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

Tellus (2009), 61A, 112–128

TELLUS

Different approaches to model error formulation in 4D-Var: a study with high-resolution advection schemes

By S. AKELLA[1,2] and I. M. NAVON[3]*, [1]*Department of Earth & Planetary Sciences, The Johns Hopkins University, Baltimore, MD 21218, USA;* [2]*Department of Applied Mathematics & Statistics, The Johns Hopkins University, Baltimore, MD 21218, USA;* [3]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, USA;*

(Manuscript received 3 September 2007; in final form 14 July 2008)

ABSTRACT

All numerical models are imperfect. Weak constraint variational data assimilation (VDA), which provides a treatment of the modelling errors, is studied; building on the approach of Vidard et al. (Tellus, **56**A, pp. 177–188, 2004). The evolution of model error (ME) is modelled using ordinary differential equations, which involve a scalar parameter. These approaches were tested using different high-resolution advection schemes. The first set of experiments were constructed to see if it is possible to account for (numerical) discretization error within such a framework. In other set of experiments, a systematic source of modelling error was introduced by deliberately specifying an incorrect value for the Coriolis parameter in the model. Results with observational state at half of the model state resolution, are also presented. We also discuss a method of estimating the scalar parameter in the ME through VDA. In all cases, the inclusion of ME provides reduction in forecasting errors. Also, our experiments indicate that different settings of the model (e.g. using different high-resolution advection schemes) would need different ME formulation. Results presented in this paper could be used to formulate sophisticated ME forms to account for systematic errors in higher dimensional models with complex advection schemes.

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 adjusting the initial conditions supplied for forward model integration (LeDimet and Talagrand, 1986; Navon et al., 1992). In the so-called 'strong constraint' or 'classical' 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 subgrid processes are not included. In addition, most of the physical processes and their interactions in the atmosphere are parametrized. Also, a complete mathematical modelling of the boundary conditions and forcing terms can never be achieved. Usually all of these modelling drawbacks are collectively addressed by the term, 'model error' (ME).

*Corresponding author.
e-mail: navon@scs.fsu.edu
DOI: 10.1111/j.1600-0870.2008.00362.x

Thus, giving up the assumption that the model is perfect, in the context of strong constraint VDA leads us to 'weak constraint' 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 dimensions). Instead, we prefer to use the general term VDA, because we have used a two-dimensional global shallow water model for presenting our results.

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,

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

where $\mathcal{M}[.]$ denotes all the mathematical operations involved in the NWP model, $\boldsymbol{\eta}$ represents ME and $\mathbf{T}[.]$ is an operator that accounts for the fact that only certain components of the state vector have modelling errors. ME usually varies both spatially and temporally and has both systematic and stochastic components.

Early efforts to model the systematic component of ME were pioneered by Derber (1989). He suggested a simplified approach to model $\boldsymbol{\eta}$ to be equal to $\lambda(t)\boldsymbol{\phi}$. The temporal part, $\lambda(t)$, is a

Table 1. List of acronyms used

Acronym	Definition
ECMWF	European Centre for Medium-Range Weather Forecasts
hPa	Hectopascals
RMSE	Root-mean-squared error
UTC	Universal time coordinate
PPM	Piecewise parabolic method
ME	Model error
DA	Data assimilation
VDA	Variational data assimilation
SWE	Shallow water equations
NWP	Numerical weather prediction
GCMs	General circulation models
T_{-06}	Data set from ERA-40 reanalysis project valid for 00 UTC 2 February 2001
T_{00}	Data set obtained by 6-h integration of T_{-06}
T_{+06}	12-h integration of T_{-06}
T_{+12}	18-h integration
T_{+18}	24-h integration
T_{+24}	30-h integration
T_{+30}	36-h integration

specified function of time alone, and $\boldsymbol{\phi}$ 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 (temporally constant, linearly and harmonically evolving) forms for systematic component of ME were considered by Griffith and Nichols (2000) and Nichols (2003).

Similar approaches for systematic ME as a discrete in time process were 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). Discrete evolution of ME as Markov process was studied by Daley (1992), Zupanski (1997), Zupanski et al. (2005) and Trémolet (2006, 2007).

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 continuous differential equations and then descretized in space and time. If the initial ME, $\eta(t_0) = \eta_0$, then VPLD04 modelled the evolution of ME as,

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)] + \mathbf{q}(t), \tag{2}$$

where $\mathbf{q}(t)$ is the stochastic component of ME, which has been assumed to be zero. They set $\Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)] = \boldsymbol{\eta}(t)$, hence the evolution of ME term was modelled by the following simple exponential growth equation,

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \boldsymbol{\eta}(t). \tag{2}$$

Such a 'deterministic' approach to model the evolution of ME significantly simplifies the weak constraint VDA since only the initial ME ($\boldsymbol{\eta}_0$) is to be obtained via solution of the optimization problem (see VPLD04 and/or Section 4 for additional details).

In this paper, we extend the approach of VPLD04 to parametrize the evolution of ME, using a continuous model. To this end, we propose three models and study their impact on the reduction for forecast errors, through data assimilation. Each of these models for ME has been tested with different model settings, by changing the advection scheme used in forward and adjoint models. A consistent method of decreasing the discretization errors (principally truncation errors) is to refine the model resolution, such an approach in VDA was studied by LeDimet and Shutyaev (2005), henceforth referred to as LDS05.[1] If discretization error is assumed to be the only source of ME, we analyse whether any of the proposed ME forms can indeed account for it via weak constraint VDA. Instead of following the model grid refinement approach of LDS05, here, we consider an alternative procedure. The observations are obtained using a high accurate scheme, such as the PPM scheme, but the model uses low accurate advection schemes (unconstrained or constrained van Leer) that

[1] 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.

differ in their diffusivity properties via application of different slope limiters, see Akella and Navon (2006) for further details. These high-resolution advection schemes are highly efficient in conserving (numerical) flux, especially, using the finite volume formulation (highly relevant e.g. in chemical tracer advection, see Hourdin et al., 2006, the Community Multiscale Air Quality (CMAQ); Byun and Schere (2006) and references therein).

We also study the performance of these approaches, when a systematic error in the coriolis parameter is introduced in the model. As in the study of VPLD04, we do not account for the stochastic component of ME; however, if the random errors cannot be neglected, ensemble data assimilation methods, such as variants of the Kalman filter (Dee and Da Silva, 1998; Zupanski and Zupanski, 2006) provide efficient framework for modelling and estimation of stochastic ME. Finally, we generalize the proposed forms of the ME formulation, using a scalar parameter. An optimal value of this parameter is estimated via minimization of the weak constraint VDA cost functional, hence the model error parameter adjusts to fit the observations at run-time.

Following is the plan of the present paper. In Section 2.1, we introduce the numerical model; a twin experiment set-up, whereby model generated trajectory is used as observations (Section 2.2), is also described. In Section 3, we briefly present the strong constraint VDA, which is followed by the weak VDA. In the same section, we formulate the different forms of ME. The experiments, results and their discussion is provided in Section 5. We summarize our results with conclusions in Section 6.

2. Numerical model and observations

2.1. Shallow water finite volume model

In this paper, a global two-dimensional shallow water equations (SWE) model has been used for numerical experiments. 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 general circulation models (GCMs). In spherical coordinates, the vorticity divergence form of the SWE can be written as,

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{V}h) = 0, \tag{3}$$

$$\frac{\partial u}{\partial t} = \Omega v - \frac{1}{a\ cos\theta}\frac{\partial}{\partial \lambda}(\kappa + \varphi), \tag{4}$$

$$\frac{\partial v}{\partial t} = -\Omega u - \frac{1}{a}\frac{\partial}{\partial \theta}(\kappa + \varphi), \tag{5}$$

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 and a is radius of the earth. The free surface geopotential is given by

$$\varphi = \varphi_s + g\,h, \tag{5}$$

$\varphi_s = g h_s$, $\kappa = \frac{1}{2}\mathbf{V}\cdot\mathbf{V}$ is the kinetic energy and $\Omega = 2\omega \sin\theta + \nabla \times \mathbf{V}$ is the absolute vorticity.

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 dynamic 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- 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 centred 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.

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 neighbouring 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 Akella and Navon (2006), hereafter referred to as AN06. 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 e.g. constrained van Leer and PPM schemes). Whereas on the D-grid, we will be using the unconstrained, constrained van Leer, and the PPM schemes. From now onwards, we will use the following convention to refer to our advection schemes.

(i) 'Unconstrained van Leer scheme': unconstrained van Leer on both C- and D-grids;

(ii) 'Constrained van Leer scheme': constrained van Leer on D-grid and unconstrained van Leer scheme on C-grid and

(iii) '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 (1996). Further details of the model can be found in LR97 and references therein.

