

Using Singular Value Decomposition to Parameterize State-Dependent Model Errors

Christopher M. Danforth*

Department of Mathematics and Statistics, University of Vermont

Burlington, VT 05401

Eugenia Kalnay

Department of Atmospheric and Oceanic Science, University of Maryland

College Park, MD 20742

May 11, 2007

*chris.danforth@uvm.edu

Abstract

The purpose of the present study is to use a new method of empirical model error correction, developed by Danforth et al. (2007), based on estimating the systematic component of the non-periodic errors linearly dependent on the anomalous state. The method uses Singular Value Decomposition (SVD) to generate a basis of model errors and states. It requires only a time series of errors to estimate covariances and uses negligible additional computation during a forecast integration. As a result, it should be suitable for operational use at relatively small computational expense.

We test the method with the Lorenz '96 coupled system as the truth, and an uncoupled version of the same system as a model. We demonstrate that the SVD method represents a significant component of the effects of the model's unresolved state on the resolved state and show that the results are better than those obtained with the empirical correction operator of Leith (1978). The improvement is attributed to the fact that the SVD truncation effectively reduces sampling errors. Forecast improvements of up to 1000% are seen when compared with the original model. The improvements come at the expense of weakening ensemble spread.

1. Introduction

No matter how well understood a physical process is, predictions of that process derived from numerical integration of models are likely to suffer from two factors. First, nonlinearities amplify uncertainties in the initial conditions, causing similar states of the system to diverge quickly on small scales. Second, deficiencies in the numerical model introduce errors during integration. These deficiencies may be structural problems (wrong equations) and are induced by inaccurate forcings and parameterizations used to represent the effect of sub-grid scale physical processes and result in large scale systematic forecast errors.

Leith (1978) proposed a statistical method to account for model bias and systematic errors linearly dependent on state anomalies. Leith derived a state-dependent empirical correction to a simple dynamical weather model by minimizing the tendency errors relative to a reference time series. The resulting correction operator attempts to predict the error in the model tendency as a function of the model state. While Leith's empirically estimated state-dependent correction term is only optimal for a linear model, it was shown to reduce the nonlinear model's bias.

DelSole and Hou (1999) perturbed the parameters of a 2-layer quasi-geostrophic (QG) model on a 8×10 grid ($N_d = 160$ degrees of freedom) to generate a 'nature' run and then modified it to create a 'model' containing a primarily state-dependent error. They found that a state-independent error correction did *not* improve the forecast skill. By adding a state-dependent empirical correction to the model, inspired by the procedure proposed by Leith, they were able to extend forecast skill up to the limits imposed by observation error. However, Leith's technique requires the solution of a N_d -dimensional linear system. As a result, before the procedure can

be considered useful for operational use, a low-dimensional representation of Leith's empirical correction operator is required.

Wilks (2005) used the Lorenz '96 coupled system as the truth, and an uncoupled version of the same system as a model, and developed a stochastic parameterization of the effects of the unresolved variables. The correction resulted in improved agreement between model and system climatologies, as well as improved ensemble mean and spread for short range forecasts. Individually deterministic forecasts were degraded by the stochastic parameterization methods. Wilks found the improvement resulting from stochastic forcing to depend strongly on both the standard deviation and time-scale of the stochastic term, and weakly on its spatial scale.

In what follows, we use the same experimental setup as Wilks with a low-dimensional representation of Leith's empirical correction operator using Singular Value Decomposition (SVD) (Golub and Van Loan, 1996) developed by Danforth et al. (2007). We use the resulting SVD modes as a basis for deterministic parameterization of the tendencies of the Lorenz '96 system unresolved by the uncoupled model. Empirical correction of the uncoupled model using the SVD modes results in significant forecast improvement (anomaly correlation and RMSE) when compared with Leith's operator, at the expense of weakening ensemble spread. The SVD method can be extremely computationally efficient, requiring only an inner product and the solution of a low-dimensional linear system. The paper concludes with a discussion of applications to numerical weather prediction.

2. Empirical Correction

Following Leith (1978), consider an arbitrary dynamical system

$$\dot{\mathbf{x}}(t) = \mathbf{M}(\mathbf{x}(t)) \quad (1)$$

where $\mathbf{x}(t)$ and $\mathbf{M}(\mathbf{x}(t))$ are the model state vector and model tendency at step t , respectively. \mathbf{M} is the best available representation of the governing dynamics of the physical process whose future behavior we are attempting to predict. Let $\mathbf{x}^a(t)$ denote the true state of the dynamical system at step t (estimated, for example, from an analysis) and $\mathbf{x}_{\Delta t}^f(t)$ denote a prediction of $\mathbf{x}^a(t)$ generated by integrating \mathbf{M} for time Δt from the state $\mathbf{x}^a(t - \Delta t)$. Leith considered the difference between $\mathbf{x}^a(t)$ and $\mathbf{x}_{\Delta t}^f(t)$ for small Δt to be an approximation of the model tendency error. The residual at step t is given by the difference between the truth $\mathbf{x}^a(t)$ and the model forecast state $\mathbf{x}_{\Delta t}^f(t)$, namely

$$\delta \mathbf{x}_{\Delta t}^a(t) = \mathbf{x}^a(t) - \mathbf{x}_{\Delta t}^f(t) \quad (2)$$

where Δt is the forecast lead time. The smaller Δt , the better $\delta \mathbf{x}_{\Delta t}^a(t)$ is as an approximation of the model error. The time-average of the residuals is an estimate of the model bias, or state-independent error. It is given by

$$\langle \delta \mathbf{x}_{\Delta t}^a \rangle = \frac{1}{N_a} \sum_{t=1}^{N_a} \delta \mathbf{x}_{\Delta t}^a(t) \quad (3)$$

where N_a denotes number of individual verifications that can be made comparing forecast and truth. The truth, model predictions, and corresponding residuals are then separated into their anomalous and time average components, namely

$$\mathbf{x}^{a'}(t) = \mathbf{x}^a(t) - \langle \mathbf{x}^a \rangle \quad (4)$$

$$\mathbf{x}_{\Delta t}^{f'}(t) = \mathbf{x}_{\Delta t}^f(t) - \langle \mathbf{x}^a \rangle \quad (5)$$

$$\delta \mathbf{x}_{\Delta t}^{a'}(t) = \delta \mathbf{x}_{\Delta t}^a(t) - \langle \delta \mathbf{x}_{\Delta t}^a \rangle \quad (6)$$

so that deviations from the mean can be analyzed.

2a. Leith's Empirical Correction Operator

Online bias corrected or *debiased* model predictions can be generated using an improved model M^+ , defined by the tendency equation

$$\dot{\mathbf{x}}(t) = M(\mathbf{x}(t)) + \frac{\langle \delta \mathbf{x}_{\Delta t}^a \rangle}{\Delta t} \equiv M^+(\mathbf{x}(t)) \quad (7)$$

The time series of anomalous residuals of M^+ , obtained by comparing predictions made by M^+ with $\mathbf{x}^a(t)$, provides an estimate of the linear state-dependent model error. Leith (1978) suggested that these residuals could be used to form a state-*dependent* correction. Leith sought an improved model of the form

$$\dot{\mathbf{x}}(t) = M^+(\mathbf{x}(t)) + L\mathbf{x}'(t) \quad (8)$$

where $\mathbf{x}'(t) = \mathbf{x}(t) - \langle \mathbf{x}^a \rangle$ is the anomalous model state at time t and $L\mathbf{x}'(t)$ is the state-dependent error correction. The tendency error of the improved model is given by

$$\mathbf{g}(t) = \dot{\mathbf{x}}^a(t) - \left(M^+(\mathbf{x}^a(t)) + L\mathbf{x}'^a(t) \right) \quad (9)$$

where $\dot{\mathbf{x}}^a(t)$ is the instantaneous time derivative of the true state. Of course, the true time tendency is unknown, so $\dot{\mathbf{x}}^a(t)$ can only be approximated by finite differences using the reference time series. The mean square tendency error of the improved model is given by $\langle \mathbf{g}^\top(t)\mathbf{g}(t) \rangle$. Minimizing this tendency error with respect to L , Leith's state-dependent correction operator is given by

$$L = \left\langle \left(\dot{\mathbf{x}}^a(t) - M^+(\mathbf{x}^a(t)) \right)' \mathbf{x}'^a(t)^\top \right\rangle \left\langle \mathbf{x}'^a(t) \mathbf{x}'^a(t)^\top \right\rangle^{-1} \quad (10)$$

where $\dot{\mathbf{x}}^a(t) - M^+(\mathbf{x}^a(t))$ is approximated by the residuals,

$$\dot{\mathbf{x}}^a(t) - M^+(\mathbf{x}^a(t)) \approx \frac{\mathbf{x}^a(t + \Delta t) - \mathbf{x}^a(t)}{\Delta t} - \frac{\mathbf{x}_{\Delta t}^f(t + \Delta t) - \mathbf{x}^a(t)}{\Delta t} = \frac{\delta \mathbf{x}_{\Delta t}^a(t + \Delta t)}{\Delta t} \quad (11)$$

which for an operational weather model are typically available from pre-implementation testing.

As a result, the operator L may be estimated with no additional model integrations.

