MODEL ERROR ESTIMATION EMPLOYING ENSEMBLE
DATA ASSIMILATION APPROACH

Dusanka Zupanski
and
Milija Zupanski

Cooperative Institute for Research in the Atmosphere (CIRA)
Colorado State University, Fort Collins, Colorado, 80523-1375

Submitted 18 May 2003
Revised 8 March 2004
Revised 10 November 2004
Revised 1 February 2005

(5 tables, 10 figures)

(Submitted to Monthly Weather Review)

Corresponding author address:
Dusanka Zupanski, Cooperative Institute for Research in the Atmosphere/Colorado State University, Fort Collins, Colorado, 80523-1375; E-mail: Zupanski@cira.colostate.edu
Abstract

A methodology for model error estimation is proposed and examined in this study. It provides estimates of the dynamical model state, the bias, and the empirical parameters by combining three approaches: (1) ensemble data assimilation, (2) state augmentation, and (3) parameter and model bias estimation. Uncertainties of these estimates are also determined - in terms of the analysis and forecast error covariances - employing the same methodology.

The model error estimation approach is evaluated in application to Korteweg-de Vries-Burgers (KdVB) numerical model within the framework of Maximum Likelihood Ensemble Filter (MLEF, Zupanski 2005). Experimental results indicate improved filter performance (in terms of analysis errors and innovation statistics) due to model error estimation. On the other hand, neglecting model errors - either in the form of an incorrect model parameter, or a model bias - has detrimental effects on data assimilation, in some cases resulting in filter divergence. The innovation statistics also indicate that the estimated uncertainties are reliable.

Although the method is examined in a simplified model framework, the results are encouraging. It remains to be seen how the methodology performs in applications to more complex models.
1. Introduction:

Atmospheric and other similar models (e.g., oceanic, hydrological, ecological) still have large but unknown deficiencies. This is one of the major obstacles for further research progress, since it can seriously undermine research results by a large unknown error. Even though it is currently not possible to fully eliminate model deficiencies, they can be appropriately taken into account using advanced data assimilation methods. These methods are capable of extracting significant information about the model error from the observations. Data assimilation methods, based on classical Kalman filtering (Kalman 1960) and weak constraint variational approach (Sasaki 1970), are now reaching enough maturity to successfully estimate and correct model errors in realistic applications. For example, a bias correction approach was successfully applied within Kalman filtering by Dee (1995), Dee and da Silva (1998), Martin et al. (1999), Kalnay (2002), Nichols (2003). Weak constraint four-dimensional variational (4DVAR) method was employed in Derber (1989), DeMaria, and Jones (1993), Bennett et al. (1993, 1996), Zupanski (1993), Zupanski (1997), Uboldi and Kamachi (2000), Vidard et al. 2000, D’Andrea and Vautard (2000), Sasaki (2003), etc. State augmentation method (e.g., Jazwinski 1970) for estimating a systematic model error has been used successfully in both Kalman filter and 4DVAR data assimilation schemes (Dee and da Silva 1998; Griffith and Nichols 1996, 2001; Martin et al. 2002). Recent studies indicated substantial benefits of model error adjustment in cases of assimilation of precipitation observations (e.g., Hou et al. 2000, 2001; Zupanski et al. 2002a,b; Lee and Lee 2003). The bias correction technique of Derber (1989) is also being successfully applied to radiative transfer models (e.g., Derber and Wu 1998).

Ensemble data assimilation (EnsDA) methods are rapidly emerging as new generation data assimilation techniques. The EnsDA methodologies are also referred as Ensemble Kalman Filter (EnKF)-like approaches (e.g., Hansen 2002). Starting from pioneering work of Evensen (1994), many techniques have been proposed as different variants of EnsDA (e.g., Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Keppenne 2000; Mitchell and Houtekamer 2000; Anderson 2001; Bishop et al. 2001; van Leeuwen 2001; Reichle et al. 2002a,b; Whitaker and Hamill 2002; Tippett et al. 2003; Zhang et al. 2004; Ott et al. 2005; Zupanski 2005). More discussion about the differences in the current approaches can be found in Evensen (2003), Anderson (2003) and Tippett et al. (2003). Independently of the specific variant of the EnsDA,
or the EnKF-like approach, the state estimate (or analysis) is obtained in an ensemble-spanned subspace, defined by a limited number of forecast model realizations (ensembles). In addition, ensemble based data assimilation techniques provide a consistent mathematical formalism for updating the analysis and the forecast error covariance matrices, employing information from the observations. Due to the ease of the ensemble framework, a prediction model (e. g., atmospheric, oceanic, carbon, and other similar models) of any complexity can be used in EnsDA, employing essentially the same algorithm.

EnsDA methodologies, however, often assume that forecast models are perfect (i. e., model error is neglected). There are problems with this assumption, even in applications to relatively simple models, using simulated observations, as it will be shown in this study. One possible way to relax the perfect model assumption is to simulate model error as a random or a serially correlated noise with prescribed error covariances (e. g., Hemink et al. 2001). A similar approach is used in Mitchell et al. (2002), where a balanced random noise, generated from a multivariate probability distribution function, is included into the model. Hansen (2002) applied an EnKF and a probabilistic 4DVAR to account for model error in a chaotic model. In Reichle et al. (2002a) a serially correlated model error was estimated employing an EnKF approach and the state augmentation method using a prescribed time-correlated model error correlation matrix.

Parameter estimation approaches are simplifications of the more complex model error estimation methods. The parameter estimation methods have previously been examined mostly within the frameworks of variational or Kalman filtering methods (e. g., Bennett and McIntosh 1982; Cohn and Dee 1988; Le Dimet and Navon 1988; Panchang and O’ Brian 1989; Zou et al. 1992; Muccino and Bemnnet 2002; also see the review by Navon 1997 for more information). The EnKF-like approaches are also emerging as new parameter estimation techniques (Mitchell and Houtekamer 2000; Anderson 2001; Kivman 2003; and Jones et al. 2004). These studies indicated that the parameter estimation is an effective approach to model error estimation when the amount of observed information and the ensemble size are not sufficient to constrain a more complex model error (e. g., serially correlated error). There is a theoretical framework, examined mostly within variational and Kalman filter approaches, which can be used for better understanding how and why the parameter estimation method should work. The framework is based on determining observability, that is, the amount of observed information about a parameter being estimated, and identifiability, that is, the uniqueness of the parameter being
estimated (e.g., Cohn and Dee 1988; Navon 1997). Only in cases when the estimated parameters are observable and identifiable the parameter estimation method is expected to work properly. When dealing with EnsDA approaches the observability and the identifiability criteria are dependent on the ensemble size, thus requiring deeper understanding of the links between the ensemble size and the number of degrees of freedom (DOF) in the control variable and the observations.

The studies mentioned above have been just initial steps toward resolving the problem of model error estimation. This problem remains fundamental for future progress in variational, Kalman filter and EnsDA methods, in ensemble forecasting (Buizza et al. 1999) and in predictability (Orrell 2003).

The methodology proposed here incorporates three approaches: (1) EnsDA, (2) state augmentation, and (3) parameter and model bias estimation. While there are more studies dealing with each of the three approaches separately, experience available so far in dealing with all three components as parts of the same methodology is much more limited. Here we list some of the pioneering studies incorporating these three components. For example, in Mitchell and Houtekamer (2000), Anderson (2001), Kivman (2003), and Jones et al. (2004), parameter estimation approaches were examined. In Reichle et al. (2002a), a serially correlated model error of a soil hydrology model was estimated.

The EnsDA approach used in this study is the Maximum Likelihood Ensemble Filter (MLEF) proposed by Zupanski (2005). This article can be considered as a continuation of the previous work by Zupanski (2005), since the same ensemble data assimilation approach (MLEF) and the same model, Korteweg-de Vries-Burgers (KdVB) are used. Although the MLEF approach is selected for this study, its use is not of essential importance. This is because the minimization, introduced by the MLEF, converges in a single iteration, since the observation operator is linear. Therefore, the proposed model error estimation is applicable to other EnKF-like approaches as well.