2.2. Observations

Reanalysed data on a 2.5° × 2.5° grid (500 hPa pressure level; geopotential height and velocity fields)

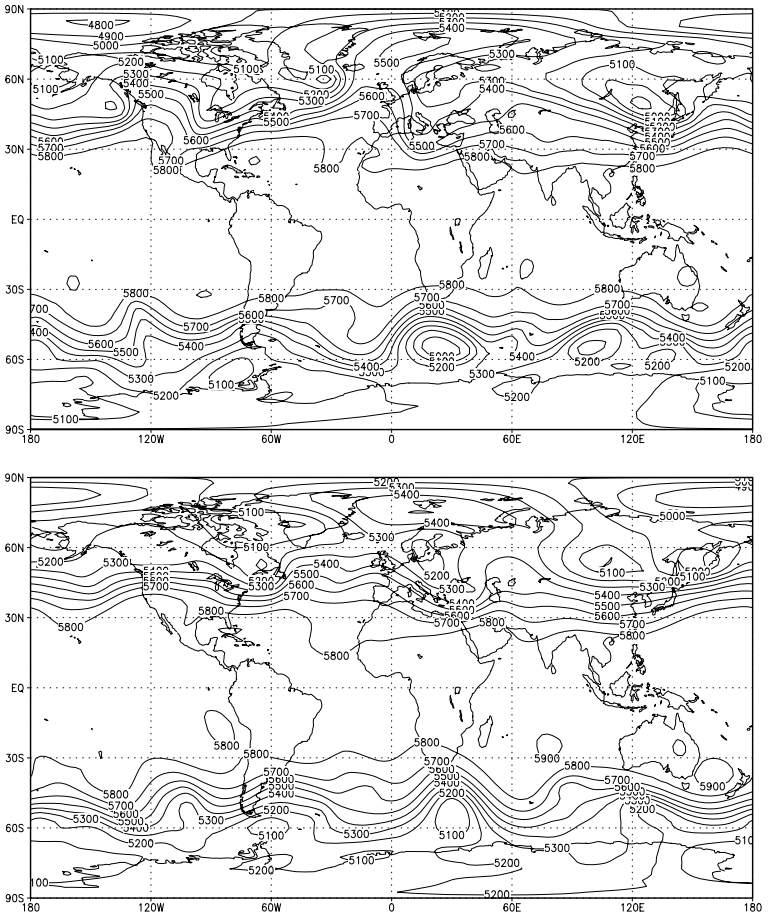


Fig. 1. Isolines of height field (m), top panel: T_{-06} (reanalysed 500 hPa pressure level data valid at 00 UTC 2 February 2001) from the ECMWF ERA-40 data set; bottom panel: a 36 forward model integration using the PPM scheme.

from the ERA-40, ECMWF 40-year reanalysis system (http://www.ecmwf.int/research/era/), valid at 0000 UTC 2 February 2001 (henceforth denoted by T_{-06}), was used to specify the initial conditions for forward model integration. The PPM scheme with a $2.5^\circ \times 2.5^\circ$ (144×72 cells) grid resolution and time step of $\Delta t = 450$ s, has been used in this article, to generate a 'reference trajectory'. Synthetic observations are obtained by randomly perturbing the reference trajectory (please see Section 5 for further details). In Fig. 1 we display configuration of the reference trajectory; height field at T_{-06} and a 36-h forecast (with the PPM scheme).

3. Strong constraint VDA

Data assimilation schemes determine the analysed atmospheric state as an 'optimal' combination of a priori background information and observational information. Let $\mathbf{x}^{\mathrm{t}}$ be a projection of 'true' state of the atmosphere onto model space, $\mathbf{x}^{\mathrm{b}}$ be the background field and $\mathbf{y}^{\mathrm{o}}$ denote observations. Usually a short-range forecast provides $\mathbf{x}^{\mathrm{b}}$. Then the error in the background field is equal to $\mathbf{e}^{\mathrm{b}} = \mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}$, and the error in the observations field is given by $\mathbf{e}^{\mathrm{o}} = \mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{t}})$; H is the observation operator. Denoting the mathematical expectation operator by $E[.]$, the background error and observation error covariances are given by $\mathbf{B} = E[(\mathbf{e}^{\mathrm{b}} - E[\mathbf{e}^{\mathrm{b}}])(\mathbf{e}^{\mathrm{b}} - E[\mathbf{e}^{\mathrm{b}}])^T]$ and $\mathbf{R} = E[(\mathbf{e}^{\mathrm{o}} - E[\mathbf{e}^{\mathrm{o}}])(\mathbf{e}^{\mathrm{o}} - E[\mathbf{e}^{\mathrm{o}}])^T]$, respectively. Error covariances measure the uncertainty, involved with both of these data sources, hence they determine the quality of data assimilation. Mostly, due to lack of knowledge of the true state of the atmosphere, $\mathbf{x}^{\mathrm{t}}$, 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 ME, we minimize the following nonlinear quadratic cost functional, $\mathcal{J}_o$, with respect to the initial state, $\mathbf{x}(t_0)$ as a control variable (Kalnay, 2003),

$$\mathcal{J}[\mathbf{x}(t_0)] = \underbrace{\frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}]^T \, \mathbf{B}^{-1} \, [\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}]}_{\mathcal{J}_{\mathrm{b}}} + \underbrace{\frac{1}{2} \sum_{i=0}^{n} [H(\mathbf{x}(t_i)) - \mathbf{y}^{\mathrm{o}}(t_i)]^T \, \mathbf{R}^{-1} \, [H(\mathbf{x}(t_i)) - \mathbf{y}^{\mathrm{o}}(t_i)]}_{\mathcal{J}_{\mathrm{o}}}, \tag{6}$$

subject to the following model equations as strong constraint,

$$\begin{aligned} \mathbf{x}(t_0) &= \mathbf{x}_0, \\ \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} &= \mathcal{M}[\mathbf{x}(t)]. \end{aligned} \tag{7}$$

This 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). Note that in the above model equations, we did not account for ME, that is, $\eta\ (t) \equiv 0, \forall\ t$.

The formulation of the so-called background cost functional, $\mathcal{J}_\mathrm{b}$, 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 neighbouring gridpoints 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, $\mathbf{x}^\mathrm{b}$ 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). Our focus is on ME in this article, hence we follow the approach of VPLD04 to construct ('static' in time) **B** as a multivariate and cross-correlated operator (Derber and Bouttier, 1999; Weaver and Courtier, 2001). The observation error covariance matrix has been taken to be a block diagonal matrix, $R = [10^4\ I,\ 100\ I,\ 100\ I]$.

4. Weak constraint VDA

In strong constraint VDA, model equations are assumed to be 'perfect'; therefore, MEs (whose causes have been described earlier in the introduction) are not taken into account. The weak constraint VDA provides a framework for incorporating ME in the model equations via explicit introduction of an extra term, $\boldsymbol{\eta}\ (t)$,

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathcal{M}[\mathbf{x}(t)] + \mathbf{T}[\boldsymbol{\eta}(t)]. \tag{8}$$

The operator $\mathbf{T}\ [\cdot]$ maps the space of the ME to the space of the model state, **x**. If one has a priori knowledge that the numerical model has some severe drawbacks, for example, modelling of the atmosphere in certain regions of the globe, (say, one of the poles) then the operator, $\mathbf{T}\ [\cdot]$ should be specified in such a way that only those model gridpoints (at that pole) have MEs, and in comparison, the rest of the model states do not have any ME. In the literature (for instance, see Griffith and Nichols, 2000 and VPLD04), it has been assumed that the model state at every grid point has an associated error, which implies that $\mathbf{T}\ [\cdot]$ is identically equal to the unit matrix, I, and the dimension of $\boldsymbol{\eta}$ is equal to that of the model state, $\mathbf{x}$; we assume, in the present article, $\mathbf{T} = I$.

Following Derber (1989) and VPLD04 in the spirit of variational continuous assimilation', we will model the evolution of ME as a continuous process, as the following initial value problem (IVP),

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)] + \mathbf{q}(t). \tag{9}$$

In this paper, we are concerned only with the systematic part of ME, hence, the stochastic component, $\mathbf{q}\ (t)$, is neglected. Therefore, the above differential equation simplifies to

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)]. \tag{10}$$

For closure of the above IVP, we need to specify the initial value of $\boldsymbol{\eta}\ (t_0) = \boldsymbol{\eta}_0$ and the nature of the mapping $\Phi[.]$.

First, we describe the methodology used to calculate the initial value of ME and then address the issue of different approaches for modelling the evolution of ME, using different forms of $\Phi[.]$. To obtain $\boldsymbol{\eta}_0$, the following weak constraint VDA cost functional is minimized (note that it is similar to the cost functional in eq. (6), but includes an extra term, $\mathcal{J}_\eta$),

$$\begin{aligned} \mathcal{J}[\mathbf{x}(t_0), \boldsymbol{\eta}(t_0)] = & \underbrace{\frac{1}{2}[\mathbf{x}_0 - \mathbf{x}^\mathrm{b}]^T\, \mathbf{B}^{-1}\, [\mathbf{x}_0 - \mathbf{x}^\mathrm{b}]}_{\mathcal{J}_\mathrm{b}} \\ & + \underbrace{\frac{1}{2}\sum_{i=0}^{n}[H(\mathbf{x}(t_i)) - \mathbf{y}^\mathrm{o}(t_i)]^T\, \mathbf{R}^{-1}\, [H(\mathbf{x}(t_i)) - \mathbf{y}^\mathrm{o}(t_i)]}_{\mathcal{J}_\mathrm{o}}, \\ & + \underbrace{\frac{1}{2}[\boldsymbol{\eta}_0 - \boldsymbol{\eta}^\mathrm{b}]^T\, \mathbf{Q}^{-1}\, [\boldsymbol{\eta}_0 - \boldsymbol{\eta}^\mathrm{b}]}_{\mathcal{J}_\eta}, \end{aligned} \tag{11}$$

where **Q** is the model error covariance matrix.