To estimate L , the time series of residuals $\delta \mathbf{x}_{\Delta t}^a(t)$ is computed using the online debiased

model M^+ . The *cross* covariance (Bretherton et al. 1992) of the residuals with their corresponding true states, the *lagged* cross covariance, and the true state covariance are given by

$$C_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}^a} \equiv \langle \delta \mathbf{x}_{\Delta t}^{a'}(t) \mathbf{x}^{a'}(t)^\top \rangle \quad (12)$$

$$C_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}_{lag}^a} \equiv \langle \delta \mathbf{x}_{\Delta t}^{a'}(t) \mathbf{x}^{a'}(t-1)^\top \rangle \quad (13)$$

$$C_{\mathbf{x}^a \mathbf{x}^a} \equiv \langle \mathbf{x}^{a'}(t) \mathbf{x}^{a'}(t)^\top \rangle \quad (14)$$

respectively. The empirical correction operator (10) is then given by

$$L = C_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}_{lag}^a} C_{\mathbf{x}^a \mathbf{x}^a}^{-1} \quad (15)$$

We define $\mathbf{w}(t) = C_{\mathbf{x}^a \mathbf{x}^a}^{-1} \cdot \mathbf{x}'(t)$ to be the anomalous state normalized by its empirically derived covariance so that the matrix-vector product $L\mathbf{x}'(t) = C_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}_{lag}^a} \mathbf{w}(t)$ in Eqn. (8) is the best estimate of the anomalous residual corresponding to the anomalous model state \mathbf{x}' over the dependent sample. Assuming that sampling errors are small and that model errors evolve linearly with respect to lead time, this correction should improve the forecast model M^+ . Errors in initial state grow exponentially with lead time, but those forced by model error tend to grow linearly (e.g. Dalcher and Kalnay 1987, Reynolds et al. 1994). Therefore, the Leith operator should provide a useful estimate of the state-dependent model error.

Using a model with very few degrees of freedom and errors that were strongly state-dependent, DelSole and Hou (1999) found that the Leith operator was very successful in correcting state-

dependent errors relative to a nature run. However, the direct computation of Lx' requires $O(N_d^3)$ floating point operations every time step. For operational models, $N_d = O(10^7)$; it is clear that this operation would be prohibitive. Approaches to reduce the dimensionality of the Leith correction are now described.

2b. Low-Dimensional Approximation

An alternative formulation of Leith's correction operator is described here, based on the correlation of the leading SVD modes. For a more detailed derivation, see Danforth et al. (2007). The dependent sample of anomalous residuals and model predictions are normalized at each grid point by their standard deviation so that they have unit variance, they are denoted $\overline{\delta x_{\Delta t}^{a'}}(t)$ and $\overline{x_{\Delta t}^{f'}}(t)$. They are then used to compute the cross correlation

$$\mathbf{C}_{\delta x_{\Delta t}^a x_{\Delta t}^f} \equiv \langle \overline{\delta x_{\Delta t}^{a'}}(t) \overline{x_{\Delta t}^{f'}}(t)^\top \rangle \quad (16)$$

where normalization is required to make $\mathbf{C}_{\delta x_{\Delta t}^a x_{\Delta t}^f}$ a correlation matrix. The matrix is then decomposed to identify pairs of spatial patterns that explain as much of possible of the mean-squared temporal covariance between the fields $\overline{\delta x_{\Delta t}^{a'}}$ and $\overline{x_{\Delta t}^{f'}}$. The SVD is given by

$$\mathbf{C}_{\delta x_{\Delta t}^a x_{\Delta t}^f} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top = \sum_{k=1}^{N_d} \mathbf{u}_k \sigma_k \mathbf{v}_k^\top \quad (17)$$

where the columns of the orthonormal matrices \mathbf{U} and \mathbf{V} are the left and right singular vectors \mathbf{u}_k and \mathbf{v}_k . $\mathbf{\Sigma}$ is a diagonal matrix containing singular values σ_k whose magnitude decreases with

increasing k . The leading patterns \mathbf{u}_1 and \mathbf{v}_1 associated with the largest singular value σ_1 are the dominant coupled signals in the time series $\overline{\delta \mathbf{x}_{\Delta t}^{a'}}$ and $\overline{\mathbf{x}_{\Delta t}^{f'}}$ respectively (Bretherton et al. 1992). Patterns \mathbf{u}_k and \mathbf{v}_k represent the k th most significant coupled signals. Expansion coefficients or Principal Components (PCs) $a_k(t)$, $b_k(t)$ describe the magnitude and time dependence of the projection of the coupled signals onto the reference time series. They are given by

$$a_k(t) = \mathbf{u}_k^\top \cdot \overline{\delta \mathbf{x}_{\Delta t}^{a'}}(t) \quad (18)$$

$$b_k(t) = \mathbf{v}_k^\top \cdot \overline{\mathbf{x}_{\Delta t}^{f'}}(t) \quad (19)$$

3. Low-Dimensional Correction

The most significant computational expense required by Leith's empirical correction (8) involves solving the N_d -dimensional linear system $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a} \mathbf{w}(T) = \mathbf{x}'(T)$ for \mathbf{w} at each time T during a forecast integration. Assuming that Δt is small (error growth is approximately linear during the short forecasts used for training), we can approximate $\mathbf{C}_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}_{lag}^a} \approx \mathbf{C}_{\delta \mathbf{x}_{\Delta t}^a \mathbf{x}_{\Delta t}^f}$ and $\mathbf{C}_{\mathbf{x}^a \mathbf{x}^a} \approx \mathbf{C}_{\mathbf{x}_{\Delta t}^f \mathbf{x}_{\Delta t}^f}$. Then, a substantial reduction in computation for this operation can be achieved by expressing $\mathbf{w} = \mathbf{C}_{\mathbf{x}_{\Delta t}^f \mathbf{x}_{\Delta t}^f}^{-1} \mathbf{x}'$ as a linear combination of the leading orthonormal right singular vectors \mathbf{v}_k , namely

$$\begin{aligned}
L\mathbf{x}' &= \mathbf{C}_{\delta\mathbf{x}_{\Delta t}^a \mathbf{x}_{\Delta t}^f} \mathbf{C}_{\mathbf{x}_{\Delta t}^f \mathbf{x}_{\Delta t}^f}^{-1} \mathbf{x}' = \mathbf{C}_{\delta\mathbf{x}_{\Delta t}^a \mathbf{x}_{\Delta t}^f} \mathbf{w} \\
&= \mathbf{U}\Sigma\mathbf{V}^T \mathbf{w} = \sum_{k=1}^{N_d} \mathbf{u}_k \sigma_k \mathbf{v}_k^T \cdot \mathbf{w} \approx \sum_{k=1}^K \mathbf{u}_k \sigma_k \mathbf{v}_k^T \cdot \mathbf{w} \quad (20)
\end{aligned}$$

where $K \ll N_d$ should be chosen such that the explained variance, given by

$$r(k) = \frac{\sum_{i=1}^k \sigma_i}{\sum_{i=1}^{N_d} \sigma_i} \quad (21)$$

exceeds a system dependent threshold for $k = K$. From trial and error, it appears that an explained variance of $r(K) \geq .95$ results in the best forecast improvement for the simple model discussed in the following section. As a result, K should be chosen to fulfill this, or a similar, inequality. For the SPEEDY model, Danforth et al. (2007) found the best results for the anomaly correlation using a truncation of $K = 10$.

It is important to note that only the component of \mathbf{w} in the K -dimensional space spanned by the right singular vectors \mathbf{v}_k can contribute to this empirical correction defined by (20). This dependence can be exploited as follows. Assume the model state at time T during a numerical model integration is given by $\mathbf{x}(T)$. The normalized state anomaly $\overline{\mathbf{x}'(T)}$ is given by the vector $\mathbf{x}'(T) = \mathbf{x}(T) - \langle \mathbf{x}^a \rangle$ normalized by the standard deviation of $\mathbf{x}_{\Delta t}^f$ over the dependent sample. The component of $\overline{\mathbf{x}'(T)}$ explained by the signal \mathbf{v}_k may then be estimated by computing the new principal component $b_k(T) = \mathbf{v}_k^T \cdot \overline{\mathbf{x}'(T)}$. The right PC covariance over the dependent sample is given by $\mathbf{C}_{bb} = \langle \mathbf{b}\mathbf{b}^T \rangle$, calculated using b_k from (19). Due to the orthogonality of the right

singular vectors \mathbf{v}_k , assuming an infinite sample, PCs b_k and b_j are uncorrelated for $k \neq j$. As a result, we restrict the finite sample covariance \mathbf{C}_{bb} to be diagonal. The linear system

$$\mathbf{C}_{bb}\boldsymbol{\gamma}(T) = \mathbf{b}(T) \quad (22)$$

may then be solved for $\boldsymbol{\gamma}$ at time T . As a result, the cost of solving (22) is $O(K)$ where K is the number of SVD modes retained, as opposed to the $(N_d \times N_d)$ linear system required by Leith's full dimensional empirical correction. The solution of (22) gives an approximation of $\mathbf{w}(T)$, namely

$$\mathbf{w}(T) = \mathbf{C}_{\mathbf{x}^f \mathbf{x}^f}^{-1} \overline{\mathbf{x}'(T)} \quad (23)$$

$$\approx \sum_{k=1}^K \gamma_k(T) \mathbf{v}_k = \sum_{k=1}^K \frac{b_k(T)}{\langle b_k^2 \rangle} \mathbf{v}_k = \tilde{\mathbf{w}}(T) \quad (24)$$

