P2.4

BALANCE DYNAMICS AND FOUR-DIMENSIONAL DATA ASSIMILATION

Lisa J. Neef[1]*, Theodore G. Shepherd[1], Saroja M. Polavarapu[2]

[1]Department of Physics, University of Toronto

[2]Meteorological Service of Canada

1. INTRODUCTION

It has been a long-standing problem of atmospheric prediction that the insertion of observations into primitive equations (PE) models excites spurious inertia-gravity waves. Conventionally, this problem has been treated by initialization methods, which separate the motion into its normal components, and then project the initial model state onto the hypothetical manifold of purely slow motion. The so-called "slow manifold" is defined as the subspace of possible motion where the fast components of the motion are slaved to the slow components.

Daley and Puri (1980) argued that direct insertion of observations into models would almost always destroy the dynamical balance in a PE model. They pointed out that advanced data assimilation schemes should seek the most likely *balanced* state, given the observations, and suggested making balance a constraint on the fit between model and observations.

When observations are assimilated into models, the *analysis* $\mathbf{x}^a$ is given as a linear combination of the model forecast $\mathbf{x}^f$ and observation increment or innovation (where $\mathbf{z}$ is the observation):

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{z} - \mathbf{H}\mathbf{x}^f). \tag{1}$$

H is an operator which maps the model variables to the observation locations. Assimilation schemes differ in how they estimate the combination of the forecast and innovations. The weights K given to the innovations at each analysis variable and gridpoint, depend upon the relative observation and forecast errors. The accuracy of an assimilation scheme therefore depends on its ability to model the error covariances.

Conventionally, the analysis step is followed by a separate initialization step, with the assumption that the resulting forecast will stay balanced until the next observation. Artificial balance constraints can also be added to the covariance model (e.g., Kalnay, 2003, and Daley, 1991).

***Corresponding author address*: Lisa J. Neef, Dept. of Physics, University of Toronto, 60 St. George Street, Toronto, Ontario, M5S 1A7, Canada; E-mail: lisa@atmosp.physics.utoronto.ca

Four-dimensional assimilation schemes, however, weave the analysis step into the dynamics of the forecast model, in effect calculating the most likely state *in time*, and evolving the covariance model according to the model dynamics.

We therefore expect that, if the true state is balanced and the forecast model is sufficiently accurate, the four-dimensional analysis will be balanced as well. However, even if the model is perfect, errors in the observations will project onto all timescales, and it is not clear whether the analysis will indeed remain balanced. In this study, we apply the Extended Kalman Filter (EKF) analysis to a perfect model, imperfect observations, and a balanced truth state, in order to test how well the assimilation scheme represents dynamical balance.

2. SIMULATIONS

2.1 *The Lorenz Model*

Lorenz's (1986) model (and the subsequent extension by Wirosoetisno and Shepherd, 2000) is derived from a triad expansion of the shallow water equations: It has a two independent vortical modes and a linear inertia-gravity wave.

$$\begin{aligned}
\frac{d\phi}{dt} &= w' + bz' \\
\frac{dw'}{dt} &= -\frac{C}{2}\sin 2(\phi + \epsilon b x) - \frac{\alpha^2 b}{\epsilon}x \\
\frac{dx}{dt} &= \frac{bw' - z'}{\epsilon} \\
\frac{dz'}{dt} &= \frac{\alpha^2 x}{\epsilon}.
\end{aligned} \tag{2}$$

Here ϕ is related to the potential vorticity of one mode, w' represents vorticity of the other mode, x divergence, and z' free-surface height. The model admits two modes of motion: a vortical mode that varies on a timescale of $O(1)$, and an inertia-gravity wave with frequency $1/\epsilon$. The primes on the divergence and height terms denote that these variables are mixtures of the gravity and vortical modes, while ϕ is a purely slow, and x a purely fast variable. They are clearly separated when $\epsilon = Rob/\sqrt{(1+b^2)}$, where Ro is the Rossby number and b the rotational Froude number for the wavelength of the gravity wave, is small. For convenience, $\alpha = (1+b^2)^{-1/2}$. C has an artifical time-dependence, added by Wirosoetisno and Shepherd in order to make the slow dynamics chaotic, and thus simulate the presence of other vortical modes.

