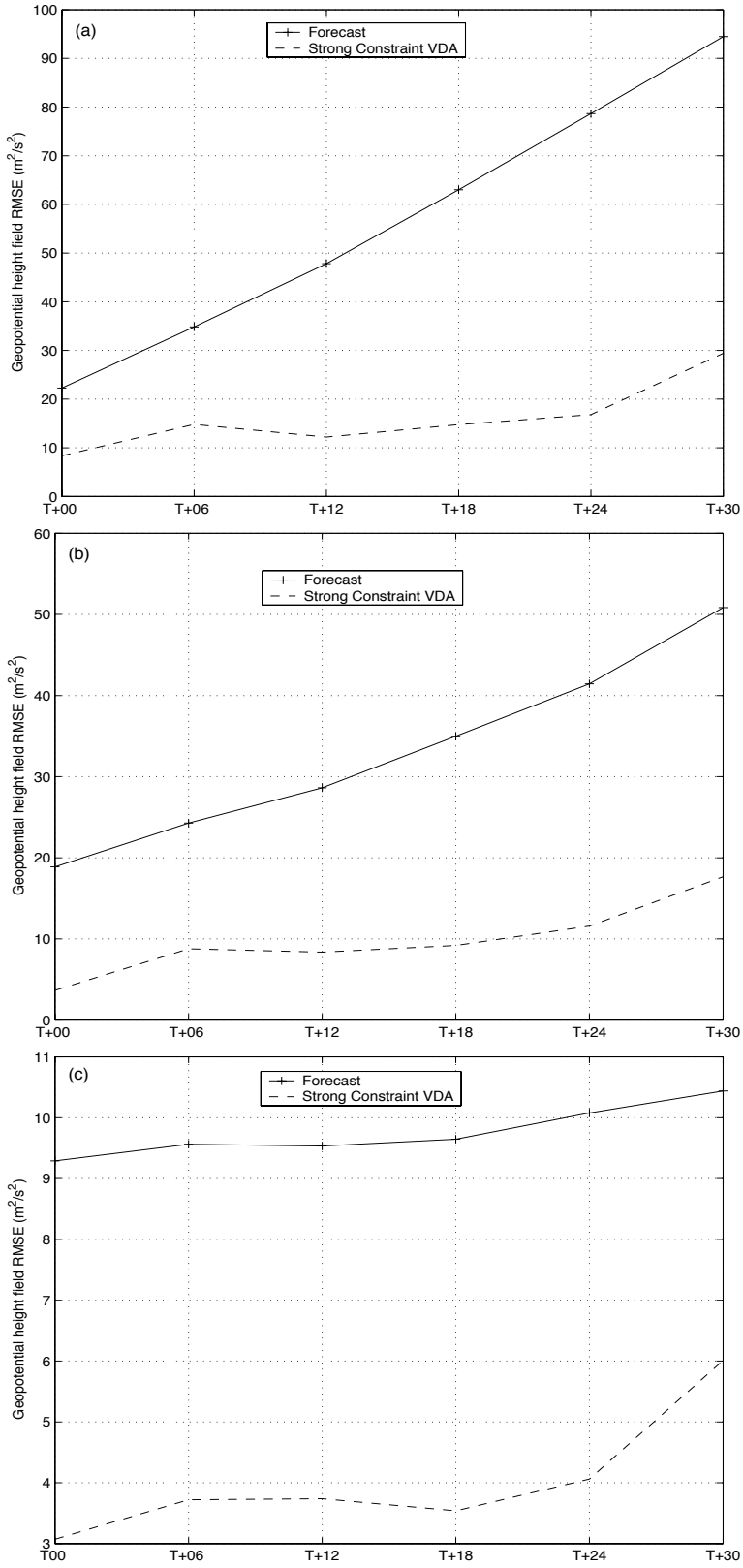

Figure 3. RMSE in the geopotential height fields for different advection schemes before and after DA in strong constraint form (a) Unconstrained Van Leer (b), Constrained van Leer (c), PPM schemes