where $\tilde{\mathbf{w}}(T)$ is generated by solving the linear system (22) in the space of the leading K singular vectors, while $\mathbf{w}(T)$ requires solving the N_d -dimensional linear system (23) in the space of the model grid. Writing \mathbf{u}_k^c for the error signal \mathbf{u}_k weighted by the standard deviation of $\delta \mathbf{x}_{\Delta t}^{a'}$ over the dependent sample, the k th component of the state-dependent error correction at time T is given by

$$\mathbf{z}_k(T) = \mathbf{u}_k^c \sigma_k \gamma_k(T) \quad (25)$$

where σ_k is the coupling strength over the dependent sample (given by the SVD (17)). The weight $\gamma_k(T)$ assigned to residual signal \mathbf{u}_k^c indicates the sign and magnitude of the correction

which may amplify, dampen, or shift the flow anomaly local to the pattern \mathbf{u}_k^c . Then the SVD corrected model is given at time T by

$$\dot{\mathbf{x}}(T) = M^+(\mathbf{x}(T)) + \frac{1}{\Delta t} \sum_{k=1}^K \mathbf{z}_k(T) \equiv M^{++}(\mathbf{x}(T)) \quad (26)$$

so that *during* a prediction, a few (K) dominant model state signals \mathbf{v}_k can be projected onto the anomalous, normalized model state vector. The resulting sum $\sum_{k=1}^K \mathbf{z}_k$ is the best representation of the original residual anomalies $\delta \mathbf{x}_{\Delta t}^{a'}$ in terms of the current forecast state $\mathbf{x}(T)$. If the correlation between the normalized state anomaly $\overline{\mathbf{x}'(T)}$ and the local pattern \mathbf{v}_k is small, the new expansion coefficient $b_k(T)$ will be negligible, no correction by \mathbf{u}_k^c will be made at time T , and therefore no harm will be done to the prediction. This fact is particularly important with respect to predicting behavior which may vary on a time scale longer than the training period, for example El Niño Southern Oscillation (ENSO) events (Barnston et al. 1999).

The SVD representation of the error is advantageous compared to Leith's correction operator for several reasons. First, it reduces the sampling errors by identifying the most robust coupled signals between the residual and forecast state anomalies. Second, the added computation is trivial; it requires solving a K -dimensional linear system and computing K inner products. Finally, the SVD signals identified by the technique can be used by modelers to isolate flow dependent model deficiencies. In ranking these signals by strength, SVD gives modelers the ability to evaluate the relative importance of various model errors.

4. Numerical Experiments

4a. Lorenz '96 model

In this section we demonstrate the empirical correction procedures using a simple nonlinear system to define the truth (the quantity that will be predicted by a model). The N_d -dimensional governing equations, given by Lorenz (1996) are

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j \text{ for } i = 1, 2, \dots, I \quad (27)$$

$$\frac{dy_j}{dt} = -cby_{j+1}(y_{j+2} - y_{j-1}) - cy_j + \frac{hc}{b} x_{\text{floor}[(j-1)/J]+1} \text{ for } j = 1, 2, \dots, JI \quad (28)$$

where $N_d = (J + 1)I$ and the subscripts i and j are treated as periodic with period I and J respectively. For example, $x_{I+1} \equiv x_1$ so that the variables form a cyclic chain. Equation (27) describes the behavior of a set of slowly changing, large-amplitude unspecified scalar meteorological quantities, such as temperature, at I equally spaced grid sites on a latitude circle. Each x_i is coupled to J quickly changing, small-amplitude variables y_j whose dynamics are described by (28). The notation $\text{floor}[(j-1)/J]$ describes integer truncation of the bracketed term and indicates that each of the small-amplitude y variables in a group is equally affected by the large-amplitude x variable to which it belongs. In our experiments, we have used the same parameter values as Wilks (2005); namely $I = 8$ and $J = 32$ for a total of $N_d = 264$ state variables, $h = 1$, $c = 10$, and $b = 10$ (which has the effect of making the small-amplitude variables y_i oscillate ten times more

rapidly than the large-amplitude variables x_i), and the forcing is chosen to be either $F = 8, 14,$ or 18 . Wilks (2005) chose $F = 18$ and $F = 20$ to ensure that the deterministic parameterizations would be competitive with the stochastic.

This system shares certain properties with many atmospheric models: a nonlinear advection-like term, a linear term representing loss of energy to thermal dissipation, and a constant forcing term F to provide energy. It has been used in several previous studies to represent atmospheric behavior in predictability studies (e.g. Lorenz 1998, Wilks 2005, Danforth and Yorke 2006), and for data assimilation (e.g. Anderson 2001, Whitaker and Hamill 2002, Ott et al. 2004). The time unit represents the dissipative decay time of 5 days (Emanuel and Lorenz 1998) and there are 13 positive Lyapunov exponents.

We use equations (27) and (28) to generate a time series \mathbf{x}^a of “true” values of the slow variables. We then set $h = 0$ in equation (27) and add a bias term with weight α to obtain the model

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \alpha \sin\left(\frac{2\pi i}{I}\right) \quad (29)$$

which fails to resolve any of the small-amplitude behavior. The sinusoidal bias term, weighted by $\alpha = 1$, is included as an additional source of model error with respect to model (27), meant to represent a longitudinally dependent misrepresentation of the dissipation or forcing. Forecasts $\mathbf{x}_{\Delta t}^f$ generated by this model exhibit model error with respect to \mathbf{x}^a that is sinusoidally dependent on the grid point, but independent of the state. Training data is then compiled by generating 10^7 short forecasts ($\Delta t = 0.1$ time units ≈ 12 hours) of model (29) and comparing these forecasts

to \mathbf{x}^a . Throughout the experiment, a fourth-order Runge-Kutta integration scheme is used with a time step of 0.001. The model bias is then given by the time average of the difference between \mathbf{x}^a and \mathbf{x}_{12}^f , namely $\langle \delta \mathbf{x}_{12}^a \rangle$ (see (2)). The anomalous errors and forecasts are then used to generate Leith's correction operator L (15), and the corresponding modes (17) for SVD correction.

We then experiment with empirical correction of (29) using the improved model

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + \alpha \sin\left(\frac{2\pi i}{I}\right) + D(\mathbf{x}) \quad (30)$$

where the term $D(\mathbf{x})$ attempts to correct the sinusoidal bias *and* represent the behavior unresolved by (29), namely the coupling to the small-amplitude variables described by (28).

4b. Empirical Correction Experiments

Four different versions of model (30) are used to forecast a set of 10,000 uncorrelated initial states chosen from the true time series \mathbf{x}^a . These initial states are distinct from those used for training and consecutive initial states are separated by 50 time units (250 days). Methods are distinguished by the explicit form of the empirical correction term $D(\mathbf{x})$ in (30) which is meant to represent the small-amplitude behavior and reduce the bias, as indicated below:

$$\begin{aligned} D^{(1)}(\mathbf{x}) &= 0 & D^{(2)}(\mathbf{x}) &= \langle \delta \mathbf{x}_{12}^a \rangle & D^{(3)}(\mathbf{x}) &= \langle \delta \mathbf{x}_{12}^a \rangle + L\mathbf{x}' & (31) \\ D^{(4)}(\mathbf{x}) &= \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^K \mathbf{z}_k(T) & D^{(5)}(\mathbf{x}) &= -\alpha \sin\left(\frac{2\pi i}{I}\right) - \frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j = \text{perfect model} \end{aligned}$$

Term $D^{(1)}(\mathbf{x})$ represents the original model (29), forecasts made with no empirical correction will be used to gauge the success of other methods. Term $D^{(2)}(\mathbf{x})$ is the time average residual observed in forecasts made by model (29) and represents a state-independent correction (see Fig 1). Term $D^{(3)}(\mathbf{x})$ represents Leith's empirical correction operator (8), and term $D^{(4)}(\mathbf{x})$ represents the SVD correction described by (25). Term $D^{(5)}(\mathbf{x})$ represents forecasts made by system (27), (28) with observational noise (see next subsection), but with no model error. Skill scores are made for the ensemble mean using Anomaly Correlation and RMS error and verifying against the time series \mathbf{x}^a .

4c. Ensemble Initialization

Our numerical experiments are initialized in a manner inspired by Wilks (2005); each ensemble forecast is initialized by choosing random perturbations from a distribution which approximates the shape of the attractor local to the initial state. The distribution corresponding to $\mathbf{x}^a(T)$, for example, is found by gathering analogues from long integrations of (27), (28). Analogues are defined to be states within an I -dimensional hypercube (with side i having length equal to 5% of the climatological span of x_i) centered on $\mathbf{x}^a(T)$. As in Wilks (2005), the analogue integrations were performed until each of the 10,000 cubes contained a minimum of 100 states. The $I \times I$ covariance matrix for the analogues of the state $\mathbf{x}^a(T)$ is denoted $\mathbf{C}(T)$. The distribution from which the initial ensemble for the forecast of $\mathbf{x}^a(T)$ is chosen is given by

$$\mathbf{C}^{\text{init}}(T) = \frac{0.05^2 \sigma_{\text{clim}}^2}{\lambda} \mathbf{C}(T) \quad (32)$$

where λ is the average eigenvalue of $C(T)$ and σ_{clim} is the climatological standard deviation of \mathbf{x}^a . The covariance $C^{\text{init}}(T)$ has the same eigenvectors and correlations as $C(T)$, but is scaled so that the average standard deviation is 5% of the climatology of the true time series \mathbf{x}^a (Wilks, 2005). Control states for each ensemble forecast are generated by adding appropriately shaped random noise to each of the 10,000 true states

$$\mathbf{x}_0^f(1, k) = \mathbf{x}^a(k) + \sqrt{C^{\text{init}}(k)}\mathbf{y}(1, k) \quad k = 1, 2, \dots, 10000. \quad (33)$$

where $\mathbf{y}(1, k)$ is an I -dimensional vector whose entries are independent random numbers chosen from a Gaussian distribution. The square-root of C^{init} is computed offline for each initial state using the Cholesky decomposition (Golub and Van Loan, 1996). Ensemble members are then generated from a multivariate Gaussian distribution by performing the same operation on $\mathbf{x}_0^f(1, k)$, namely

$$\mathbf{x}_0^f(j, k) = \mathbf{x}_0^f(1, k) + \sqrt{C^{\text{init}}(k)}\mathbf{y}(j, k) \quad j = 2, 3, \dots, N_e. \quad (34)$$

where \mathbf{y} is different for each of the 10,000 initial states and $N_e = 20$ ensemble members.