The paper is organized as follows. Serially correlated model error is described in section 2. State augmentation method, and its application within the framework of the EnsDA are explained in section 3. Experimental design is presented in section 4, and the results are discussed in section 5. The conclusions and future work are given in section 6.
2. Serially correlated model error (bias)

We assume that the model error is a serially correlated first order Markov process variable (e.g., Daley 1992). Different variants of a serially correlated model error have been successfully used in variational (Derber 1989; Bennett et al. 1993, 1996; DeMaria, and Jones 1993; Zupanski 1993; Griffith and Nichols 2001; Zupanski 1997) and Kalman filter (e.g., Dee 1995; Dee and da Silva 1998; Martin et al. 1999; Nichols 2003) approaches.

Time propagation, from time $t_{n-1}$ to time $t_n$, of the model state variable $x_{n,i}$ can be defined as

$$x_n = M(t_n, t_{n-1})(x_{n-1}) + \Phi_n + q_n ; \quad n = 1, N_{\text{max}} ,$$

where $M$ is a non-linear forecast model operator, $n$ is the model time step index, and $N_{\text{max}}$ is the maximum number of model time steps in a data assimilation interval (cycle). Vector $\Phi_n$ is an unknown model error correction term, a variable to be estimated. Since we do not know the future estimate of $\Phi_n$ when integrating the model (1), we use the most recent estimate available (explained in appendix A). Vector $q_n$ is a white random noise with Gaussian distribution. In common EnsDA applications $q_n$ is either neglected, or kept, but used as a prescribed vector. The effect of $q_n$ is neglected in this study, and in the subsequent derivations, since the focus is on estimating $\Phi_n$. In addition, we assume that $\Phi_n$ is a special case of a first order Markov process variable, defined on a coarse time scale (more details are given in appendix A):

$$\Phi_n = \alpha^n \Phi_0 + (1 - \alpha^n) G(b) ; \quad n = 1, N_{\text{max}} ,$$

where vector $b$ is the model bias parameter and the non-linear mapping, which transforms the vector $b$ into the model state space, is denoted $G$. To distinguish between $\Phi_n$ and $b$, we adopt notations “bias $\Phi_n$” and “bias parameter $b$”, respectively. The parameter $\alpha$ is a constant measuring relative influences of the current bias and the previous bias.

Assuming that $\Phi_n$ is known, and that $b$ is estimated, (1) and (2) describe time evolution of both the model state and the model error. Thus, the estimation of the bias $\Phi_n$ reduces to the
estimation of $b$, a vector assumed constant over a period of time (data assimilation interval). In this study we assume that both $\Phi_o$ and the guess of $b$ (i. e., the value of $b$ before optimization) are zero in the first data assimilation cycle, since no better guess is available. In any other cycle these values are determined from the previous data assimilation cycle (explained in more detail in appendix A).

The model bias parameter $b$ can be defined in the same space as the model state variable $x$. This allows for possible corrections of model state variables in each grid point and for each component (e. g., temperature, pressure, wind, humidity). This is of special importance for limited area models, where assumption (2) can be used to estimate and correct the error in the horizontal boundary conditions (e. g., Zupanski 1997). On the other hand, defining the serially correlated model error in the same space as the model state vector could be impractical for complex atmospheric models, since the number of DOF in the control variable would become very large (of the order of $10^8$-$10^{10}$). In such cases, an appropriate choice of the operator $G$ may be needed to reduce the number of DOF of the bias estimation problem. These problems will be examined in this study.

3. State augmentation method

State augmentation approach will be used in this study to estimate serially correlated model error as well as to estimate unknown parameters. Let us define a new state vector $z$ by augmenting the standard model state $x$ with the model bias parameter $b$ and with the vector $\nu$ of empirical parameters. Assuming also that both $b$ and $\nu$ remain constant over a period of time, we can include time subscript $n$ and write (e. g., Jazwinski 1970)

$$b_n = b_{n-1} = b_0 = b, \quad \text{and} \quad \nu_n = \nu_{n-1} = \nu_0 = \nu,$$

and combine (1)-(3) into the single augmented dynamical system

$$z_n = \begin{bmatrix} x_n \\ b_n \\ \nu_n \end{bmatrix} = \begin{bmatrix} M_{n-1} & (1-\alpha^n)G & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} x_{n-1} \\ b_{n-1} \\ \nu_{n-1} \end{bmatrix} + \begin{bmatrix} \alpha^n \nu \\ 0 \\ 0 \end{bmatrix} \Phi_0 = S_{n-1}(z_{n-1}) + \alpha^n \Phi_0,$$
where the nonlinear operator $S$ describes time evolution from $t_{n-1}$ to $t_n$ of the augmented state variable $z_{n-1}$, and $\alpha^n \Phi_0$ is a forcing term. The bias parameter $(b_n)$ is defined as $b$ in (2) and in appendix A, which is the most recent estimate of the bias parameter. The initial bias $\Phi_0$ is defined using the bias parameter $b^*$ (Appendix A), which is the bias parameter estimated previously (in the previous data assimilation cycle). Similarly, parameter $\nu_n$ is defined as the most recent estimate of this parameter.

By introducing operator $F_{n,0}$, that describes model state evolution from the initial time $t_0$ to time $t_n$, we can write (4) in a more compact form as

$$z_n = F_{n,0} (z_0, \alpha, \Phi_0).$$

(5)

Let us also assume that $y_n$ is the vector of observations at time $t_n$, and that the observations have white Gaussian errors $\varepsilon_n$ with covariance $R_n = \langle \varepsilon_n \varepsilon_n^T \rangle$, where $\langle \rangle$ denotes mathematical expectation. Assuming $H_n^x$ is a standard non-augmented observation operator, mapping the vector $x_n$ into the observation vector $y_n$, we can define the augmented observation operator $H_n$ and use

$$y_n = [H_n^x \quad 0 \quad 0] z_n + \varepsilon_n = H_n (z_n) + \varepsilon_n$$

(6)

to transform the model state vector $z_n$ into the observation space. Eqs. (5) and (6) describe the model and the data (observations), the two necessary ingredients of standard data assimilation (e.g., Tarantola 1987). In addition, by omitting index 0 and defining $z = z_a = (x, b, \nu)$ as a control variable, the data assimilation problem can be solved as in the case without state augmentation. In further derivations we employ the formalism of the MLEF approach of Zupanski (2005).

Assuming $P_f$ and $P_a$ are the augmented forecast and analysis error covariance matrices, respectively, and $F$ is a linearization (i.e., Jacobian) of $F$, we can write

$$P_{f+1}^{z+1} = \left[ F(P^z_f)^{1/2} \right] \left[ F(P^z_f)^{1/2} \right]^T \approx \left[ F(z + p_t) - F(z) \right] \left[ F(z + p_t) - F(z) \right]^T,$$

(7)
where \( k \) denotes a data assimilation interval, and vectors \( p_i \) are columns of \( \left(P^k_a\right)^{1/2} \). Subscripts in \( F \) have been excluded for simplicity. Since the linear operator \( F \) is replaced by the difference of perturbed and un-perturbed non-linear forecasts \( F \), the equality in (7) is only approximate, valid for small perturbations (of magnitudes comparable to typical analysis errors).

We minimize functional \( J \) defined as

\[
J(z) = \frac{1}{2} [z - z_h]^T P^{-1}_f [z - z_h] + \frac{1}{2} [y - H(z)]^T R^{-1} [y - H(z)],
\]  

(8)

by introducing the following change of variable

\[
z - z_h = P^{1/2}_f (I + A)^{-T/2} \zeta,
\]

(9)

where the matrix \( A \) is defined in ensemble subspace as

\[
A = P^{T/2}_f H^T R^{-1} H P^{1/2}_f = (R^{-1/2} H P^{1/2}_f)^T (R^{-1/2} H P^{1/2}_f). 
\]

(10)

Finally, the minimization of \( J \) is performed employing the gradient, calculated as

\[
g_{\zeta} = (I + A)^{-1} \zeta - (I + A)^{-1/2} [R^{-1/2} H P^{1/2}_f]^T R^{-1/2} [y - H(z_h + P^{1/2}_f (I + A)^{-T/2} \zeta)].
\]

(11)

The vector \( \zeta \) is the control variable defined in the minimization space, \( z_h \) is a background estimate of \( z \), and \( R = R_n \) (we assume that the observation errors do not change in time, for simplicity).

