

# *Four-dimensional data assimilation and balanced dynamics*

Article

Published Version

Neef, L. J., Polavarapu, S. M. and Shepherd, T. G. ORCID: <https://orcid.org/0000-0002-6631-9968> (2006) Four-dimensional data assimilation and balanced dynamics. *Journal of the Atmospheric Sciences*, 63 (7). pp. 1840-1858. ISSN 1520-0469 doi: <https://doi.org/10.1175/JAS3714.1> Available at <https://centaur.reading.ac.uk/32073/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://dx.doi.org/10.1175/JAS3714.1>

To link to this article DOI: <http://dx.doi.org/10.1175/JAS3714.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



## Four-Dimensional Data Assimilation and Balanced Dynamics

LISA J. NEEF

*Department of Physics, University of Toronto, Toronto, Ontario, Canada*

SAROJA M. POLAVARAPU

*Environment Canada, Toronto, Ontario, Canada*

THEODORE G. SHEPHERD

*Department of Physics, University of Toronto, Toronto, Ontario, Canada*

(Manuscript submitted 14 April 2005, in final form 9 November 2005)

### ABSTRACT

The problem of spurious excitation of gravity waves in the context of four-dimensional data assimilation is investigated using a simple model of balanced dynamics. The model admits a chaotic vortical mode coupled to a comparatively fast gravity wave mode, and can be initialized such that the model evolves on a so-called slow manifold, where the fast motion is suppressed. Identical twin assimilation experiments are performed, comparing the extended and ensemble Kalman filters (EKF and EnKF, respectively). The EKF uses a tangent linear model (TLM) to estimate the evolution of forecast error statistics in time, whereas the EnKF uses the statistics of an ensemble of nonlinear model integrations.

Specifically, the case is examined where the true state is balanced, but observation errors project onto all degrees of freedom, including the fast modes. It is shown that the EKF and EnKF will assimilate observations in a balanced way only if certain assumptions hold, and that, outside of ideal cases (i.e., with very frequent observations), dynamical balance can easily be lost in the assimilation. For the EKF, the repeated adjustment of the covariances by the assimilation of observations can easily unbalance the TLM, and destroy the assumptions on which balanced assimilation rests. It is shown that an important factor is the choice of initial forecast error covariance matrix. A balance-constrained EKF is described and compared to the standard EKF, and shown to offer significant improvement for observation frequencies where balance in the standard EKF is lost. The EnKF is advantageous in that balance in the error covariances relies only on a balanced forecast ensemble, and that the analysis step is an ensemble-mean operation. Numerical experiments show that the EnKF may be preferable to the EKF in terms of balance, though its validity is limited by ensemble size. It is also found that overobserving can lead to a more unbalanced forecast ensemble and thus to an unbalanced analysis.

### 1. Introduction

Four-dimensional (4D) data assimilation schemes are ones that use background knowledge from the assimilating model to derive flow-dependent forecast error statistics. Such flow-dependent forecast error statistics should also, in principle, contain information about any dynamical balance that might exist between the mass and velocity fields. This is important because, since ob-

servations project onto all degrees of freedom, dynamical balance can easily be destroyed by the insertion of observations, causing the excitation of spurious, unrealistic inertia-gravity waves in the subsequent model evolution.

In 3D assimilation, imbalance is handled with an initialization step following the insertion of an observation, wherein some approximation is made to remove spurious fast motion (e.g., Daley 1991). However, such a projection of the analysis onto a so-called slow manifold could result in an analysis that is farther from the observations than the original forecast, rather than a balanced state that is the best fit to the observations (Daley and Puri 1980).

---

*Corresponding author address:* Dr. T. G. Shepherd, Dept. of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada.  
E-mail: tgs@atmosp.physics.utoronto.ca

One might think that the dynamical consistency of 4D error covariance models insures dynamical balance between model variables. However, the extent to which 4D schemes retain balance within the forecast error covariance model, given the limitations of a specific algorithm, is not clear. In fact, every 4D assimilation algorithm relies on a set of assumptions and approximations, and while these may be valid for the time scale of interest, they may not be justified when motions of different time scales are possible. All 4D schemes use either a tangent linear model (TLM), or forecast ensembles, to spread error statistics in time and space. It has not yet been entirely clarified how TLM-based and ensemble methods compare in terms of their ability to reflect dynamical balance.

In fact, existing studies have shown that balance indeed remains a problem in 4D assimilation. While Cohn and Parrish (1991) showed that the linear Kalman filter will compute a balanced analysis state if the model error term is specified to consist of slow variables only, the problem becomes much more complicated if the dynamics are nonlinear.

In the context of four-dimensional variational assimilation (4DVAR), Courtier and Talagrand (1990) showed that, because a standard 4DVAR algorithm uses all degrees of freedom of the problem to minimize the cost function, it generates as many gravity waves as needed in order to best fit the observations. Thus, a balanced analysis requires the addition of a balance constraint to the cost function minimization. This issue has been further studied by Thépaut and Courtier (1991), Polavarapu et al. (2000), and Gauthier and Thépaut (2001).

Tanguay et al. (1995) found that the accuracy of a TLM depends strongly on scale, declining most quickly at the smallest scales. They also found, however, that the assimilation algorithm can transfer information from the spatial scale of the observations to smaller scales, in effect filling in the unobserved scales. In the context of ensemble methods, Mitchell et al. (2002) found that imbalance in the analysis is a direct consequence of spatial localization of error covariances (which may be necessary in order to avoid spurious noise in covariances at large radii).

As 4D data assimilation becomes operationally feasible, the balance problem is changing from one of initialization to one of representing dynamical balance within the evolving covariance model. This study aims to establish the balance properties of 4D schemes in the most basic way possible, by comparing TLM-based evolution of error covariances to ensemble-based techniques. This will be done by comparing the two most basic nonlinear approximations to the standard Kalman

filter: the extended Kalman filter (EKF), which uses the TLM, and the ensemble Kalman filter (EnKF), which evolves an ensemble of forecasts between observations.

We are thus restricting this study to sequential assimilation. Variational algorithms are likely to behave differently in terms of balance, but since 4DVAR still uses a TLM, the results shown below for the EKF may have implications for 4DVAR.

The model used is that of Lorenz (1986), as modified by Wirosoetisno and Shepherd (2000, hereafter WS00), and will be referred to here as the extended Lorenz (1986) model or exL86. It has only 4 degrees of freedom, but admits both a chaotic vortical mode and a gravity wave, with an asymptotic, nonlinear balance between slow and fast variables. The advantage of such a model is that the balance between fast and slow variables is well understood, and the assimilated analysis can thus be easily interpreted in terms of the balanced and unbalanced components of the motion.

The fact that this model is conservative does not pose a great difficulty, since the intention here is to use it to study assimilation algorithms in the context of the free atmosphere gravity wave/initialization problem, rather than dissipative processes. As pointed out by Lorenz (1986) and WS00, dissipation of gravity waves is not the cause of the existence of a slow manifold, and therefore models such as this one can be quite representative of realistic balance dynamics. Another reason for using a conservative model is that representing dissipation—especially in a low-order model—requires some form of parameterization, which further complicates the assimilation process.

This study follows in the spirit of Miller et al. (1994) and Evensen (1997), who used the celebrated Lorenz (1963) model to test the EKF and EnKF in the context of highly nonlinear, chaotic dynamics. These studies found that the accuracy of both filters depends strongly on the frequency and accuracy of observations. Evensen (1997) showed that the EnKF may have an advantage over the EKF, since it includes nonlinear evolution of the forecast error distribution, whereas the EKF assumes that errors evolve linearly in time. The important difference between those studies and the present one is the presence of nonlinear balance dynamics and the possibility of gravity wave generation.

An important caveat to mention is that we perform so-called identical twin experiments, where the forecast and truth are evolved with the same “perfect” model. We are thus dealing with a case where the dynamics are completely understood, such that the difference between a forecast and the truth is due only to initial-state error, observation error, and accumulated analysis error. This makes the behavior of each assimilation

scheme more transparent, and is thus a good place to begin study. We note, however, that the presence of model error—and a corresponding model error term in the error covariance evolution—as will be discussed in section 3a) will likely have a significant effect on the resulting balance results. A study dealing with model error is a point of future research.

The model and its basic properties are discussed in section 2, with a detailed derivation provided in the appendix. Section 3 briefly reviews the Kalman filter equations, and casts them into the language of balanced dynamics. In sections 4 and 5, identical twin experiments are performed with the EKF and EnKF, respectively, in order to examine how well the TLM- or ensemble-predicted forecast error covariances reflect the true statistics of the balanced state. A discussion and conclusions are offered in section 6.

## 2. Model

The exL86 model is given by

$$\frac{d\phi}{dt} = w' + bz' \quad (2.1)$$

$$\frac{dw'}{dt} = -\frac{C}{2} \sin 2(\phi + \epsilon bx) - \frac{\alpha^2 b}{\epsilon} x \quad (2.2)$$

$$\frac{dx}{dt} = \frac{bw' - z'}{\epsilon} \quad (2.3)$$

$$\frac{dz'}{dt} = \frac{\alpha^2 x}{\epsilon}. \quad (2.4)$$

This system admits two kinds of motion (WS00): a chaotic slow mode with a time scale of  $O(1)$ , and a linear gravity wave with frequency  $\epsilon^{-1}$ . The time-scale separation between these two modes is governed by the smallness of  $\epsilon$ ;  $b$  corresponds to a rotational Froude number in the original triad expansion (see appendix), and  $\alpha = (1 + b^2)^{-1/2}$ .

The full derivation of (2.1)–(2.4) spans four papers (Lorenz 1980, 1986; Bokhove and Shepherd 1996; WS00), each with different notation. A summary derivation, using the notation of the present paper, is given in the appendix.

The variables  $\phi$  and  $w'$  represent vorticity, and  $z'$  and  $x$ , respectively, geopotential height and divergence. In the above system,  $\phi$  is an entirely slow variable, corresponding to a geostrophic vortical mode. It is coupled to a mixed time-scale mode defined by the three remaining variables, where (at leading order in  $\epsilon$ )  $w'$  and  $z'$  exhibit both time scales, while  $x$  is entirely a fast variable. The parameter  $C$  is made artificially time-

dependent (WS00), in order to mimic the presence of additional vortical modes, and to make the model's slow mode chaotic. In WS00, and again in this study, this time dependence is given by  $C = a_0 + a_1 \cos(\gamma t)$ , with  $a_0 = 1$ ,  $a_1 = 0.8$ , and  $\gamma = 0.92$ .

Transforming  $w \equiv w' + bz'$  (corresponding to potential vorticity) and  $z \equiv z' - bw'$  (corresponding to geostrophic imbalance), the linear model separates into normal modes, where  $x$  and  $z$  contain the gravity wave, and  $\phi$  and  $w$  are entirely slow variables. The normal mode system is given in the appendix [(A.11)–(A.14)].

One can approximate motion on a slow manifold, where the evolution of the system depends on the slow variables. The fast variables are found diagnostically as functions of the slow variables, and the gravity wave is suppressed. The lowest order approximation to such a manifold is found by setting  $x = z = 0$  and evolving only  $\phi$  and  $w$ . For  $\epsilon = 0$ , this manifold is exact.