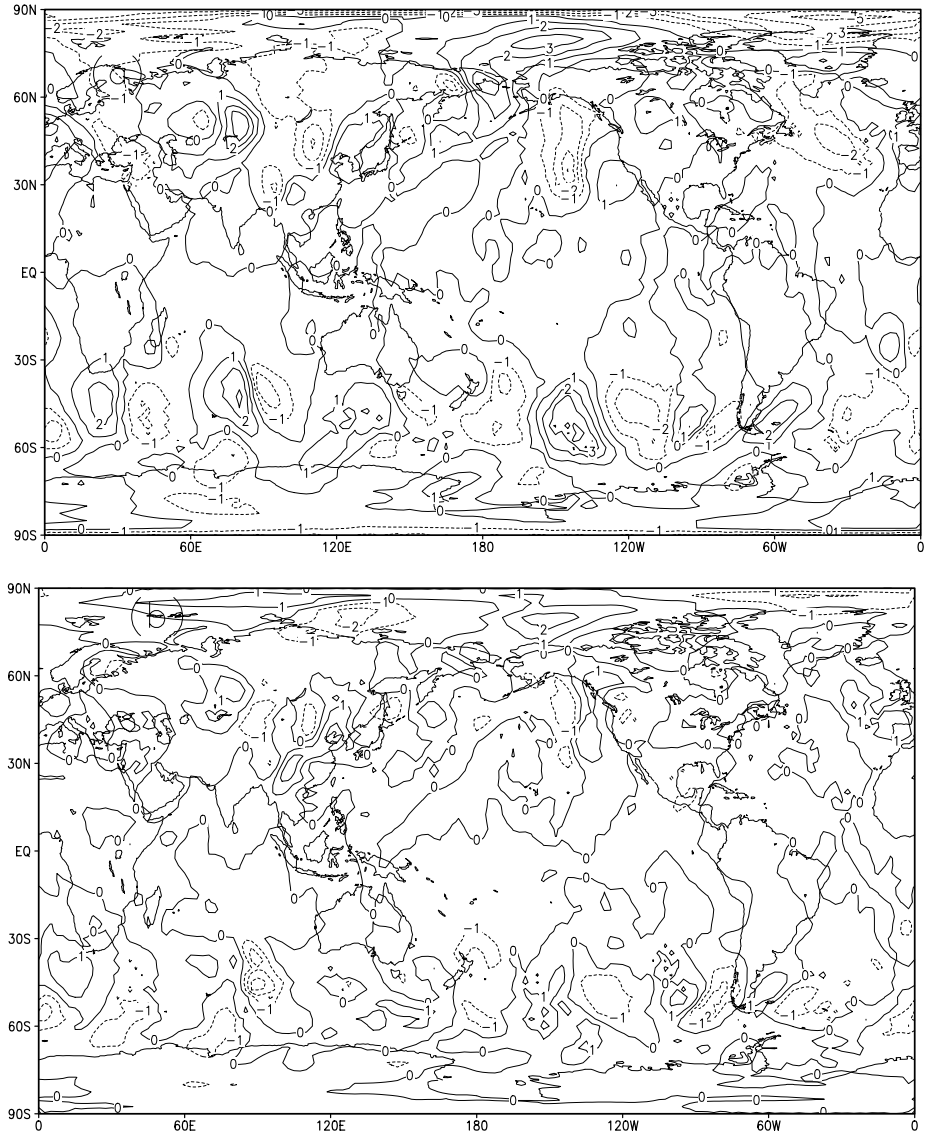

Figure 4. Isolines of differences in geopotential height field between model forecast using PPM scheme for advection and observations at T_{+30}, forecast verification time (a), Using $\mathbf{x}^b$ (b), using initial condition obtained after strong constraint data assimilation

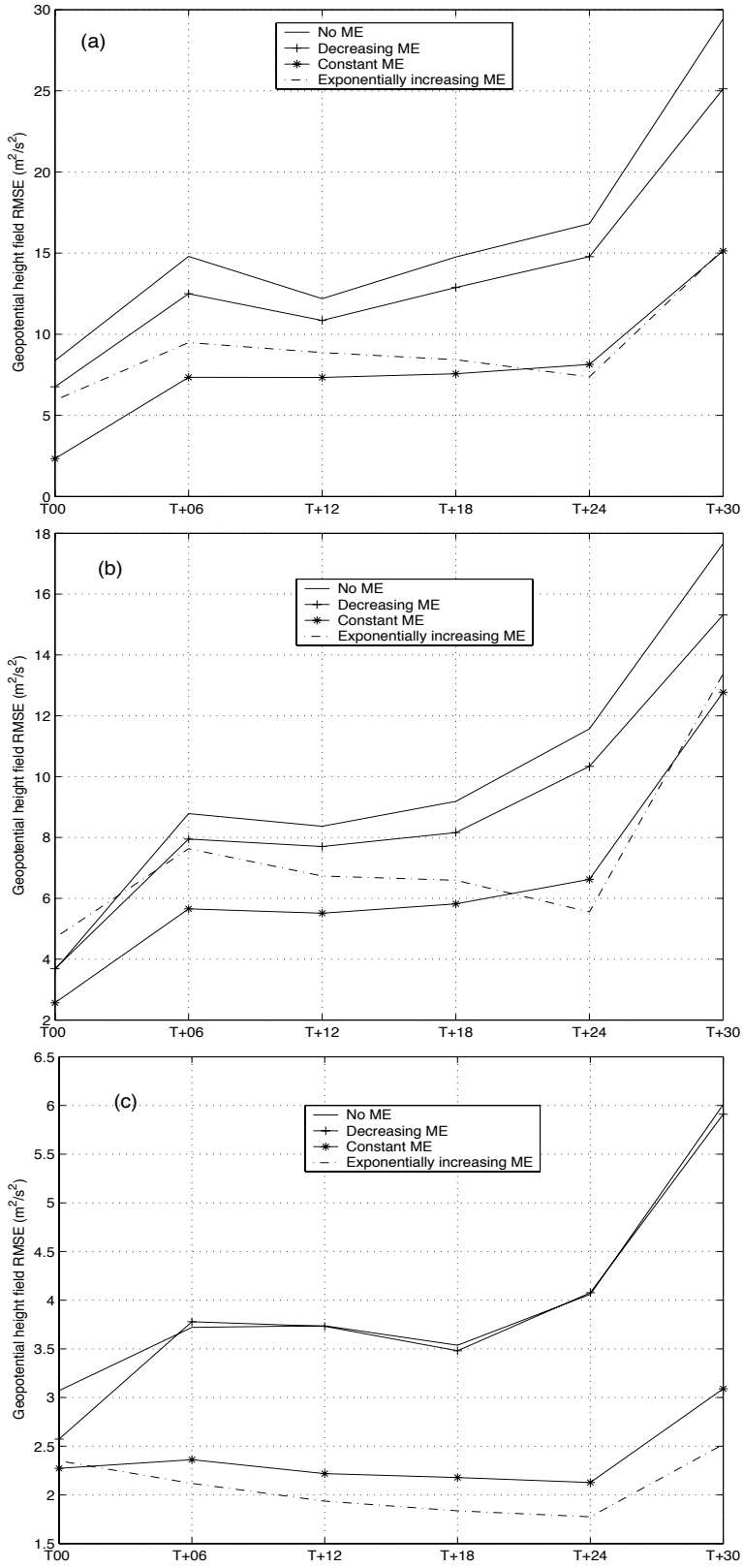

Figure 5. RMSE in the geopotential height fields for different advection schemes with different forms of model error (a) Unconstrained Van Leer (b), Constrained van Leer (c), PPM schemes

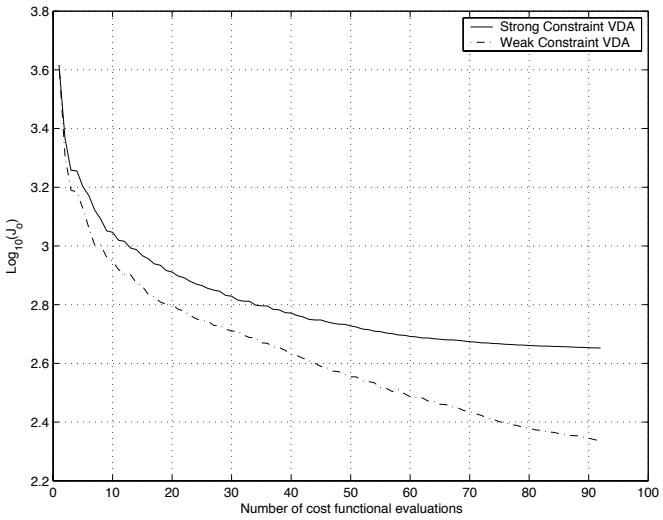

Figure 6. Observational component variation using constant ME, with unconstrained van Leer, weak and strong constraint VDA