Just as in the background state, $\mathbf{x}^\mathrm{b}$ was used as an initial guess for $\mathbf{x}_0$ to minimize the strong constraint VDA cost functional, we use $\boldsymbol{\eta}^\mathrm{b}$ as an initial guess for $\boldsymbol{\eta}_0$ to minimize the above $\mathcal{J}_\eta$ in the weak constraint VDA. Hence the above cost functional, $\mathcal{J}(\mathbf{x}_0, \boldsymbol{\eta}_0)$ is minimized subject to the following equations as constraints:

$$\left.\begin{aligned} &\mathbf{x}(t_0) = \mathbf{x}_0;\ \boldsymbol{\eta}(t_0) = \boldsymbol{\eta}_0, \\ &\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathcal{M}[\mathbf{x}(t)] + \boldsymbol{\eta}(t); \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)]. \end{aligned}\right\} \tag{12}$$

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

unconstrained problem,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\eta}, \mathbf{x}^*, \boldsymbol{\eta}^*) = \mathcal{J}(\mathbf{x}_0, \boldsymbol{\eta}_0) + \int_{t_0}^{t_n} \left\langle \mathbf{x}^*, \left\{ \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} - \mathcal{M}[\mathbf{x}(t)] - \boldsymbol{\eta}(t) \right\} \right\rangle \mathrm{d}t + \int_{t_0}^{t_n} \left\langle \boldsymbol{\eta}^*, \left\{ \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} - \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)] \right\} \right\rangle \mathrm{d}t, \quad (13)$$

where $\mathbf{x}^*$, $\boldsymbol{\eta}^*$ are the Lagrange multiplier vectors corresponding to $\mathbf{x}$, $\boldsymbol{\eta}$, respectively, 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 $\mathcal{J}(\mathbf{x}_0, \boldsymbol{\eta}_0)$). Using the first-order optimality criteria, at the extrema of the Lagrangian, $\mathcal{L}$, following equations are satisfied.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0, \frac{\partial \mathcal{L}}{\partial \boldsymbol{\eta}} = 0, \quad (14a)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}^*} = 0, \frac{\partial \mathcal{L}}{\partial \boldsymbol{\eta}^*} = 0. \quad (14b)$$

Equations (14b) yield the equations describing the evolution of model state and ME,

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathcal{M}[\mathbf{x}(t)] + \boldsymbol{\eta}(t), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t), \mathbf{x}(t)], \quad (14b)$$

respectively. Whereas eqs. (14a) yield the following adjoint equations that describe the evolution of the adjoint variables $\mathbf{x}^*$, η^*,

$$\mathbf{x}^*(t_n) = 0, \ \boldsymbol{\eta}^*(t_n) = 0,$$
$$-\frac{\mathrm{d}\mathbf{x}^*(t)}{\mathrm{d}t} = \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right]^T \mathbf{x}^* + \left[\frac{\partial \Phi}{\partial \mathbf{x}}\right]^T \boldsymbol{\eta}^* + \delta(t - t_i) \sum_{i=0}^{n} \left[\frac{\partial H}{\partial \mathbf{x}}\right]^T \mathbf{R}^{-1} [H(\mathbf{x}(t_i)) - \mathbf{y}^{\mathrm{o}}(t_i)], \quad (15a)$$

$$-\frac{\mathrm{d}\boldsymbol{\eta}^*(t)}{\mathrm{d}t} = \left[\frac{\partial \Phi}{\partial \boldsymbol{\eta}}\right]^T \boldsymbol{\eta}^* + \mathbf{x}^*. \quad (15b)$$

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

$$\nabla_{\mathbf{x}_0} \mathcal{J} = \nabla_{\mathbf{x}_0} \mathcal{J}_{\mathrm{b}} + \nabla_{\mathbf{x}_0} \mathcal{J}_{\mathrm{o}} = \mathbf{B}^{-1}[\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}] + \mathbf{x}^*(t_0), \quad (16a)$$

$$\nabla_{\boldsymbol{\eta}_0} \mathcal{J} = \nabla_{\boldsymbol{\eta}_0} \mathcal{J}_{\boldsymbol{\eta}} + \nabla_{\boldsymbol{\eta}_0} \mathcal{J}_{\mathrm{O}} = \mathbf{Q}^{-1}[\boldsymbol{\eta}_0 - \boldsymbol{\eta}^{\mathrm{b}}] + \boldsymbol{\eta}^*(t_0). \quad (16b)$$

As usual, backward integration of the adjoint models (15a) and (15b) from time $t_n \rightarrow t_0$, provides us the values of initial adjoint states $\mathbf{x}^*(t_0)$ and $\boldsymbol{\eta}^*(t_0)$. Therefore, the gradient in weak constraint VDA is given by $(\nabla_{\mathbf{x}_0} \mathcal{J}, \nabla_{\boldsymbol{\eta}_0} \mathcal{J})^T$. Comparing this with the gradient in strong constraint VDA, which was only $\nabla_{\mathbf{x}_0} \mathcal{J}$, the size of the optimization problem is doubled.

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 $\mathbf{B}^{-1}$. The ME covariance matrix, $\mathbf{Q}$, has been handled in a similar fashion, details were provided by VPLD04.

Now we address possible approaches to model the evolution of ME, using different forms of the mapping, $\Phi[\boldsymbol{\eta}\ (t), \mathbf{x}\ (t)]$, which maps the space of state variables, $\mathbf{x}$ and the space of ME, $\boldsymbol{\eta}$, onto $\boldsymbol{\eta}$ only. As noted above, this mapping couples the evolution of the adjoint variables corresponding to the model states and ME, and it also increases the complexity involved in the backward integration of the adjoint models, eq. (15). '*To the best of our knowledge, the issue of* MEs *in solutions of inverse problems using high resolution advection schemes has not been addressed as yet*'; hence to begin with, in this paper, we assume that $\frac{\partial \Phi}{\partial \mathbf{x}} = 0$, that is, $\Phi[.]$ maps ME onto itself. This assumption significantly simplifies the adjoint model equations, since the evolution of $\mathbf{x}^*$ is unchanged, and we can concentrate only on 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. Since the ME evolution is assumed to be governed by the following equation,

$$\boldsymbol{\eta}(t_0) = \boldsymbol{\eta}_0, \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \Phi[\boldsymbol{\eta}(t)], \quad (16b)$$

the rate of growth (or decay) of $\boldsymbol{\eta}\ (t)$ in time is given by the particular form of $\Phi[\boldsymbol{\eta}]$. If $\Phi[\boldsymbol{\eta}] < 0, \forall \boldsymbol{\eta}$, then ME monotonically decreases in time; if $\Phi[\boldsymbol{\eta}] = 0, \forall \boldsymbol{\eta}$ then it is constant, and if $\Phi[\boldsymbol{\eta}] > 0, \forall \boldsymbol{\eta}$, then it grows in time. We considered all these possibilities and investigated the following three forms of Φ,

(i) Decaying ME, $\Phi[\boldsymbol{\eta}] = -\beta \boldsymbol{\eta}$,
(ii) Constant ME, $\Phi[\boldsymbol{\eta}] = 0$,
(iii) Growing ME, $\Phi[\boldsymbol{\eta}] = \gamma \boldsymbol{\eta}$,

where β and γ are constants, and in our numerical results, we specified $\beta \Delta t = 0.2$ and $\gamma \Delta t = 0.01$ (henceforth, the time step $\Delta t = 450$ s as in Section 2.2). These 'values' of β and γ were chosen arbitrarily, so that the qualitative aspects of different forms (decaying, constant and, growing) of ME can be studied in the context of the various advection schemes (see Sections 5.1–5.3 for details). In the next section, we generalize the above forms and provide a methodology to obtain the parameter (β or γ) value via 'optimal' parameter estimation.

4.1. Model error parameter estimation

We start by writing the following general form for the evolution of ME,

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}t} = \alpha\, \boldsymbol{\eta}(t), \quad (17)$$

where the scalar parameter, α, is determined via minimization of the weak constraint VDA cost functional. Therefore, we add

the following quadratic penalty term:

$$\mathcal{J}_\alpha = \frac{1}{2}\frac{1}{b_\alpha}(\alpha - \alpha_\mathrm{b})^2 \tag{18}$$

to the cost functional in eq. (11), where the initial guess or background value of the parameter is denoted by α_b and the quadratic penalty coefficient for deviations from that value by b_α. Hence after every kth minimization iteration in addition to the initial conditions $\mathbf{x}_0^k$ and $\boldsymbol{\eta}_0^k$, we also obtain an optimal parameter value α^k, such that there is a correspondence between the optimal ME parameter α and initial condition $\boldsymbol{\eta}_0$ and model initial conditions $\mathbf{x}_0$. To briefly sketch the details of this procedure, if at the end of kth minimization iteration, we have the set $\{\mathbf{x}_0^k, \boldsymbol{\eta}_0^k, \alpha^k\}$, with a cost functional and gradient values $\mathcal{J}_k, \nabla\mathcal{J}_k$, respectively. We compute a new parameter value, α^{k+1}, in a sequence of i inner iterations, through a parameter estimation procedure (for details on parameter estimation in VDA, see e.g. Navon, 1998). These inner iterations are terminated when

$$\frac{\|\nabla\mathcal{J}(\mathbf{x}_0^k, \boldsymbol{\eta}_0^k, \alpha_i^{k+1})\|_2}{\|\nabla\mathcal{J}(\mathbf{x}_0^k, \boldsymbol{\eta}_0^k, \alpha^k)\|_2} \leqslant TOL_\mathrm{inner},$$

or if $i \geqslant nMAX_\mathrm{inner}$, where $nMAX_\mathrm{inner}$ is a preset number of maximum iterations. Having found an optimal parameter value at $k+1$ iteration, we 'fix' it and use it in eq. (17), to obtain new values of $\mathbf{x}_0^{k+1}$ and $\boldsymbol{\eta}_0^{k+1}$ in a sequence of 'outer' iterations, yielding $\{\mathbf{x}_0^{k+1}, \boldsymbol{\eta}_0^{k+1}, \alpha^{k+1}\}$. The termination criteria for the outer iteration cycles will be specified in Section 5.1. To summarize, the qualitative form (decaying, constant or growing) of ME will be primarily determined based on observational misfit, and such a parameter estimation leads to a general modelling of ME. We also need to specify b_α, which will be discussed in Section 5.4.

5. Results with various experiments

5.1. PPM *perfect* 2.5 × 2.5

In the first set of experiments, using the initial condition at T_{-06}, (introduced in Section 2.2), we integrated the forward model for 36 h, with the PPM advection scheme, saving forecasts at every 6-h interval (2.5° × 2.5° grid resolution and time step of 450 s), as reference trajectory. A 24-h data assimilation time window $[T_{00} - T_{+24}]$ has been used, T_{+30} was taken to be forecast verification time. See Table 1 and Fig. 2 for nomenclature of the different time intervals.

The goal of this experiment is to investigate if it is possible to account for the discretization error of the other advection schemes by including for them a ME term in VDA (unconstrained and constrained van Leer schemes are known to be diffusive when compared with the PPM scheme, please see AN06; LR97; Lin et al. (1994)). Hence the grid resolution and time step are same for the observational (PPM advection) and numerical (other advection schemes) models; also, observations are exactly the same as the reference trajectory. Note that the PPM scheme is 'exact' for this particular experiment.