4d. Results

The bias in model (29), relative to (27), results from the unresolved behavior of the small-amplitude variables y_i and the additional state-independent error term $\alpha \sin(\frac{2\pi i}{T})$. The time-average effect of these model errors, namely $\langle \mathbf{q} \rangle$ is given by

$$\langle q_i \rangle = \frac{1}{N_a} \sum_{t=1}^{N_a} \left[-\frac{hc}{b} \sum_{j=J(i-1)+1}^{iJ} y_j(t) \right] - \alpha \sin\left(\frac{2\pi i}{I}\right) \quad \text{for } i = 1, 2, \dots, I \quad (35)$$

[Figure 1 about here.]

The time-average residual $\langle \delta x_{12}^a \rangle$ (3) is an empirical estimate of $\langle q \rangle$. Figure 1 shows the true bias $\langle q \rangle$, and $\langle \delta x_{12}^a \rangle$ weighted by a factor of -12^{-1} representing the fact that it is a correction applied every hour, but was calculated by averaging 12-hour errors. The trained empirical correction term $D^{(2)}(x) = \langle \delta x_{12}^a \rangle$ slightly underestimates and shifts the true bias of model (29).

Coupled signals between normalized, anomalous residuals and forecasts, namely $\overline{\delta x_{12}^{a'}}$ and $\overline{x_{12}^{f'}}$ respectively, are identified and ranked by Singular Value Decomposition (17). The left and right singular vectors \mathbf{u}_k and \mathbf{v}_k are shown in Figure 2 for forecasts made with model (29); they are superimposed on a shifted scale for visual simplicity. Unit vectors \mathbf{u}_1 and \mathbf{v}_1 suggest that states of model (29) of the shape \mathbf{u}_1 are typically misrepresented. Because each of the coupled signal pairs for the Lorenz '96 model with forcings $F = 14$ and $F = 18$ roughly satisfies $\mathbf{u}_k \sim -\mathbf{v}_k$, the state anomalies will be damped by the SVD empirical correction. The SVD method is also capable of suggesting amplification of anomalies if $\mathbf{u}_k \sim \mathbf{v}_k$, as seen for $F = 8$ mode 3 and as demonstrated by Danforth et al. (2007).

[Figure 2 about here.]

The explained variance (21) for the spectrum of singular values of the cross-covariance matrix $C_{\delta x_{12}^a x_{12}^f}$ is shown in Figure 3. It indicates that the most of the empirically estimated state-

dependent model error can be captured with the first few modes for $F = 8$, but not for $F = 18$. The additional computational expense of including the few modes required to reach 95% is negligible for this model, where the number of degrees of freedom that one can attempt to correct is a maximum of $K = I = 8$. For an operational weather model, the spectrum is likely to be significantly flatter than that observed for $F = 8$. As a result, we may be forced to correct many forecast patterns (large K) to see improvement. Fortunately, the SVD technique we are describing is very cheap, even for large K .

[Figure 3 about here.]

A sample of 10^7 short forecasts was used to train the operators in order to predict a maximum of $I = 8$ degrees of freedom. In practice, such a large sample size is unavailable for training. In the case of a small training set, the singular value spectrum may be steep not due to the importance of the leading modes, but due to the smaller sample size (Hamill et al. 2001). The larger the sample size, the more likely the operator will represent the true covariance and, hopefully, the greater the number of forecast patterns that can be corrected.

Typical 10-day, 20-member ensemble forecasts of x_1 using model (30) and $F = 14$, with empirical correction terms described by (31), are shown in Figure 4. Forecasts empirically corrected by the observed bias of model (29), $D^{(2)}$, perform slightly better than forecasts not corrected at all, $D^{(1)}$. Ensemble divergence is typically significant by day 5 for both $D^{(1)}$ and $D^{(2)}$. State-dependent empirical correction significantly improves forecasts. Ensemble spread is weak for both Leith's empirical correction $D^{(3)}$ and the SVD correction $D^{(4)}$ with mode truncation $K = 5$. However, small spread is seen for perfect model forecasts $D^{(5)}$, and the effect is less evident for

$F = 8$ and $F = 18$. Since the ensemble spread represents the uncertainty in the forecast, and since the forecast skill is clearly improved by the Leith and SVD empirical corrections, this result should be expected.

[Figure 4 about here.]

Figure 5 shows the average anomaly correlation and RMSE of the ensemble mean of 10,000 independent 20-member ensemble forecasts. The state-independent correction adds approximately 1 time unit (5 days) to the usefulness of $F = 8$ forecasts, and 0.1 time units (12-hours) to the usefulness of $F = 14$ and $F = 18$ forecasts. For $F = 14$, Leith's operator improves forecasts by 710% (27.2 days), and the SVD correction results in an improvement of 1176% (45 days). The SVD correction term $D^{(4)}(\mathbf{x})$ is chosen to have $K = 7, 5,$ and 2 modes for forcings $F = 18, 14,$ and 8 respectively; the truncation was chosen to explain 95% of the variance in the cross-covariance matrix $C_{\delta x_{12}^a, x_{12}^f}$ (see Fig. 3). Table 1 summarizes the improvement in AC scores. While we present results for $N_e = 20$, AC scores for $N_e = 1$ and $N_e = 50$ are qualitatively similar, indicating that the performance of the SVD method is insensitive to ensemble size.

[Figure 5 about here.]

Wilks (2005) used differences between the tendencies of the resolved variables in model (29), with $\alpha = 0$, $F = 18$, and the actual tendencies of system (27), (28), to approximate model error. The collection of tendency errors for each resolved variable x_i were then fit with a degree four polynomial

$$D^{(6)}(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + e_i \quad (36)$$

where the term e_i was a stochastic component and β_0 corresponds to $\langle \delta \mathbf{x}_{\Delta t}^a \rangle$ from (3). For $e_i = 0$, Wilks found that on average, 20-member ensemble forecasts crossed the 0.6 anomaly correlation line at a time of 4 days. This improvement is illustrated by the **X** in the bottom left window of Figure 5. Wilks demonstrated a particular choice for the stochastic term e_i to increase the crossing time of 20-member ensemble forecasts by 180% of the crossing time for single integrations with a deterministic parameterization of model error. Ensemble spread was also shown to improve as a result of the stochastic parameterization, with deterministic parameterizations resulting in smaller spread. Figure 5 demonstrates Leith's state-dependent linear correction to improve on the crossing time of 20-member ensemble forecasts by 292% for $F = 18$, and the SVD correction to improve on the same measure by 338%. For $F = 8$ and $F = 14$, the SVD method outperforms Leith's method by a much larger margin (see Table 1).

[Table 1 about here.]

Figure 6 shows the average ensemble spread vs. time and vs. RMSE. Weak ensemble dispersion is seen for the SVD correction $D^{(4)}$ for both $F = 14$ and $F = 18$. Since $K = 2$ modes were used for SVD correction of $F = 8$, the ensemble spread is quite good. Figure 2 indicates that for $F = 14$ and $F = 18$, all of the modes used to correct forecasts result in damping of anomalies, and consequently damping ensemble spread as well. It is also possible that a perfect parameterization of the small-amplitude behavior should result in convergence of the ensemble to within observational noise, similar to convergence of a data assimilation scheme. The weakening of ensemble spread may also be a result of the choice of $F = 18$, which is much larger than is typically chosen for the Lorenz '96 model. Both SVD and Leith's empirical correction methods

are essentially finding the maximum likelihood estimate of the probability distribution of corrections observed during the training period, given the current state. Improvement of the ensemble spread will be sought in a future study by adding random corrections drawn from this distribution to each ensemble member. Ideally, a low-order method for estimating the uncertainty associated with each correction should be built into the SVD scheme.

[Figure 6 about here.]

5. Discussion

Leith's method consistently improves forecasts for short lead times, outperforming the SVD method for the first 10 days of $F = 14$ and $F = 18$ forecasts. After 10 days, the ensemble spread of forecasts made using Leith's method grows rapidly, while the spread in SVD method forecasts remains small. $F = 14$ and $F = 18$ forecasts made with the SVD correction deteriorate rapidly for the first few days, after which time they degrade at essentially the same rate as forecasts made with a perfect model. This second dynamic behavior is an indication that after the first few days, the SVD method is an excellent parameterization of the behavior of the small-amplitude variables. In fact, the SVD method performs as well or better than the perfect model for the first 10 days of $F = 8$ forecasts. However, we see in Figure 3 that as F increases, the SVD method requires a greater number of modes to represent the cross-correlation matrix utilized by Leith's method. As a result, in the SVD method, $F = 18$ forecasts are corrected by modes whose coupling is less statistically significant than $F = 8$ and $F = 14$. This is demonstrated by mode $k = 8$

in Figure 2 which significantly harms SVD corrected forecasts (see final $D^{(4)}(\mathbf{x})$ row in Table 1) relative to truncation at mode $K = 7$.