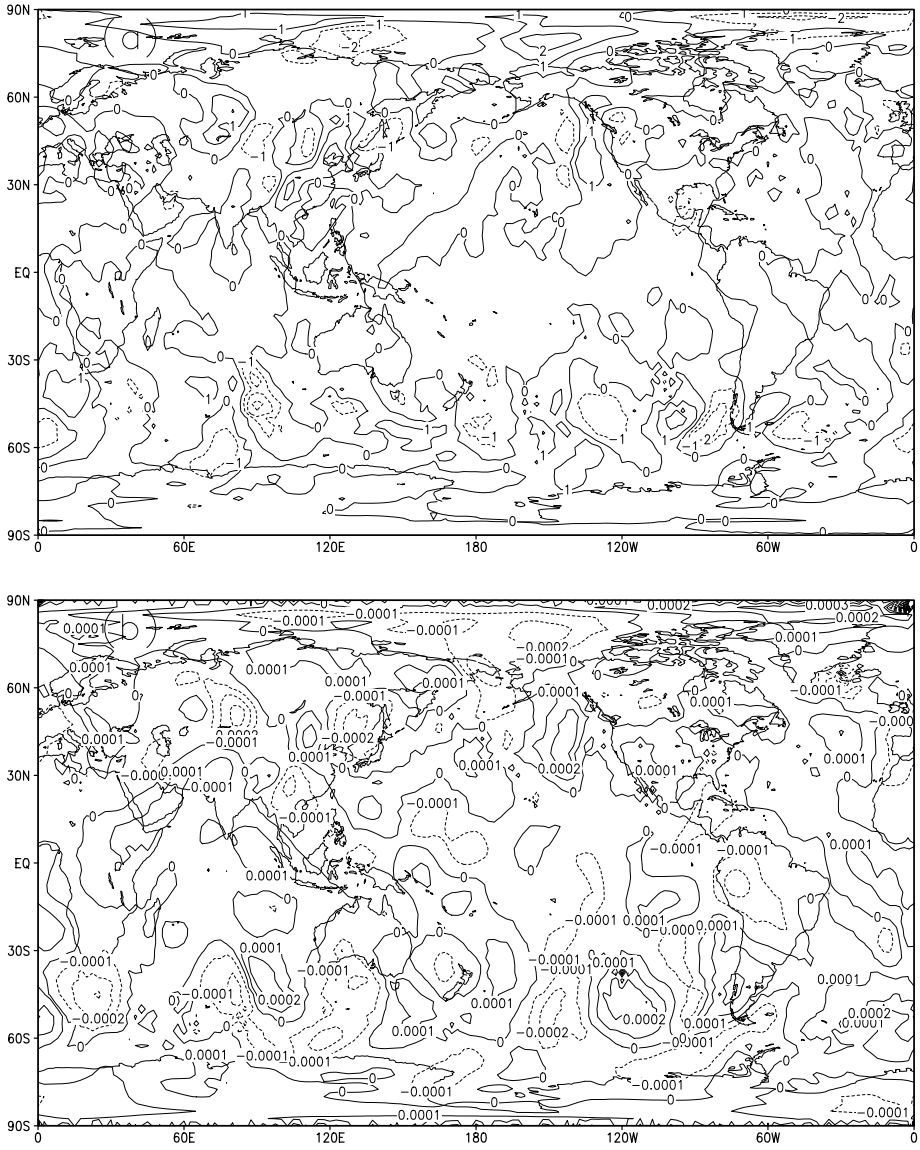

Figure 7. (a) Isolines of differences in geopotential height field between model forecast using optimized initial conditions obtained after weak constraint VDA (decreasing in time form of ME) and the PPM advection scheme, (b) initial model error state corresponding to the geopotential height field

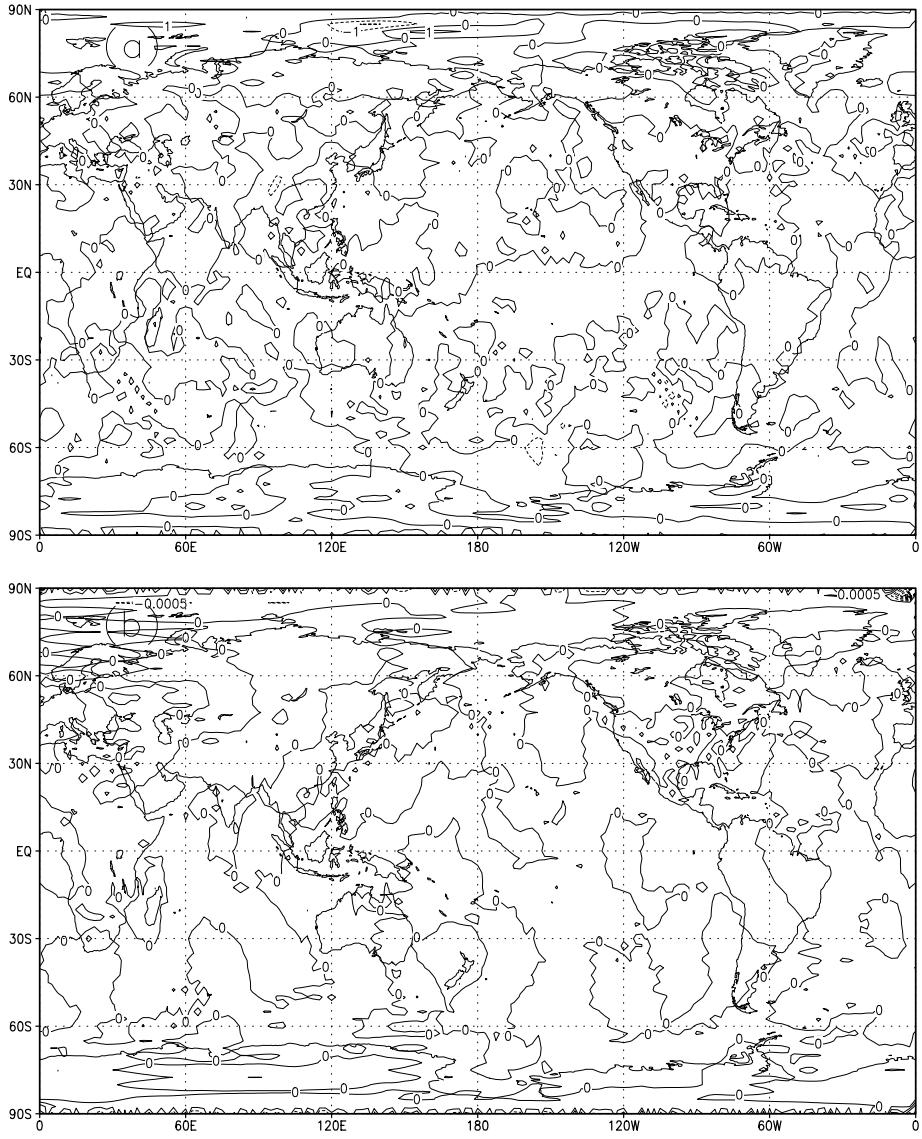

Figure 8. (a) Isolines of differences in geopotential height field between model forecast using optimized initial conditions obtained after weak constraint VDA (constant ME) and the PPM advection scheme, (b) initial model error state corresponding to the geopotential height field