If we transform $w = w' + bz'$ (corresponding to potential vorticity) and $z = z' - bw'$ (corresponding to geostrophic imbalance), we attain a normal-mode version of (2):

$$\begin{aligned}
\frac{d\phi}{dt} &= w \\
\frac{dw}{dt} &= -\frac{C}{2}\sin 2(\phi + \epsilon bx) \\
\frac{dx}{dt} &= -\frac{z}{\epsilon} \\
\frac{dz}{dt} &= \frac{x}{\epsilon} + \frac{bC}{2}\sin 2(\phi + \epsilon bx). \qquad (3)
\end{aligned}$$

In the above model, x and z contain the gravity wave, and ϕ and w are entirely slow variables. Wirosoetisno and Shepherd approximated balanced dynamics by slaving the fast variables to the slow. To second order, these are given by

$$\begin{aligned}
U_x &= -\frac{\epsilon}{2}Cb\sin 2\phi + O(\epsilon^3) \\
U_z &= \epsilon^2(Cbw\cos 2\phi + \frac{C'}{2}b\sin 2\phi) + O(\epsilon^3). \qquad (4)
\end{aligned}$$

where C' is the time-derivative of C.

When the model is initialized such that $x = U_x$ and $z = U_z$, its trajectory stays on the slow manifold (with $O(\epsilon^2)$ accuracy) for long times (see Wirosoetisno and Shepherd for details).

We chose the particular solution of (2) with initial conditions $\phi = -0.50$ and $w' = 0.37$ (with x and z' given by their corresponding values using (4)), as the "truth", which is on the slow manifold. We choose $\epsilon = 0.1$ throughout this paper. We then perturb the initial conditions slightly to produce the initial forecast. Since the slow dynamics are chaotic, the model will quickly diverge from the reference solution unless observations are added regularly.

We generate observations of one of the four model variables by perturbing the truth state by a random gaussian error. Writing the observation vector $\mathbf{z}$ as $\mathbf{x}^{obs}$,

$$\mathbf{x}^{obs} = \mathbf{H}\mathbf{x}_k^t + \sigma^{obs}\mathbf{b}_k \qquad (5)$$

where $\mathbf{b}_k$ is $N(0,1)$, and the brackets denote an average over the statistical distribution. $\mathbf{H}$ is here simply a matrix of zeros and ones, which selects the variable to be observed. The observations are then assimilated in time using the EKF.

2.2 *The Extended Kalman Filter*

Writing the model (3) as a nonlinear operator $M_k(\mathbf{x}_k)$, the forecast at each time step is evolved from the analysis of the previous timesp:

$$\mathbf{x}_{k+1}^f = M_k(\mathbf{x}_k^a). \qquad (6)$$

If an observation was made at the k^{th} time step, $\mathbf{x}_k^a$ is given by (1). If there was no observation, then $\mathbf{x}_k^a = \mathbf{x}_k^f$.

The forecast covariance matrix is defined as

$$\mathbf{P}_k^f = \left\langle (\mathbf{x}_k^f - \mathbf{x}_k^t)(\mathbf{x}_k^f - \mathbf{x}_k^t)^{\mathrm{T}} \right\rangle, . \qquad (7)$$

To find the evolution of the error, we expand (6) as a Taylor series about the analysis state:

$$\begin{aligned}
\mathbf{x}_{k+1}^f - \mathbf{x}_{k+1}^t &= M_k(\mathbf{x}_k^f) - M_k(\mathbf{x}_k^a) \\
&+ (\mathbf{x}_k^f - \mathbf{x}_k^t)\frac{\partial M_k(\mathbf{x}_k^a)}{\partial \mathbf{x}} \\
&+ \frac{1}{2}(\mathbf{x}_k^f - \mathbf{x}_k^t)^2\frac{\partial^2 M_k(\mathbf{x}_k^a)}{\partial \mathbf{x}^2} \ldots \qquad (8)
\end{aligned}$$

Since (7) is a statistical average, the Taylor expansion for $\mathbf{P}^f$ will involve higher and higher statistical moments, and will be very computationally expensive. We therefore truncate (8) at first order and, defining the tangent linear model,

$$\mathbf{M}_k = \frac{\partial M_k(\mathbf{x}_k^a)}{\partial \mathbf{x}}, \qquad (9)$$

we approximate the evolution of the covariance matrix as

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k\mathbf{P}_k^a\mathbf{M}_k^{\mathrm{T}} \qquad (10)$$

The Kalman gain matrix, **K**, weights the innovations by the relative magnitudes of the observation and forecast error covariances:

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}, \qquad (11)$$

where $\mathbf{R}$ is the observation error covariance matrix.