Clearly, these results are overoptimistic in that the model error in (29) relative to system (27) is highly state-dependent. However, Figure 5 indicates that both Leith's empirical correction operator and the SVD approximation do an excellent job representing the state-dependent component of the unresolved small-amplitude behavior. In fact, the SVD method isolates and ranks the most relevant spatial correlations described by Leith's operator. As a result, truncation can actually improve performance. This was verified by using $K = I = 8$ modes for term $D^{(4)}(\mathbf{x})$; forecasts were slightly worse than those made using Leith's operator for forcings $F = 14$ and $F = 18$.

The methods presented here have relied on an exact characterization of the true state for a very long training period in order to understand the best possible impact of empirical correction. While the analysis increments for an operational weather model are typically available from pre-implementation testing, they are computed as the difference between an analysis which suffers from deficiencies in the model used to create it, and are only available for short training periods. Future studies will examine the effectiveness of model error parameterization by SVD using less accurate estimates of the true state and shorter training periods.

6. Conclusion

A new method of state-dependent error correction was introduced, based on Singular Value Decomposition of coupled residual and forecast state anomalies. The cross covariance is the same as that which appears in Leith's formulation, but it would be prohibitive to compute for the grid

density required by operational weather models. The new method uses the SVD modes as a basis for linear regression and results in significant forecast improvement. The new method is also many orders of magnitude faster than Leith's empirical correction. The method can be applied at a rather low cost, both in the training and in the correction phases, and yields significant forecast improvements, at least for the Lorenz '96 model and the simple but realistic global QG and SPEEDY models (Danforth et al., 2007). It could be applied with low computational cost and minimal sampling problems to data assimilation and ensemble numerical weather prediction, applications where accounting for model errors has been found to be important. Investigation of the weakening of ensemble spread is needed. The method may be particularly useful for forecasting of severe weather events where a posteriori bias correction will typically weaken anomalies. Furthermore, the patterns identified by SVD could also be used to identify sources of model deficiencies and thereby guide future model improvements.

Acknowledgement This research was supported by a NOAA THORPEX grant NOAA/NA040AR4310103 and a NASA Phase-II grant # NNG 06GE87G to the Vermont Advanced Computing Center.

References

- Anderson, J., 2001: An ensemble adjustment filter for data assimilation. *Monthly Weather Review*, **129**.
- Barnston, A., M. Glantz, and Y. He, 1999: Predictive skill of statistical and dynamical climate models in sst forecasts during the 1997-98 el niño episode and the 1998 la niña onset. *Bulletin of the American Meteorological Society*, **80**.
- Bishop, C., B. Etherton, and S. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter. part i: The theoretical aspects. *Monthly Weather Review*, **129**, 420–436.
- Bretherton, C., C. Smith, and J. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, **5**, 541–560.
- Carter, G., J. Dallavalle, and H. Glahn, 1989: Statistical forecasts based on the national meteorological center's numerical weather prediction system. *Weather Forecasting*, **4**, 401–412.
- Dalcher, A. and E. Kalnay, 1987: Error growth and predictability in operational ecmwf forecasts. *Tellus*, **39**, 474–491.
- D'Andrea, F. and R. Vautard, 2000: Reducing systematic errors by empirically correcting model errors. *Tellus*, **52A**, 21–41.
- Danforth, C. M., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Monthly Weather Review*, **135**, 281–299.

- Danforth, C. M. and J. A. Yorke, 2006: Making forecasts for chaotic physical processes. *Physical Review Letters*, **96**, 144102.
- DelSole, T. and A. Y. Hou, 1999: Empirical correction of a dynamical model. part 1: Fundamental issues. *Monthly Weather Review*, **127**, 2533–2545.
- Emanuel, K. and E. Lorenz, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, **55**.
- Faller, A. and D. Lee, 1975: Statistical corrections to numerical prediction equations. *Monthly Weather Review*, **103**, 845–855.
- Faller, A. and C. Schemm, 1977: Statistical corrections to numerical prediction equations. ii. *Monthly Weather Review*, **105**, 37–56.
- Ferranti, L., E. Klinker, A. Hollingsworth, and B. Hoskins, 2002: Diagnosis of systematic forecast errors dependent on flow anomalies. *Quarterly Journal of the Royal Meteorological Society*, **128**, 1623–1640.
- Glahn, H. and D. Lowry, 1972: The use of model output statistics in objective weather forecasting. *Journal of Applied Meteorology*, **11**, 1203–1211.
- Golub, G. and C. V. Loan, 1996: *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD.
- Hamill, T. and C. Snyder, 2000: A hybrid ensemble kalman filter-3d variational analysis scheme. *Monthly Weather Review*, **128**, 2905–2919.

- Hamill, T., J. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, **129**.
- Harlim, J., M. Oczkowski, J. Yorke, E. Kalnay, and B. Hunt, 2005: Convex error growth patterns in a global weather model. *Physical Review Letters*, **94**.
- Hoke, J. E. and R. A. Anthes, 1976: The initialization of numerical models by a dynamic initialization technique. *Monthly Weather Review*, **104**, 1551–1556.
- Houtekamer, P. and H. Mitchell, 2001: A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, 796–811.
- Hunt, B., E. Kalnay, E. Kostelich, E. Ott, D. Patil, T.Sauer, I. Szunyogh, J. Yorke, and A. Zimin, 2004: Four-dimensional ensemble kalman filtering. *Tellus*, **56**, 273–277.
- Kaas, E., A. Guldborg, W. May, and M. Deque, 1999: Using tendency errors to tune parameterization of unresolved dynamical scale interactions in atmospheric general circulation models. *Tellus*, **51**, 612–629.
- Kalnay, E., 2003: *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press.
- Klinker, E. and P. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *Journal of the Atmospheric Sciences*, **49**, 608–627.

- Leith, C. E., 1978: Objective methods for weather prediction. *Annual Review of Fluid Mechanics*, **10**, 107–128.
- 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics*, **33**, 141–266.
- Lorenz, E. N., 1996: Predictability: a problem partly solved. *Proceedings of the ECMWF Seminar on Predictability*, **1**.
- Marshall, C. H., K. C. Crawford, K. E. Mitchell, and D. J. Stensrud, 2003: The impact of the land surface physics in the operational ncep eta model on simulating the diurnal cycle: Evaluation and testing using oklahoma mesonet data. *Weather and Forecasting*, **18**, 748–768.
- Marshall, J. and F. Molteni, 1993: Toward a dynamical understanding of planetary-scale flow regimes. *Journal of the Atmospheric Sciences*, **50**, 1792–1818.
- Miyoshi, T., 2005: Ensemble kalman filter experiments with a primitive-equation global model. *Doctoral Dissertation*.
- Miyoshi, T., E. Kalnay, J. Ballabrera, C. M. Danforth, and H. Li, 2006: How model errors affect 3d-var and ensemble kalman filter? *In preparation*.
- Molteni, F., 2003: Atmospheric simulations using a gcm with simplified physical parametrizations. i: model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20**, 175–191.

- Ott, E., B. Hunt, I. Szunyogh, A. Zimin, E. Kostelich, M. C. T. Sauer, E. Kalnay, D. Patil, and J. Yorke, 2004: A local ensemble kalman filter for atmospheric data assimilation. *Tellus*, **56**, 415–428.
- Parrish, D. and J. Derber, 1992: The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, **120**, 1747–1763.
- Renwick, J. and J. Wallace, 1995: Predictable anomaly patterns and the forecast skill of northern hemisphere wintertime 500-mb height fields. *Monthly Weather Review*, **123**, 2114–2131.
- Reynolds, C., P. J. Webster, and E. Kalnay, 1994: Random error growth in nmc's global forecasts. *Monthly Weather Review*, **122**, 1281–1305.
- Saha, S., 1992: Response of the nmc mrf model to systematic error correction within integration. *Monthly Weather Review*, **120**, 345–360.
- Schemm, C., D. Unger, and A. Faller, 1981: Statistical corrections to numerical predictions. iii. *Monthly Weather Review*, **109**, 96–109.
- Schemm, J.-K. and A. Faller, 1986: Statistical corrections to numerical predictions. iv. *Monthly Weather Review*, **114**, 2402–2417.
- Wallace, J., C. Smith, and C. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of Climate*, **5**, 561–576.
- Whitaker, J. and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, **130**.

Widmann, M., 2005: One-dimensional cca and svd, and their relationship to regression maps.

Journal of Climate, **18**, 2785–2792.

Wilks, D. S., 2005: Effects of stochastic parameterizations in the lorenz '96 system. *Quarterly*

Journal of the Royal Meteorological Society, **131**, 389–407.