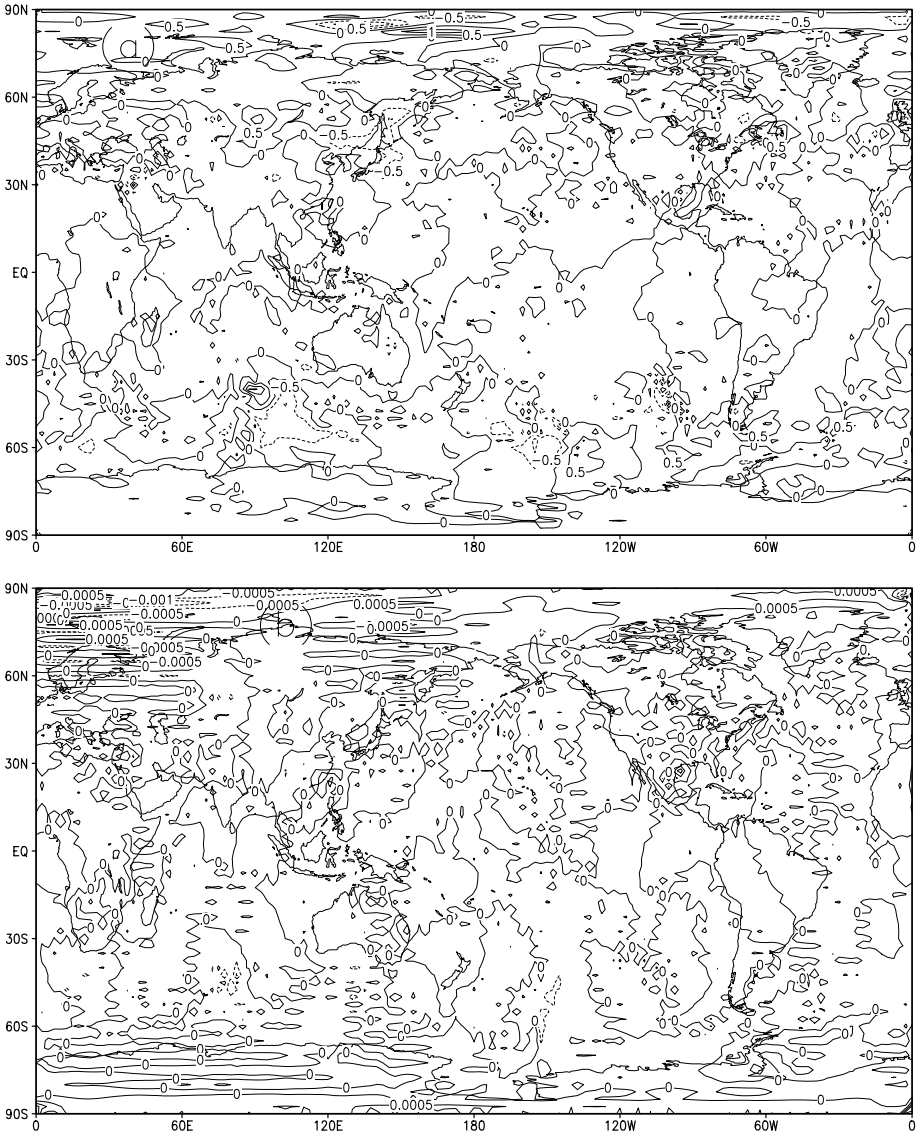

Figure 9. (a) Isolines of differences in geopotential height field between model forecast using optimized initial conditions obtained after weak constraint VDA (increasing in time form of ME) and the PPM advection scheme, (b) initial model error state corresponding to the geopotential height field

ACKNOWLEDGEMENTS

The authors acknowledge support from the NSF, ATM-9731472. We would like to thank Drs. S.-J. Lin and R. Rood for providing us their finite volume shallow water equations model and the ECMWF for providing us the ERA-40 data, which was obtained from the ECMWF Data Server.

APPENDIX A

Description of the background error covariance operator

Following Eq. (8) the strong constraint VDA cost functional is given by,

$$\mathcal{J} = \mathcal{J}_b + \mathcal{J}_o.$$

A static-in-time $\mathbf{B}$ is constructed in the grid point space as an operator, which is based on the formulation provided by Weaver and Courtier (2001); Derber and Bouttier (1999).

Let $\delta\mathbf{x} = \mathbf{x}(t_0) - \mathbf{x}^b$, and define a transformation, $\mathbf{v} = \mathbf{B}^{-1/2}\,\delta\mathbf{x}$, which implies that $\delta\mathbf{x} = \mathbf{B}^{1/2}\mathbf{v}$. Where the $\mathbf{B}^{1/2}$ is taken to be any square-root matrix, such that $\mathbf{B} = \mathbf{B}^{1/2}\,\mathbf{B}^{T/2}$; $\mathbf{B}^{T/2}$ denotes the transpose of $\mathbf{B}^{1/2}$. Therefore the background cost functional can be rewritten as,

$$\mathcal{J}_b = \frac{1}{2}\delta\mathbf{x}^T\,\mathbf{B}^{-1}\,\delta\mathbf{x} = \frac{1}{2}\delta\mathbf{x}^T\,\left(\mathbf{B}^{1/2}\,\mathbf{B}^{T/2}\right)^{-1}\,\delta\mathbf{x} = \frac{1}{2}\mathbf{v}^T\,\mathbf{v}.$$