For finite  $\epsilon$ , higher-order slaving relations can be derived, which then asymptotically define a fuzzy manifold (WS00). To second order, the slaving relations are given by

$$x = U_x(\phi; \epsilon) = -\frac{\epsilon}{2} Cb \sin 2\phi + O(\epsilon^3) \quad (2.5)$$

$$z = U_z(\phi, w; \epsilon) = \epsilon^2 \left( Cbw \cos 2\phi + \frac{C'}{2} b \sin 2\phi \right) + O(\epsilon^3), \quad (2.6)$$

where  $C'$  is the time derivative of  $C$ .

We define a nonlinear mapping from the slow-manifold variables  $\mathbf{y} = (\phi, w)^T$  to the full model state  $\mathbf{x} = (\phi, w', x, z')^T$ :

$$\mathbf{x} = \mathbf{f}(\mathbf{y}) = \begin{Bmatrix} \phi \\ \alpha^2[w - bU_z(\phi, w; \epsilon)] \\ U_x(\phi; \epsilon) \\ \alpha^2[U_z(\phi, w; \epsilon) + bw] \end{Bmatrix}. \quad (2.7)$$

Note that  $\mathbf{f}$  contains the slaving relations defined to some order in  $\epsilon$ , and is thus not invertible. We also define a mapping  $\mathbf{y} = \mathbf{F}\mathbf{x}$ , which projects some mixed-variable state, whether it is balanced or not, onto the slow manifold. For the exL86 model,

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & b \end{pmatrix}. \quad (2.8)$$

The nonlinearity of  $\mathbf{f}$  is important for data assimilation, as will be shown in subsequent sections, because of linearity approximations made in modeling the evolu-

tion of errors. The relative importance of the nonlinear terms in  $\mathbf{f}$  depends on the size of  $\epsilon$ . For increasing  $\epsilon$ , the balance relationship becomes more strongly nonlinear, while the separation between slow and fast variables becomes concomitantly less well defined.

If a state is initialized using the slaving relations to some order in  $\epsilon$ , it will evolve with a gravity wave of amplitude  $\epsilon^n$ , where  $n - 1$  is the order of initialization. A measure of the imbalance of the state can be defined as  $\text{Imb} \equiv (x - U_x)^2 + (z - U_z)^2$ . WS00 showed that the time scale over which Imb grows is proportional to  $\exp(-2.2/\epsilon)$ .

It is important to note that a free gravity wave in this model can neither propagate away nor interact with other waves. This study therefore does not address the effects of phenomena such as geostrophic adjustment in data assimilation, but focuses instead on the excitation of spurious gravity waves due to the assimilation, and, once induced, their effect on the assimilation scheme.

One can derive a single time-scale model by setting  $\epsilon = 0$  (thus assuming that the two modes are perfectly separated) and setting  $x = z = 0$ . This gives the following system:

$$\frac{d\phi}{dt} = w \quad (2.9)$$

$$\frac{dw}{dt} = -\frac{C}{2} \sin 2\phi. \quad (2.10)$$

In this system the gravity wave is simply not admitted; it corresponds physically to the quasigeostrophic equations. Keeping the parameter  $C$  time-dependent, this system is still chaotic.

### 3. The nonlinear Kalman filter

The Kalman filter (Kalman 1960; Kalman and Bucy 1961; Ghil et al. 1981; Cohn and Parrish 1991) is a discrete time algorithm. It is similar to optimal interpolation (OI), with the special feature that it evolves the forecast and analysis error covariances in time, using a dynamical model.

#### a. Basic equations

If an observation is made at time  $t = k\Delta t$ , the analysis,  $\mathbf{x}_k^a$ , is computed as a linear combination of the model forecast,  $\mathbf{x}_k^f$ , and observation increment (the distance between the observation,  $\mathbf{z}_k$ , and a transformation  $\mathbf{H}_k \mathbf{x}_k^f$  of the model space to the observation space):

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f). \quad (3.1)$$

Here all vectors are column vectors;  $\mathbf{x}_k^a$  and  $\mathbf{x}_k^f$  are  $n$  vectors, where  $n$  is the number of grid points times the number of variables in the model, and  $\mathbf{z}_k$  is an  $m$  vector, where  $m$  is the number of observations.

The forecast at the next time step is given by the evolution of the analysis state by the model:

$$\mathbf{x}_{k+1}^f = \mathcal{M}(\mathbf{x}_k^a) + \mathbf{q}_k, \quad (3.2)$$

where  $\mathcal{M}$  denotes the discretized nonlinear model, and  $\mathbf{q}_k$  represents model error.

The  $(n \times m)$  gain matrix  $\mathbf{K}_k$  weights the innovations by the relative magnitudes of the observation and forecast error covariances. It is optimal if chosen as

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (3.3)$$

where  $\mathbf{R}_k$  is the  $(m \times m)$  observation error covariance matrix, and  $\mathbf{P}_k^f$  is the  $(n \times n)$  forecast error covariance matrix. The observation errors are typically assumed to be uncorrelated and white in time (such that  $\mathbf{R}_k = \mathbf{R}$  is diagonal). Using an approximation of how the model dynamics cause the error probability distribution function (PDF) to evolve,  $\mathbf{P}_k^f$  is evolved in time. The particular way in which this is done characterizes each type of Kalman filter.

Note that (3.3) includes a mapping  $\mathbf{H}_k^T$  from the observation space to the analysis space;  $\mathbf{K}_k$  distributes, over the forecast state, the information from all observations made at one time, according to the dynamical relationships between the variables, which are contained in  $\mathbf{P}_k^f$ .

Concurrently with the analysis step (3.1), forecast error covariances are adjusted to contain the new information that is given by the observation. The resulting analysis error covariance matrix then follows from (3.1):

$$\mathbf{P}_k^a = \langle (\mathbf{x}_k^a - \langle \mathbf{x}_k^a \rangle) (\mathbf{x}_k^a - \langle \mathbf{x}_k^a \rangle)^T \rangle \quad (3.4)$$

$$= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^T, \quad (3.5)$$

where  $\mathbf{I}$  is the  $n \times n$  identity matrix (e.g., Daley 1991, chapter 4) and the angle brackets indicate the ensemble mean value. For the optimal gain matrix (3.3), it can be shown that (3.5) reduces to

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f. \quad (3.6)$$

Since the observation errors project onto all degrees of freedom, (3.1) can return an unbalanced analysis state, unless  $\mathbf{P}_k^f$  contains sufficient information about balance such that the observation increment is distributed in a balanced way.

#### 1) THE EXTENDED KALMAN FILTER

The initial-guess covariance field,  $\mathbf{P}_0^f$ , can either be chosen as some arbitrary matrix, with the assumption

that it will adjust to more physical values as observations are inserted, or it can be estimated using initial background knowledge of the covariance field.

The forecast error at each time step is then approximated using a Taylor expansion of the model about the previous time step. First defining the error vector at time step  $k + 1$  as

$$\mathbf{e}_{k+1}^f \equiv \mathbf{x}_{k+1}^f - \mathbf{x}_{k+1}^t \quad (3.7)$$

(where  $\mathbf{x}_{k+1}^t$  is the truth at time step  $k + 1$ ), then substituting (3.2), we have

$$\begin{aligned} \mathbf{e}_{k+1}^f &= \mathcal{M}(\mathbf{x}_k^a) - \mathcal{M}(\mathbf{x}_k^t) + \mathbf{q}_k \\ &\simeq \mathbf{M}_k \mathbf{e}_k^a + \mathbf{q}_k. \end{aligned} \quad (3.8)$$

Here,  $\mathbf{M}_k = \partial \mathcal{M}_k(\mathbf{x}_k^a) / \partial \mathbf{x}$  is called the tangent linear model, which involves a linearization about the analysis state at time step  $k$ . The forecast error covariance matrix is then found by multiplying (3.9) by its transpose and ensemble averaging, which gives

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k, \quad (3.10)$$

where  $\langle \mathbf{q}_k \mathbf{q}_k^T \rangle \equiv \mathbf{Q}_k$  is the model error covariance matrix. Note that (3.10) assumes that  $\mathbf{e}_k^a$  and  $\mathbf{q}_k$  are uncorrelated, which implies that the  $\mathbf{q}_k$  themselves are sequentially uncorrelated. This is an extreme assumption, since we would expect that a forecast that incorrectly represents the true dynamical balances will have errors that are smoothly correlated in space and time. The representation of balance within  $\mathbf{Q}$  has been investigated by Cohn and Parrish (1991) for the linear case. In this study, however, we will restrict ourselves to a perfect model, and thus set  $\mathbf{Q}_k = 0$  in the evolution of forecast errors, leaving the effect of model error on balance to future studies.

The forward evolution of  $\mathbf{P}^f$ , as given by (3.10), neglects third- and higher order moments in the error covariance field. The ability of the TLM to estimate the correct forecast error PDF over the interval of time between two observations, depends both on the accumulated analysis error preceding it, and the nonlinearity of the model evolution between observations. The repeated addition of observations, followed by forward evolution of the analysis, is expected to make forecast errors smaller, which makes (3.10) a more viable assumption.

Divergence of the EKF happens when the estimated forecast error (and forecast error covariance) is much smaller than the actual distance from the forecast to the truth. In this case, observations will be much farther from the forecast than expected, and will be given too little weight.

## 2) THE ENSEMBLE KALMAN FILTER

In the EnKF, the initial-guess covariance field is simply given by an ensemble of perturbations about the initial state, and its evolution is approximated by the evolving ensemble statistics, with  $\mathbf{P}^f$  at time step  $k + 1$  given by

$$\mathbf{P}_{k+1}^f = \langle (\mathbf{x}_{i,k+1}^f - \langle \mathbf{x}_{i,k+1}^f \rangle) (\mathbf{x}_{i,k+1}^f - \langle \mathbf{x}_{i,k+1}^f \rangle)^T \rangle, \quad (3.11)$$

where the  $i$ th member in the ensemble is given by

$$\mathbf{x}_{i,k+1}^f = \mathcal{M}(\mathbf{x}_{i,k}^a) + \mathbf{q}_{i,k}, \quad (3.12)$$

where  $\mathbf{q}_{i,k}$  represents a perturbation to  $\mathbf{q}_k$ .

Each ensemble member is adjusted according to (3.1), with a perturbed observation,  $\mathbf{z}_{i,k} = \mathbf{z}_k + \mathbf{r}_{i,k}$ , with random perturbations  $\mathbf{r}_{i,k}$  representing observation error. The EnKF analysis is then simply the ensemble mean:

$$\mathbf{x}_k^a = \langle \mathbf{x}_{i,k}^a \rangle, \quad (3.13)$$

with analysis error covariance matrix

$$\mathbf{P}_k^a = \langle (\mathbf{x}_{i,k}^a - \langle \mathbf{x}_{i,k}^a \rangle) (\mathbf{x}_{i,k}^a - \langle \mathbf{x}_{i,k}^a \rangle)^T \rangle. \quad (3.14)$$

The EnKF thus preserves nonlinearity in the evolution of forecast error statistics. However, the accuracy of the EnKF rests on the assumption that a finite-size forecast ensemble, and its forward evolution, captures the true error statistics both between observation times and after the analysis step. If the ensemble is unable to sufficiently capture the error correlations of a specific problem, while the linearity assumptions made in the EKF are justified, the EKF may have the advantage.