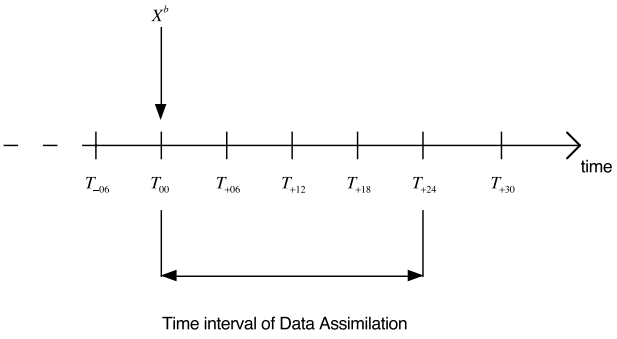


Fig. 2. Illustration of data assimilation time window; $\mathbf{x}^b$ was generated by a 6-h integration, using data at T_{-06}. T_{+30} is forecast verification time. Observations are used at T_{00}, T_{+06}, T_{+12}, T_{+18} and T_{+24}.

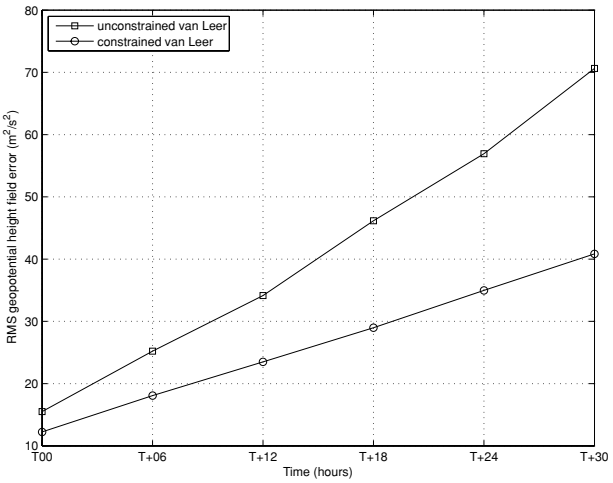


Fig. 3. RMS geopotential height field (m^2/s^2) forecast errors with different advection schemes for the PPM perfect experiments described in Section 5.1.

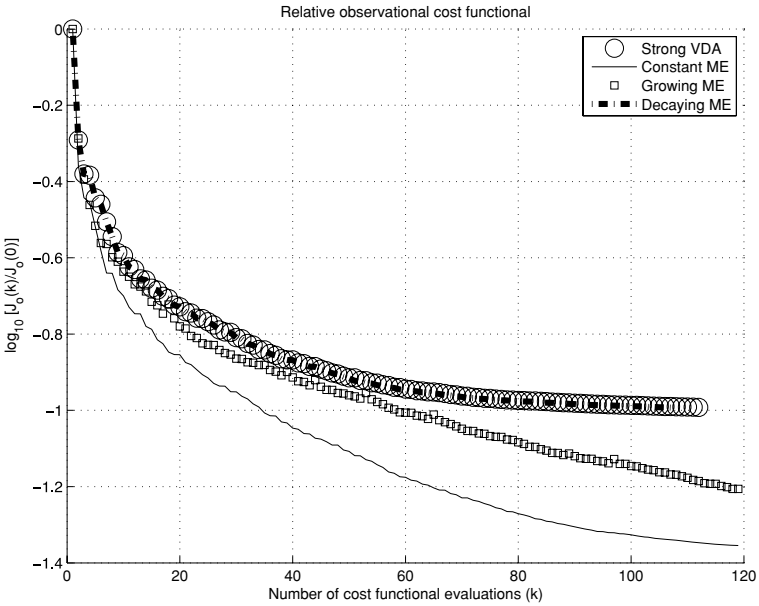
To motivate our results, in Fig. 3, we show the geopotential height field RMSE[2] between model forecasts and observations for the two van Leer advection schemes. Among the van Leer schemes, the constrained van Leer scheme is better than the unconstrained van Leer scheme due to the monotonicity constraint applied in the former, hence it yields lesser forecast error.

For the assimilation experiments, 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 functionals given in eqs. (6) and (11). Following termination criteria was used in all DA experiments:

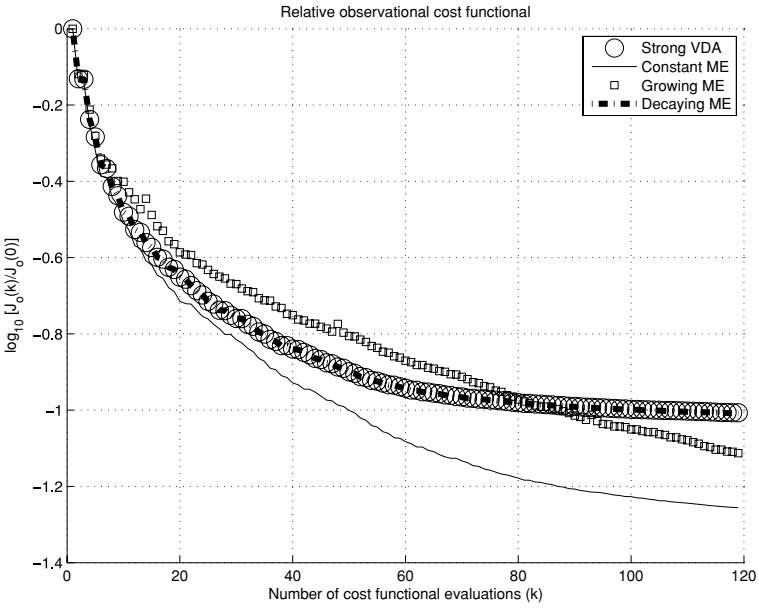
$$\frac{\|(\nabla\mathcal{J})_k\|_2}{\|(\nabla\mathcal{J})_0\|_2} \leq EPS, \tag{19}$$

$(\nabla\mathcal{J})_k$ is the gradient at the kth minimization iteration and EPS was set to 1.0×10^{-2}. During the minimization process, due

[2] RMSE between fields $\mathbf{x}$ and $\mathbf{y}$ was calculated as $\sqrt{\frac{1}{N\times M}\sum_{i=1}^{N}\sum_{j=1}^{M}(x_{i,j} - y_{i,j})^2}$



(a) Unconstrained van Leer advection scheme

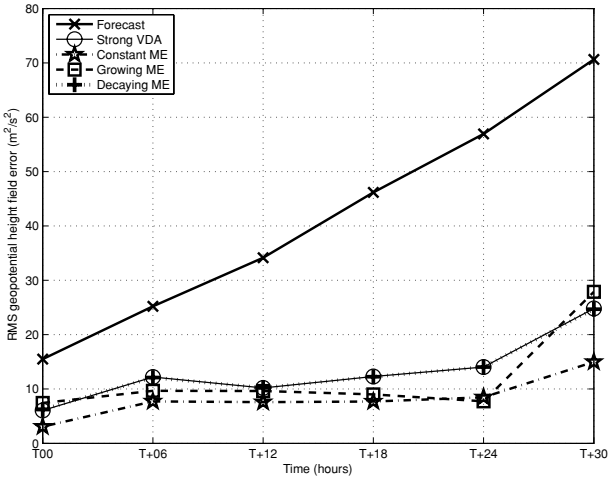


(b) Constrained van Leer advection scheme

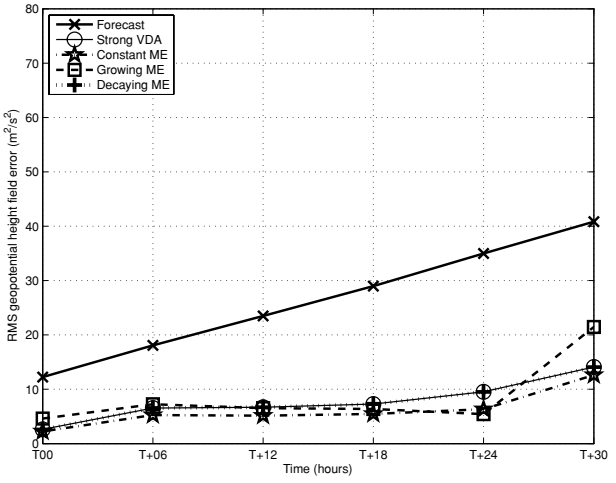
Fig. 4. Plot of the observational cost functional, $\mathcal{J}_o$ (normalized using value at first minimization iteration), for the strong and weak constraint VDA, with various forms of ME for the PPM perfect 2.5×2.5 experiment, described in Section 5.1. Panel (a) uses the unconstrained van Leer advection scheme and panel (b), constrained van Leer.

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. Hence, all VDA minimization iterations were terminated using above criteria, or after 120 function evaluations, which ever happens first.

Figures 4(a) and (b) provide a plot of the normalized misfit between observations and model prediction, $\mathcal{J}_o$ in eqs. (6) or (11), versus number of cost functional evaluations during VDA for both unconstrained and constrained van Leer schemes. Note the improved fit of model states and observations when ME term is included, when compared with no ME in strong constraint VDA. Among the various forms (decaying, constant, growing) of ME proposed in Section 4, the constant ME exhibits the best fit with the observations, whereas the decaying ME performs negligibly better than the strong constraint VDA. In the earlier minimization iterations, for the constrained van Leer, the growing ME form



(a) Unconstrained van Leer



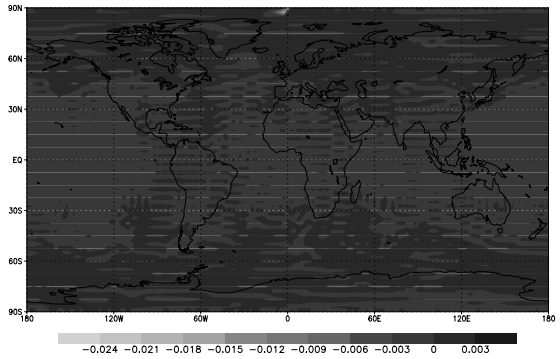
(b) Constrained van Leer

Fig. 5. RMS geopotential height field (m^2/s^2) errors before and after strong constraint, as well as weak constraint VDA for the PPM perfect 2.5×2.5 experiment (Section 5.1). Panel (a) uses the unconstrained van Leer advection scheme and Panel (b) constrained van Leer.