Hence the contribution to the gradient of the cost functional, $\mathcal{J}$ from the background cost functional is equal to $\nabla_{\mathbf{v}}\,\mathcal{J}_b = \mathbf{v}$, and to the Hessian of the cost functional, $\nabla^2{}_{\mathbf{v}}\,\mathcal{J}_b = I$. At the beginning of the minimization, $\mathbf{v} = \delta\mathbf{x} = 0$, such that the initial guess for $\mathbf{x}(t_0)$ is $\mathbf{x}^b$. This transformation of variables preconditions the minimization problem for faster convergence of the minimization algorithm. An ideal preconditioning is obtained if the Hessian matrix is an identity matrix. A good approximation to this is to ensure that the Hessian of $\mathcal{J}_b$ is equal to I, which is indeed the case here, since the minimization is performed in the $\mathbf{v}$ space. To summarize,

$$\mathcal{J} = \mathcal{J}_b + \mathcal{J}_o = \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^b]^T\mathbf{B}^{-1}[\mathbf{x}(t_0) - \mathbf{x}^b] + \mathcal{J}_o = \frac{1}{2}\mathbf{v}^T\,\mathbf{v} + \mathcal{J}_o,$$

where $\delta\mathbf{x} = \mathbf{x}(t_0) - \mathbf{x}^b$, and $\mathbf{v} = \mathbf{B}^{-1/2}\,\delta\mathbf{x}$, which implies $\delta\mathbf{x} = \mathbf{B}^{1/2}\mathbf{v}$. Therefore gradient of the cost functional with respect to $\mathbf{v}$ is given by,

$$\nabla_{\mathbf{v}}\,\mathcal{J} = \mathbf{v} + \nabla_{\mathbf{v}}\,\mathcal{J}_o = \mathbf{v} + \mathbf{B}^{T/2}\,\nabla_{\mathbf{x}_0}\,\mathcal{J}_o.$$

Thus every minimization iteration requires application of $\mathbf{B}^{1/2}$ to obtain the analysis increment $\delta\mathbf{x}$ from $\mathbf{v}$ and $\mathbf{B}^{T/2}$ to get the gradient $\nabla_{\mathbf{v}}\,\mathcal{J}_o$ from $\nabla_{\mathbf{x}_0}\,\mathcal{J}_o$ (which is computed by a single integration of the adjoint model backward in time). As evident, we do not require inverse of $\mathbf{B}$ in the above formulation.

The model variables $(\mathbf{h}, \mathbf{u}, \mathbf{v})$ are partitioned into balanced and unbalanced components. The so-called balancing operator, $\mathbf{K}_b$ acts on the unbalanced components of the model variables and in-turn, $\mathbf{K}_b = \mathbf{K'}_b + I$. Following Vidard et al. (2004), $\mathbf{K'}_b$ is formulated using the linear balance equations, based on geostrophic balance (written in spherical coordinates) and hydrostatic hypothesis.

Geostrophic balance:

$$\begin{aligned} u &= -\frac{1}{\rho\,f}\,[\frac{1}{a}\frac{\partial p}{\partial\theta}], \\ v &= \frac{1}{\rho\,f}\,[\frac{1}{a\,\cos\theta}\frac{\partial p}{\partial\lambda}]. \end{aligned}$$

Hydrostatic hypothesis: $p = \rho\,g\,h$.

Which implies,

$$\begin{aligned} u &= -\frac{g}{f}\,[\frac{1}{a}\frac{\partial h}{\partial \theta}], \\ v &= \frac{g}{f}\,[\frac{1}{a\,\cos\theta}\frac{\partial h}{\partial \lambda}]. \end{aligned}$$

Therefore

$$\mathbf{K}_b = \mathbf{K'}_b + I = \begin{pmatrix} I & 0 & 0 \\ -\frac{g}{a\,f}\frac{\partial}{\partial \theta} & I & 0 \\ \frac{g}{a\,f\,\cos\theta}\frac{\partial}{\partial \lambda} & 0 & I \end{pmatrix}$$

which is a lower triangular matrix, since our control vector is of the form $(\mathbf{h},\ \mathbf{u},\ \mathbf{v})^T$.

Remark: At the North and South poles, one sided differences have been used for computing the above derivative with respect to the latitude and at the equator, where $\theta = \pi/2$, we have used the average values of the derivative (with respect to the longitude) from the two neighboring latitude circles, above and below the equator.

Using the balance operator, we can write $\mathbf{B} = \mathbf{K}_b\,\mathbf{B}_u\,\mathbf{K}_b^T$, where $\mathbf{B}_u$ is a block diagonal error covariance matrix for the unbalanced component of the variables (see Weaver and Courtier (2001)), which implies that the cross-covariances between the unbalanced variables is taken to be negligible. Thus $\mathbf{B}_u = \mathbf{\Sigma_b}\,\mathbf{C}\,\mathbf{\Sigma_b}$, where $\mathbf{\Sigma_b}$ is a block-diagonal matrix of the background-error variances in the grid point space, such that the diagonal entries represent error variances at every grid point (in this work, we prescribed $\Sigma_b = [2000\ I, 100\ I, 100\ I]$).

$\mathbf{C}$ is a symmetric matrix of background-error correlations for the unbalanced component of the variables. Assuming that $\mathbf{C}$ is block-diagonal, which is a valid assumption, since $\mathbf{B}_u$ has already been assumed to be block-diagonal, we obtain the square-root factorization $\mathbf{C} = \mathbf{C}^{1/2}\,\mathbf{C}^{T/2}$.

Thus the square-root factorization of the background error covariance can be written as,

$$\begin{aligned} \mathbf{B} &= \mathbf{K}_b\,\mathbf{B}_u\,\mathbf{K}_b^T = \mathbf{K}_b\,(\mathbf{\Sigma_b}\,\mathbf{C}\,\mathbf{\Sigma_b})\,\mathbf{K_b^T} = \mathbf{K_b}\,(\mathbf{\Sigma_b}\,\mathbf{C^{1/2}}\,\mathbf{C^{T/2}}\mathbf{\Sigma_b})\,\mathbf{K_b^T} \quad \text{(A.1)} \\ &= (\mathbf{K}_b\,\mathbf{\Sigma_b}\,\mathbf{C^{1/2}})\,(\mathbf{C^{T/2}}\mathbf{\Sigma_b}\,\mathbf{K_b^T}) \\ &= \mathbf{B}^{1/2}\,\mathbf{B}^{T/2}. \end{aligned}$$

Notice that the above formulation ensures that $\mathbf{B}$ is symmetric and positive definite, both of these properties are usually required to be satisfied by any preconditioning matrix. The analysis increment is given by $\delta\mathbf{x} = \mathbf{B}^{1/2}\mathbf{v} = \mathbf{K}_b\,\mathbf{\Sigma_b}\,\mathbf{C^{1/2}}\mathbf{v}$. Since $\mathbf{C}$ is block-diagonal, the operation $\mathbf{C}^{1/2}\,\mathbf{v}$ can be split into individual operators $\mathbf{C}_\alpha^{1/2}\,\mathbf{v}_\alpha$, that act independently on different components of the variable $\mathbf{v}$, such as $\mathbf{v}_\alpha$. For each variable, the univariate operator can be factorized into $\mathbf{C}_\alpha = \mathbf{C}_\alpha^{1/2}\,\mathbf{C}_\alpha^{T/2}$. The procedure suggested by Weaver and Courtier (2001) has been implemented to model the univariate correlation operator, has been implemented to model the univariate correlation operator, $\mathbf{C}_\alpha$ as an isotropic diffusion operator, assuming Gaussianity with a decorrelation length equal to $500\ km$.

