Nonglobal Parameter Estimation Using Local Ensemble Kalman Filtering

THOMAS BELLSKY
School of Mathematics and Statistical Sciences, Arizona State University, Tempe, Arizona

JESSE BERWALD
Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, Minnesota

LEWIS MITCHELL
Department of Mathematics and Statistics, University of Vermont, Burlington, Vermont

(Manuscript received 15 June 2013, in final form 21 January 2014)

ABSTRACT

The authors study parameter estimation for nonglobal parameters in a low-dimensional chaotic model using the local ensemble transform Kalman filter (LETKF). By modifying existing techniques for using observational data to estimate global parameters, they present a methodology whereby spatially varying parameters can be estimated using observations only within a localized region of space. Taking a low-dimensional nonlinear chaotic conceptual model for atmospheric dynamics as a numerical test bed, the authors show that this parameter estimation methodology accurately estimates parameters that vary in both space and time, as well as parameters representing physics absent from the model.

1. Introduction

Accurately estimating model parameters represents a significant challenge in operational climate modeling. Model parameters are numerical values typically encoding physical information about a dynamical system: fundamental constants such as the acceleration due to gravity, or parameterizations of subgrid-scale processes such as cloud formation (Grabowski 2001; Randall et al. 2003). Modern numerical climate models contain hundreds of such parameters, all of which must be tuned in some manner so as to ensure numerical stability of the code, as well as guarantee the best possible match with historic observational data, while also providing an adequate parameterization of subgrid-scale physics (Frederiksen et al. 2012; Majda et al. 2001; Palmer 2001). Current techniques for estimating such parameters include comparing free runs of a model with historical observational data (Peacock 2012) or Markov chain Monte Carlo (MCMC) methods (Solonen et al. 2012; Urban and Keller 2010). However, both approaches have limitations: the first does not provide a systematic way for using data to inform the model, and the second is unsuitable for large models (Annan and Hargreaves 2007).

Meanwhile, data assimilation (DA) techniques have been used operationally for more than 20 years to combine a numerical model with observational data in order to provide optimal state estimates for periodic reinitialization (Kalnay 2002). In recent years, DA techniques have been used to estimate model parameters in conjunction with updating the state estimate. Typically, there are observable data for the state, but no direct observable data for the parameter. These DA parameter estimation techniques can be divided into two types: the first simultaneously updates the estimated parameter when updating the estimated state (see, e.g., Anderson 2001; Aksoy et al. 2006; Hacker and Angevine 2013), and the second separately updates the state estimate and then the parameter estimate (Koyama and Watanabe 2010; Ruiz et al. 2013). The latter approach utilizes an ensemble Kalman filter (EnKF) to first update an ensemble of state vectors and then separately update an ensemble of global parameters using an EnKF without localization. A nonlocalized EnKF is found to be sufficient in this

Corresponding author address: Thomas Bellsky, School of Mathematics and Statistical Sciences, Arizona State University, Physical Sciences, A-Wing, P.O. Box 871804, Tempe, AZ 85287-1804.
E-mail: bellskyt@asu.edu

DOI: 10.1175/MWR-D-13-00200.1

© 2014 American Meteorological Society
case for estimating global parameters (Koyama and Watanabe 2010). Ruiz et al. (2013) modify the above, implementing the local ensemble transform Kalman filter (LETKF) to update an ensemble of state vectors and then use an EnKF without localization to again estimate an ensemble of global parameters. We note that there are examples of other Kalman filtering–based parameter estimation methods in the literature, for example, an extended Kalman and ensemble adjustment Kalman filter approach in Annan et al. (2005) and an EnKF–MCMC hybrid approach in Hakkarainen et al. (2013).

While the techniques of Koyama and Watanabe (2010) and Ruiz et al. (2013) have been shown to work effectively for estimating global parameters, estimating parameters that vary in both space and time remains a significant challenge. In Annan (2005), the authors use an augmented ensemble Kalman filter to estimate time-varying parameters for a modified Lorenz-63 model (Palmer 1999). Alternatively, Kang et al. (2011) use the LETKF to estimate a spatially varying parameter, where they use a technique called “variable localization,” which zeros out any covariance between state variables that do not have a physically significant contribution to the estimated parameter.