Equation (6) indicates that the observation operator \( H \) is rank deficient, with non-zero values corresponding to the \( x \) component only. This is because observations for the model error components of \( z \) (components \( b \) and \( v \)) are not available. A consequence is that the model error is only influenced by the observations through the cross covariance, describing correlations between the error of the model state estimate and the error of the model bias estimate, that is \( P^{f}_{ab} \).
part of the full covariance \( P_f \). Therefore, the actual estimation of model bias is only possible if these cross-correlations are non-zero. Note that \( P_{sf}^f \) is not neglected in this study. Neglecting \( P_{sb}^f \) has a consequence of being able to only simulate the impact of model error through a prescribed model error covariance matrix (often denoted \( Q \)).

As mentioned before, the model error estimation procedure, described above, should be applicable not only to the MLEF, but to also to other EnKF-like approaches. These include the Monte Carlo EnKF (e.g., Evensen 1994; Houtekamer and Mitchell 1998), Ensemble Transform Kalman Filter (ETKF) of Bishop et al. (2001), Ensemble Adjustment Kalman Filter (EAKF) of Anderson (2001), the Local Ensemble Kalman Filter (LEKF) approach of Ott et al. (2005), etc. The differences between various EnKF-like approaches do not impact the model error estimation methodology, since the equivalence between the Kalman gain and inverse Hessian/gradient calculation (as shown in Zupanski 2005) is still valid in the augmented state variable approach.

4. Experimental design

a. Model

We will perform data assimilation experiments with a numerical version of Korteweg-de Vries-Burgers (KdVB) model

\[
\frac{\partial u}{\partial t} + uu + 6u^3 = \nu \frac{\partial^2 u}{\partial x^2},
\]

where \( u \) is a non-dimensional model state vector (used to define control variable component \( x \)), and \( \nu \) is a diffusion coefficient. The 4-th order Runge-Kutta scheme is used for time integration (Marchant and Smyth 2002). The model domain has dimension 101, with the grid-spacing \( \Delta x = 0.5 \) non-dimensional units, and the time-step is \( \Delta t = 0.01 \) non-dimensional units. The periodic boundary conditions are used. In the experiments presented the diffusion coefficient is given one of the two values \( \nu = 0.07 \) or \( \nu = 0.20 \).
b. Observations

The observations are defined by perturbing the true state (12), where perturbations are drawn from a white noise Gaussian distribution. The observation error covariance \( R \) is diagonal (e.g., variance), with prescribed elements equal to a constant \( \sigma_{\text{obs}}^2 = 0.05^2 \). The observation number varies in the experiments. In most experiments only 10 irregularly spaced observations are assimilated, targeted to sample the crests of the two waves as in Zupanski 2005. Full set of observations (101 per data assimilation cycle) is also used in some experiments to examine the impact of additional information given by the observations. The observations are always defined at the end of the data assimilation interval, corresponding to 200 model time steps (the integration time in each cycle is 2 non-dimensional units). In all experiments presented 100 data assimilation cycles are performed; however, when appropriate, the results from later data assimilation cycles (up to 200) are discussed. In Figure 1 we illustrate the true model solution and observations used in the experiments with the reduced number of observations. The experiments with full set of observations use the same true model solution, but observations at each grid point are used (figure not shown).

The value of diffusion coefficient in Figure 1 is \( \nu = 0.07 \). As Figure 1 indicates, the amplitude of the model state solution is decreasing with time, due to diffusion, which might contribute to stabilizing the filter performance by reducing the RMS errors of the estimated solution, in all experiments. Relative comparisons between the experiments (if performed under the same conditions) are still valid.

The observation operator chosen for this study is a linear (identity) transformation operator, defined as \( H(u) = u \). The choice of observation operator does not significantly impact the model error estimation results; for the cases of non-linear observation operators, more minimization iterations are needed (2-3 instead of 1). The impact of different observation operators on the minimization performance is examined in Zupanski (2005).

c. Experiments

1) COLD START PROCEDURE
Control variable (initial conditions, model bias parameter, and empirical parameters) and ensemble perturbations have to be initialized in the first data assimilation cycle, that is, a cold start procedure has to be performed. This initialization is not needed in subsequent cycles, since the data assimilation scheme updates both the control variable and the ensemble perturbations. The cold start procedure used here is similar to the procedures commonly used in other EnKF-like approaches. Due to the specific, less common, extension of the EnsDA methodology to include model bias and parameter estimation, we will explain the cold start procedure in more detail.

Regarding the initial conditions component of the control variable, two possibilities are used: perfect initial conditions and imperfect initial conditions. Perfect initial conditions are used to initialize the true KdVB model. The perfect initial conditions are defined as a two-soliton analytic solution of the Korteweg-de Vries (KdV) equation (Vvedensky 1993). For data assimilation purposes, we assume that the knowledge about the truth is not available, and use imperfect initial conditions. The imperfect initial conditions are created using a slightly different KdV analytical solution. Namely, the difference between the perfect and the imperfect initial conditions is roughly 10% in the amplitudes of the soliton waves and there is a lag in time corresponding to 100 time steps of the numerical KdVB model (half of the data assimilation interval).

The model bias component of the control variable is initialized with zero (the actual bias is, however, different from zero in the bias estimation experiments). The third component of the control variable, the diffusion parameter, is initialized with values 0.07 or 0.20, depending on the particular experiment, as explained later.

Initial ensemble perturbations are imposed upon the initial values of the control variable. These perturbations are defined as white noise Gaussian random variables for all components of the control variable. The standard deviation of the perturbations is set to 10% - 20% of the typical initial magnitude of the corresponding control variable. For the model bias, the standard deviation of the ensemble perturbations could not be set to a fraction of the typical value of the control variable, since its initial magnitude is zero. In this case, the standard deviation is defined as a fraction (1%) of the standard deviation used for generating perturbations for the initial conditions. The initial ensemble perturbations are only defined for the control variable components needed in the particular experiment. For example, in the data assimilation
experiments with the perfect model (referred as \textit{correct_model} experiments) and in the experiments with neglecting model error (referred as \textit{neglect_err} experiments), only the initial conditions are perturbed. Likewise, in the experiments with model bias estimation the initial conditions and the model bias are perturbed; in the experiments with parameter estimation the initial conditions and the diffusion parameter are perturbed. Finally, the initial forecast error covariance is defined by integrating the forecast model (5) from a set of the initial ensemble perturbations two cycles prior to the first observation time. In this way the dynamical constraint of the forecast model is added on the initial random perturbations.

Two sets of data assimilation experiments are performed to examine the impact of (\textit{i}) \textit{parameter estimation}, as one of the simplest, and the impact of (\textit{ii}) \textit{model bias estimation}, as one of the most complex approaches to model error estimation.

2) \textbf{PARAMETER ESTIMATION}

Parameter estimation experiments are designed to assess the impact of incorrect diffusion in the KdVB model. In the correct model, which is used to define observations, as well as in one of the control experiments explained below, the diffusion coefficient is $\nu=0.20$. In the erroneous model the diffusion coefficient is $\nu=0.07$. The following experiments are performed: (\textit{i}) \textit{correct_model}, where the model with correct diffusion is used (defines the upper limit of success of model error estimation), (\textit{ii}) \textit{neglect_err}, where erroneous model is used, but the error was neglected (used as the lower limit of success of model error estimation method), and (\textit{iii}) \textit{param_estim}, where diffusion parameter is estimated, staring from an incorrect value. The diffusion parameter is estimated by augmenting the control variable to include the unknown parameter. This results in the increase of the control variable dimension from 101 to 102. All experiments are performed employing 10 ensemble members. In basic experiments, 10 observations are assimilated in each data assimilation interval. To examine the impact of increased number of observations, additional experiments are performed using 101 observations. In all experiments, observations are defined at the end of each data assimilation interval. The parameter estimation experiments are summarized in Table 1.