Information is lost if the ensemble spreads so much between observation times that the distribution of forecasts does not have a sharp mean. In this case, however, the ensemble variance reflects the uncertainty in the forecast and is thus still useful. If some components of the forecast are better known than others, this information is retained.

Divergence of the EnKF occurs when the ensemble mean is much further from the truth than any of the ensemble members are from the mean; in this case, the predicted forecast error variance would reflect a high confidence in the forecast, and the actual forecast error would be much larger than the predicted error.

Some argument exists in the literature about how an ensemble should be initialized (e.g., Houtekamer and Mitchell 1998). Many variations of the standard EnKF have been proposed in the literature (Evensen 2003), and it is likely that the treatment of balance could vary



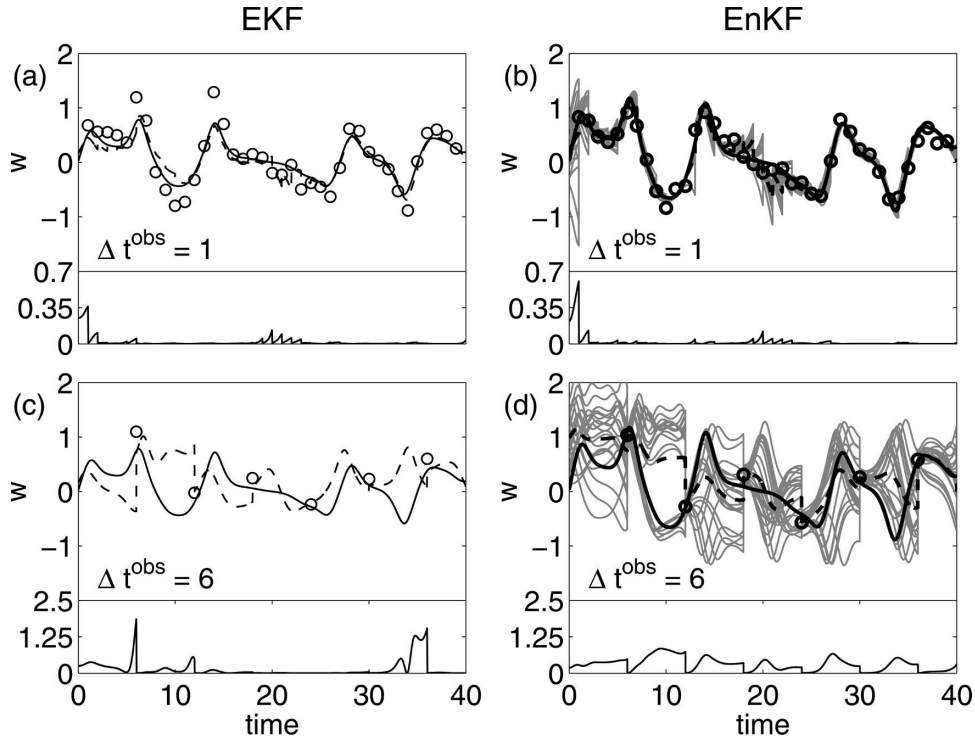


FIG. 1. Sample analyses of the single time-scale model, for the (left) EKF and (right) EnKF, at two observation frequencies, with observations of  $w$  (circles) and the same initial perturbation. In all plots, the true state of  $w$  (solid) is compared to the analysis (dashed) and, in the EnKF case, the ensemble (gray). The forecast error covariance  $\langle (e_w^f)^2 \rangle$  is shown below the respective plots for each case.

greatly between them. As a starting point, this study uses only the basic, perturbed-observation EnKF, as introduced by Evensen (1994) and modified by Burgers et al. (1998) and Houtekamer and Mitchell (1998), and we leave the comparison of other schemes to future research.

*b. Identical twin experiments*

As mentioned above, in this study the truth and forecast are evolved with the same model, and thus  $\mathbf{Q}_k = 0$ . We return to this point in the final section of the paper.

For one particular solution of the model chosen to be the truth,  $\mathbf{x}_0^t$ , an initial forecast is given by

$$\mathbf{x}_0^f = \mathbf{x}_0^t + \delta\mathbf{x}_0, \tag{3.15}$$

and observations are generated at time intervals  $\Delta t^{\text{obs}}$ , as

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k^f + \delta\mathbf{x}_k^{\text{obs}}, \tag{3.16}$$

where  $\delta\mathbf{x}_0$  and  $\delta\mathbf{x}_k^{\text{obs}}$  are random vectors chosen from normal distributions,  $\mathcal{N}(0, \sigma_0^2 \mathbf{I})$  and  $\mathcal{N}(0, \sigma_{\text{obs}}^2 \mathbf{I})$ , with zero means and  $\sigma_0^2$  and  $\sigma_{\text{obs}}^2$  specified initial-state and observation error variances, respectively.

The performance of either filter is evaluated by computing the true error following each analysis. To separate the balance problem from general filter divergence, we separate the total squared error into fast and slow components:

$$e_{\text{fast}}^2 = (x^a - x^f)^2 + (z^a - z^f)^2 \tag{3.17}$$

$$e_{\text{slow}}^2 = (\phi^a - \phi^f)^2 + (w^a - w^f)^2. \tag{3.18}$$

In all examples shown in this paper, we set  $\sigma_{\text{obs}}^2 = 0.01$  and  $\sigma_0^2 = 0.25$ . Unless otherwise stated, we set  $\epsilon = 0.1$ , and compute the slaving relations (2.5)–(2.6) up to the  $O(\epsilon^2)$  terms. In all experiments shown, the observed variable is  $w$  (in the single time-scale case) or  $w'$  (in the two time-scale case).

*c. Assimilation for a single time scale*

To separate the problem of balance from that of general nonlinearity and chaos, we first establish how well the EKF and EnKF estimate the model state in the single time scale model [(2.9)–(2.10)]. This analysis is qualitatively similar to that of Miller et al. (1994) and Evensen (1997).

Figure 1 shows two example assimilation runs for the



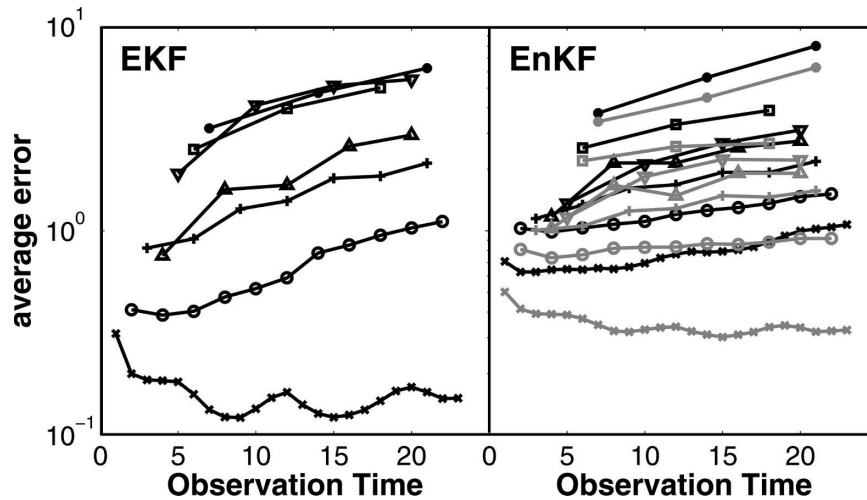


FIG. 2. Comparison of the average state error (over 600 runs) immediately following the analysis steps with (left) the EKF and (right) the EnKF, for the single time-scale model. For the EnKF, we also compare ensemble sizes  $N = 4$  (black) and  $N = 10$  (gray). Observation frequencies shown are  $\Delta t^{\text{obs}} = 1$  (x), 2 (circles), 3 (+), 4 (triangles), 5 (inverted triangles), 6 (squares), and 7 (dots). In all runs, the initial true state is randomly generated and then the forecast and observations are generated using (3.15) and (3.16).

single-time scale model, using the EKF (left) and the EnKF (right). Observations of  $w$  are taken every time unit (top row) and every six time units (i.e., roughly once during a typical cycle of the slow mode, bottom row) in each case, with the same truth for all cases. Each plot compares the true state to the assimilated analysis, along with the forecast error variance  $\langle (e_w^f)^2 \rangle$ . For the EKF cases in this example, the initial forecast error covariance matrix is estimated by a diagonal matrix,

$$\mathbf{P}_0^f = \sigma_0^2 \mathbf{I}, \quad (3.19)$$

with the expectation that the series of analysis steps will adjust  $\mathbf{P}^f$  to a more physical value (the criticality of this assumption, in terms of balance, is explored more thoroughly below). An ensemble of  $N = 20$  forecasts is used for the EnKF. The examples in Fig. 1 illustrate several key properties of Kalman filter assimilation for nonlinear, chaotic dynamics.

Good analysis increments are made when growth in the forecast error covariance for each variable reflects the rate at which the forecast is diverging from the true state. The forecast error grows when the forecast and true state diverge, and then decreases sharply when an observation is made. For the EnKF (Figs. 1b,d), growth in the forecast error variance term for each variable reflects the visible spread in the ensemble.

The accuracy of the analysis, in both schemes, decreases for  $\Delta t^{\text{obs}} = 6$ . A difference between the EKF and EnKF also starts to become clear at larger  $\Delta t^{\text{obs}}$ : the

spread of the ensemble between observations, in the EnKF case, becomes less Gaussian, which means that TLM-predicted error covariances will become less accurate.

Note also that the analyses of  $w$  in Fig. 1 depend on the ability of each Kalman filter to estimate the entire state ( $w$  and  $\phi$ ) from observations of just one component of it ( $w$ ). This is the crux of the overall data assimilation problem, and will form the basis of the more specific problem of balance.

A more quantitative comparison of the two filters is shown in Fig. 2, which shows the analysis error immediately following the insertion of each observation, averaged over 600 runs, for the EKF (left panel) and the EnKF (right panel), at  $\Delta t^{\text{obs}} = 1-7$ . For the EnKF case, errors are shown for runs with ensembles of 4 and 10 forecasts. Comparison of these two figures reveals that the performance of the EnKF is similar to the EKF if the forecast ensemble consists of four forecasts, though the analysis error in the EnKF is somewhat reduced when the ensemble size is increased to 10 forecasts. Moreover, analysis errors grow in time for all cases shown, except for  $\Delta t^{\text{obs}} = 1$  for the EKF, and  $\Delta t^{\text{obs}} = 1$  and 2 for the 10-forecast EnKF.

#### d. Balance in the assimilation problem

In light of the single time-scale results, we now ask whether, and under what conditions, the EKF and EnKF evolve forecast error covariances that reflect the dynamical balance in the true state. The degree of bal-

ance captured by  $\mathbf{P}^f$ , for both Kalman filters, will depend on each scheme's treatment of balance in (i) the initial-guess error covariances, (ii) the evolution of error covariances, and (iii) the repeated adjustment of error covariances by the Kalman gain.