We propose a robust methodology to estimate both the state vector as well as spatially and temporally varying model parameters. By only using state observations in a local region of a fixed radius about each grid location, our method utilizes local information on the state to update the estimated parameter at each grid location. We then analyze the fitness of our choice of localization over a range of radii to arrive at a localization radius that is optimal for updating a spatially and temporally varying parameter. Distinct from previous EnKF parameter estimation methods, we use the LETKF parameter estimation methodology for both state and parameter estimation, where we investigate a localization scheme in the parameter estimation distinct from the localization scheme in the state estimation. This offers significant improvement in capturing local features of both the state and parameter spaces compared with previous approaches.

We estimate three different types of nonglobal parameters:

1) spatially varying parameters,
2) spatially and temporally varying parameters, and
3) parameters representing unobserved physics in the model.

Spatially varying parameters are a common occurrence in weather and climate prediction. For example, it is well known that convection statistics in the tropics are very different from those in the midlatitudes (Xu and Randall 2001) and so must be parameterized differently in each region. Furthermore, climate models are sensitive to the chosen parameterization of convection (Stainforth et al. 2005; Rodwell and Palmer 2007). Spatially and temporally varying parameters are particularly relevant in climate studies, for example when estimating or assimilating carbon forcing into models (Engelen et al. 2009). We modify the existing global parameter within our studied conceptual model problem so that it is both a spatially and temporally varying parameter.

The use of parameters that represent unobserved physics is a generic problem across many fields involving prediction. In this case decisions must be made about how to represent unknown and unobservable processes in a model. We will examine two types of parameters representing unobserved physics that are relevant to weather and climate prediction. The first type of parameter represents an unobserved deterministic subsystem evolving on a faster time scale than the observable slow process (Palmer 2001). The second type of parameter presents a more difficult challenge. Now, a stochastic process, whose strength is determined by a deterministic parameter, forces the true system dynamics, implying that very little spatial structure can be leveraged by the data assimilation algorithm.

The outline of the remainder of the paper is as follows: in section 2, we describe the LETKF data assimilation methodology and the proposed parameter estimation methodology, as well as the conceptual atmospheric model; in section 3, we detail results for a conceptual atmospheric model when estimating spatially and temporally varying parameters, as well as for the model error problem when the estimated parameters describe missing physics in the model; and we conclude with a discussion in section 4.

2. Methodology

a. Kalman filter data assimilation

The Kalman filter (Kalman 1960) is a recursive state estimation algorithm, which is used to estimate a physical phenomenon by combining an average of a model forecasting the state with observations of the state. For any positive integer N, let \( z(t) \in \mathbb{R}^N \) denote the true state vector and \( M \) be a map that propagates the state vector at time \( t_i \) to a new state at time \( t_{i+1} \), where

\[
    z(t_{i+1}) = Mz(t_i) + \eta. \tag{1}
\]

Above, \( \eta \) is a Gaussian random vector called the model process noise and \( Q \) is the covariance matrix for \( \eta \), where \( \eta \sim \mathcal{N}(0, Q) \). The linear Kalman filter supposes
at time step $t_{l+1}$ that we have a previous analysis estimate of the state $\mathbf{z}^a(t_i)$ and a covariance matrix for this analysis state $\mathbf{P}^a(t_i)$, which describes the uncertainty in the analysis state. The linear Kalman filter provides an update of the state estimate at $t_{l+1}$ as the background solution:

$$\mathbf{z}^b(t_{l+1}) = \mathbf{Mz}^a(t_i).$$

(2)

Assuming we are at the $l + 1$ time step, the covariance matrix for the background solution is given as

$$\mathbf{P}^b = \mathbf{MP}^a\mathbf{M}^\top + \mathbf{Q}.$$  

(3)