List of Figures

- 1 The empirically generated bias $\langle \delta \mathbf{x}_{12}^a \rangle$ (time-average residual) in model (29) relative to (27) slightly underestimates and shifts the true bias $\langle \mathbf{q} \rangle$. The true bias is a combination of the sinusoidal state-independent error and the bulk effect of ignoring the small-amplitude modes. It is described by equation (35). The Lorenz '96 model with forcing $F = 18$ exhibits a slightly larger bias due the effect the large-amplitude variables (with increased energy) have on the small-amplitude variables. 34
- 2 Coupled signals between normalized, anomalous residuals and forecasts, namely $\overline{\delta \mathbf{x}_{12}^{a'}}$ and $\overline{\mathbf{x}_{12}^{f'}}$ respectively, are identified and ranked by Singular Value Decomposition (17). The left and right singular vectors \mathbf{u}_k (solid) and \mathbf{v}_k (dashed) are shown for forecasts made with model (29); they are superimposed for visual simplicity. For example, unit vectors \mathbf{u}_1 and \mathbf{v}_1 represent the most dominant coupled signal between errors and forecasts respectively. Qualitative similarities are seen between the modes found for the $F = 14$ and $F = 18$ Lorenz '96 model. Mode 3 for $F = 8$ suggests a strengthening of anomalies of the shape \mathbf{v}_3 , while mode 3 for both $F = 14$ and $F = 18$ suggests a weakening of anomalies of the shape \mathbf{v}_3 35

- 3 The explained variance (21) for the spectrum of singular values of the cross-covariance matrix $C_{\delta x_{12}^a x_{12}^f}$ shows how much of the empirically estimated state-dependent model error can be captured with the leading modes. To explain 95% of the variance, $K = 7, 5,$ and 2 modes are required for the Lorenz '96 model with forcings $F = 18, 14,$ and 8 respectively. Steep spectrums, like that seen for $F = 8,$ indicate that the SVD representation is likely to be able to capture the relevant model error information with very few degrees of freedom. 36
- 4 Typical 10-day ensemble forecasts of x_1 using model (30), $F = 14,$ with empirical correction terms described by (31). The dashed curve is a true solution of system (27), (28). The solid curves are a 20-member ensemble forecast of model (30), initialized according to equation (34). Forecasts empirically corrected by the observed bias of model (29), namely $D^{(2)},$ perform slightly better than forecasts not corrected at all, $D^{(1)}.$ Ensemble divergence is typically significant by day 5 for both $D^{(1)}$ and $D^{(2)}.$ Ensemble spread is weak for both Leith's empirical correction $D^{(3)}$ and the SVD correction $D^{(4)}$ with mode truncation $K = 5.$ However, small spread is seen for perfect model forecasts $D^{(5)},$ and the effect is less evident for $F = 8$ and $F = 18.$ 37

- 5 Average anomaly correlation and RMSE the ensemble mean of 10,000 independent 20-member ensemble forecasts. The state-independent correction $D^{(2)}$ adds approximately 1 time unit (5 days) to the usefulness of forecasts with no correction ($D^{(1)}$) for $F = 8$, and 0.1 time units (12-hours) to the usefulness of $F = 14$ and $F = 18$ forecasts. With a parameterization of model error in $F = 18$ forecasts, Wilks (2005) improved forecasts by a similar length of time (see the **X** in the lower left hand window). For $F = 14$, Leith's operator $D^{(3)}$ improves forecasts by an average of 710%, and the SVD correction $D^{(4)}$ results in an average improvement of 1176%. The SVD correction is chosen to have $K = 7, 5,$ and 2 modes for forcings $F = 18, 14,$ and 8 respectively; the truncation was chosen to explain 95% of the variance in the cross-covariance matrix $C_{\delta x_{12}^a x_{12}^f}$ (see Fig. 3). The only source of error in forecasts made with the perfect model $D^{(5)}$ is observational noise. See Table 1 for a complete list of improvements. 38
- 6 Average ensemble spread is shown vs. time and vs. RMSE for 10,000 independent 20-member ensemble forecasts. Terms $D^{(1)}$ and $D^{(2)}$ have been removed for visual clarity. Weak ensemble dispersion is seen for $D^{(4)}$ for $F = 14$ and $F = 18$. Since $K = 2$ modes were used for SVD correction of $F = 8$, the ensemble spread is quite good. As more modes are used to correct the forecast, the empirical correction appears to overpower the model dynamics. This is most likely a result of the use of coupled modes which are less statistically significant. 39

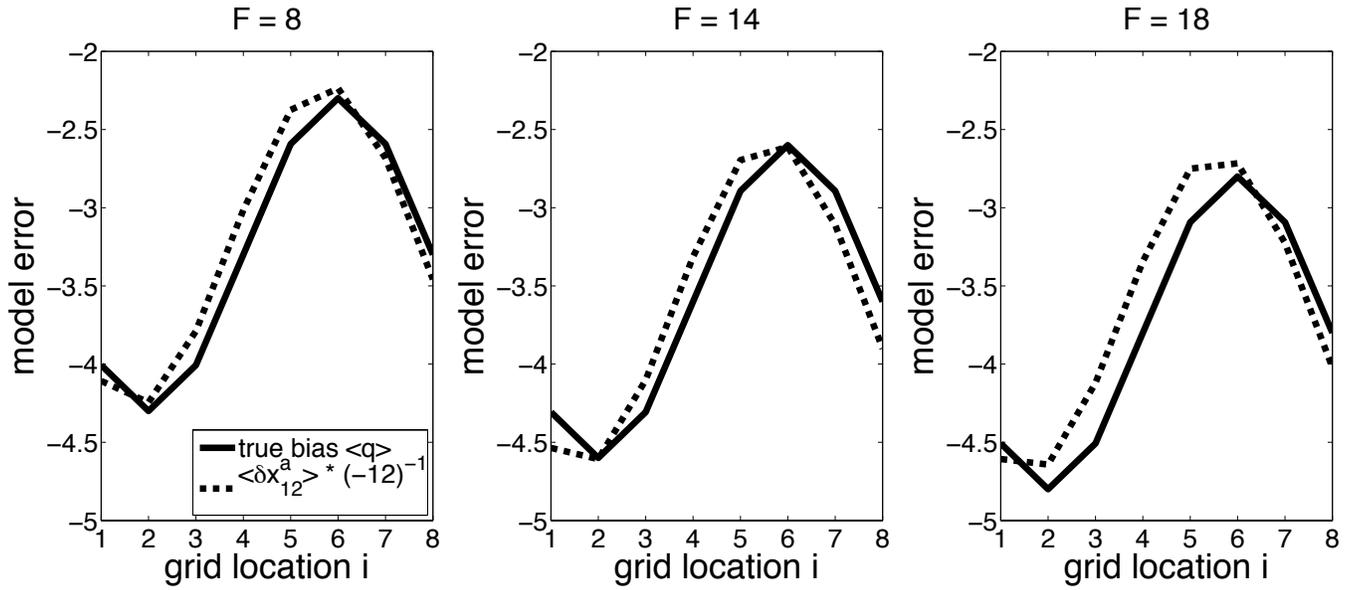


Figure 1: The empirically generated bias $\langle \delta \mathbf{x}_{12}^a \rangle$ (time-average residual) in model (29) relative to (27) slightly underestimates and shifts the true bias $\langle \mathbf{q} \rangle$. The true bias is a combination of the sinusoidal state-independent error and the bulk effect of ignoring the small-amplitude modes. It is described by equation (35). The Lorenz '96 model with forcing $F = 18$ exhibits a slightly larger bias due the effect the large-amplitude variables (with increased energy) have on the small-amplitude variables.

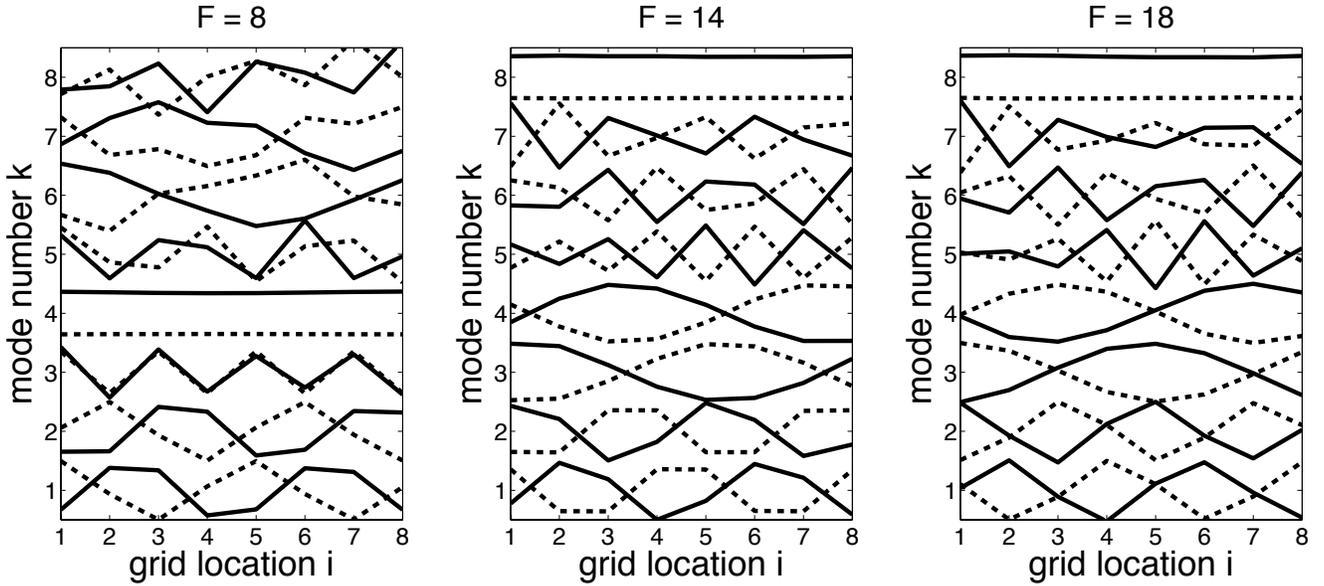


Figure 2: Coupled signals between normalized, anomalous residuals and forecasts, namely $\overline{\delta x_{12}^{a'}}$ and $\overline{x_{12}^{f'}}$ respectively, are identified and ranked by Singular Value Decomposition (17). The left and right singular vectors \mathbf{u}_k (solid) and \mathbf{v}_k (dashed) are shown for forecasts made with model (29); they are superimposed for visual simplicity. For example, unit vectors \mathbf{u}_1 and \mathbf{v}_1 represent the most dominant coupled signal between errors and forecasts respectively. Qualitative similarities are seen between the modes found for the $F = 14$ and $F = 18$ Lorenz '96 model. Mode 3 for $F = 8$ suggests a strengthening of anomalies of the shape \mathbf{v}_3 , while mode 3 for both $F = 14$ and $F = 18$ suggests a weakening of anomalies of the shape \mathbf{v}_3 .