3) \textbf{MODEL BIAS ESTIMATION}
For this set of experiments the model error is created as an additive model error term
\(-\eta \mathbf{u}_0\) and it is included into the KdVB model as:

\[
\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \frac{\partial \mathbf{u}}{\partial x} + 6 \frac{\partial^3 \mathbf{u}}{\partial x^3} = \nu \frac{\partial^2 \mathbf{u}}{\partial x^2} - \eta \mathbf{u}_0 ,
\]

(13)

where \(\mathbf{u}_0\) is the model state at the end of the previous data assimilation interval. The bias in (13) varies in space and from one data assimilation cycle to another. The magnitude of the bias is controlled by the constant \(\eta\). The following experiments are performed: (i) **correct_model**, where \(\eta=0\) is used (model error does not exist), (ii) **neglect_err**, where model error exists \((-\eta \mathbf{u}_0\)) but it was neglected in the experiment, and (iii) **bias_estim**, where the model error is defined as \(-\eta \mathbf{u}_0\), and the model bias parameter \(b\) is estimated in order to account for this model error term. Thus, when estimating model bias, no knowledge about the actual model error is used. In all experiments, the observations are created using the non-biased (i.e., correct) model. Only 10 ensembles and 10 observations, per data assimilation cycle, are used in all experiments. In the experiments with the erroneous model (**neglect_err** and **bias_estim**) the value of \(\eta=0.0001\) was selected. This choice of model bias magnitude had a significant effect on the data assimilation results, ultimately causing filter divergence if neglected. While it is possible that such significant effects of model bias may not be obtained in applications to atmospheric models, previous experience with realistic model bias estimation, employing variational methods, indicates that the biases of similar, or even larger magnitudes, were also detected in atmospheric models (Zupanski et al. 2002a,b). Similar bias magnitudes (approximately \(10^{-3}\) of the model state magnitude) are also used in Uboldi and Kamachi (2000).

For the case of the experiment **bias_estim**, the estimated model bias parameter \(b\) is used to correct the erroneous model (13) via (4). The value of constant \(\alpha=0.8\) is used in all bias estimation experiments, defining a smooth change from the bias of the previous cycle to the bias of the current cycle. (Note that \(\alpha\) is not used in the parameter estimation experiments, since \(b\) is not included in the control variable.) The experiments with different values of \(\alpha\) indicated that
small variations (10%-20%) of this parameter did not have a significant impact on the experimental results.

Regarding the operator $G$, two different choices, representing two limiting cases, are examined: (i) $G=I$, the identity operator, and (ii) $G=G(u)$, a state dependent operator. The first choice is the most straightforward approach, implying that the model error correction does not depend on the state $u$ and the dimension of the control vector $b$ is equal to the dimension of the model state vector. The second choice of $G$ is used to examine the impact of the control vector $b$ that has a smaller size than the model state $u$, which may be desirable in application to complex atmospheric models. One simple way to accomplish this is to define $b$ in the coarse spatial resolution, and use an interpolation operator $G$ to transform $b$ into the full space of the forecast model. In this study, rather than making $b$ tied to the grid points of the model domain, it is defined in the phase space, tied to the magnitudes of the vector $u$. Operator $G$, defined as our second choice $G=G(u)$, is described in appendix B. In the experiments presented $G$ is used to map vector $b$ (of dimension 10) into 10 categories of magnitudes of the model state vector $u$ (of dimension 101). This implies that similar magnitudes of $u$ have similar model bias corrections. For the case of the KdVB model this is a desirable property, since it makes the bias traveling with the waves. On the other hand, since the solution of the KdVB in our experiments consists of two waves (see Figure 1), the specific choice of $G$ would occasionally result in applying the same bias correction to the $u$ points having the same magnitude, but belonging to two different waves, which could potentially degrade the bias estimation results. Having these considerations in mind, the choices of the operator $G$ used in this study should be interpreted only as means to examine the impact of varying size of vector $b$. The optimal choice of the operator $G$ would depend on the properties of the forecast model and would require further studies. Note that in all experiments the number of DOF in the “real” bias ($\eta u_\eta$) was kept as large as the model state vector (dim=101) in all experiments, to simulate the conditions of complex atmospheric models. The bias estimation experiments are summarized in Table 2.

d. Validation
The experimental results will be compared and validated in terms of verification scores used in Kalman filter verifications, such as the root-mean-square (RMS) errors of the state estimate, Kolmogorov-Smirnov (K-S) test of the probability density function (PDF) of the innovations, and $\chi^2$ – test of the innovations. The RMS error measures how close is the calculated estimate to the true state, while the innovation statistics provide information about the performance of the filter itself.

The innovation statistics are calculated as follows. Normalized innovations $Y$ are defined as (e.g., Reichle et al. 2002b)

$$ Y = \frac{1}{\sqrt{N_{\text{obs}}}} [HP_f H^T + R]^{-\frac{1}{2}} [y_k - H(x_k)] , \quad (14) $$

where $y_k - H(x_k)$ are innovations, normalized by the analysis error covariance in the observational space $[HP_f H^T + R]$, and by the number of observations $N_{\text{obs}}$. Assuming $R$ has a Gaussian distribution and $P_f$ is also Gaussian at initial time (first data assimilation cycle), then for a linear forecast model $F$ and a linear observation operator $H$ the error statistics will remain Gaussian, thus resulting in the normalized innovations (14) having normal distribution $\mathcal{N}(0,1)$. In the experiments of this study we test PDFs of the components $\{y_i\}$ of $Y$ against normal distribution $\mathcal{N}(0,1)$ by calculating (K-S) statistic (e.g., Hollander and Wolfe 1973)

$$ D = \max_{1 \leq i \leq N_{\text{obs}}} \left| \frac{F(y_i)}{N_{\text{obs}}} - \frac{i}{N_{\text{obs}}} \right| , \quad (15a) $$

where $| \cdot |$ denotes the absolute value. The significance levels $\alpha$ are determined as

$$ \alpha(10\%) = 1.22 \sqrt{N_{\text{obs}}} , \quad \alpha(5\%) = 1.36 \sqrt{N_{\text{obs}}} , \quad \text{and} \quad \alpha(1\%) = 1.63 \sqrt{N_{\text{obs}}} . \quad (15b) $$

The K-S test (15) compares maximum absolute value of the difference between the empirical distribution function $\frac{i}{N_{\text{obs}}}$ and the theoretical distribution function $F(y_i)$ against a critical value, determined by the significance level $\alpha$, in order to accept or reject the hypothesis $H_0$, stating that the empirical and the theoretical distribution are the same. Even though we do not expect
dramatic departures from Gaussian statistics, due to the use of a nonlinear model \( F \) in calculation of \( P_f \), and a statistically small sample (i.e., relatively few observations) it is possible that the innovations (14) do not strictly follow the theoretical distribution \( \mathcal{N}(0,1) \). In addition, parameter and bias errors are unlikely to be Gaussian; for example we do not expect negative diffusion. Even though the estimated diffusion coefficient in the experiments of this study remained positive, it is still possible that the errors of the diffusion coefficient are not strictly Gaussian.

The \( \chi^2 \) – validation diagnostic is adopted from the definition used in Dee et al. (1995) and Menard et al. (2000). Namely, a random variable \( \chi^2 \) is defined as

\[
\chi^2 = \text{trace} YY^T.
\]

For Gaussian distribution of non-normalized innovations, a linear forecast model \( F \) and a linear observation operator \( H \), the random variable \( \chi^2 \) should have a chi-square distribution with a conditional mean equal to 1, and a variance equal to 2. As for the case of K-S statistic, one can expect only mean values of \( \chi^2 \) close to one, not necessarily equal to one. The values consistently diverging from one would indicate poor filter performance (i.e., filter divergence). As in Menard et al. (2000), and in Zupanski (2005) the conditional mean is substituted by a time mean. A 10-cycle moving average is computed, as well as the instant values of \( \chi^2 \), calculated at each assimilation cycle.