We considered height field which was comprised of a single Dirac delta pulse located at equator and longitude 180^o, and prescribed no wind field, the action of $\mathbf{B}$ on such a field is shown in Fig. A.1 (a). We see the effect of the correlation operator on the

Dirac pulse and also on the wind field obtained under geostrophic balance assumption (Fig. A.1 (b)), which is parallel to the isobars of the pressure. Since there is a *high pressure* at the center, the direction of the wind is clockwise in the Northern hemisphere and anti-clockwise in the Southern hemisphere; at the equator due to the balancing of the pressure gradient and Coriolis forces, the wind blows straight.

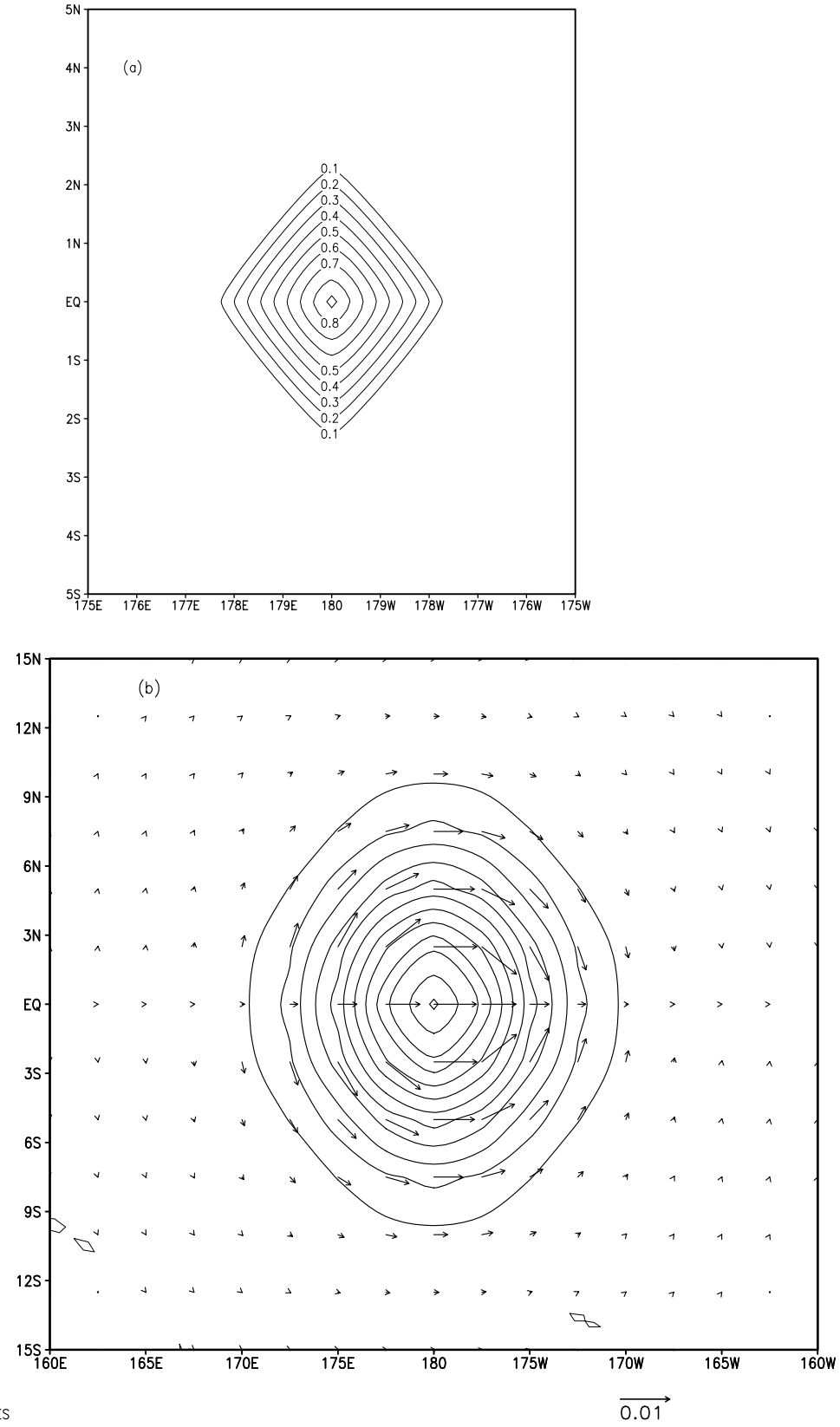

Figure A.1. Result obtained by operating with **B** on a single Dirac delta pulse in the height field (a), isolines of the height field (b), geostrophic wind plotted along with the isolines of the height field.

APPENDIX B

*Square-root transformation used for **B** and **Q***

We have extended the procedure described in the previous Appendix A to include model error control vector, thereby we achieve all the preconditioning properties for the minimization process we discussed earlier (this procedure has also been described by VPLD04). This approach circumvents the need to specify inverses of the background and model error covariance matrices via a transformation which is similar to the one described earlier for the strong constraint VDA which included only $\mathcal{J}_b$ and $\mathcal{J}_o$ in the cost functional. Let

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \eta \end{bmatrix},$$

such that

$$\delta\mathbf{z} = \begin{bmatrix} \mathbf{x}_0 - \mathbf{x}^b \\ \eta_0 - \eta^b \end{bmatrix}.$$

Recall that the sum of background and model error cost functionals ($\mathcal{J}_b$ and $\mathcal{J}_\eta$, respectively) is given by,

$$\begin{aligned} \mathcal{J}_b + \mathcal{J}_\eta &= \frac{1}{2}[\mathbf{x}_0 - \mathbf{x}^b]^T \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b] + [\eta_0 - \eta^b]^T \mathbf{Q}^{-1} [\eta_0 - \eta^b] \\ &= \frac{1}{2}\delta\mathbf{z}^T \begin{bmatrix} \mathbf{B}^{-1} & 0 \\ 0 & \mathbf{Q}^{-1} \end{bmatrix} \delta\mathbf{z} \end{aligned} \tag{B.1}$$