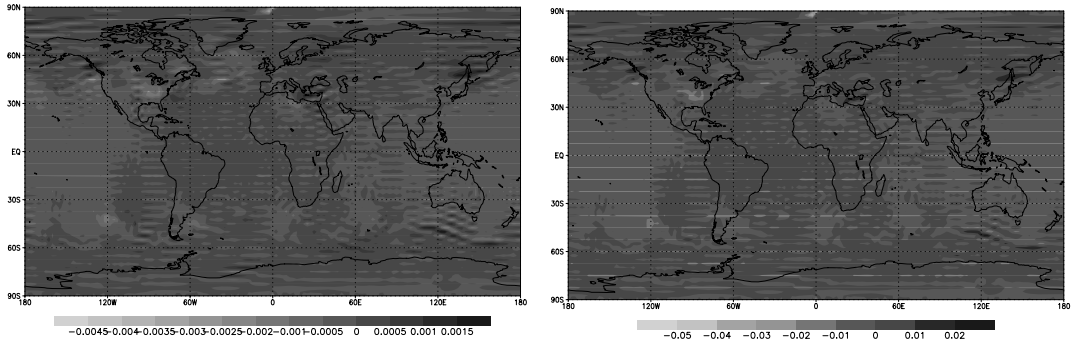
does not perform well, but does better in fitting the small-scale differences towards the latter iterations (after 80), though not as well as the constant ME.

In Fig. 5, we compare the RMSE in geopotential height field before and after VDA for the unconstrained and constrained van Leer schemes (though not shown, the RMSE in wind fields was significantly reduced using different forms of ME). In all cases, the RMSE has been reduced by 50% or more via VDA. Also the optimized initial condition is able to provide a better forecast at T_{+30} (30-h forward integration from T_{00}). Consistent with the best observational fit as in Fig. 4, the constant ME provides the least RMSE. As earlier, results with the decaying ME form are similar to the strong VDA. The growing ME does provide lesser RMSE (compared with strong VDA, specifically, the unconstrained van Leer, except at T_{00}) within the window of data assimilation, whereas the forecast at T_{+30} is deteriorated. We note that in the case of the constrained van Leer scheme, all the proposed forms of ME exhibit similar behaviour in RMSE reduction after VDA, particularly within the window of assimilation. This loss of contrasting performance of various ME forms, compared with the unconstrained van Leer scheme could be due the fact that the constrained van Leer scheme fits the observations better than the unconstrained scheme (Fig. 3) even without VDA. Also, we recall that the cause of the observational misfit is purely due to the discretization error of the PPM and constrained van Leer schemes, which is smaller compared with the error between the PPM and unconstrained van Leer schemes. This smaller disctreziation error could explain the reason for various ME forms to be peforming slightly better than the strong VDA in the case of more accurate constrained van Leer scheme; however, this issue will be further examined in Section 5.3.

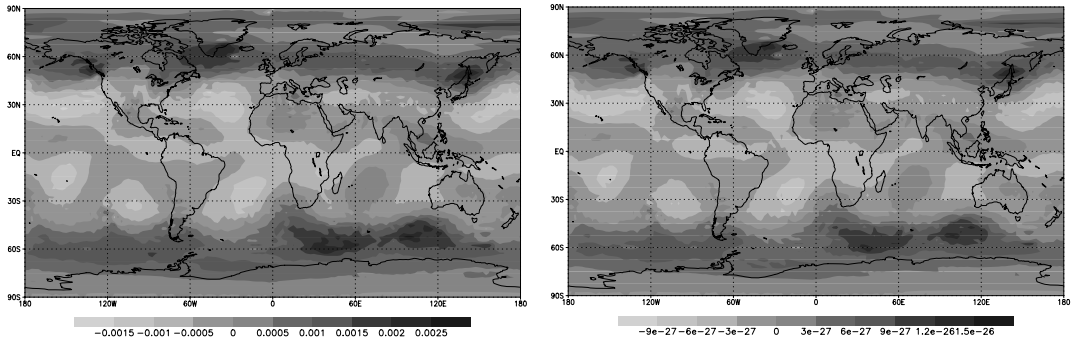
The above results indicate that discretization error for the lesser accurate van Leer schemes, compared with the PPM scheme, could be accounted for, by including a ME term. However, the forecast RMSE, obtained using growing ME, is higher than that obtained in the strong VDA case, for both advection schemes. Also the decaying ME form performed very similar to the strong VDA, which provides no treatment of model error. In Fig. 6, we plot the initial (at time T_{00}) model error (height field component) after minimization and the forecast (at time T_{+30}) for the constant, growing and decaying ME forms, with the unconstrained van Leer scheme. We note that the model error at T_{00} has identically similar features for the constant and growing forms, except for the difference in magnitudes. For the growing ME, because of the growth rate $\gamma = 0.01/\Delta t$ specified in Section 4 and following the ordinary differential equation in time, the model error at T_{+30} looks exactly similar to that at T_{00} but increased by about one order of magnitude; this increase could explain why the forecast RMSE is higher than the strong VDA. The initial model error for the decaying ME at T_{00} looks like a smoothed version of the initial model error for the constant and growing ME forms and, due to the specified decay rate $\beta = 0.2/\Delta\ t$, decreases to insignificantly small value very quickly, which explains why its performance is similar to the strong VDA. Further, the analysis error at T_{00} and forecast error at T_{+30}, after minimization, plotted in Fig. 7 supports the above argument. The strong VDA and the decaying ME form exhibit almost identical (both in magnitude and features) errors at T_{00} and T_{+30}. The growing ME shows the highest error (larger than strong VDA, as was shown in the RMSE plot, Fig. 5). Indeed the locations of maximum error seem to be the same for all the methods after VDA, in particular, the constant ME reduces the error magnitude at those locations by about one-fifth, comparing before and after VDA (the plots for the constrained van Leer scheme lead to similar conclusion, hence not shown). These results suggest that the specification of growth and decay rates is very important for



(a) Constant ME



(b) Growing ME



(c) Decaying ME

Fig. 6. Spatial distribution of model error after weak constraint VDA, height field (m) component at initial and final time for the PPM perfect 2.5 × 2.5 experiments described in Section 5.1, using the unconstrained van Leer advection scheme. Left-hand column, at time T_{00}; right-hand column, T_{+30}.

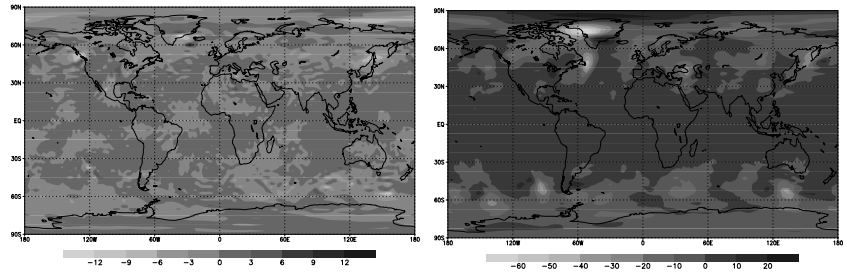
better performance of proposed forms of ME, this issue will be further discussed in Section 5.4.

In the following experiments, we examine the impact of differing ME in a more 'realistic' setting. The number of observations used for assimilation is decreased and therefore only a part of the state vector is observed. [3]

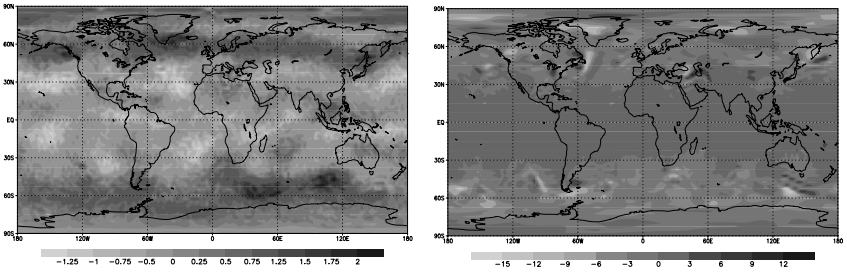
5.2. PPM *perfect* 5 × 5

This experiment is similar to above experiment, except that the observations used in assimilation were specified at a grid resolution of 5° × 5°. Hence the spatial degree of freedom for minimization is reduced by one-fourth; however, the time frequency has been kept the same, that is, observations used every 6 h as indicated in Fig. 2. Also the criteria for termination of minimization iterations has been kept the same as earlier (1% decrease in gradient L_2 norm or 120 cost function evaluations, whichever came first, details were given in Section 5.1).

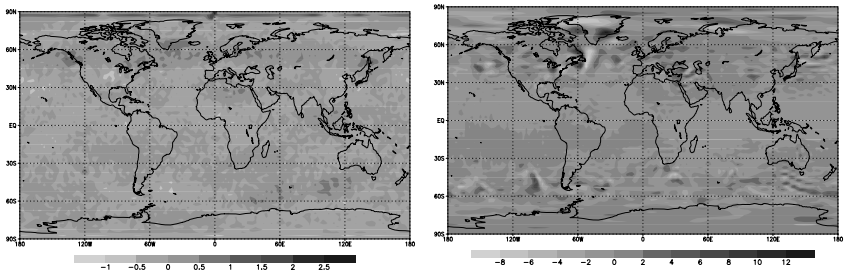
[3] we use a regular grid resolution for the observations; more complicated, non-uniform observational network would be tackled in future work.