To examine what it means for error covariances to reflect balance, we define a balanced error covariance matrix as one where the errors in the "slaved" fast variables are dependent on the errors in the slow "master" variables. For a perfectly balanced state with zero free fast motion, the errors in the fast variables will then be entirely functions of the errors in the slow variables.

Defining  $\mathbf{e}_y^f = \mathbf{y}^f - \mathbf{y}^t$  as the forecast error vector in terms of the slow variables, the forecast error in terms of mixed time-scale variables can be approximated as a Taylor expansion of (2.7):

$$\mathbf{e}_x^f = \mathbf{x}^f - \mathbf{x}^t = \mathbf{L}\mathbf{e}_y^f + \text{nonlinear terms}, \quad (3.20)$$

where  $\mathbf{L} = (\partial\mathbf{f}/\partial\mathbf{y})|_y$  is the first derivative of the balance relationship, evaluated about the slow-manifold forecast state at some point in time. The forecast error covariances are found by multiplying (3.20) by its transpose, and computing the ensemble mean. Truncating (3.20) at the linear term, the forecast error covariance matrix, in terms of the full model state, can be approximated as

$$\begin{aligned} \mathbf{P}_x^f &= \langle \mathbf{e}_x^f (\mathbf{e}_x^f)^T \rangle \approx \langle (\mathbf{L}\mathbf{e}_y^f) (\mathbf{L}\mathbf{e}_y^f)^T \rangle \\ &= \mathbf{L} \langle \mathbf{e}_y^f (\mathbf{e}_y^f)^T \rangle \mathbf{L}^T = \mathbf{L}\mathbf{P}_y^f \mathbf{L}^T, \end{aligned} \quad (3.21)$$

where we have defined

$$\mathbf{P}_y^f = \langle \mathbf{e}_y^f (\mathbf{e}_y^f)^T \rangle \quad (3.22)$$

as the error covariance matrix in terms of the slow variables only. Since (3.21) is a tangent linear operation, a covariance matrix  $\mathbf{P}_x^f$  that is formulated in this way can be thought of as tangent to the balance manifold. Thus, (3.21) will hereafter be referred to as a tangent linear balance transformation, or TLBT, in analogy to the TLM (3.10). As in the TLM, this approximation neglects dependence on higher order statistical moments in the forecast error distribution.

For the exL86 model,

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ -\alpha^2 b \frac{\partial U_z}{\partial \phi} & \alpha^2 \left( 1 - b \frac{\partial U_z}{\partial w} \right) \\ \frac{\partial U_x}{\partial \phi} & \frac{\partial U_x}{\partial w} = 0 \\ \alpha^2 \frac{\partial U_z}{\partial \phi} & \alpha^2 \left( \frac{\partial U_z}{\partial w} + b \right) \end{bmatrix}, \quad (3.23)$$

and is a function of the slow variable state. The accuracy of (3.21) depends (as in the TLM) on the magnitude of  $\mathbf{e}_y^f$ , but also on the size of  $\epsilon$ , and the order of accuracy in  $\epsilon$  to which the balance approximation is made.

#### 4. Balance in the extended Kalman filter

##### a. Example

Figure 3 shows four example EKF assimilation experiments for the exL86 model, all with the same (balanced) true state and initial perturbation. Observations in all four cases are taken of  $w'$ , and thus contain both time scales. The initial and final values of  $\text{Imb}$  (for the analysis) are shown for each case, and the forecast error variance  $\langle (e_w^f)^2 \rangle$  is shown underneath each analysis. The forecasts in Figs. 3a,b are initialized using (2.5)–(2.6), following the initial perturbation, thus reflecting a case where there is a priori knowledge of the absence of gravity waves. In Figs. 3c,d, the initial forecasts are unbalanced, with initial imbalance due only to initial forecast error.

Very frequent observations ( $\Delta t^{\text{obs}} = 1$ , left column of plots), are again compared to observations taken on the order of a typical cycle of the slow mode ( $\Delta t^{\text{obs}} = 6$ , right column of plots). In all four cases shown, the initial covariance field is estimated by (3.19), where  $\mathbf{I}$  is now the  $4 \times 4$  identity matrix.

The primary result of this example is that balance in the EKF analysis depends largely on observation frequency. We note also that very frequent observations not only retain balance in a balanced initial forecast (Fig. 3a), but can also establish balance in an unbalanced initial forecast (Fig. 3c), despite the fact that no balance information is contained in the initial forecast error covariances. For  $\Delta t^{\text{obs}} = 6$ , on the other hand, balance deteriorates after a few observations, and most information from subsequent observations is consequently rejected. In these cases, the covariance model itself takes on a significant fast oscillation, which is not removed by the addition of observations.

Though this example depends in detail on the particular instabilities and realizations of the random errors used, it points out two important factors for balance in the EKF. The first is observation frequency: For longer  $\Delta t^{\text{obs}}$ , the forecast drifts more from the truth, resulting in a larger analysis increment, which in turn amplifies any misestimations of the forecast errors and, more importantly, of the balance relationship in the error covariances. The second factor is the initialization of the covariance model, via  $\mathbf{P}_0^f$ . Since the forecast error covariances in Figs. 3b,d clearly do not adjust to reflect the balance relationship, it is worth investigating to

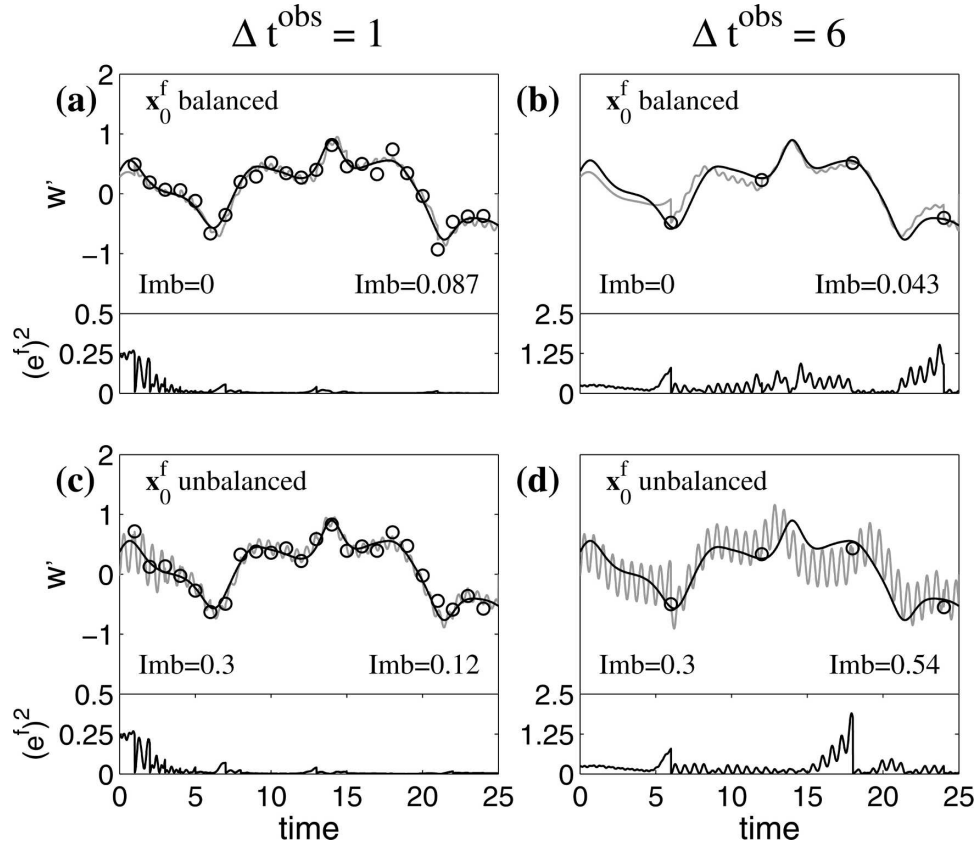


FIG. 3. Four sample EKF assimilations for the full (multiple time scale) exL86 model. For each case, the truth (black) and analysis (gray) of  $w'$  are shown, along with the observations (circles). Underneath each plot is shown the forecast error variance for  $w'$ . (a) Balanced initial forecast, with  $\Delta t^{\text{obs}} = 1$ . (b) Balanced initial forecast, with  $\Delta t^{\text{obs}} = 6$ . (c) Unbalanced initial forecast, with  $\Delta t^{\text{obs}} = 1$ . (d) Unbalanced initial forecast, with  $\Delta t^{\text{obs}} = 6$ . The initial perturbation is  $\delta \mathbf{x} = (0.450, -0.800, -0.465, 0.725)^T$ . Initial and final values of Imb are shown on each plot.

what extent the covariance model may be improved by providing it with an initial knowledge of balance, as in (3.21), and the extent to which this information is retained as the assimilation progresses.

### b. Balance in the EKF covariance model

How does the EKF covariance model capture the balance relationship? This problem has three general components.

#### 1) INITIAL ESTIMATE OF $\mathbf{P}^f$

Instead of specifying an initial time error covariance matrix using (3.19)—that is, with no correlations between variables—one might instead initialize  $\mathbf{P}^f$  using the TLBT (3.21), or some approximation to it. The accuracy of such an initialization depends on how well the balance relation is known, and the validity of the linearization. For small  $\epsilon$ , the higher order balance terms

become less important, and the drift of an initialized state from the slow manifold becomes slower. The nonlinear terms in  $\mathbf{f}$  also become less important for small  $\epsilon$ . Thus, the usefulness of balance-initializing  $\mathbf{P}_0^f$  depends on the smallness of  $\epsilon$ .

#### 2) TLM EVOLUTION OF $\mathbf{P}^f$

It is not clear whether a forecast error covariance matrix initialized using the TLBT will remain tangent to the slow manifold as it is evolved in the TLM (3.10). This question can be examined by comparing the TLM evolution of the error vector (3.9) to the true error (3.8). If, at time step  $k$ , both the forecast and the true state are balanced, such that

$$\mathbf{x}'_k = \mathbf{f}(\mathbf{y}'_k) \quad (4.1)$$

$$\mathbf{x}^a_k = \mathbf{f}(\mathbf{y}^a_k), \quad (4.2)$$

and if the model error is zero, then the true error at the next time step will be given by

$$\mathbf{e}_{\mathbf{x},k+1}^f = \mathcal{M}[\mathbf{f}(\mathbf{y}_k^a)] - \mathcal{M}[\mathbf{f}(\mathbf{y}_k^f)]. \quad (4.3)$$

If the forecast model evolves a balanced state to produce a balanced state, then

$$\mathbf{e}_{\mathbf{x},k+1}^f = \mathbf{f}(\mathbf{y}_{k+1}^f) - \mathbf{f}(\mathbf{y}_{k+1}^a). \quad (4.4)$$

In contrast, if the forecast errors at time step  $k$  are balanced according to (3.21), then the forecast errors at the next time step are given by

$$\mathbf{e}_{\mathbf{x},k+1}^f = \mathbf{M}_k \mathbf{L}_k \mathbf{e}_{\mathbf{y},k}^a. \quad (4.5)$$

They remain tangent to the slow manifold if

$$\mathbf{M}_k \mathbf{L}_k \mathbf{e}_{\mathbf{y},k}^a = \mathbf{L}_{k+1} \mathbf{e}_{\mathbf{y},k+1}^f. \quad (4.6)$$

### 3) ANALYSIS STEP

For the standard EKF, use of the TLM is justified if the information brought in from observations in (3.6) is able to keep the evolving covariance model close to the true error statistics. Extending this to balance, we expect that the assimilation of observations can also make the degree of balance represented in the covariance model more accurate.