Assuming that observation error covariance matrix \( R \) is known, and the Gaussian assumptions are considered appropriate, the statistics (15) and (16) can be used as verification tools for the forecast error covariance \( P_f \).

5. Results

a. Parameter estimation

The results of the experiments \textit{correct\_model}, \textit{neglect\_err} and \textit{param\_estim}, in terms of RMS errors with respect to the true solution - defined by the numerical solution of the KdVB model (12), starting from the perfect initial conditions - are presented in Figure 2. In Figure 2a
only 10 observations are used in the experiments, while in Figure 2b, the experimental results employing 101 observations in each data assimilation cycle are presented. As the figures indicate, the model error estimation experiment performs as expected, having the RMS errors between the lower (neglect_err) and upper limit of success (correct_model). Note that the RMS errors are generally greater at the initial data assimilation cycles, due to initial adjustments in all experiments, and the RMS errors are decreasing as the number of the data assimilation cycles increases. Comparison of the two figures indicates that problems caused by the error in the diffusion coefficient cannot be eliminated by increasing the number of observations, since the experiment neglect_err shows little improvement. On the other hand, increasing the number of observations is more effective in the case when model error is estimated and corrected (param_estim), and even more in the case when model error does not exist (correct_model).

These results, if proven valid for realistic atmospheric models, may have a significant consequence on designing the future observing systems, since including additional observations, without taking a proper care of the model error, would not necessarily result in improved data assimilation results.

The innovation statistics, in terms of the PDF of the innovations, calculated for the experiments correct_model, neglect_err and param_estim are presented in Figs. 3 and 4. Figure 3 shows data assimilation results employing 10 observations per data assimilation cycle, and Figure 4, employing 101 observations per data assimilation cycle. The innovations are calculated in all observation points during 90 data assimilation cycles (from cycle 10 to cycle 100) in all experiments. As Figs. 3 and 4 indicate, the results of the correct_model experiment are in best agreement with the theoretical Gaussian distribution, the param_estim experiment is also in a reasonably good agreement, while neglecting model error (neglect_err) resulted in largest departure from the expected PDF. This is an indication of a questionable quality of the analysis and forecast error covariances in the case of an erroneous model. As for the RMS errors, comparison of Figs. 3 and 4 indicates that the model error problem cannot be eliminated by increasing the number of observations. On the other hand, the additional observations do provide an improvement in the cases when model error does not exist (correct_model), or it has been taken into account (param_estim).

For quantitative comparisons between the PDFs in Fig. 3 and 4 the results of the K-S test statistic $D$, given in (15), calculated against theoretical distribution $\mathcal{N}(0,1)$, are summarized in
Table 3. The statistic $D$ indicates the largest departures from the standard normal distribution for experiments `neglect_err` and the smallest departures for the experiments `correct_model`; the K-S statistic for the experiment `param_estim` is between the two limiting cases. Due to larger ensemble size, experiments with 101 observations per cycle generally have smaller values of $D$ than the experiments with 10 observations per cycle; this is also reflected in the smaller critical value $\alpha$ at the same significance level (5%). Further examination of Table 3 indicates that the hypothesis $H_0$ is accepted only for the experiments with the correct model. Even though one should expect only approximate Gaussian distributions, as explained in section 4.d, the fact that $H_0$ is accepted for the experiments with correct model, but rejected for the parameter estimation experiments (and also for the experiments `neglect_err`), indicates that the parameter estimation process is less than perfect. In addition, Figure 7a, discussed in more detail later, shows that the estimated values of $\nu$ are different from the true values, except for the parameter estimation experiment with 102 ensemble members and 101 observations. In order to find out if the incorrectly estimated value of $\nu$ is causing the rejection of the hypothesis $H_0$, additional K-S tests are performed and summarized in Table 4. Namely, the experimental results of two `param_estim` experiments are given in Table 4; both experiments are using 102 ensemble members and 101 observations, but in one experiment true $\nu=0.20$, and in the other true $\nu=0.07$ (Figure 7a). For comparison, K-S test results of the experiment `neglect_err` with 101 ensemble members, 101 observations, and $\nu=0.07$ (true $\nu=0.20$) are also included in Table 4 (figure is not shown for this experiment). In other words, the results of the experiments solving full rank problem and employing full set of observations are presented in Table 4. These experiments should provide the best possible results. As shown in Table 4, the hypothesis $H_0$ is now accepted for both `param_estim` experiments, but it is still rejected for the experiment `neglect_err`. These results indicate that the results of the parameter estimation can be significantly improved (in terms of the K-S statistics) by including more observations and more ensemble members, while the opposite is true for the experiment `neglect_err`.

Finally, the K-S tests, as in Tabs. 3 and 4 (and also in Tab. 5 discussed later) have also been performed for significance levels 10% and 1%. The experimental results, regarding accepting or rejecting the hypothesis $H_0$, remained the same.
The results of the $\chi^2$– innovation test are presented in Figs. 5 and 6, for the same three experiments. As before, the results employing 10 observations (Figure 5) and 101 observations (Figure 6) are presented. Note that this statistical measure, if significantly departing from the value of one, indicates the divergence of the filter. One should be aware however, that since a significance test for the $\chi^2$– statistic have not been performed, the meaning of the term “significant” has a degree of arbitrariness. Nevertheless, relative comparisons between the experiments are still valid. As in previous measures, correct_model experiment has the best performance, and neglect_err, the worst. As expected, the parameter estimation (param_estim) results are in between the two limiting cases. One can also notice in Figure 6a, a pronounced improvement of the filter performance for the experiment neglect_err, and to some extent of the experiment param_estim, as the number of cycles increases. This could be a consequence of using a highly diffusive model ($\nu=0.20$) when creating the observations. Note, that we consider the KdVB model highly diffusive only in the experiments with $\nu=0.20$ (not in all experiments).

The experimental results presented so far indicated a good performance of the proposed parameter estimation method. Let us now examine in more detail the value of the estimated diffusion parameter itself. The diffusion coefficients obtained in various parameter estimation experiments are shown in Figure 7, plotted as functions of data assimilation cycles. Figure 7a shows the experimental results employing (i) 102 ensembles and 101 observations (full rank problem, full set of observations), (ii) 10 ensembles and 101 observations, and (iii) 10 ensembles, 10 observations; the true value of the parameter (true $\nu=0.20$) is also shown. As expected, the experiment with the maximum number of ensemble members and observations recovers the value that is closest to the true value. It is somewhat surprising that the experiment with 10 observations was able to recover a value that is closer to the truth than the experiment with more observations (101). Understanding this issue needs further investigation since it is related to projection of the observed information onto the ensemble subspace. Thus, a better understanding of the observability (e. g., Jazwinski 1970; Cohn and Dee 1988; Navon 1997) in the EnsDA framework is needed. Note that this experiment, even though not perfect, was clearly better than the experiment without parameter estimation (see Figs. 4 and 6, and Tab. 3).

Examination of Figure 7b indicates that increasing the number of ensemble members (102 ensembles are used) helps to further improve the estimate of the unknown parameter (the
experiments with two true values of the coefficient, 0.07 and 0.20, are presented). As shown in Table 4, this improvement is also reflected in the K-S test results.