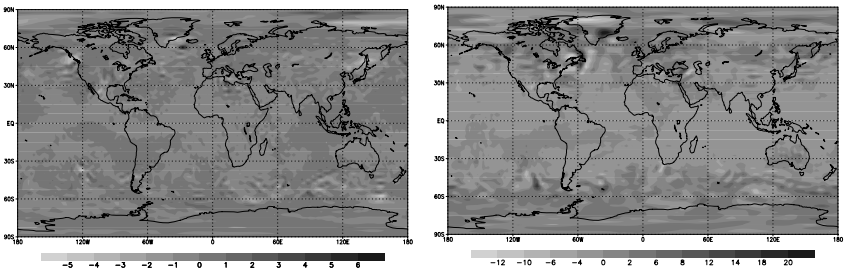
(a) Before VDA



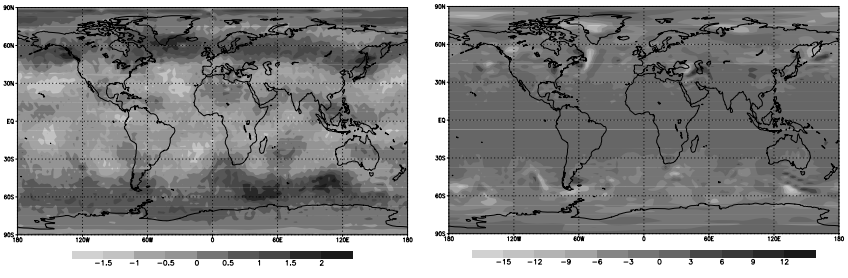
(b) Strong constraint VDA



(c) Weak constraint VDA, constant ME

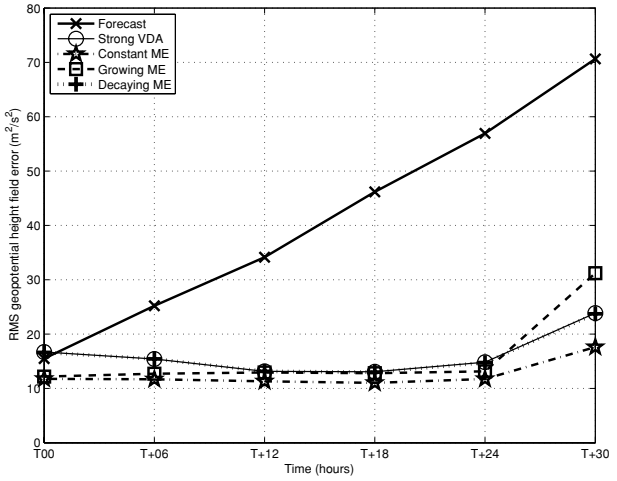


(d) Weak constraint VDA, growing ME

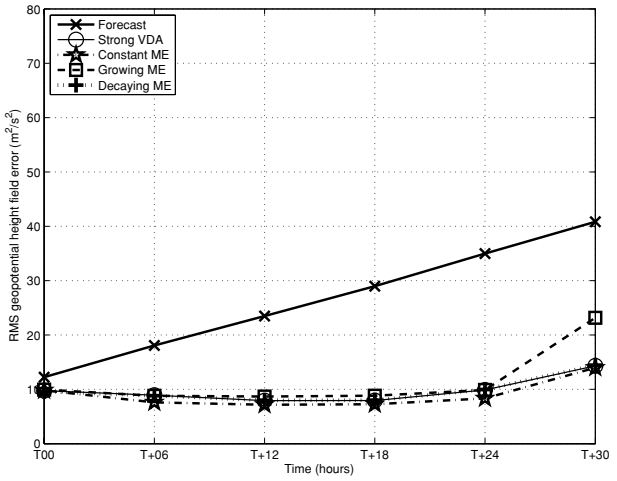


(e) Weak constraint VDA, decaying ME

Fig. 7. Spatial distribution of error in height field (m) at initial and final time for the PPM perfect 2.5 × 2.5 experiments described in Section 5.1, using the unconstrained van Leer advection scheme. Left-hand column, at time T_{00}; right-hand column, T_{+30}.



(a) Unconstrained van Leer



(b) Constrained van Leer

Fig. 8. Same as in Fig. 5, but with the observations at $5^\circ \times 5^\circ$ grid resolution, PPM perfect 5×5 experiments described in Section 5.2.

Forecast RMSE obtained using the assimilated initial conditions, is plotted in Fig. 8 (the fit with observations, $\mathcal{J}_o$, mirrors the information provided in forecast RMSE plot, hence, not shown). As the number of observations is decreased ($2.5^\circ \times 2.5^\circ$– $5^\circ \times 5^\circ$ grid resolution), the RMSE after assimilation increased when compared with results presented in the above section. We note that the error at T_{00} is comparable to the error before VDA; however, when we relaxed one of our termination criteria (in this case, 120 cost function evaluations, which was reached first), the error at T_{00} was also decreased. However, for the sake of a fair comparison, we used the same termination criteria as before.

As earlier, the constant ME yields the smallest RMSE among all the VDA methods considered, for both of the van Leer schemes. Once again, for the more accurate constrained van Leer scheme, the proposed ME forms give performance comparable to the strong VDA. However, the results with the unconstrained van Leer scheme do suggest that when sources of error are—(a) discretization error: modelled via different advection schemes (observations: PPM, model: unconstrained or, constrained van Leer); (b) lack of observations—supplied at a lower resolution than model grid, can be accounted for by a constant model error. In the case of the constrained van Leer scheme, the above forms of ME seem not to be peforming much better than the strong VDA. To investigate this behavior, we present the following set of experiments, in which we introduced another source of ME, by specifying an incorrect value of the Coriolis parameter.[4]

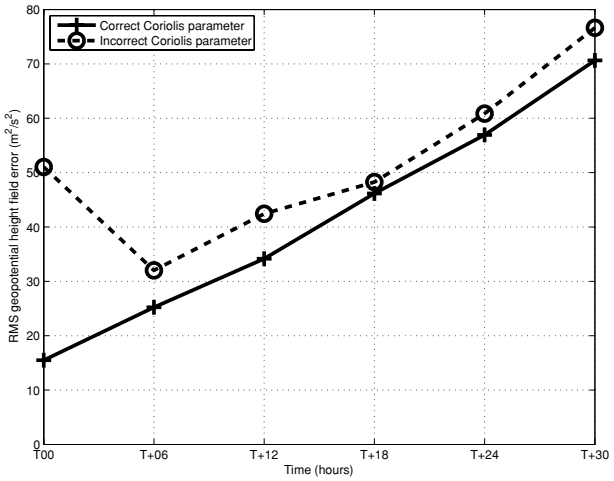
5.3. *Wrong Coriolis*

In the following experiments, the Coriolis parameter specified in the forecast and adjoint models was wrongly specified by setting the angular velocity of the earth to 0.98 times the correct value of 7.292×10^{-5} s^{-1}, which is a consistent source of modelling error. In Fig. 9, a comparison of the forecasting errors using the correct and incorrect Coriolis parameter is shown for the two van Leer schemes. Note that the constrained van Leer scheme is more sensitive to this systematic error than the unconstrained van Leer scheme and therefore, hopefully, a good test case to see if any of the proposed ME forms will work better than the strong VDA.

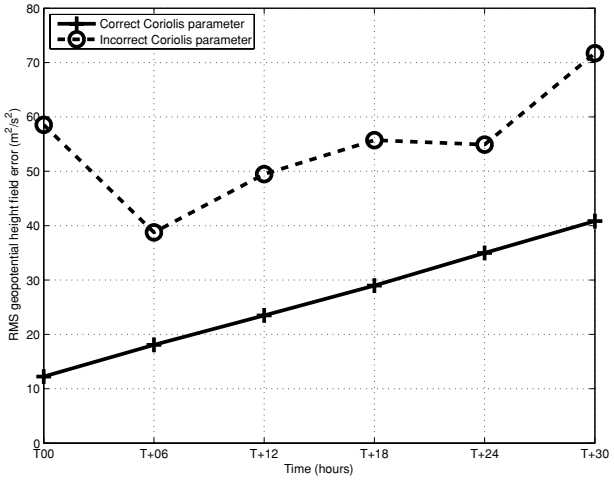
As in the above PPM perfect experiments (Section 5.1), observations were obtained by using the PPM advection scheme, with the correct Coriolis parameter and adding a 1% random perturbation. Therefore, the observations were used every 6 h, on a 2.5×2.5 grid resolution. Forecast RMSE obtained after data assimilation is plotted in Figs. 10(a)–(c). Though we report results obtained using the PPM scheme (in the forecast and adjoint) model, one should not compare the PPM results with the van Leer scheme results, recall that the (reference trajectory and hence) observations were generated using the PPM scheme. However, in these experiments, the PPM has been included to see how it behaves in the above proposed weak constraint VDA framework, when there is a consistent source of ME, in particular, can we use any of the MEs to account for the error from misspecified Coriolis parameter alone? On the other hand, for the van Leer schemes the sources of error are: incorrect Coriolis parameter, less accurate advection scheme.

Indeed, the consistent source of model error through misspecified Coriolis parameter contrasts the peformance of various ME forms, when compared with the strong VDA for the higher accurate constrained van Leer advection scheme. However, as discussed earlier in PPM perfect 2.5×2.5 (Section 5.1), the strong VDA and decaying ME form yield similar RMSE errors. Since the misspecification of Coriolis parameter is a 'constant' source of

[4] similar approaches to introduce consistent model error were used, see for instance, VPLD04, Dee and Da Silva (1998).