As in appendix A, let $\mathbf{B}^{1/2}$ and $\mathbf{Q}^{1/2}$ be any square-root matrices such that $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$, $\mathbf{Q} = \mathbf{Q}^{1/2}\mathbf{Q}^{T/2}$, and let

$$\begin{aligned} \mathbf{w} &= \begin{bmatrix} \mathbf{B}^{-1/2} & 0 \\ 0 & \mathbf{Q}^{-1/2} \end{bmatrix} \delta\mathbf{z} \\ \Rightarrow \delta\mathbf{z} &= \begin{bmatrix} \mathbf{B}^{1/2} & 0 \\ 0 & \mathbf{Q}^{1/2} \end{bmatrix} \mathbf{w}. \end{aligned} \tag{B.2}$$

The above transformation from $(\mathbf{x}, \eta) \rightarrow \mathbf{w}$ is similar to the previously described transformation: $\mathbf{x} \rightarrow \mathbf{v}$, which involved $\mathbf{B}^{1/2}$ only. Using the above equations (B.1) and (B.2),

$$\mathcal{J}_b + \mathcal{J}_\eta = \frac{1}{2}\mathbf{w}^T\mathbf{w},$$

hence the entire cost functional,

$$\begin{aligned} \mathcal{J} &= \mathcal{J}_b + \mathcal{J}_\eta + \mathcal{J}_o \\ &= \frac{1}{2}\mathbf{w}^T\mathbf{w} + \mathcal{J}_o, \end{aligned} \tag{B.3}$$

and the gradient of the cost functional with respect to the transformed variable, $\mathbf{w}$ is given by

$$\begin{aligned} \nabla_{\mathbf{w}}\,\mathcal{J} &= \mathbf{w} + \nabla_{\mathbf{w}}\,\mathcal{J}_o \\ &= \mathbf{w} + \begin{bmatrix} \mathbf{B}^{T/2} & 0 \\ 0 & \mathbf{Q}^{T/2} \end{bmatrix} \begin{pmatrix} \nabla_{\mathbf{x}_0}\,\mathcal{J}_o \\ \nabla_{\eta_0}\,\mathcal{J}_o \end{pmatrix}. \end{aligned} \tag{B.4}$$

We have modeled $\mathbf{Q}$ as a block diagonal matrix,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{hh} & 0 & 0 \\ 0 & \mathbf{Q}_{uu} & 0 \\ 0 & 0 & \mathbf{Q}_{vv} \end{bmatrix},$$

such that each of the blocks is an univariate Gaussian correlation operator and has a square-root decomposition given by,

$$\mathbf{Q}_{\alpha\alpha} = \mathbf{\Sigma_q C}_\alpha \mathbf{\Sigma_q} = \mathbf{\Sigma_q C}_\alpha{}^{\mathbf{1/2}} \mathbf{C}_\alpha{}^{\mathbf{T/2}} \mathbf{\Sigma_q} = \left(\mathbf{\Sigma_q C}_\alpha{}^{\mathbf{1/2}}\right) \left(\mathbf{\Sigma_q C}_\alpha{}^{\mathbf{1/2}}\right)^{\mathbf{T}} = \mathbf{Q}_{\alpha\alpha}^{\mathbf{1/2}} \mathbf{Q}_{\alpha\alpha}^{\mathbf{T/2}},$$

where $\alpha = \mathbf{h}, \mathbf{u}, \mathbf{v}$; $\mathbf{\Sigma_q}$ is a diagonal matrix of variances (we prescribed $\mathbf{\Sigma_q} = \mathbf{10^{-2}\Sigma_b}$) and $\mathbf{C}_\alpha$ is an isotropic diffusion operator, construction of which was described in the previous appendix. The most simplistic model error covariance matrix is a diagonal matrix, which implies that the analyzed model error increment at any specific grid point does have any influence on the increments at the neighboring grid points. Alternatively a diffusion operator provides such an increment in a localized region (given by the length scale of diffusion) inexpensively. We have not used the balance operator, $\mathbf{K}_b$ in the above construction of $\mathbf{Q}$. The goestrophic balance and hydrostatic hypothesis which were used in the construction of $\mathbf{K}_b$ are not required for $\mathbf{Q}$ since we do not have the same information about model error covariances as we have for background error covariances; therefore usage of $\mathbf{K}_b$ in specification of $\mathbf{Q}$ will only involve extra computational work.

At the beginning of minimization, the initial guess for $\mathbf{x}_0 = \mathbf{x}^b$ and $\eta_0 = \eta^b$, therefore $\mathbf{w} = \delta\mathbf{z} = 0$. Every minimization iteration (carried out in $\mathbf{w}$ space) requires application of $\mathbf{B}^{1/2}$, $\mathbf{Q}^{1/2}$ to obtain the analysis increment $\delta\mathbf{z}$ from $\mathbf{w}$ (Eq. (B.2)) and the adjoint operators, $\mathbf{B}^{T/2}$, $\mathbf{Q}^{T/2}$ to get the gradient $\nabla_{\mathbf{w}}\mathcal{J}_o$ from $\left(\nabla_{\mathbf{x}_0}\mathcal{J}_o, \nabla_{\eta_0}\mathcal{J}_o\right)^T$ (Eq. (B.4)).

REFERENCES

Akella, S. and Navon, I. M. 2005 A comparative study of the performance of high resolution advection schemes in the context of data assimilation, In Press with *International Journal for Numerical Methods in Fluids* URL [http://www.csit.fsu/~navon/pubs/Santha_navon.pdf]

Bennett, A. 1992 *Inverse methods in physical oceanography.* Cambridge University Press.

Bennett, A., Chua, B. and Leslie, L. 1996 Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.*, **60,** 165–178

1997 Generalized inversion of a global numerical weather prediction model (II). *Meteorol. Atmos. Phys.*, **62,** 129–140

Bennett, A., Leslie, L., Hagelberg, C. and Powers P. 1993 Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Weather Rev.*, **121,** 1714–1729

Bloom, S. C. and Shubert, S. 1990 'The influence of Monte-Carlo estimates of model error growth on the GLA OI assimilation system'. Pp. 467–470 in International symposium on assimilation of observations in meteorology and oceanography, WMO, Geneva, Switzerland

Boer, G. J. 1984 A spectral analysis of predictability and error in an operational forecast system. *Mon. Weather Rev.*, **112,** 1183–1197

Chepurin, G., Carton, J. and Dee, D. 2005 Forecast model bias correction in ocean data assimilation. *Mon. Weather Rev.*, **133,** 1328–1342

Dalcher, A. and Kalnay, E. 1987 Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39 A,** 474–491

Daley, R. 1992a The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Weather Rev.*, **120,** 164–177