The covariance matrix for the new analysis is then given by

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^f. \qquad (12)$$

The EKF analysis (1) is performed every time an observation is made, and the resulting analysis is then advected forward in time according to (6). Since we also advect the covariances (equations 9 and 10), the analysis at each observation step is a best fit between the current estimate and all prior observation increments.

In effect, the model trajectory is forced off the slow manifold every time an observation is made, because the observation errors project onto all scales, including the fastest ones. Since this intermittent analysis is always an *estimate*, the resulting analysis trajectory does not necessarily have to be balanced, even if the true state is balanced.

3. RESULTS AND IMPLICATIONS

3.1 *Loss of Balance in the EKF Analysis*

Figure 1 shows the result of the EKF analysis for the mixed-timescale vorticity variable w' and the pure fast variable x, given observations of w', for two different observation frequencies and two different observation errors. We could observe any of the model variables, but choose w' as an example here, since it is a mixed variable, and real observations are of mixed timescales. Figure 1a shows the analysis for observations every time unit (corresponding to every 100 timesteps), and a mean observation error of 0.1 (corresponding to roughly 10% the amplitude of w'). While the analysis tracks the low-frequency evolution of w' and x, it also contains a fast oscillation, which is not present in the balanced true state. Increasing the observation frequency (Figure 1b) seems to control the unbalanced motion for some time, but at later times, the high-frequency oscillation reappears, and with greater amplitude. However, increasing the observation accuracy to 0.01 (Figure 1c) is sufficient to restore balance. The initial imbalance in the forecast state for $t < 10$ is quickly removed by the analysis.

3.2 *Tracking the Vortical Mode*

Figure (2) shows the same three analyses for the pure slow potential vorticity variable ϕ. For cases (a) and (b), the EKF tracks the slow mode for some time, but eventually loses the model's transition into a slightly different evolution. This is somewhat similar to the results of Miller and Ghil (1994), who found that the EKF was unable to track transitions between the two attractor points in the 3-component Lorenz model, if observations were not frequent or accurate enough.

In our case, the EKF loses the ϕ trajectory because the incorrect analysis of ϕ still fits the observations of w'. Increasing observation frequency, therefore, does not improve the analysis. More accurate observations (Figure 2c) again solve this problem.

4. A Modified Analysis

We now discuss possible modifications to the assimilation scheme in which balance is made a strong constraint on the analysis equation. In order to constrain the unbalanced motion, we transform the model (2) to its normal-mode form (3), and define the analysis variables as the slow variables ϕ and w, and the unbalanced components of the motion,

$$\tilde{x} = x - U_x \tag{13}$$

$$\tilde{z} = z - U_z \tag{14}$$

where U_x and U_z are given by (4). For a completely balanced state, we have

$$\tilde{x} = \tilde{z} = 0. \tag{15}$$

The analysis step (1) can now be changed such that we analyze $\mathbf{y} = (\phi, w, \tilde{x}, \tilde{z})^{\mathrm{T}}$, and then transform the resulting analysis back to the original model variables, $\mathbf{x} = (\phi, w', x, z')^{\mathrm{T}}$. The analysis step (1) becomes

$$\mathbf{y}^a = \mathbf{y}^f + \mathbf{K}_y[\mathbf{x}^{obs} - f(\mathbf{y}^f)], \tag{16}$$

and is followed by the the transformation to model variables,

$$\mathbf{x}^a = f(\mathbf{y}^a) = \begin{pmatrix} \phi^a \\ \alpha^2(w^a - b(\tilde{z}^a + U_z)) \\ \tilde{x}^a + U_x \\ \alpha^2(\tilde{z}^a + U_z + bw^a) \end{pmatrix}. \tag{17}$$

$\mathbf{K}_y$ is the Kalman gain matrix in analysis space, and in order to calculate it we need to transfrom model-space error covariances to analysis-space error covariances, which requires a linearization of $f(\mathbf{y})$ similar to (9). Defining

$$\mathbf{L} = \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}}, \tag{18}$$

we have

$$\mathbf{P}_x \approx \mathbf{L}\mathbf{P}_y\mathbf{L}^{\mathrm{T}} \tag{19}$$

and

$$\mathbf{K}_y = \mathbf{P}_y^f \mathbf{G}^{\mathrm{T}}(\mathbf{H}\mathbf{P}_x^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{20}$$