If the forecast error covariance matrix (in terms of mixed variables) is balanced—that is, if  $\mathbf{P}_{\mathbf{x},k}^f = \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{L}_k^T$ —then the gain matrix becomes

$$\mathbf{K}_{\mathbf{x},k} = \mathbf{P}_{\mathbf{x},k}^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{\mathbf{x},k}^f \mathbf{H}_k^T + \mathbf{R})^{-1} \quad (4.7)$$

$$= \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{L}_k^T \mathbf{H}_k^T (\mathbf{H}_k \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{L}_k^T \mathbf{H}_k^T + \mathbf{R})^{-1} \quad (4.8)$$

$$= \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{G}_k^T (\mathbf{G}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{G}_k^T + \mathbf{R})^{-1} \quad (4.9)$$

$$\equiv \mathbf{L}_k \mathbf{K}_{\mathbf{y},k}, \quad (4.10)$$

where we have defined  $\mathbf{K}_{\mathbf{y},k}$  as the gain matrix in terms of the slow variables, and  $\mathbf{G}_k = \mathbf{H}_k \mathbf{L}_k$  as a generalized observation operator. Note that  $\mathbf{G}_k$  selects only the slow-manifold projection of the observed variable, and  $\mathbf{K}_{\mathbf{x},k}$ , consequently, includes the TLBT.

The analysis error covariance matrix is then given by

$$\mathbf{P}_{\mathbf{x},k}^a = (\mathbf{I} - \mathbf{K}_{\mathbf{x},k} \mathbf{H}_k) \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{L}_k^T \quad (4.11)$$

$$= \mathbf{L}_k (\mathbf{P}_{\mathbf{y},k}^f - \mathbf{L}_k^{-1} \mathbf{K}_{\mathbf{x},k} \mathbf{H}_k \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^f) \mathbf{L}_k^T \quad (4.12)$$

$$= \mathbf{L}_k (\mathbf{I} - \mathbf{K}_{\mathbf{y},k} \mathbf{G}_k) \mathbf{P}_{\mathbf{y},k}^f \mathbf{L}_k^T \quad (4.13)$$

$$= \mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^a \mathbf{L}_k^T. \quad (4.14)$$

Since  $\mathbf{P}_{\mathbf{x},k}^a$  can be written as  $\mathbf{L}_k \mathbf{P}_{\mathbf{y},k}^a \mathbf{L}_k^T$ , it is still tangent to the slow manifold. Thus, (3.6) retains the TLBT, while still being reduced to include information from the observations.

The above transformations are equivalent to those derived by Cohn and Parrish (1991) for the linear case.

For the present (nonlinear) case, balanced forecast errors in the EKF rely on four assumptions. First, the TLM must be a valid approximation over  $\Delta t^{\text{obs}}$ , which depends on the size of the forecast error. Second, the TLBT must be a valid approximation. This also depends on the size of the forecast error, as well as on the size of  $\epsilon$  and the order of accuracy of the slaving relations. Third, the initial truth and initial forecast must be balanced. Fourth, the model evolution must preserve balance. In the exL86 model, this is only true to the order in  $\epsilon$  to which the model was initialized.

Therefore, while it makes sense to initialize  $\mathbf{P}_0^f$  using the TLBT, it is not obvious that such an initialization will ensure a balanced assimilation.

### c. Numerical evaluation of the EKF

Figure 4 shows the average, over 600 assimilation runs, of the fast and slow analysis errors [(3.17)–(3.18)], immediately following each analysis, as a function of observation time. The observed variable in all runs is  $w'$ . Both figures compare runs where  $\mathbf{P}_0^f$  is initialized using the TLBT, to ones where  $\mathbf{P}_0^f$  is chosen as a diagonal matrix (3.19). Also shown are runs where it is assumed that the balance relations are not known, but are guessed to be functions that are proportional to  $\epsilon$ , such that the tangent error covariances are approximated as

$$\mathbf{P}_0^f = \sigma_0^2 \begin{pmatrix} 1 & 0 & \epsilon & 0 \\ 0 & 1 & \epsilon & 1 \\ \epsilon & \epsilon & \epsilon^2 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \quad (4.15)$$

For each case, three observation frequencies ( $\Delta t^{\text{obs}} = 2, 4, \text{ and } 6$ ) are shown.

The first thing to note is that the analysis of the slow mode is unstable for  $\Delta t^{\text{obs}} \geq 2$ ; the average slow analysis errors all grow as the assimilation progresses. Average fast error does not depend strongly on observation frequency, and in fact tends to level off as the assimilation progresses. This indicates that the failure of the EKF at these observation frequencies is due to filter divergence, which can be verified by comparing the forecast errors and true errors for these cases. Forecast errors become much smaller than the true errors (not shown), and observations consequently have no impact on the forecast. For the chaotic slow mode, this means that the distance between truth and analysis will then grow in time, whereas the spurious fast wave, and thus fast error, neither decays nor grows.

However, initializing  $\mathbf{P}_0^f$  with a knowledge of the balance relationship clearly decreases the overall fast er-

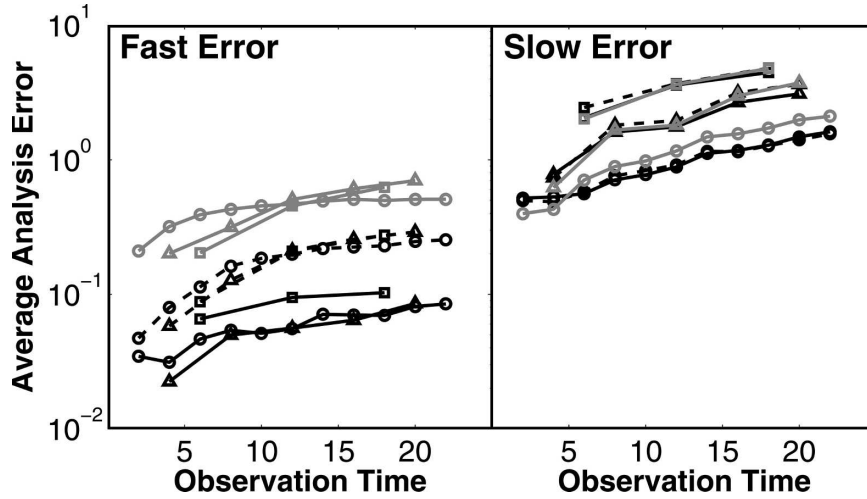


FIG. 4. Comparison of the average (left) fast and (right) slow analysis error, following each analysis step, between three ways of initializing the error covariance matrix: either using the balance approximation [(3.21), black], or by initializing  $\mathbf{P}_0^f$  as a diagonal matrix [(3.19), gray], or by guessing at a balance relation [(4.15), dashed]. Three observation frequencies are compared for each case:  $\Delta t^{\text{obs}} = 2$  (circles), 4 (triangles), and 6 (squares). The average is over 600 assimilation experiments.

ror. When  $\mathbf{P}_0^f$  is initialized without correlations between variables, the average fast error immediately exceeds the magnitude of the slaved mode in the true state (2.5)–(2.6), which is  $O(\epsilon)$ . If  $\mathbf{P}^f$  is initialized with the TLBT, the average fast error (for the assimilation periods considered) stays below the magnitude of the slaved fast mode, even for  $\Delta t^{\text{obs}} = 6$ . Even an educated guess at balanced error covariances (4.15) improves the fast error substantially. While the growth of fast error in each case is similar, the overall imbalance is smaller if  $\mathbf{P}_0^f$  is properly initialized, simply because the spurious imbalance induced by the first few analysis steps is less.

We note also that the average analysis error for the slow mode does not depend significantly on the initialization of  $\mathbf{P}_0^f$ . This means that an EKF initialized with (3.19) may still succeed in terms of the slow mode, even while the analysis contains a spurious fast mode.

Figure 5 examines the effectiveness of initializing  $\mathbf{P}_0^f$  using the TLBT, as a function of  $\epsilon$ . It shows the fast error following the single insertion of an observation at  $t = 4$  (for 600 different assimilation runs) with varying values of  $\epsilon$ , and with  $\mathbf{P}_0^f$  initialized either using (3.19) or (3.21). A curve corresponding to  $\epsilon^2$  is added to Fig. 5 in order to emphasize the asymptotic nature of the TLBT. For  $\epsilon$  smaller than about 0.3, the initialization of  $\mathbf{P}^f$  with the TLBT tends to yield an analysis with smaller fast error, while as  $\epsilon$  increases, the initialization of  $\mathbf{P}_0^f$  no longer makes a difference in the fast error. This makes sense, since the separation of fast and slow modes becomes asymptotically less well defined as  $\epsilon \rightarrow 1$ .

d. EKF with a balance constraint

A balance constraint can be incorporated into the EKF, by performing the analysis step in terms of the projection of the model state onto the slow variables  $\mathbf{y} = (\phi, w)^T$  only. Starting from a mixed time-scale forecast  $\mathbf{x}_k^f$ , the projection onto the slow manifold is given by

$$\mathbf{y}_k^f = \mathbf{F}\mathbf{x}_k^f \tag{4.16}$$

$$\mathbf{P}_y^f = \mathbf{F}\mathbf{P}_x^f\mathbf{F}^T. \tag{4.17}$$

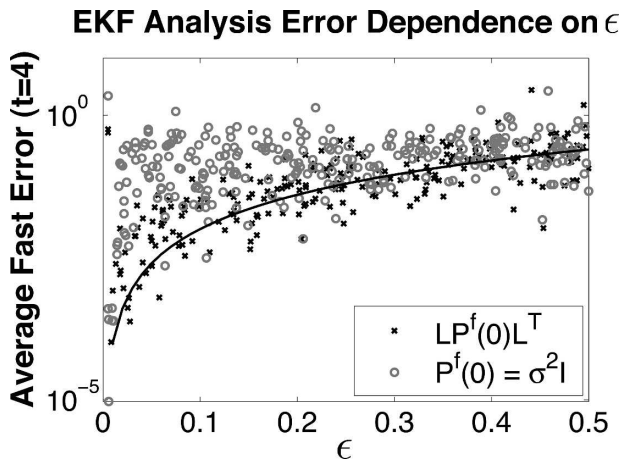


FIG. 5. Fast error immediately following the insertion of an observation at  $t = 4$ , for 600 randomly initialized runs (as in Fig. 4), as a function of  $\epsilon$ . The x's are runs where  $\mathbf{P}_0^f$  is initialized using the balance approximation (3.21), and circles where  $\mathbf{P}_0^f$  is initialized as a diagonal matrix (3.19). The  $\epsilon^2$  curve is also shown (solid line).