(a) Unconstrained van Leer

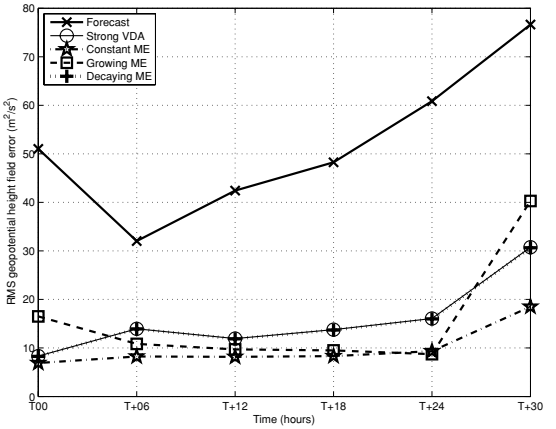


(b) Constrained van Leer

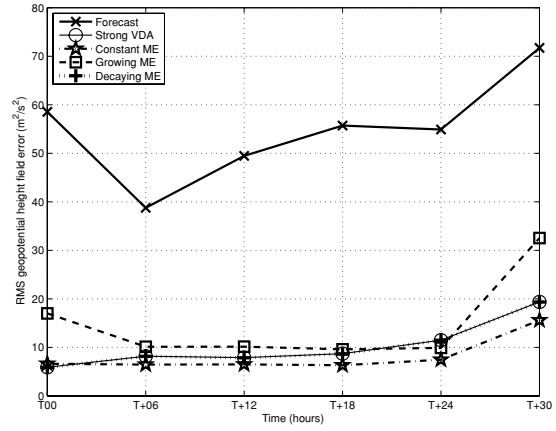
Fig. 9. A comparison of the forecast RMSE, using the correct and incorrect (2% error) Coriolis parameters. Panel (a), unconstrained van Leer scheme; panel (b), constrained van Leer scheme.

ME, it is expected that constant ME form would work well, exactly as the results reveal, even for the PPM scheme. On the other hand, the growing ME exhibits the largest errors. We note that for the constrained van Leer and PPM scheme the error at T_{00} is slightly higher than that for the strong VDA, even with the constant ME, as mentioned earlier; this could be alleviated by relaxing the minimization termination criteria and doing more function evaluations. However, notable gain is obtained in lesser error at all other times, including at T_{+30}.

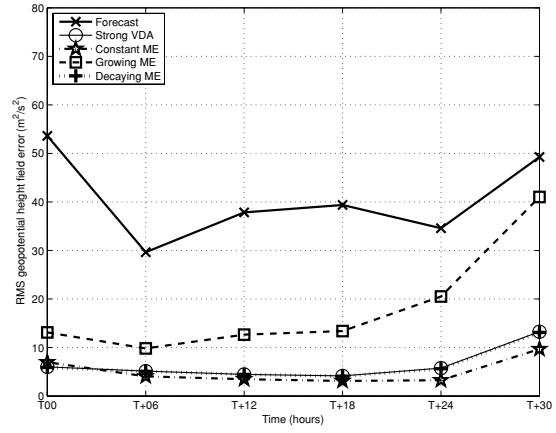
All of the above results do indicate that the choice of ME is not only highly dependent on the source of the error, but also the advection scheme, that is, the model (both forward and adjoint) used. In other words, higher accurate models, such as those with constrained van Leer and PPM advection schemes could be less



(a) Unconstrained van Leer



(b) Constrained van Leer



(c) PPM scheme

Fig. 10. RMS geopotential height field (m^2/s^2) errors before and after strong constraint, as well as weak constraint VDA for the Wrong Coriolis experiment, described in Section 5.3. (a) and (b) use the unconstrained and constrained van Leer schemes and (c) uses the PPM scheme.

sensitive to certain forms of modelling ME. Though our results suggest that the constant ME could be used to model systematic errors as considered above, but depending on the nature the model, the performance with different ME forms (when compared to strong VDA) could significantly vary.

5.4. Estimation of model error parameter

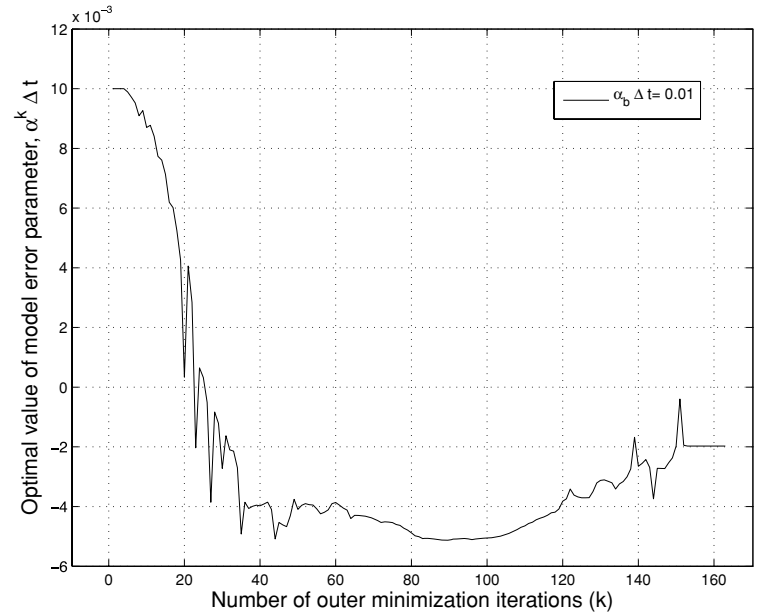
In the weak constraint VDA results presented so far, we arbitrarliy assigned certain values to the model error parameters β, γ, to model the evolution of ME as decaying, constant or growing ME. We note that in practice, it would be extremely difficult to determine both the exact nature of the ME (decaying/ constant/ growing) and values of the parameters. The above results with fixed set of parameter values indicated that usage of such values could result in no positive gains, for example, the decaying ME form leads to model error values of almost zero and therefore same results as strong VDA. On the other hand, the growing ME form leads to higher errors than the strong constraint VDA. Therefore, the choice of particular forms of ME considered needs to be generalized, such that the parameter value is 'adaptively' found while minimizing the weak constraint VDA cost functional. In Section 4.1, we mentioned such a procedure of optimal model error parameter estimation. To that end, we modified the cost functional by addition of a quadratic penalty term given in eq. (18). This also implies that the form of ME 'varies' as minimization is carried out, therefore we iteratively find 'optimal' initial conditions $\mathbf{x}_0$ and $\boldsymbol{\eta}_0$, as well as optimal parameter value α after every inner and outer iterations. We should note that this parameter is not part of the model equations (in our case, the SWE equations system), rather it is part of the ME formulation in the context of weak constraint VDA, therefore the process is 'dynamically' estimating a new optimal parameter value, in order to fit the model trajectory with the observations. Hence as the observational misfit decreases with minimization iterations, a 'saturation' occurs, that is, a steady value for the parameter could be obtained, after every cycle of inner and outer iterations.

Finding an optimal value for the penalty coefficient b_α in the cost functional $\mathcal{J}_\alpha$, eq. (18), could be laborious, as described in Zhu and Navon (1999). If from the start of minimization, the parameter values are far from the feasible region, the minimization is highly sensitive to the penalty coefficient. However, if from the beginning, we are already in the feasible region, the minimization could be insensitive to the choice of the penalty coefficient. In such a situation, if we increase the penalty term, for example, by an order of magnitude, there could be no change in parameter estimate, except for a slowing down of minimization convergence rate.

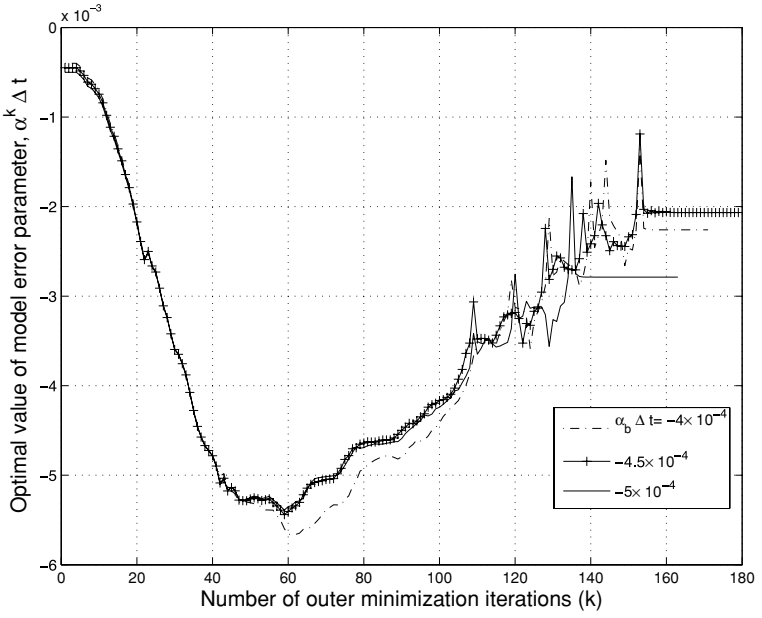
For the initial guess α_b, we specified the same value of the ME parameter as was used in the growing ME, that is, $\gamma\,\Delta t = 0.01$ (with $\Delta t = 450$ s as mentioned earlier in Sections 2.2 and 4). This value was indeed in the feasible region as our cost functional was insensitive to the choice of the penalty coefficient b_α, which we set to 10^{-2}. Further decreasing b_α, that is, increasing the penalty cost functional $\mathcal{J}_\alpha$, from this value only slowed down the minimization, but yielded the same results. The inner iteration termination criteria in Section 4.1, TOL_{inner} and $nMAX_{\mathrm{inner}}$ were respectively specified to be equal to 0.8 and 20. In Fig. 11(a) we plot the variation of the ME parameter α, versus number of iterations, the experiment setting is same as in the PPM perfect 2.5×2.5, Section 5.1 (here we show results with the constrained van Leer, similar results were obtained with the unconstrained van Leer). Note that the parameter value monotonically decreases before saturating, at around -2×10^{-3}. Based on this result, and given the previous observation that the constant ME form, for which the parameter value is simply equal to zero, we chose initial guess values $\alpha_b\ \Delta\ t$ that are 'smaller' in magnitude than both $\gamma\,\Delta t$ and currently obtained 'optimal' estimate of -2×10^{-3}, in the vicinity of -4.5×10^{-4}. The goal of these additional VDA experiments is to see if indeed we obtain the same saturated value for the ME parameter, and is there any impact of the choice of different α_b on the estimated parameter value. In Fig. 11(b), we compare the evolution of the parameter estimates in the three additional VDA experiments. We observe that indeed the final saturated value is 'close' to the previously obtained saturated value of -2×10^{-3}. We do not expect the same value to be obtained from all the experiments, due to the round off errors in minimization, particularly towards the termination of minimization. Further, when we specified initial guess $\alpha_\mathrm{b}\Delta t$ larger in magnitude than those mentioned, for example, $\alpha_\mathrm{b}\Delta t = |\beta\,\Delta| = 0.2$ which was used in the decaying ME form (Section 4), we obtained a similar saturated value, but at the expense of a few hundred more minimization iterations. The RMSE from all these various adaptive parameter estimation experiments are plotted in Fig. 12, these results compare very well with the constant ME results. In the current experimental setting, the source of model error, being only the difference in advection scheme, we would expect that a constant form for the ME to work best, hence the saturated value for the model error parameter estimate would be small. This generalized approach to obtain the value of the ME parameter provides a flexible approach to modelling ME, which deserves further investigation.