The balance constraint is applied by combining (17) with (15), such that

$$\mathbf{x}^a = g(\mathbf{y}) = \begin{pmatrix} \phi^a \\ \alpha^2(w^a - bU_z) \\ U_x \\ \alpha^2(U_z + bw^a) \end{pmatrix}. \tag{21}$$

Figure 3 shows the resulting analyses of w' and x, given observations of w' every time unit, with a mean observation error of 0.1. The analysis is now explicitly balanced for all times. The corresponding analysis of ϕ (figure 4), however, still loses the true trajectory after some time.

Note that we have inserted the balance assumption only into the analysis-transformation step (17), but *not* into the transformation of the covariances (19). It is also possible to modify (18) and (19) to include (15); this is equivalent to analysing *only* ϕ and w and neglecting all covariance terms that involve $\tilde{x}$ and $\tilde{z}$ in (19).

However, in our case this approach yields a highly unstable analysis trajectory, because $\mathbf{P}^f$ now neglects part of the information from the observation increment, and thereby causes insufficient and unphysical terms to appear in the Kalman gain (20).

5. Conclusions

This study points out, in the context of a simple model, a possible weakness in 4D assimilation: because the error covariance model is produced by the dyanmics, the resulting analysis can be unbalanced even when the true state is not. In section 4 we have enforced balance in our analysis diagnostically, immediately following each analysis step. We point out, however, that a simple slaving approximation such as (4) may not be known in realistic models. It will therefore be very important to understand, for realistic models and data, how well the EKF and its covariance model can represent the dynamical separation of timescales. It may also be instructive to compare the above results to similar analyses using other 4D schemes, such as as the Ensemble Kalman Filter and 4DVAR.

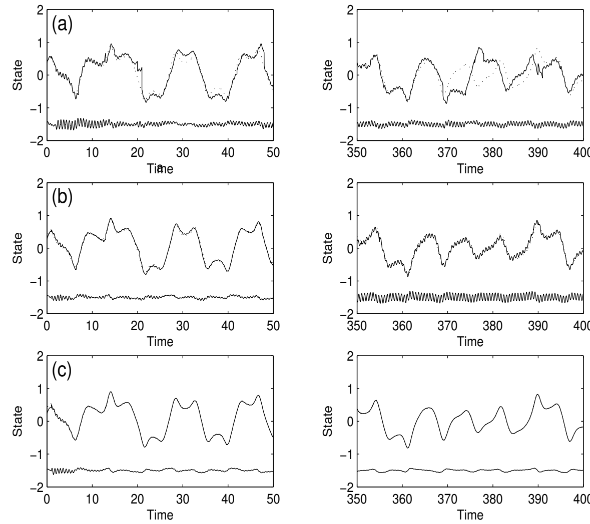

Figure 1: Comparison between the "truth" trajectories of w' and x (dotted line), and the EKF analysis, observing w'. x is offset by -1.5 for clarity. (a) With observations every time unit and $\sigma^{obs} = 0.1$, (b) with observations every half time unit and $\sigma^{obs} = 0.1$, and (c) with observations every time unit and $\sigma^{obs} = 0.01$.

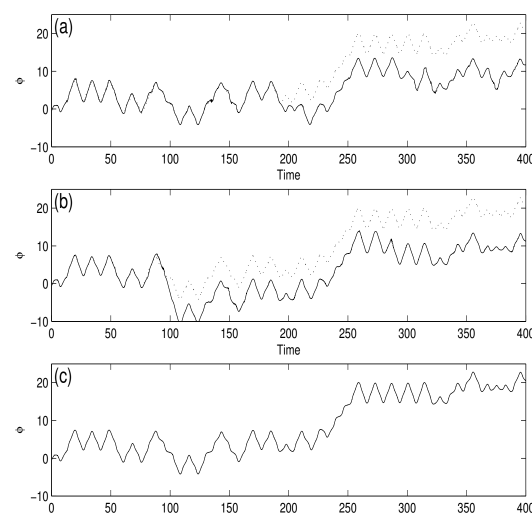

Figure 2: Comparison between the "truth" trajectory of ϕ (dotted line) and the EKF analysis for the same three cases (see previous plot).