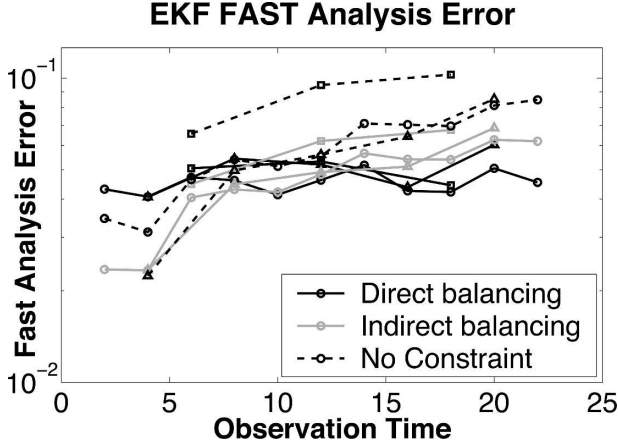


FIG. 6. As in Fig. 4, but now comparing the two balance-constraint modifications to the EKF: either by (black, solid) directly balancing the analysis via (4.18) followed by (2.7), or by (gray, solid) indirectly mapping the analysis increment as in (4.19). These modified schemes are compared to the regular EKF (black, dashed), where  $\mathbf{P}_0^f$  is initialized using the tangent approximation (3.21). As in previous figures, observation frequencies shown are  $\Delta t^{\text{obs}} = 2$  (circles), 4 (triangles), and 6 (squares).

It is followed by the analysis step:

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{K}_{\mathbf{y},k}[\mathbf{z}_k - \mathbf{H}_k \mathbf{f}(\mathbf{y}_k^f)], \quad (4.18)$$

where  $\mathbf{K}_{\mathbf{y},k} = \mathbf{P}_{\mathbf{y},k}^f \mathbf{G}_k^T (\mathbf{G}_k \mathbf{P}_{\mathbf{y},k}^f \mathbf{G}_k^T + \mathbf{R})^{-1}$ . Note that  $\mathbf{K}_{\mathbf{y},k}$  uses the TLBT in order to select the slow-manifold projection of the observation error. The analysis in terms of the full model state can then be computed by mapping the slow state back to a balanced mixed-variable state, using (2.7).

Alternatively, one could map the gain matrix back to the mixed-variable state, and update the mixed time-scale state as

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_{\mathbf{x},k}[\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f] = \mathbf{x}_k^f + \mathbf{L}_k \mathbf{K}_{\mathbf{y},k}[\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f], \\ &= \mathbf{x}_k^f + \mathbf{L}_k \Delta \mathbf{y}_k, \end{aligned} \quad (4.19)$$

where  $\Delta \mathbf{y}_k = \mathbf{y}_k^a - \mathbf{y}_k^f = \mathbf{K}_{\mathbf{y},k}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f)$  is the analysis increment on the slow manifold. However, this requires an additional use of the TLBT.

A comparison between these two modified analysis steps and the regular EKF is shown as before, in the average fast error (Fig. 6) immediately following each analysis step, for observation frequencies  $\Delta t^{\text{obs}} = 2, 4,$  and  $6$ . For increasing  $\Delta t^{\text{obs}}$ , the modified schemes offer a substantial improvement over the TLBT-EKF, with the direct-transformation analysis [(4.18)] yielding on average more balanced analyses than the indirect analysis [(4.19)]. This is not surprising because the indirect transformation uses an additional approximation. However, for frequent observations, the indirect relationship may be sufficient.

### 5. Balance in the ensemble Kalman filter

The single time-scale example in section 3 showed that the EnKF, by preserving the nonlinearity of the evolving error distribution, may be preferable to the EKF. The EnKF is a (weakly nonlinear) combination of model states (Evensen 2003), which suggests that the EnKF analysis might naturally be more balanced. Also, the averaging nature of the analysis step in itself implies a kind of balancing. A balanced analysis state is still not guaranteed, however, as long as fast waves are allowed in the individual ensemble members.

How can the EnKF covariance model capture balance? If  $\mathbf{P}^f$  is given by an ensemble of balanced forecasts, no TLBT approximation is required. Balanced perturbations for the exL86 model can be generated by transforming the central forecast to normal modes, adding random perturbations to  $\phi^f$  and  $w^f$ , and then transforming back to mixed variables using (2.7). [In realistic applications, this step is a bit more complicated, but similar. For example, one can randomly perturb streamfunction, and then derive wind, temperature, and pressure perturbations following some balance assumption (see, e.g., Mitchell et al. 2002). In lieu of an explicit slow manifold initialization, one might also integrate the ensemble forward while filtering out fast waves with, say, a digital filter (Evensen 1997).]

Whether or not the forward evolution of such an ensemble and the subsequent ensemble analysis indeed yields a balanced analysis, however, depends on three assumptions, which in turn depend on observation frequency and the size of the forecast ensemble: First, that the evolving  $N$ -member forecast ensemble sufficiently represents the full statistics of the true system, including balance; second, that the analysis step does not destroy the ensemble's representativeness; and, third, that the ensemble analysis step (3.13) yields a balanced state.

#### a. Example

Figure 7 shows two sample EnKF assimilation runs, with observations of  $w'$ , and all other assimilation parameters as in the EKF example (section 4). The observation frequency is  $\Delta t^{\text{obs}} = 1$  in Fig. 7a, and  $\Delta t^{\text{obs}} = 6$  in Fig. 7b. To illustrate more clearly how the forecast ensemble captures the balance relationship, the analyses of  $x$  are shown for each case along with the analyses of  $w'$ . In both cases shown, the ensemble size is  $N = 10$  forecasts.

For  $\Delta t^{\text{obs}} = 1$ , the ensemble collapses toward a single forecast which is often indistinguishable from the true state, whereas for  $\Delta t^{\text{obs}} = 6$ , the ensemble diverges significantly between observations—though, in both cases,



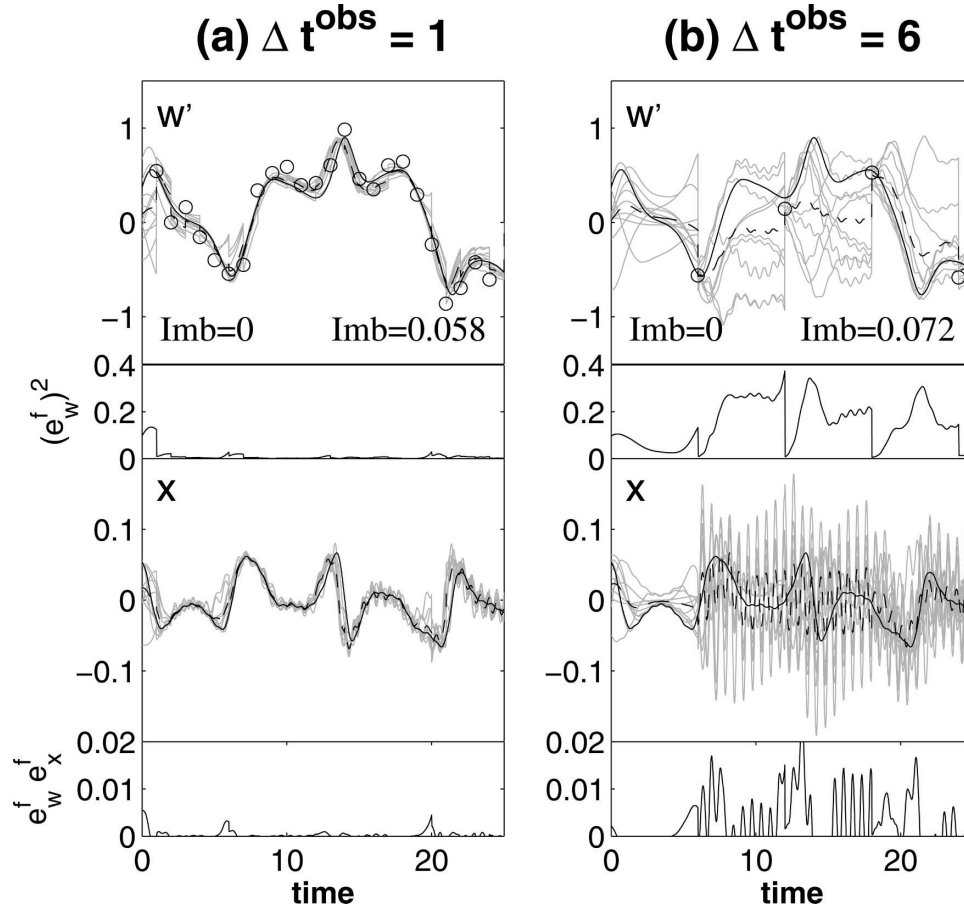


FIG. 7. Two sample assimilation runs with the EnKF, comparing the truth (black), the forecast ensemble (gray), and the ensemble mean (dashed) of (top)  $w'$  and (bottom)  $x$  for each case. (a) The observation frequency is  $\Delta t^{\text{obs}} = 1$ , and (b)  $\Delta t^{\text{obs}} = 6$ . Underneath each plot is shown the corresponding forecast error covariance. Observations of  $w'$  are indicated with circles. The initial perturbation is, as in the EKF example,  $\delta \mathbf{x} = (0.450, -0.800, -0.465, 0.725)^T$ .

the analysis of  $w'$  is consistently brought back to the truth state.

Note, however, that some amount of imbalance is generated in the individual ensemble members (this can be seen best in the plots of  $x$  in both cases). While the forecast error covariances contain this imbalance, it is much smaller than in the EKF example (Fig. 3). Moreover, the net amount of imbalance in the ensemble-mean analysis is small compared to the magnitude of the slow variables, simply because the average of the many unbalanced forecasts is a more balanced state. Compared to the degrees of freedom of the model,  $N = 10$  is large; the effect of ensemble size is more closely investigated in the next subsection.

#### b. Balance in the ensemble covariance model

The example shown illustrates three main components of the balance problem in the EnKF.

#### 1) ENSEMBLE REPRESENTATION OF BALANCE

It can be argued that the existence of a balance relation might simplify the EnKF problem, since a balanced model state has fewer degrees of freedom than an unbalanced state, and therefore fewer ensemble members will actually be required to represent the error statistics of a balanced state—if the forecasts in the ensemble are all balanced. If the assimilation of observations unbalances the forecasts (as in Fig. 7), however, then the existence of a true balance will bring about the additional difficulty that the actual error statistics could differ greatly from those contained within the ensemble.