Kalman filter techniques suppose there exists a continuous function $\mathbf{H}: \mathbb{R}^N \rightarrow \mathbb{R}^m$, where $m$ is also a positive integer and typically $m \ll N$, for which the observations $\mathbf{y} \in \mathbb{R}^m$ satisfy $\mathbf{y} = \mathbf{H}(\mathbf{z}) + \mathbf{e}$; here the observation noise $\mathbf{e}$ is a Gaussian random vector of dimension $m$, where $\mathbf{R}$ denotes the covariance matrix for $\mathbf{e}$ such that $\mathbf{e} \sim \mathcal{N}(0, \mathbf{R})$. The Kalman linear filter analysis formulates a maximum likelihood estimate that is found by minimizing the quadratic cost function:

$$J(\mathbf{z}) = (\mathbf{z} - \mathbf{z}^b)^\top (\mathbf{P}^b)^{-1}(\mathbf{z} - \mathbf{z}^b)$$

$$+ [\mathbf{y} - \mathbf{H}(\mathbf{z})]^\top \mathbf{R}^{-1} [\mathbf{y} - \mathbf{H}(\mathbf{z})].$$

(4)

When both the model $\mathbf{M}$ and the observation operator $\mathbf{H}$ are linear, the minimizer of Eq. (4) is a unique, unbiased, minimum variance estimate of $\mathbf{z}$ (Hunt et al. 2007). The Kalman filter algorithm has many equivalent formulations; we use the following form for the Kalman filter update of the analysis state and the analysis covariance, respectively, at time $t_{l+1}$,

$$\mathbf{z}^a = \mathbf{z}^b + \mathbf{K} (\mathbf{y} - \mathbf{Hz}^b),$$

(5)

$$\mathbf{P}^a = (\mathbf{I} + \mathbf{P}^b\mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{P}^b.$$  

(6)

The Kalman gain matrix $\mathbf{K}$ in Eq. (5) is defined as

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^\top \mathbf{R}^{-1}.$$  

(7)

We see in Eq. (5) that this update step is a weighted average between the model estimate background state $\mathbf{z}^b$ and the observations $\mathbf{y}$.

The Kalman filter has been applied to various complex geophysical models (Ghil and Malanotte-Rizzoli 1991). A particular difficulty found in first-generation Kalman filters is the problem of estimating large state vectors resulting from more complex models. To reduce computational expense, approximations of the background covariance matrix $\mathbf{P}^b$ are often made. In three-dimensional variational data assimilation (3D-Var) and four-dimensional variational data assimilation (4D-Var) schemes, the background covariance matrix is replaced by an online constant or slowly time-varying matrix representing typical forecast uncertainties (Corazza et al. 2007).

EnKF methods have been successful (Evensen 1994; Hamill et al. 2001; Houtekamer and Mitchell 1998) in determining online, low-rank approximations of $\mathbf{P}^b$ from an ensemble of $k$ model forecasts where $k \ll N$. At a fixed time step $t_{l+1}$, ensemble Kalman filter methods begin with the previous ensemble of $k$ analysis states $\{\mathbf{z}_i^{a(i)}; i = 1, 2, \ldots, k\}$. The model $\mathbf{M}$, now possibly nonlinear, provides $k$ background states as

$$\mathbf{z}^b_i = \mathbf{Mz}_i^{a(i)},$$

(8)

for $i = 1, 2, \ldots, k$. Traditional ensemble Kalman filter methods update the model covariance matrix $\mathbf{P}^b$ in ensemble space, thus making computations feasible when the ensemble size $k \ll N$ is small. Dropping the time-step notation, these methods produce the following forecast ensemble covariance:

$$\mathbf{P}^b_{en} = \frac{1}{k-1}\mathbf{z}^b(\mathbf{z}^b)^\top,$$

(9)

where $\mathbf{z}^b \in \mathbb{R}^{k \times k}$, and the $i$th column of $\mathbf{z}^b$ is

$$\mathbf{z}^b(i) = \mathbf{z}^b - \bar{\mathbf{z}}^b.$$

(10)

Here, $\mathbf{z}^b$ is the mean of the ensemble of background states from Eq. (8). Additionally, traditional EnKF methods produce an analysis ensemble covariance:

$$\mathbf{P}^a_{en} = \frac{1}{k-1}\mathbf{z}^a(\mathbf{z}^a)^\top,$$

(11)

where $\mathbf{z}^a \in \mathbb{R}^{k \times k}$, and the $i$th column of $\mathbf{z}^a$ is

$$\mathbf{z}^a(i) = \mathbf{z}^{a(i)} - \bar{\mathbf{z}}^a.$$

(12)

Likewise, $\mathbf{z}^b$ is the mean of the $k$ analysis states at time $t_{l+1}$.