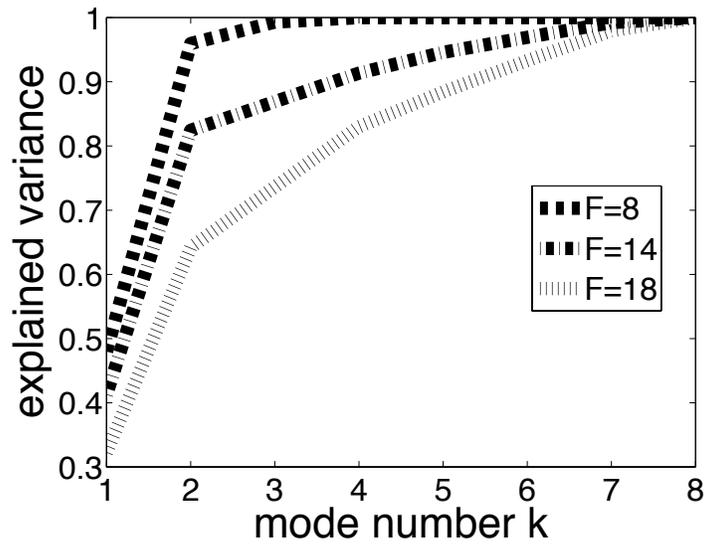


Figure 3: The explained variance (21) for the spectrum of singular values of the cross-covariance matrix $\mathbf{C}_{\delta\mathbf{x}_{12}^a, \mathbf{x}_{12}^f}$ shows how much of the empirically estimated state-dependent model error can be captured with the leading modes. To explain 95% of the variance, $K = 7, 5,$ and 2 modes are required for the Lorenz '96 model with forcings $F = 18, 14,$ and 8 respectively. Steep spectrums, like that seen for $F = 8$, indicate that the SVD representation is likely to be able to capture the relevant model error information with very few degrees of freedom.

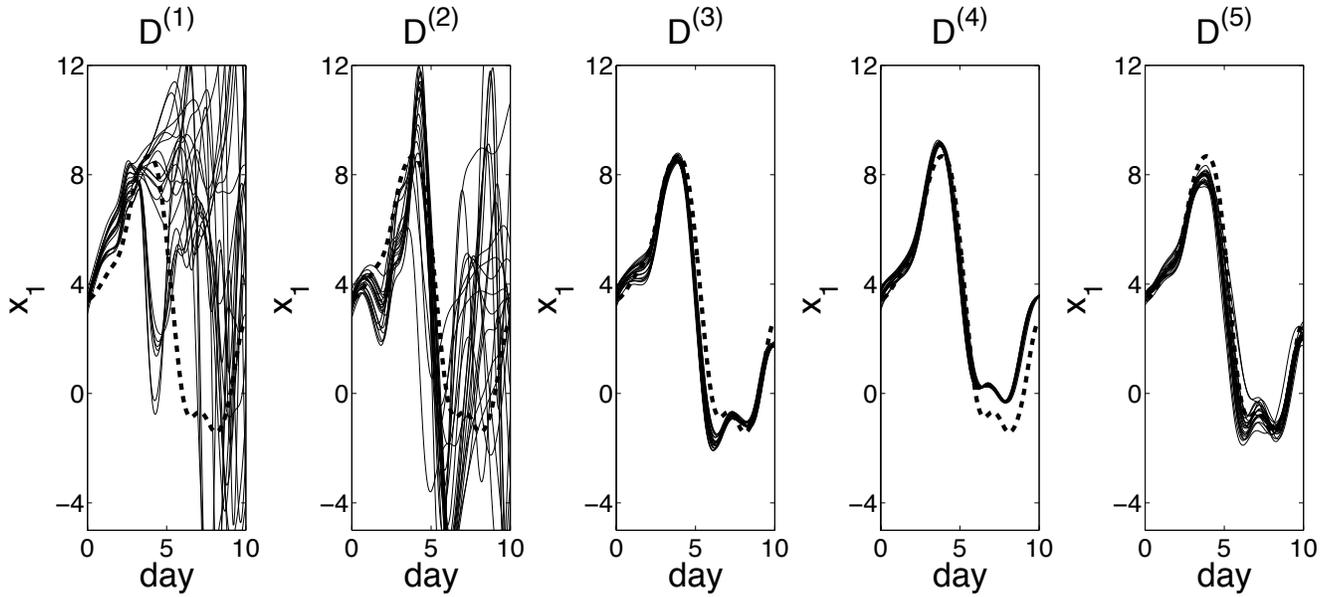


Figure 4: Typical 10-day ensemble forecasts of x_1 using model (30), $F = 14$, with empirical correction terms described by (31). The dashed curve is a true solution of system (27), (28). The solid curves are a 20-member ensemble forecast of model (30), initialized according to equation (34). Forecasts empirically corrected by the observed bias of model (29), namely $D^{(2)}$, perform slightly better than forecasts not corrected at all, $D^{(1)}$. Ensemble divergence is typically significant by day 5 for both $D^{(1)}$ and $D^{(2)}$. Ensemble spread is weak for both Leith's empirical correction $D^{(3)}$ and the SVD correction $D^{(4)}$ with mode truncation $K = 5$. However, small spread is seen for perfect model forecasts $D^{(5)}$, and the effect is less evident for $F = 8$ and $F = 18$.