#### 2) EVOLUTION OF ENSEMBLE STATISTICS BETWEEN OBSERVATION TIMES

As the forecast ensemble evolves between observation times, imbalance in the error covariances will only

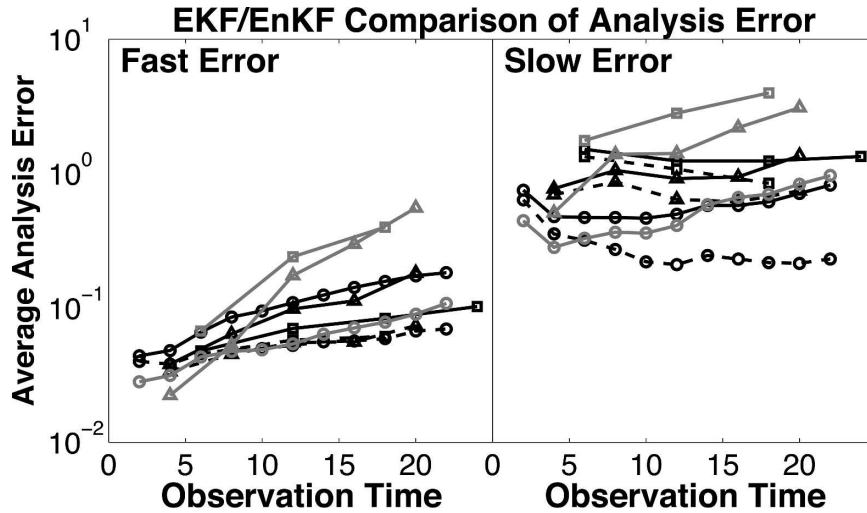


FIG. 8. Comparison of the average (left) fast and (right) slow analysis error, now comparing the EKF (gray) to the EnKF, for ensembles of  $N = 4$  forecasts (black, solid) and  $N = 10$  forecasts (black, dashed). For the EKF,  $\mathbf{P}_0^f$  is initialized using the TLBT (3.21). In these experiments, observations of  $w'$  were randomly alternated with observations of  $\phi$ .

grow as much as the mean imbalance in the ensemble (as opposed to the unbounded growth that would happen with the TLM evolution of errors). Even if the ensemble spread becomes saturated in the slow mode, it could still contain information about the (slaved) fast mode (though this is not shown in Fig. 7).

### 3) ENSEMBLE ANALYSIS STEP

It was shown above that even an ensemble of balanced forecasts will to some extent result in individual analyses that are unbalanced. The amount of imbalance remaining in the ensemble mean will then depend on the magnitude and relative phases of the fast motion in individual analyses. It may thus be argued that, for models with a spatial dimension and a spectrum of possible gravity waves, spurious imbalance in individual ensemble members will easily disappear in the ensemble average. The results of Houtekamer and Mitchell (1998) suggest, however, that spurious imbalance in the ensemble mean analysis is difficult to avoid for realistic ensemble sizes.

It is also important to note that the ensemble can develop a very small variance (reflecting high confidence in the forecast) around a significantly unbalanced mean forecast. Thus, a form of filter divergence, in terms of balance, can happen, wherein the forecast ensemble has a very narrow distribution, but predicts the wrong type of motion—in this case, a gravity wave. Here we note that filter divergence is related to the details of how the ensemble is generated and how the analysis step is carried out. While we are aware of the known caveats of the Burgers et al. (1998) and

Houtekamer and Mitchell (1998) formulation of the EnKF, it is used here as a first step to illustrating the balance properties of the EnKF.

The example in Fig. 7 shows that, while the EnKF may be advantageous for at least two reasons (by allowing for the nonlinear evolution of error statistics, and because spurious imbalance may be filtered out in the ensemble mean), even if the assimilation is begun with an ensemble of balanced forecasts, the repeated adjustment of these forecasts may not yield an analysis that is balanced.

### c. Numerical evaluation of the EnKF

A comparison of the EnKF and EKF is shown in Fig. 8, again comparing the mean (over 600 runs) fast and slow errors, for observation frequencies of  $\Delta t^{\text{obs}} = 2, 4, 6$ . In these experiments, observations of  $w'$  were randomly alternated with observations of  $\phi$ . These experiments thus simulate a case in which we have a variety of observations, with some of the slow mode only. EnKF cases with  $N = 4$  and  $N = 10$  forecasts are compared to the EKF case where  $\mathbf{P}_0^f$  is initialized using the TLBT.

The performance of the EnKF is comparable to that of the TLBT-initialized EKF for the first few observations, but has a slower growth of analysis error as the assimilation progresses. The EnKF is also, on average, more accurate for longer  $\Delta t^{\text{obs}}$ : At  $t \geq 10$ , the fast error for the EnKF, for both ensemble sizes shown, is consistently smaller than the EKF fast error at corresponding observation frequencies (except at  $\Delta t^{\text{obs}} = 2$ ). The EnKF slow error levels off for both ensemble sizes, whereas the corresponding EKF slow error grows in time.

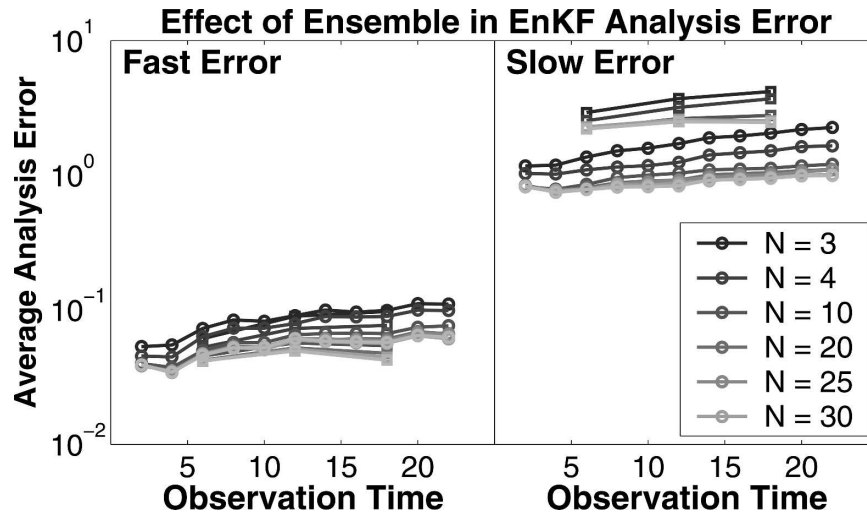


FIG. 9. Comparison of the average (left) fast and (right) slow analysis error, now comparing the EnKF for six different ensemble sizes, and comparing  $\Delta t^{\text{obs}} = 2$  (circles) and  $\Delta t^{\text{obs}} = 6$  (squares), with observations of  $w'$  only. The same axis has been kept for both plots in order to facilitate comparison to the other figures.

It is also interesting to note that increasing observation frequency improves the EnKF estimate for the slow mode, but actually seems to worsen the estimate of the slaved fast mode. This is a robust result for similar experiments with different ensemble sizes, and is explainable: more frequent forcing of the forecasts in the ensemble means that there are more chances to excite spurious fast motion in individual forecasts and, consequently, balance in the ensemble covariance model deteriorates earlier. If observations are less frequent, the gravity wave in each ensemble member, and hence the net imbalance in the covariance model, cannot grow as quickly, and the overall analysis is consequently more balanced.

The effect of ensemble size is examined in Fig. 9, which compares the average fast and slow analysis errors for  $\Delta t^{\text{obs}} = 2$  and 6, at six different ensemble sizes, and observations of  $w'$  only. For both observation frequencies, both the fast and slow mode analyses are improved by increasing  $N$ , but do not improve significantly for ensemble sizes beyond about 20. It is also interesting that fast error in the EnKF is controlled more by ensemble size than by observation frequency.

## 6. Discussion and conclusions

We have examined the performance of the EKF and EnKF for problems where there exists a separation of time scales between relatively fast and slow motions, where the free fast motion is oscillatory, and where the evolution of the fast variables in the true state is slaved

to that of the slow. Experiments with the exL86 model showed that the EKF and EnKF have quite different properties when it comes to preserving balance in the assimilated analysis. For both Kalman filters, the degree to which balance is preserved in the analysis increment depends on how closely the error covariances contained in the matrix  $\mathbf{P}^f$  reflect the slaving relationship between the fast and slow variables.

In the EKF, this is dependent on the initial formulation of  $\mathbf{P}^f$ , followed by its forward evolution via the TLM and subsequent adjustment with new information from observations. In the ideal case where observations are very frequent, it is possible that  $\mathbf{P}_0^f$  can be estimated without any knowledge of balance, with the information from observations alone ensuring that the analysis state becomes balanced (e.g., Figs. 3a,c).

Outside of this ideal case, however, the initial covariance field does matter, and should ideally involve some form of tangent linear balance transformation (TLBT). Given this, the length of time between observations determines: (i) the extent to which the forecast drifts from the truth (which in turn determines the size of the necessary analysis increment, and hence the impact of a misestimation of the balance relationship within  $\mathbf{P}^f$ ), and (ii) the validity of the TLM evolution of  $\mathbf{P}^f$ . The length of time between observations that is feasible for the EKF is therefore constrained, not just by the validity of the TLM, but also by that of the TLBT. A theoretical prediction of what this time scale might be for realistic models is a very interesting question for further research.

Alternatively, a balance constraint can be added to

the EKF algorithm. This can be done by projecting the model state onto the slow manifold, performing the analysis in terms of the slow variables only, and then using the slaving relationship to compute the analysis in terms of mixed variables. Numerical experiments showed that such a constraint can significantly improve the EKF analysis at observation frequencies where balance is otherwise lost, but relies upon accurate knowledge of the balance relationship.

The EnKF has two possible advantages for balance: First, it retains the balance relationship to the extent that ensemble members themselves are balanced, because the forecasts in the ensemble use the full nonlinear model and the gravity wave is therefore bounded. Secondly, the ensemble-averaging in the analysis step means that a spurious gravity wave in a single forecast is compared to every other forecast, so that the resulting analysis has less net imbalance. Note, however, that this assumes that the ensemble is sufficiently phase mixed. If several ensemble members contain the same gravity wave, this wave will not average out (this can be seen in Fig. 7).

The accuracy of the EnKF is nonetheless limited by ensemble size. Numerical experiments showed that, while the knowledge of a balance relationship reduces the dimension of the problem, enough ensemble members are still required to capture, within the covariance matrix, the fact that the fast mode is slaved to the slow. For the exL86 model, the best-possible balanced analysis required an ensemble of at least 20 forecasts, which is much larger than the degrees of freedom of the problem (which is four).

A numerical comparison of the EKF and EnKF (Fig. 8) showed that the two schemes seem to perform similarly well, in terms of balance, when the initial covariance model in the EKF is computed using a TLBT, and when the ensemble in the EnKF is large enough to capture the slaving of the fast motion. The consequence of the EnKF's two additional properties is that (i) growth of imbalance in the analysis, as the assimilation progresses, is slower in the EnKF, and (ii) the EnKF allows for longer time intervals between observations without a great loss of balance.

This stability of the EnKF reflects the stability of gravity waves in the exL86 model; whereas, in contrast, gravity waves can grow in the TLM. The way in which gravity waves grow in the TLM has been illustrated numerically but not yet investigated analytically. Future work will further investigate the development of imbalance in the TLM.

Another property to note about the EnKF is that observing too frequently can overforce the ensemble, and actually cause greater imbalance, if there are not

enough other gravity waves in the ensemble to eliminate the spurious gravity wave (via phase mixing) in the ensemble mean.

This study is intended as a complement to similar studies involving larger, more complicated models, such as Mitchell et al. (2002). We have highlighted, in a simple context, several key points of the balance problem in the still-evolving field of 4D data assimilation. This extends beyond the problem of spurious gravity waves, to any problem where there is a time scale of interest, and a comparatively fast, unobserved scale (e.g., Lorenz 1995). Our results suggest several interesting points for further research:

*More degrees of freedom.* An analysis similar to this one could be carried out with a model that is more complex than the exL86 model, yet still simple enough to retain the transparency of this analysis. A model that admits more than one gravity wave frequency, or more spatial degrees of freedom, would make it possible to address the effects of, say, geostrophic adjustment or localization of error covariances.