1992b Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Weather Rev.*, **120,** 1735–1746

1993 Estimating observation error statistics for atmospheric data assimilation. *Annales Geophysicae-Atmospheres Hydrospheres and Space Sciences.* **11,** 634–647

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

Dee, D. and Da Silva, A. 1998 Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.*, **124,** 269–295

1999 Maximum-likelihood estimation of forecast and observational error covariance parameters. Part 1: Methodology. *Mon. Weather Rev.*, **127,** 1822–1834

Dee, D., Gaspari, G., Redder, C., Rukhovets, L. and Da Silva, A. 1999 Maximum-likelihood estimation of forecast and observational error covariance parameters. Part 2: Applications. *Mon. Weather Rev.*, **127,** 1835–1849

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

Derber, J. C. 1989 A variational continuous assimilation technique. *Mon. Weather Rev.*, **117,** 2437–2446

Derber, J. C. and Bouttier, F. 1999 A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51 A,** 195–221

ECMWF 2002 ECMWF Re-Analysis Project, ERA-40. URL [http://www.ecmwf.int/research/era/]

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

Griffith, A. K. and Nichols, N. K. 1996 'Accounting for model error in data assimilation using adjoint methods'. Pp 195–205 in *Computational Differentiation: Techniques, Applications, and Tools.* Ed. M. Berz, C. Bischof, G. Corliss, and A. Greiwank, SIAM, Philadelphia

2000 Adjoint methods in data assimilation for estimating model error. *Flow, Turbulence and Combustion* **65,** 469–488

Haltiner, G. J. and Williams, R. T. 1980 *Numerical Prediction and Dynamic Meteorology.* John Wiley & Sons, New York, 2nd. edition.

Jarvinnen, H., Thépaut, J.-N. and Courtier, P. 1996 Quasi-continuous variational data assimilation. *Q. J. R. Meteorol. Soc.*, **122,** 515–534

Kalnay, E. 2003 *Atmospheric modeling, data assimilation and predictability.* Cambridge University Press.

Le Dimet, F.-X., Navon, I. M. and Daescu, D. N. 2002 Second-order information in data assimilation *Mon. Weather Rev.*, **130,** 629–648

Le Dimet, F.-X. and Shutyaev, V. 2005 On deterministic error analysis in variational data assimilation. *Nonlinear Processes in Geophysics* **12,** 481–490

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

Li, Z. and Navon, I. M. 2001 Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.*, **127,** 661–683

Lin, S.-J., Chao, W. C., Sud, Y. C., and Walker, G. K. 1994 A class of van Leer transport schemes and its applications to the moisture transport in a general circulation model. *Mon. Weather Rev.*, **122,** 1575–1593

Lin, S.-J. and Rood, R. 1997 An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Q. J. R. Meteorol. Soc.*, **123,** 2477–2498

Liu, D. C. and Nocedal, J. 1989 On the limited memory BFGS method for large scale minimization. *Mathematical Programming* **45,** 503–528

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

Martin, J., Bell, M., and Nichols, N. 2002 Estimation of systematic error in an equatorial ocean model using data assimilation. *International Journal for Numerical Methods in Fluids* **40,** 435–444

Navon, I. M., Daescu, D. N., and Liu, Z. 2005 'The impact of background error on incomplete observations for 4d-var data assimilation with the FSU GSM'. Pp 837–844 in *Computational Science- ICCS 2005, Part 2, Lecture Notes in Computer Science 3515.* Ed. V. S. Sunderam, Springer-Verlag Berlin

Navon, I. M., Zou, X.,Derber, J. and Sela, J. 1992 Variational data assimilation with an adiabatic NMC spectral model. *Mon. Weather Rev.*, **120,** 1433–1446

Nash, S. G. and Nocedal, J. 1991 A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization. *SIAM Journal on Optimization* **1,** 358–372

Riishøjgaard, L. P. 1998 A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50 A,** 42–57

Saha, S. 1992 Response of the NMC MRF model to systematic error correction within the integration. *Mon. Weather Rev.*, **120,** 345–360

Suarez, M. J. and Takacs, L. L. 1995 Documentation of the ARIES/GEOS dynamical core: version 2. *Technical Memorandum* 104606, volume 5, NASA

Thiébaux, H. J. and Morone, L. L. 1990 Short-term systematic errors in global forecasts: their estimation and removal. *Tellus*, **42 A,** 209–229

Uboldi, F. and Kamachi, M. 2000 Time-space weak-constraint data assimilation for nonlinear models. *Tellus*, **52 A,** 412–421

Vidard, P. A., Piacentini, A. and Le Dimet, F.-X. 2004 Variational data analysis with control of the forecast bias *Tellus*, **56 A,** 177–188

Vidard, P. A.,Blayo, E., Le Dimet, F.-X. and Piacentini, A. 2000 4D variational data analysis with imperfect model. *Flow, Turbulence and Combustion* **65,** 489–504

Weaver, A. and Courtier, P. 2001 Correlation modeling on the sphere using a generalized diffusion equation *Q. J. R. Meteorol. Soc.*, **127,** 1815–1846

Wergen, W. 1992 The effect of model errors in variational assimilation. *Tellus*, **44 A,** 297–313

Williamson, D. L., Drake, J. B., Hack, J. J., Jakob, R. and Swarztrauber, P. N. 1992 A standard test set for numerical approximations to the shallow water equations on the sphere. *J. Comput. Phys.*, **102,** 211–224

Zhu, J. and Kamachi, M. 2000 An adaptive variational method for data assimilation with imperfect models. *Tellus*, **52 A,** 265–279

Zou, X., Navon, I. M., Berger, M., Phua, M. K., Schlick, T. and LeDimet, F.-X. 1993 Numerical experience with limited-memory, quasi-newton methods for large-scale unconstrained nonlinear minimization. *SIAM Journal on Optimization* **3,** 582–608

Zupanski, D. 1997 A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Weather Rev.*, **125,** 2274–2292

Zupanski, D. and Zupanski, M. 2002 Fine-resolution 4D-VAR data assimilation for the great plains tornado outbreak of 3 May 1999. *Weather and Forecasting* **121,** 506–525

Zupanski, M. 1993 Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Weather Rev.*, **121,** 2396–2408

2005 Maximum-likelihood ensemble filter: theoretical aspects. *Mon. Weather Rev.*, **133,** 1710–1726

Zupanski, M., Zupanski, D., Vukicevic, T., Eis, K. and Haar, T. 2005 CIRA/CSU Four- Dimensional variational data assimilation system. *Mon. Weather Rev.*, **123,** 829–843