REFERENCES

Daley, R., 1991: *Atmospheric Data Analysis*, Cambridge University Press, 457 pp.

Daley, R., and Puri, K., 1980: Four-dimensional data assimilation and the slow manifold, *Mon. Wea. Rev.*, **108**, 85–99.

Kalnay, E., 2003: *Atmospheric Modelling, Data Assimilation and Predictability*, Cambridge University Press, 341 pp.

Lorenz, E.N., 1986: On the existence of a slow manifold, *J. Atmos. Sci.*, **43**, 1547–1557.

Miller, R.N., and Ghil, M., and Gauthiez, F., 1994: Advanced data assimilation for strongly nonlinear dynamical systems, *J. Atmos. Sci.*, **51**, 1037–1056.

Wirosoetisno, D., and Shepherd, T.G., 2000: Averaging, slaving and balanced dynamics in a simple atmospheric model, *Physica*, **D 141**, 37–53.

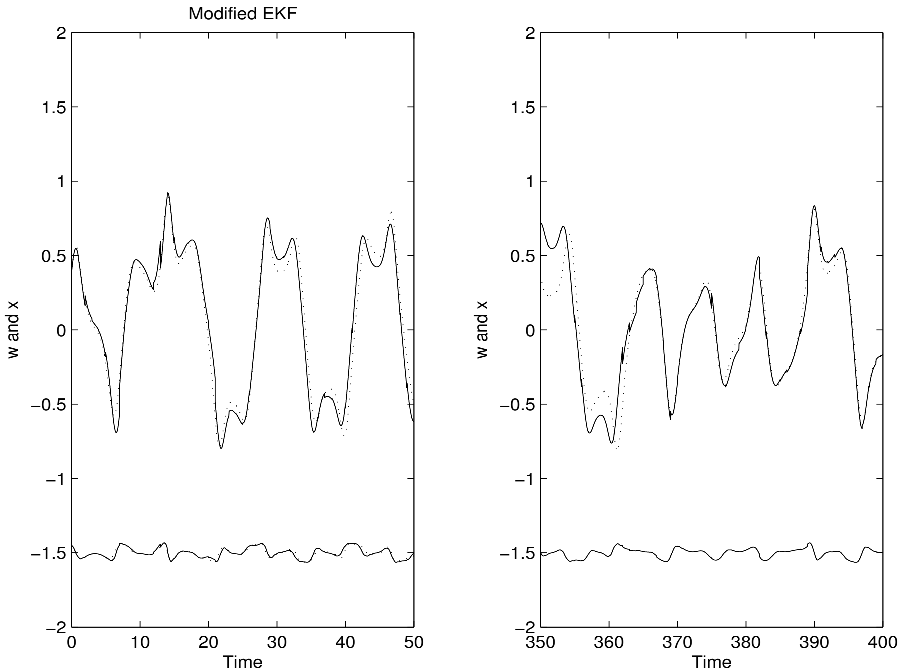

Figure 3: Comparison between the "truth" trajectory of w' and x (dotted line), and the modified EKF analysis, with observations of w' every time step, and $\sigma^{obs} = 0.1$. x is offset by -1.5 for clarity. The resulting analysis remains balanced.

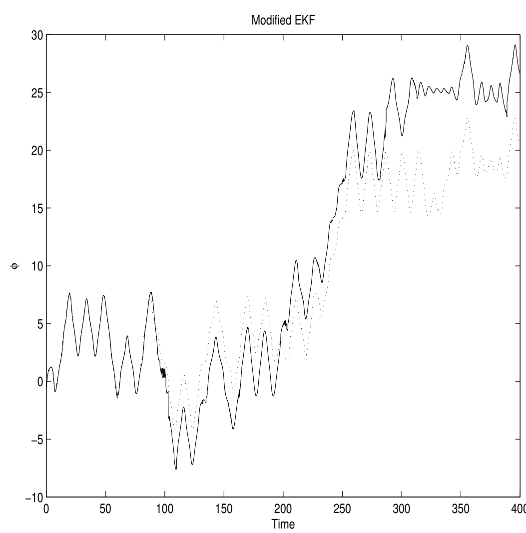

Figure 4: Comparison between the "truth" trajectory of ϕ (dotted line) and the modified EKF analysis, with observations of w' every time step, and $\sigma^{obs} = 0.1$.