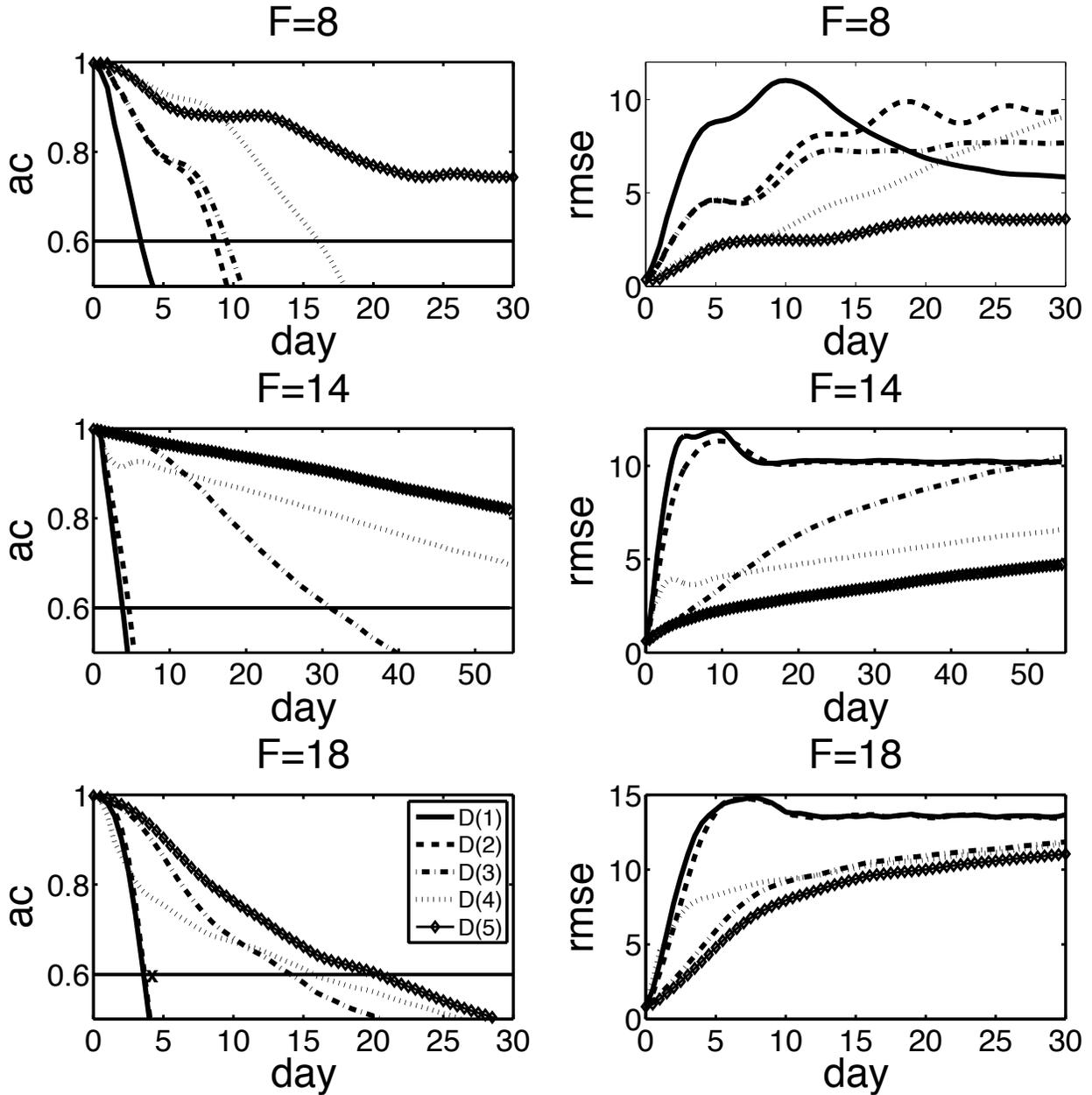


Figure 5: Average anomaly correlation and RMSE the ensemble mean of 10,000 independent 20-member ensemble forecasts. The state-independent correction $D^{(2)}$ adds approximately 1 time unit (5 days) to the usefulness of forecasts with no correction ($D^{(1)}$) for $F = 8$, and 0.1 time units (12-hours) to the usefulness of $F = 14$ and $F = 18$ forecasts. With a parameterization of model error in $F = 18$ forecasts, Wilks (2005) improved forecasts by a similar length of time (see the X in the lower left hand window). For $F = 14$, Leith's operator $D^{(3)}$ improves forecasts by an average of 710%, and the SVD correction $D^{(4)}$ results in an average improvement of 1176%. The SVD correction is chosen to have $K = 7, 5,$ and 2 modes for forcings $F = 18, 14,$ and 8 respectively; the truncation was chosen to explain 95% of the variance in the cross-covariance matrix $C_{\delta x_{12}^a x_{12}^f}$ (see Fig. 3). The only source of error in forecasts made with the perfect model $D^{(5)}$ is observational noise. See Table 1 for a complete list of improvements.

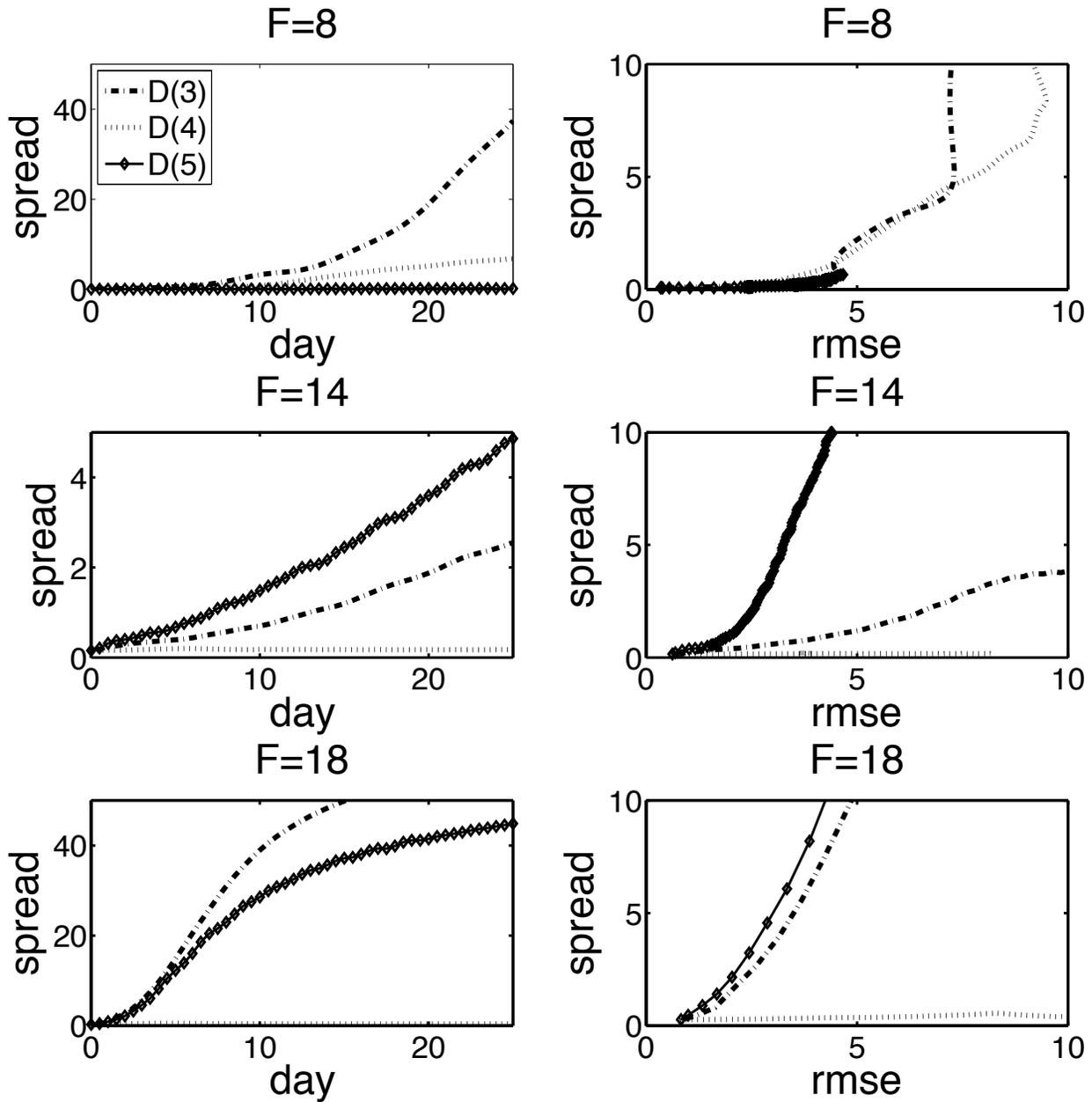


Figure 6: Average ensemble spread is shown vs. time and vs. RMSE for 10,000 independent 20-member ensemble forecasts. Terms $D^{(1)}$ and $D^{(2)}$ have been removed for visual clarity. Weak ensemble dispersion is seen for $D^{(4)}$ for $F = 14$ and $F = 18$. Since $K = 2$ modes were used for SVD correction of $F = 8$, the ensemble spread is quite good. As more modes are used to correct the forecast, the empirical correction appears to overpower the model dynamics. This is most likely a result of the use of coupled modes which are less statistically significant.

List of Tables

- 1 Improvement in crossing time of anomaly correlation scores with 0.6 for different empirical correction schemes relative to $D^{(1)}(\mathbf{x}) = 0$. For the anomaly correlations, see Figure 5 where $D^{(4)}(\mathbf{x})$ is truncated at mode $K = 2, 5,$ and 7 for the Lorenz '96 model with forcings $F = 8, 14,$ and 18 respectively. These improvements are shown in bold in the above chart. The truncation was chosen to explain 95% of the variance in the cross-covariance matrix $C_{\delta\mathbf{x}_{12}^a\mathbf{x}_{12}^f}$, see Fig 3. 41

Lorenz '96 Forcing	$F = 8$	$F = 14$	$F = 18$
$D^{(1)}(\mathbf{x}) = 0$	0% (0 days)	0% (0 days)	0% (0 days)
$D^{(2)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle$	156% (5.3 days)	22% (0.8 days)	2% (.1 days)
$D^{(3)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + L\mathbf{x}'$	181% (6.1 days)	710% (27.2 days)	292% (10.5 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^1 \mathbf{z}_k(\mathbf{T})$	509% (17.3 days)	-26% (-1.0 days)	-6% (-0.2 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^2 \mathbf{z}_k(\mathbf{T})$	375% (12.7 days)	53% (2.0 days)	-2% (-0.1 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^3 \mathbf{z}_k(\mathbf{T})$	309% (10.5 days)	215% (8.2 days)	57% (2.0 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^4 \mathbf{z}_k(\mathbf{T})$	285% (9.7 days)	471% (18.0 days)	120% (4.3 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^5 \mathbf{z}_k(\mathbf{T})$	270% (9.2 days)	1176% (45.0 days)	213% (7.6 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^6 \mathbf{z}_k(\mathbf{T})$	276% (9.4 days)	1288% (49.3 days)	294% (10.6 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^7 \mathbf{z}_k(\mathbf{T})$	280% (9.5 days)	1325% (50.7 days)	338% (12.1 days)
$D^{(4)}(\mathbf{x}) = \langle \delta \mathbf{x}_{12}^a \rangle + \sum_{k=1}^8 \mathbf{z}_k(\mathbf{T})$	280% (9.5 days)	384% (14.7 days)	139% (5.0 days)
$D^{(5)}(\mathbf{x}) = \text{perfect model}$	1528% (51.8 days)	3026% (115.6 days)	479% (17.2 days)

Table 1: Improvement in crossing time of anomaly correlation scores with 0.6 for different empirical correction schemes relative to $D^{(1)}(\mathbf{x}) = 0$. For the anomaly correlations, see Figure 5 where $D^{(4)}(\mathbf{x})$ is truncated at mode $K = 2, 5,$ and 7 for the Lorenz '96 model with forcings $F = 8, 14,$ and 18 respectively. These improvements are shown in bold in the above chart. The truncation was chosen to explain 95% of the variance in the cross-covariance matrix $\mathbf{C}_{\delta \mathbf{x}_{12}^a \mathbf{x}_{12}^f}$, see Fig 3.