In summary, the experimental results of this section indicated clearly beneficial results of the parameter estimation algorithm, even in the cases with little observations, employing a small number of ensemble members. For the case of the KdVB model, as used in this study, 10 observations and 10 ensemble members were sufficient for beneficial results.

b. Model bias estimation

Model bias estimation experiments are performed to evaluate the potential of the EnsDA approach to estimate and correct model errors in the form of biases, commonly present in atmospheric and other similar models. Since it was anticipated that the many degrees of freedom present in model biases could pose a serious challenge to the bias estimation problem, data assimilation experiments estimating model biases of varying DOF are performed. In Figure 8, the RMS errors for the following 4 experiments are presented: (i) correct_model, (ii) neglect_err, (iii) bias_estim (with b of dimension 101), and (iv) bias_estim (with b of dimension 10). In all experiments 10 ensembles and 10 observations are used. As explained before, the bias was created by applying an additive term, defined as $-\eta u_\theta$, in each time step of the model integration. Comparison of the results in Figure 8a indicates superior performance of the experiment with the non-biased model (correct_model) and a dramatic adverse effect of the model bias on the data assimilation results when the bias was neglected (neglect_err). One can observe, however, that after cycle 90 the curves for bias_estim with 10 and 101 observations cross, suggesting a possibility that the results are opposite in later data assimilation cycles. We have performed additional 100 data assimilation cycles with all 4 experiments (Figure 8b). The results from 100 additional cycles indicate that the two bias_estim curves cross back after a few cycles, and the experiment bias_estim (dim=10) still remains close to the experiment correct_model, while the RMS errors of the experiment neglect_err continues to increase. The experiment bias_estim with bias dimension of 101, even though it is relatively close to the experiment correct_model, shows a trend of filter divergence in later data assimilation cycles. These results indicate that estimation of a bias with many DOF might not work well in cases when the ensemble size is much smaller. On the other hand, reducing the number of DOF of the
estimated bias, as in the experiment bias_estim with bias dimension of 10, helps preventing filter divergence. Thus, we conclude that neglecting the model bias could result in filter divergence; however, the filter divergence could be alleviated by appropriate treatment of the model bias (as in bias_estim, dim=10). It has also been recognized in previous studies that filter divergence could occur due to neglecting model errors (e.g., Jazwinski 1970; Hansen 2002; Mitchell et al. 2002).

In Figure 9 the PDF innovation statistics are shown for the same 4 experiments. Obviously, neglecting the model bias (Figure 9a) results is considerable degradation of the filter performance. As before, correct_model experiment (Figure 9d) shows the best performance, thus defining the upper limit of success to be expected from the bias estimation process. Bias estimation of the bias parameter with 101 DOF (Fig. 9b), even though it shows an improvement with respect to neglect_err (Fig. 9a), has relatively large discrepancies in comparison with correct_model (Fig. 9d). The bias estimation process is clearly beneficial in the bias_estim experiment with reduced number of DOF (Figure 9c), since the results are closer to the experiment correct_model than to the experiment neglect_err.

The results of the K-S tests performed for the experiments in Figure 9 are summarized in Table 5. As shown in Table 5, the largest departures from the standard normal distribution are obtained for the experiment neglect_err, and the smallest for the experiment correct_model. As before, the experimental results of the experiment bias_estim are in between the two limiting cases. The K-S test results also confirm that the bias estimation employing the control vector $b$ of reduced dimension (10 vs. 101) is beneficial. Neither of the experiments, however, passed the K-S test. This is not a surprising result, since it is not expected that the statistics should be strictly Gaussian. On the other hand, dramatic departures from the standard normal distribution are not expected either. Obviously, understanding this issue is not trivial, since it is related to the validity of the Gaussian assumption in application to biased models. Addressing non-Gaussian probability distributions is beyond the scope of this study, but it is planned for the future.

The results of the $\chi^2$– validation tests are shown in Figure 10. Figure 10a indicates filter divergence for the experiment neglect_err (also pronounced in the RMS errors in Fig.8). Bias estimation process is clearly beneficial (figures 10b,c), with the values of $\chi^2$ much closer to the results of the correct_model (Figure 10d) than to the experiment neglect_err (Figure 10a). These
results are quite encouraging, indicating the potential of the EnsDA methodologies to properly estimate and correct model biases.

6. Conclusions

A methodology for model error estimation, employing EnsDA approach and state augmentation method, was proposed and examined in this study. The model errors in the form of a serially correlated error (bias) and a parameter error are examined. The methodology is fully adaptive, providing estimates of the atmospheric state and the model error, as well as the uncertainties of these estimates as defined by the analysis and forecast error covariance matrices.

Data assimilation and model error estimation experiments were performed in application to the KdVB model. The experiments employing a correct and an erroneous model were used as control experiments to define the upper and lower limits of the expected success of the model error estimation process. All validation scores examined (RMS errors, $\chi^2$ – test, K-S test, etc.) were within the two extremes, and in general, much closer to the experiments with the correct model, thus indicating substantial improvements due to model error estimation. Neglecting model error (either in the form of incorrect diffusion, or model bias), however, had detrimental effects on the data assimilation results, in some cases resulting in filter divergence. The experimental results indicated that bias estimation of the model bias with many DOF, but employing a small ensemble size, could also result in filter divergence. It was demonstrated that in such cases, reducing the number of DOF in the estimated bias helps prevent filter divergence. Therefore, it may be of fundamental importance to appropriately estimate and correct model error when employing an EnsDA approach.

There are also some issues that need further attention, but have not been addressed due to the limited scope of this study. These issues include understanding observability in the EnsDA framework and addressing non-Gaussian probability distribution functions when dealing with biased models. This research is planned for the future.

The model error estimation approach was examined within the framework of the MLEF of Zupanski (2005), but it was not tied to this particular variant of EnsDA. Its application to other EnsDA approaches should be straightforward. The methodology can also be used within an ensemble smoother. The integration of a forecast model and assimilation of observations over a
time interval performed in a smoother, rather than using instant observations as in a filter, may further improve the results, since more observations could better constrain the model error.

Model error estimation applying an ensemble smoother approach is also planned for future work.
Acknowledgements

The authors would like to thank Nancy Nichols, Eugenia Kalnay and Dale Barker for helpful discussion during the preparation of this manuscript. We have also benefited from the helpful suggestions made by Arthur Hou and Sara Zhang. Inspiring discussions with Thomas Vonder Haar and Tomislava Vukicivic are very much appreciated. The authors are also thankful to Rolf Reichle and two anonymous reviewers for providing much helpful comments that resulted in substantial improvements of the manuscript. This work was supported by the Department of Defense Center for Geosciences/Atmospheric Research grant DAAD19-02-2-0005.
APPENDIX A

Markov Process Variable on a Coarse Time Scale

Let us start from the first order Markov process variable defined as

\[ \Phi_n = \alpha \Phi_{n-1} + (1 - \alpha) G(r_n) . \]  \hspace{1cm} (A.1)

where \( \Phi_{n-1} \) is a serially correlated error component, \( r_n \) is a random error component, and \( G \) is a mapping between the spaces of the two components. The constant \( \alpha < 1 \) defines relative contribution of the two terms to the total model error \( \Phi_n \) during the time step \( n \). Let us assume that the random error term stays constant for a number of time steps. Then we have:

\[
\begin{align*}
\Phi_1 &= \alpha \Phi_0 + (1 - \alpha) G(b) \\
\Phi_2 &= \alpha \Phi_1 + (1 - \alpha) G(b) \\
\Phi_3 &= \alpha \Phi_2 + (1 - \alpha) G(b) \\
\vdots \\
\Phi_n &= \alpha^n \Phi_0 + (1 - \alpha^n) G(b)
\end{align*}
\]  \hspace{1cm} (A.2)

With this assumption, random error becomes a bias parameter, denoted \( b \), since it has an average effect, over a period of time, not just an instant effect over a single time step as in (A.1). Since \( \alpha < 1 \), we can assume that for \( n = N_{\text{max}} \) (a reasonably large maximum number of time steps in a data assimilation interval) \( \Phi_{N_{\text{max}}} \approx G(b) \). We can than use the model bias parameter from the previous data assimilation to initialize model error in the current data assimilation cycle. Thus, we have the following:

\[ \Phi_0 = \Phi_{N_{\text{max}}} \approx G(b^*) , \]  \hspace{1cm} (A.3)

\[ \Phi_n = \alpha \Phi_{n-1} + (1 - \alpha) G(b) \approx \alpha^n G(b^*) + (1 - \alpha^n) G(b) , \]  \hspace{1cm} (A.4)
where $b$ and $b'$ are bias parameters estimated in the most recent, and the previous data assimilation cycle, respectively. As indicated in (A.2), $b$ is defined at the end of the forecast integration ($n=N_{\text{max}}$), and that $b'$ is defined at the beginning the forecast integration ($n=0$). Note, that according to (A.3), the bias $\Phi_{N_{\text{max}}}$ is defined at the end of the forecast integration in the previous data assimilation cycle. The above model error definition is a serially correlated model bias, or first order Markov process variable, defined on a coarse time scale. This model error definition was successfully used in 4DVAR data assimilation experiments (e.g., Zupanski 1997; Zupanski et al. 2002a,b). For more generality, one can include additional random noise to (A.4), defined as a new vector in each time step $n$. 
APPENDIX B

Operator \( G(u) \)

Let us assume that the model state \( u \) is a vector of dimension \( N \) with components \( u = \{u_i\} \); similarly the model bias parameter \( b \) is a vector of dimension \( M \) with components \( b = \{b_j\} \).