6. Summary and conclusions

In this paper, we study the impact of including a systematic (or, deterministic) ME term, in the framework of weak constraint VDA. Following VPLD04, a continuous (in time) differential equation form for the evolution of ME has been considered; to that end, three forms—decreasing, constant and growing models—were tested, with respect to the strong constraint VDA, which does not have any treatment of ME. Note that these are very simple forms of ME, which are independent of the model state. In addition, we proposed and tested a model error parameter estimation methodology that generalized the above mentioned three ME forms.



(a)



(b)

Fig. 11. Variation of the ME parameter, as the weak constraint VDA (using constrained van Leer advection scheme) cost functional is minimized, for parameter estimation described in Section 5.4; (a) starting with an initial guess for the parameter, that is same as the growing ME, $\gamma \Delta t = \alpha_b \Delta t = 10^{-2}$, considered in Section 4 and (b) ME parameter estimation starting with initial guesses for the parameter value that are slightly perturbed from an arbitrarily small initial value of -4.5×10^{-4}.

Discretization errors in numerical models have been considered to contribute to ME. As pointed in LDS05, one way to account for such errors would be to increase model resolution, but if the observations are only available at a certain resolution, it is questionable, whether a highly resolved model will provide lesser forecast errors. Here, we considered an alternative procedure, where by the observations are obtained using a high accurate scheme, such as the PPM scheme, but the model uses lower accurate, advection schemes (unconstrained or constrained van Leer), which differ in their diffusivity properties (via, applying different slope limiters). Indeed as expected, amongst the two van Leer schemes, the higher accurate, constrained van Leer scheme yielded lower forecast errors, with no assimilation, Fig. 3. Then in assimilation experiment, PPM perfect 2.5 × 2.5 we show that it is possible to account for discretization error of van Leer advection schemes by including for them a ME term; more than 50% reduction in forecast RMSE was obtained. Therefore, changing the formulation of numerical discretization scheme,

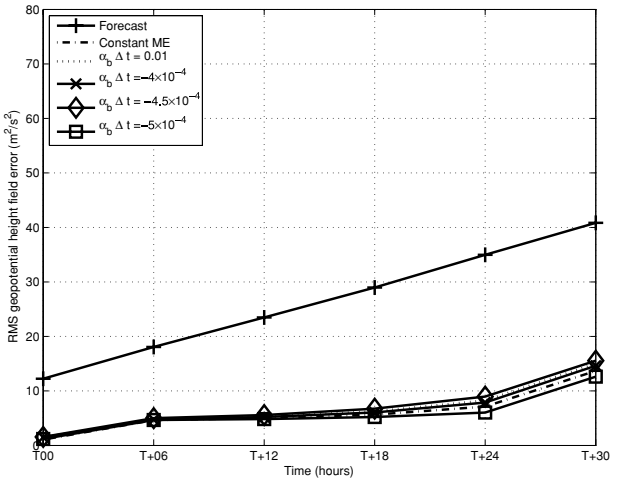


Fig. 12. RMS geopotential height field (m^2/s^2) errors before and after weak constraint VDA (using constrained van Leer advection scheme) for the adaptive model error parameter estimation experiment (Section 5.4). We compare the forecast RMSE for the constant form of the ME and that obtained with an optimal parameter values estimated after minimization. The ME parameters were adaptively estimated as shown in (a) and (b).

and weak VDA is another possible way to decrease model error. Comparable results were shown in the PPM perfect 5 × 5 experiment, where the number of observational degrees of freedom was reduced to one-fourth of the model state. The constant ME for the unconstrained van Leer scheme consistently yielded lesser forecast RMSE within every experiment (as expected, lesser observational resolution degraded the VDA results amongst different experiments). However, for the constrained van Leer scheme, the improvement using ME was not as significant as in the case unconstrained van Leer scheme, though for every experiment it was better than the unconstrained van Leer scheme. Therefore, we observed that the nature of the results obtained using different forms of ME is not only highly dependent on the source of the error, but also on the particular advection scheme, that is, the model used. Hence different settings of the model could need different ME forms to achieve higher reduction in forecasting errors.

We conducted another set of experiments (Wrong Coriolis) in which, a systematic modelling error was introduced by incorrect specification of Coriolis parameter. Such a source of ME consistently degraded forecasts. Forecast RMSE after weak constraint VDA showed contrasting differences in the performance of different ME forms. The PPM scheme was used to generate the observations, therefore it was not used for assimilation in PPM perfect experiments. However, we did use it in the wrong Coriolis experiments to study whether we can we use any of the model error forms to account for the systematic error from Coriolis parameter misspecification alone (since, in this case, the advection scheme is same for model and observation trajectories). Results with the decaying form and strong VDA were always coincident, for all the experiments presented, this was shown to be due to the specified decay rate. The growing form exhibited higher errors than the no ME, that is, strong VDA, this was shown to be due to the assigned growth rate. In all cases, the constant ME performed better than the strong VDA.

In our experiments, we had an a priori knowledge of the source of model error (discretization error, misspecified Coriolis parameter), and therefore we choose certain simple parameterized forms for the ME to study which one of them is able to account for the ME in VDA framework. To generalize these various forms of proposed model erros, we suggested a parametrization of the model error, whereby the model error form or the parameter value would be adaptively determined during minimization of the weak constraint VDA. Forecast RMSE results with the optimal parameter are very similar to the constant ME, given the nature of systematic MEs studied in this paper. Therefore, these results indicate that one could parametrize ME in the proposed framework, and also estimate the parameters while minimizing observational misfit.

In this study, the model error was assumed to be systematic; however, the random or stochastic component (Buizza et al., 1999; Tsyrulnikov, 2005) of model error could be significant, and its estimation and possible removal (Dee and Da Silva, 1998) in either an ensemble or variational data assimilation (possibly using high resolution advection schemes, such as the van Leer and/or PPM schemes presented here) system deserve further exploration, particularly, using 3-D plus time models such as the FV-GCM.

7. Acknowledgments

SA would like to thank Dr. Akhil Datta-Gupta (PE, TAMU), who supported him during 2007–08 as his post-doc. IMN acknowledges support from the National Science Foundation, through grants ATM-9731472 and ATM-0201808. 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 were obtained from the ECMWF data server. Dr. Steve Cohn (GMAO, NASA) provided helpful comments and suggestions on our numerical experiments. Comments and remarks from the two anonymous reviewers improved the formulation, analysis and presentation of results.

References

Akella, S. and Navon, I. M. 2006. A comparative study of the performance of high resolution advection schemes in the context of data assimilation. *Int. J. Num. Meth. Fluids.* **51**, 719–748.

Buizza, R., Milleer, M. and Palmer, T. N. 1999. Stochastic representation of model uncertainties in the ECMWF ensembleprediction system. *Q. J. R. Meteorol. Soc.* **560**, 2887–2908.

Byun, D. and Schere, K. L. 2006. Review of the governing equations, computational algorithms, and other components ofthe models-3

Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* **59**, 51–77.

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

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

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

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

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

Griffith, A. K. and Nichols, N. K. 2000. Adjoint methods in data assimilation for estimating model error. *Flow Turb. Comb.* **65**, 469–488.

Hourdin, F., Talagrand, O. and Idelkadi, A. 2006. Eulerian backtracking of atmospheric tracers. II, Numerical aspects. *Q. J. R. Meteorol. Soc.* **132**, 585–603.

Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK, 364 pp.

LeDimet, F.-X. and Shutyaev, V. 2005. On deterministic error analysis in variational data assimilation. *Nonlin. Proc. Geophy.* **12**, 481–490.

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

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

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

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

Lorenc, A. 1986. Analysis methods 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. *Int. J. Num. Meth. Fluids.* **40**, 435–444.

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 J. Opt.* **1**, 358–372.

Navon, I. M. 1998. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Oceans* **27**, 55–79.

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

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. In: *Computational Science- ICCS 2005, Part 2, LectureNotes in Computer Science 3515*, (ed. V. S. Sunderam), Springer-Verlag, Berlin, 837–844.

Nichols, N. K. 2003. Treating model error in 3-D and 4-D data assimilation (pp. 127–135). In: *Data Assimilation for the Earth System* (NATO science series IV: Earth and Environment Science, eds. R. Swinbank, V. Shutyaev and W. A. Lahoz), Kluwer Academic Publishers, The Netherlands, 377pp.

Suarez, M. J. and Takacs, L. L. 1996. *Documentation of the ARIES/GEOS dynamical core*. NASA Technical Memorandum 104606, NASA.

Trémolet, Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**, 2483–2504.

Trémolet, Y. 2007. Model error estimation in 4D- Var. *Q. J. R. Meteorol. Soc.* **133**, 1267–1280.

Tsyrulnikov, M. D. 2005. Stochastic modelling of model errors: a simulation study. *Q. J. R. Meteorol. Soc.* **131**, 3345–3371.

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

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

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* **44A**, 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 in spherical geometry. *J. Comp. Phys.* **102**, 211–224.

Zhu, Y. and Navon, I. M. 1999. Impact of parameter estimation on the performance of the FSU Global Spectral Model using its full- physics adjoint. *Mon. Wea. Rev.* **127**, 1497–1517.

Zupanski, D. 1999. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.* **125**, 2274–2292.

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

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

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