*Chaotic (non-gravity wave) fast mode.* A model in which the fast mode is chaotic (in which case slaving is impossible), such as the two time-scale model of Lorenz (1995), is another interesting context for examining assimilation for multiple time scales, and could address problems of mesoscale and convective-scale assimilation (e.g., Snyder and Zhang 2003).

*Variations on the standard algorithms.* Many variations of the EKF and EnKF have been proposed in recent years, with the intent of either increasing cost-efficiency, or of freeing the KF from some of the assumptions on which it rests. These are well summarized by Evensen (2003).

Square root filters (SRF) compute analysis ensembles deterministically from the analysis covariance matrix given by (3.6), which reduces sampling error, and thus helps prevent filter divergence (Tippett et al. 2003, and references therein). This might prevent undesirable phase locking such as we see in Fig. 7. The accuracy and possible differences between the various incarnations of SRFs schemes, with respect to balance, is, to our knowledge, still to be investigated.

Schemes that rely on fewer linearity assumptions may also handle balance dynamics more accurately, since (as shown in this study) balance in the assimilation often fails because of faulty assumptions of linearity. A particle filter (Anderson and Anderson 1999; Pham 2001), for example, departs from the standard KF linear analysis step (3.1), and

may thus skirt some of the problems cited in our study, such as gravity wave explosions in the EKF, or undesirable phase-locking of the ensemble.

Modifications to the EnKF that are intended to improve cost-efficiency might turn out to be naturally beneficial for balance, because such schemes take advantage of the low-dimensionality of the model attractor, and, in the real atmosphere, the slow manifold is itself a lower-dimensional attractor. Examples of such schemes are the singular evolutive interpolated Kalman (SEIK) filter of Pham (2001), and schemes in which the ensemble is constrained by bred vectors (e.g., Toth and Kalnay 1997) or singular vectors (e.g., Ehrendorfer and Tribbia 1997).

*Comparison to 4DVAR.* This analysis could be extended from the Kalman filter to 4DVAR assimilation. While 4DVAR also makes use of a TLM, its analysis cycle, and the formulation of the forecast error covariance field, is quite different, and it is unclear how these differences change the treatment of nonlinear balance.

*Unclear time-scale separation.* We have not yet examined cases where two different motions are admitted, but not well-separated in time scale. In the Tropics, for example, Žagar et al. (2004) propose an approach whereby capturing balance means interpreting the observed field as the right combination of the different types of tropical waves that are present. In the exL86 model, this problem can be approached by letting  $\epsilon \rightarrow 1$ .

*Unbalanced truth state.* It also remains to examine how the nonlinear KFs perform in the case where both fast and slow motion is present in the true state, or rather, where the fast waves carry a significant amount of energy, such as in the mesosphere.

*Effect of model error.* This study dealt with a very specialized situation where both the model dynamics and the balance relationship are perfectly known. Since this is far from true in realistic cases, the effect of model error formulation on balance needs to be investigated. Even in the perfect model limit, the addition of a model error term in the EKF forecast error evolution (3.10) may stabilize the analysis step, and prevent observations from shocking the system into highly unbalanced states, as in Fig. 3b. Similarly, adding stochastic errors to the ensemble in the EnKF (3.12) could prevent the phase-locking of the ensemble around a spurious gravity wave.

*Acknowledgments.* This research was supported in part by the Modelling of Global Chemistry for Climate

project, with support from the Natural Sciences and Engineering Research Council, the Meteorological Service of Canada through its Climate Research Network, the Canadian Space Agency, and the Canadian Foundation for Climate and Atmospheric Sciences. LN has also been supported by a University of Toronto Open Fellowship and a University of Toronto Blythe Fellowship. The authors thank Philippe Courtier and Jean Côté for helpful suggestions during the development of this work, and Peter Houtekamer, Jeff Kepert, and two anonymous reviewers, for comments and suggestions on the manuscript.

## APPENDIX

### Derivation of the exL86 Model

The following is a summary derivation of the exL86 model, tracing the development of this simple system through four papers: Lorenz (1980, 1986), Bokhove and Shepherd (1996), and WS00.

In Lorenz (1980), the  $f$ -plane shallow water equations are nondimensionalized and then simplified by expanding vorticity, divergence, and height as an interacting resonant wave triad. This yields a system of nine equations that describe the evolution of the vorticity, divergence, and height of three interacting waves. These amplitudes are then transformed into normal modes, corresponding to potential vorticity, divergence, and geostrophic imbalance. In Lorenz (1986), the latter two variables are eliminated for two of the three waves, which leaves two geostrophic vorticity modes, and a third wave which admits both vortical motion and a gravity wave. Here,  $U$  and  $V$  are the vorticity amplitudes of the two truncated waves; and  $W$ ,  $X$ , and  $Z$  are the potential vorticity, divergence, and geostrophic imbalance of the third wave, respectively. The equations that describe their evolution are given by

$$\frac{dU}{dT} = -VW + bVZ \quad (\text{A.1})$$

$$\frac{dV}{dT} = UW - bUZ \quad (\text{A.2})$$

$$\frac{dW}{dT} = -UV \quad (\text{A.3})$$

$$\frac{dX}{dT} = -Z \quad (\text{A.4})$$

$$\frac{dZ}{dT} = bUV + X \quad (\text{A.5})$$



These equations describe a nonlinearly interacting vorticity triad ( $U$ ,  $V$ , and  $W$ ), coupled to an inertia-gravity wave ( $X$  and  $Z$ ). The parameter  $b$  is the rotational Froude number of the third wave.

Bokhove and Shepherd (1996) emphasize the time-scale separation between the nonlinear vortical mode and the gravity wave by scaling the variable amplitudes  $U \equiv \epsilon u$ ,  $V \equiv \epsilon v$ ,  $W \equiv \epsilon w$ ,  $X \equiv \epsilon x$ ,  $Z \equiv \epsilon z$ , and scaling time  $T \equiv t/\epsilon$ . The scaled system is

$$\frac{du}{dt} = -vw + bvz \quad (\text{A.6})$$

$$\frac{dv}{dt} = uw - buz \quad (\text{A.7})$$

$$\frac{dw}{dt} = -uw \quad (\text{A.8})$$

$$\frac{dx}{dt} = -\frac{z}{\epsilon} \quad (\text{A.9})$$

$$\frac{dz}{dt} = buw + \frac{x}{\epsilon}. \quad (\text{A.10})$$

For  $\epsilon \ll 1$ ,  $x$  and  $z$  vary on a time scale that is fast compared to the evolution of  $u$ ,  $v$ , and  $w$ . From the dimensions of the original equations, it can be shown that  $\epsilon$ , which defines the inverse of the ratio between the advective time scale (corresponding to  $t$ ) and that of the inertia-gravity wave, is given by  $\epsilon = RF/\sqrt{R^2 + F^2}$ , where  $R$  is the Rossby number and  $F$  the Froude number.

This system is further simplified by noting the invariant  $C = u^2 + v^2$ , and defining  $u \equiv \sqrt{C} \cos\phi'$ ,  $v \equiv \sqrt{C} \sin\phi'$ , and  $\phi \equiv \phi' - \epsilon bx$ . This yields the following four-variable system:

$$\frac{d\phi}{dt} = w \quad (\text{A.11})$$

$$\frac{dw}{dt} = -\frac{C}{2} \sin(2\phi + 2\epsilon bx) \quad (\text{A.12})$$

$$\frac{dx}{dt} = -\frac{z}{\epsilon} \quad (\text{A.13})$$

$$\frac{dz}{dt} = \frac{x}{\epsilon} + \frac{bC}{2} \sin(2\phi + 2\epsilon bx). \quad (\text{A.14})$$

Bokhove and Shepherd (1996) showed that the vortical mode in (A.11)–(A.14) has periodic solutions for most initial conditions. To make the evolution of  $\phi$  and  $w$  sufficiently chaotic, WS00 let the parameter  $C$  vary in time as  $C(t) = a_0 + a_1 \cos\gamma t$ .

Since the present study focuses on the treatment of the given time-scale separation in the Kalman filter,

and since observed quantities are not clearly separated slow or fast variables, it makes sense to transform  $w$  and  $z$  in the above system back to mixed variables. This is simply done by defining  $w \equiv w' + bz'$  and  $z \equiv z' - bw'$ , which yields the following system:

$$\frac{d\phi}{dt} = w' + bz' \quad (\text{A.15})$$

$$\frac{dw'}{dt} = -\frac{C}{2} \sin 2(\phi + \epsilon bx) - \frac{\alpha^2 b}{\epsilon} x \quad (\text{A.16})$$

$$\frac{dx}{dt} = \frac{bw' - z'}{\epsilon} \quad (\text{A.17})$$

$$\frac{dz'}{dt} = \frac{\alpha^2 x}{\epsilon}, \quad (\text{A.18})$$

where  $\alpha = (1 + b^2)^{-1/2}$ . In this system,  $\phi$  describes the (geostrophic) vorticity of modes 1 and 2, and  $w'$ ,  $x$ , and  $z'$  the (nongeostrophic) vorticity, divergence, and height, respectively, of mode 3.

#### REFERENCES

- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Bokhove, O., and T. G. Shepherd, 1996: On Hamiltonian balanced dynamics and the slowest invariant manifold. *J. Atmos. Sci.*, **53**, 276–297.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Cohn, S. E., and S. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757–1785.
- Courtier, P., and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531–549.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- , and K. Puri, 1980: Four-dimensional data assimilation and the slow manifold. *Mon. Wea. Rev.*, **108**, 85–99.
- Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- , 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342–1354.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.
- Gauthier, P., and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.*, **129**, 2089–2102.
- Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather pre-



- diction. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil and E. Källén, Eds., Springer-Verlag, 139–224.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilations using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45.
- , and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95–108.
- Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- , 1980: Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.*, **37**, 1685–1699.
- , 1986: On the existence of a slow manifold. *J. Atmos. Sci.*, **43**, 1547–1557.
- , 1995: Predictability—A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.
- Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation for strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.
- Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Polavarapu, S., M. Tanguay, and L. Fillion, 2000: Four-dimensional variational data assimilation with digital filter initialization. *Mon. Wea. Rev.*, **128**, 2491–2510.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.
- Tanguay, M., P. Bartello, and P. Gauthier, 1995: Four-dimensional data assimilation with a wide range of scales. *Tellus*, **47A**, 974–997.
- Thépaut, J.-N., and P. Courtier, 1991: Four-dimensional data assimilation using the adjoint of a multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 1225–1254.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Wirosoetisno, D., and T. G. Shepherd, 2000: Averaging, slaving and balanced dynamics in a simple atmospheric model. *Physica D*, **141**, 37–53.
- Žagar, N., N. Gustafsson, and E. Källén, 2004: Variational data assimilation in the tropics: The impact of a background-error constraint. *Quart. J. Roy. Meteor. Soc.*, **130**, 103–125.