Assuming that minimum and maximum components of \( u \) are obtained as
\[
\begin{align*}
    u_{\text{min}} &= \min_i u_i, \quad \text{and} \quad u_{\text{max}} = \max_i u_i, \\
\end{align*}
\]
we can group elements \( u_i \) into \( M \) categories. Each value of \( u_i \) should belong to one of the categories numbered by the integer valued function \( f(u_i) \) with values \( 1 \leq f(u_i) \leq M \). Assuming \( \Delta u \) is category interval defined as \( \Delta u = \frac{u_{\text{max}} - u_{\text{min}}}{M-1} \) we have
\[
    f(u_i) = \text{int} \left( \frac{u_i - u_{\text{min}}}{\Delta u} \right) + 1,
\]
where \( \text{int} (x) \) is an integer number of \( x \).

The state dependent operator \( G(u) \) that transforms \( b \) into the categories of \( u \) can now be defined as matrix \( G \) with elements
\[
    G_{ij} = \begin{cases} 
    1, & \text{if } f(u_i) = j \\
    0, & \text{if } f(u_i) \neq j
    \end{cases}
\]
\[
\text{(B.3)}
\]
In other words, if \( G_{ij} = 1 \), than \( u_i \) belongs to the category \( j \) and bias parameter \( b_j \) represents the error of \( u_i \). Similarly if \( G_{ij} = 0 \), than \( u_i \) does not belong to the category \( j \) and bias parameter \( b_j \) does not represent the error of \( u_i \). Thus, same value \( b_j \) could be assigned to multiple values \( u_i \). In the experiments presented in Section 5.b the operator \( G \) is used to transform the vector \( b \) of dimension \( M=10 \) into the model state vector \( u \) of dimension \( N=101 \).
References:


Martin, M. J., N. K. Nichols, and M. J. Bell, 1999: Treatment of Systematic Errors in Sequential
Data Assimilation, UK Meteorological Office, Ocean Applications Division, Tech. Note No. 21, 45 pp.


Table Captions

Table 1. Parameter estimation experiments.

Table 2. Bias estimation experiments.

Table 3. The results of K-S test statistic $D$ (15) for the parameter estimation experiments (Figs. 3 and 4). The hypothesis $H_0$ that the normalized innovations (14) have Gaussian distribution $\mathcal{N}(0,1)$ is tested.

Table 4. As in Table 3, but for experiments param_estim and neglect_err; in all experiments full rank problem is solved employing full set of observations.

Table 5. The results of K-S test statistic $D$ (15) for the model bias estimation experiments (Fig. 9). The hypothesis $H_0$ that the normalized innovations (14) have Gaussian distribution $\mathcal{N}(0,1)$ is tested.
Table 1. Parameter estimation experiments. Note that $N_{\text{obs}}$ indicates the number of observations per data assimilation cycle.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Control variable</th>
<th>DOF in control variable</th>
<th>$\nu$</th>
<th>True $\nu$</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{ens}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct_model</td>
<td>$u$</td>
<td>101</td>
<td>0.20</td>
<td>0.20</td>
<td>10 or 101</td>
<td>10</td>
</tr>
<tr>
<td>neglect_err</td>
<td>$u$</td>
<td>101</td>
<td>0.07</td>
<td>0.20</td>
<td>10 or 101</td>
<td>10</td>
</tr>
<tr>
<td>param_estim</td>
<td>$u, \nu$</td>
<td>102</td>
<td>estimated</td>
<td>0.20 or 0.07</td>
<td>10 or 101</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 2. Bias estimation experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Control variable</th>
<th>DOF in control variable</th>
<th>$\eta$</th>
<th>DOF in $\eta u_0$</th>
<th>DOF in vector $b$</th>
<th>$\nu$</th>
<th>$N_{\text{obs}}$</th>
<th>$N_{\text{ens}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct_model</td>
<td>$u$</td>
<td>101</td>
<td>0</td>
<td>_</td>
<td>_</td>
<td>0.07</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>neglect_err</td>
<td>$u$</td>
<td>101</td>
<td>0.0001</td>
<td>101</td>
<td>_</td>
<td>0.07</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>bias_estim</td>
<td>$u$, $b$</td>
<td>202 or 111</td>
<td>0.0001</td>
<td>101</td>
<td>101 or 10</td>
<td>0.07</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 3. The results of K-S test statistic $D$ (15) for the parameter estimation experiments (Figs. 3 and 4). The hypothesis $H_0$ that the normalized innovations (14) have Gaussian distribution $\mathcal{N}(0,1)$ is tested.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$N_{\text{obs}}$</th>
<th>Test statistic $D$</th>
<th>Critical value $\alpha(5%)$</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct_model (10 ens, 10 obs, Fig.3)</td>
<td>819</td>
<td>0.025962</td>
<td>0.047522</td>
<td>Accept $H_0$</td>
</tr>
<tr>
<td>neglect_err (10 ens, 10 obs, Fig.3)</td>
<td>819</td>
<td>0.198561</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>param_estim (10 ens, 10 obs, Fig.3)</td>
<td>819</td>
<td>0.076648</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>correct_model (10 ens, 101 obs, Fig.4)</td>
<td>9191</td>
<td>0.007347</td>
<td>0.014186</td>
<td>Accept $H_0$</td>
</tr>
<tr>
<td>neglect_err (10 ens, 101 obs, Fig.4)</td>
<td>9191</td>
<td>0.102725</td>
<td>0.014186</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>param_estim (10 ens, 101 obs, Fig.4)</td>
<td>9191</td>
<td>0.036210</td>
<td>0.014186</td>
<td>Reject $H_0$</td>
</tr>
</tbody>
</table>
Table 4. As in Table 3, but for experiments *param_estim* and *neglect_err*; in all experiments full rank problem is solved employing full set of observations.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$N_{\text{obs}}$</th>
<th>Test statistic $D$</th>
<th>Critical value $\alpha$(5%)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>param_estim</em> (102 ens, 101 obs, true $\nu$=0.07)</td>
<td>9191</td>
<td>0.008695</td>
<td>0.014186</td>
<td>Accept $H_0$</td>
</tr>
<tr>
<td><em>param_estim</em> (102 ens, 101 obs, true $\nu$=0.20)</td>
<td>9191</td>
<td>0.006954</td>
<td>0.014186</td>
<td>Accept $H_0$</td>
</tr>
<tr>
<td><em>neglect_err</em> (101 ens, 101 obs, $\nu$=0.07, true $\nu$=0.20)</td>
<td>9191</td>
<td>0.101526</td>
<td>0.014186</td>
<td>Reject $H_0$</td>
</tr>
</tbody>
</table>
Table 5. The results of K-S test statistic $D$ (15) for the model bias estimation experiments (Fig. 9). The hypothesis $H_0$ that the normalized innovations (14) have Gaussian distribution $\mathcal{N}(0,1)$ is tested.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$N_{\text{obs}}$</th>
<th>Test statistic $D$</th>
<th>Critical value $\alpha(5%)$</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct_model (10 ens, 10 obs)</td>
<td>819</td>
<td>0.080781</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>neglect_err (10 ens, 10 obs)</td>
<td>819</td>
<td>0.498816</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>bias_estim (10 ens, 10 obs, dim 101)</td>
<td>819</td>
<td>0.251252</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>bias_estim (10 ens, 10 obs, dim 10)</td>
<td>819</td>
<td>0.184432</td>
<td>0.047522</td>
<td>Reject $H_0$</td>
</tr>
</tbody>
</table>
Figure Captions

Fig. 1. Time integration of the KdVB model and targeted observations. The triangles denote observations. The horizontal axis represents the model domain, and the ordinate axis is the amplitude. The shown cycles are No. 1, 5, 10 and 20. Note how the targeted observations follow the wave crests of the model solution. Also, solution decays with time.

Fig. 2. (a) RMS errors of the optimal model state solution with respect to the true KdVB numerical solution showing the impact of incorrect diffusion. The errors for 3 experiments: correct_model, neglect_err, and param_estim are presented as functions of data assimilation cycles. In all experiments 10 ensemble members and 10 observations are used. (b) As in (a), but using 101 observation in each data assimilation cycle.

Fig. 3. Histogram of the PDF of the innovations, calculated using all observations assimilated during 100 data assimilation cycles. In each cycle 10 observations are used ($N_{obs}=10$). The theoretical $\mathcal{N}(0,1)$ Gaussian PDF is also plotted. The impact of incorrect diffusion is examined. The results of the following experiments are shown: (a) neglect_err, (b) param_estim, and (c) correct_model.

Fig. 4. As in Fig. 3, but for $N_{obs}=101$.

Fig. 5. $\chi^2$ – innovation test values, calculated as instant values in each data assimilation cycle (dashed) and as a 10-cycle running averages (solid) for the experiments: (a) neglect_err, (b) param_estim, and (c) correct_model. The experimental results employing $N_{ens}=10$ and $N_{obs}=10$ are shown.

Fig. 6. As in Fig. 5, but for $N_{obs}=101$.

Fig. 7. (a) Diffusion coefficients obtained in various parameter estimation experiments: (i) $N_{ens}=102$ and $N_{obs}=101$ (solid bold), (ii) $N_{ens}=10$ and $N_{obs}=101$ (dashed bold), (iii) $N_{ens}=10$ and $N_{obs}=10$ (solid thin), and (iv) true value of the coefficient (dashed thin). (b) Diffusion coefficients
obtained in parameter estimation experiments with two different values of the true coefficient (0.07 and 0.20). Parameter estimation experiments in (b) were performed employing $N_{ens}=102$ and $N_{obs}=101$.

Fig. 8. (a) RMS errors of the optimal model state solution with respect to the true KdVB numerical solution showing the impact of model bias. The errors for 4 experiments: correct_model, neglect_err, bias_estim (bias dimension =101), and bias_estim (bias dimension =10) are presented as functions of data assimilation cycles. In all experiments 10 ensemble members and 10 observations are used. (b) As in (a), but the experiments are performed for 100 hundred additional data assimilation cycles (200 total).

Fig. 9. Histogram of the PDF of the innovations, calculated using all observations assimilated during 100 data assimilation cycles. In each cycle 10 observations are used ($N_{obs}=10$). The theoretical $N(0,1)$ Gaussian PDF is also plotted. The impact of model bias estimation is examined. The results of the following experiments are shown: (a) neglect_err, (b) bias_estim (dim=101), (c) bias_estim (dim=10) and (d) correct_model.

Fig. 10. $\chi^2$ – innovation test values, calculated as instant values in each data assimilation cycle (dashed) and as a 10-cycle running averages (solid) for the experiments: (a) neglect_err, (b) bias_estim (dim=101), (c) bias_estim (dim=10), and (c) correct_model. The experimental results employing $N_{ens}=10$ and $N_{obs}=10$ are shown.
Fig. 1. Time integration of the KdVB model and targeted observations. The triangles denote observations. The horizontal axis represents the model domain, and the ordinate axis is the amplitude. The shown cycles are No. 1, 5, 10 and 20. Note how the targeted observations follow the wave crests of the model solution. Also, solution decays with time.
Fig. 2. (a) RMS errors of the optimal model state solution with respect to the true KdVB numerical solution showing the impact of incorrect diffusion. The errors for 3 experiments: `correct_model`, `neglect_err`, and `param_estim` are presented as functions of data assimilation cycles. In all experiments 10 ensemble members and 10 observations are used. (b) As in (a), but using 101 observation in each data assimilation cycle.
Fig. 3. Histogram of the PDF of the innovations, calculated using all observations assimilated during 100 data assimilation cycles. In each cycle 10 observations are used ($N_{\text{obs}}=10$). The theoretical $\mathcal{N}(0,1)$ Gaussian PDF is also plotted. The impact of incorrect diffusion is examined. The results of the following experiments are shown: (a) $\text{neglect\_err}$, (b) $\text{param\_estim}$, and (c) $\text{correct\_model}$.
Fig. 4. As in Fig. 3, but for $N_{\text{obs}}=101$. 

---

**Innovation histogram (incorrect diffusion)**

- **neglect_err, 10 ens, 101 obs**

- **Category bins**
  - PDF

---

**Innovation histogram (incorrect diffusion)**

- **param_estim, 10 ens, 101 obs**

- **Category bins**
  - PDF

---

**Innovation histogram (correct diffusion)**

- **correct_model, 10 ens, 101 obs**

- **Category bins**
  - PDF
Fig. 5. $\chi^2$ – innovation test values, calculated as instant values in each data assimilation cycle (dashed) and as a 10-cycle running averages (solid) for the experiments: (a) neglect_err, (b) param_estim, and (c) correct_model. The experimental results employing $N_{\text{ens}}=10$ and $N_{\text{obs}}=10$ are shown.
Fig. 6. As in Fig. 5, but for $N_{obs}=101$. 

---

Innovation $\chi^2$ Test (incorrect diffusion)

Innovation $\chi^2$ Test (incorrect diffusion)

Innovation $\chi^2$ Test (correct diffusion)
Fig. 7. (a) Diffusion coefficients obtained in various parameter estimation experiments: (i) $N_{ens}=102$ and $N_{obs}=101$ (solid thick), (ii) $N_{ens}=10$ and $N_{obs}=101$ (dashed thin), (iii) $N_{ens}=10$ and $N_{obs}=10$ (solid thin), and (iv) true value of the coefficient (dashed thick). (b) Diffusion coefficients obtained in parameter estimation experiments with two different values of the true coefficient (0.07 and 0.20). Parameter estimation experiments in (b) were performed employing $N_{ens}=102$ and $N_{obs}=101$. 
Fig. 8a. RMS errors of the optimal model state solution with respect to the true KdVB numerical solution showing the impact of model bias. The errors for 4 experiments: correct_model, neglect_err, bias_estim (bias dimension =101), and bias_estim (bias dimension =10) are presented as functions of data assimilation cycles. In all experiments 10 ensemble members and 10 observations are used.
Fig. 8b. As in Fig. 8a, but the experiments are performed for 100 hundred additional data assimilation cycles (200 total).
Fig. 9. Histogram of the PDF of the innovations, calculated using all observations assimilated during 100 data assimilation cycles. In each cycle 10 observations are used ($N_{\text{obs}}=10$). The theoretical $\mathcal{N}(0,1)$ Gaussian PDF is also plotted. The impact of model bias estimation is examined. The results of the following experiments are shown: (a) $\text{neglect\_err}$, (b) $\text{bias\_estim}$ (dim=101), (c) $\text{bias\_estim}$ (dim=10) and (d) $\text{correct\_model}$.
Fig. 10. $\chi^2$ – innovation test values, calculated as instant values in each data assimilation cycle (dashed) and as a 10-cycle running averages (solid) for the experiments: (a) $\text{neglect\_err}$, (b) $\text{bias\_estim}$ (dim=101), (c) $\text{bias\_estim}$ (dim=10), and (c) $\text{correct\_model}$. The experimental results employing $N_{\text{ens}}=10$ and $N_{\text{obs}}=10$ are shown.