It is important to note that the $N \times N$ symmetric matrix $\mathbf{P}^b_{en}$ is of rank $k - 1$, so it is not invertible. It is one-to-one on its $(k - 1)$-dimensional column space $\mathbf{S}$, however, so $(\mathbf{P}^b_{en})^{-1}$ is well defined in $\mathbf{S}$, which is where the Kalman filter minimization is performed. The ensemble transform Kalman filter (ETKF) determines an appropriate coordinate system for performing the
minimization in \( \mathbf{S} \), where it uses the columns of \( \mathbf{Z}^b \), described in Eq. (10), to span \( \mathbf{S} \). Since the columns of \( \mathbf{Z}^b \) are linearly dependent (they sum to zero), these transform methods regard \( \mathbf{Z}^b \) as a transformation from some \( k \)-dimensional space \( \mathbf{S} \) to the \( k-1 \)-dimensional space \( \mathbf{S} \). Then for \( \mathbf{w} \in \mathbf{S} \), we have \( \mathbf{Z}^b \mathbf{w} \in \mathbf{S} \) and the corresponding model state is

\[
\mathbf{z} = \mathbf{z}^b + \mathbf{Z}^b \mathbf{w}. \tag{13}
\]

If \( \mathbf{w} \) is a Gaussian random vector where \( \mathbf{w} \sim \mathcal{N}[0, (k-1)\mathbf{I}] \), then the model state in Eq. (13) has mean \( \mathbf{z}^b \) with the same covariance matrix as in Eq. (9). From this, ETKF methods motivate a new cost function on \( \mathbf{w} \) in \( \mathbf{S} \). In particular, when \( \mathbf{w}^a \) is the minimizer of this new cost function, then we have that \( \mathbf{z}^b = \mathbf{z}^b + \mathbf{Z}^b \mathbf{w}^a \) minimizes the original cost function in Eq. (4) (Bishop et al. 2001; Hunt et al. 2007). Thus, ETKF methods take this \( \mathbf{z}^b \) to be the updated analysis mean.

One concern of ensemble Kalman filter methods is that the ensemble size \( k \) is often too small to provide an accurate approximation of the covariance over the entire spatial domain. The LETKF (Ott et al. 2004) uses spatial localization to help address this concern. The LETKF performs each analysis independently at each model grid point using only observations within a prescribed spatial distance. This is easily explained using the linear Kalman filter formulation above. At some \( i \)th grid point, assume we specify that we use only observations within a radius \( d \) of \( \mathbf{z}^p_i \) in Eq. (2) to update the next analysis. This results in \( r \) observations used in the update step, where \( r \approx m \) (recall \( m \) is the total number of state observations). Additionally, the corresponding covariance background matrix will be in \( \mathbb{R}^{(2d+1) \times 2d+1} \) for a new \( \mathbf{z}^b \in \mathbb{R}^{2d+1} \) and the observation operator will map \( \mathbb{R}^{2d+1} \) to \( \mathbb{R}^r \). Then the Kalman filter update step results in a new \( \mathbf{z}^b \in \mathbb{R}^{2d+1} \) and a new analysis covariance matrix in \( \mathbb{R}^{2d+1 \times 2d+1} \), where the center term in the new \( \mathbf{z}^b \) is used to update \( \mathbf{z}^p_i \) (recall the analysis at the \( i \)th grid point). Additionally, each of the localized analysis covariance matrices can be combined to construct a global covariance matrix in \( \mathbb{R}^{N \times N} \), where any entry more than distance \( d \) from the diagonal will be zero, eliminating spurious artificial correlations between distant positions. Thus, localization results in each local analysis determining different linear combinations of ensemble members and the combined global analysis explores a much larger dimensional space than the \( k \) ensemble members alone (Hunt et al. 2007).

b. Ensemble Kalman filter parameter estimation

Next, we examine how EnKF methods are used to estimate model parameters, where we use \( \mathbf{p} \) to represent the model parameter state (likewise \( \mathbf{p}^b \) represents the background model parameter state and \( \mathbf{p}^f \) represents the analysis model parameter state). We use state augmentation methods (Aksoy et al. 2006; Hacker and Angevine 2013; Koyama and Watanabe 2010; Ruiz et al. 2013), where these methods augment an ensemble of background states \( \{ \mathbf{z}^b_i; i = 1, 2, \ldots, k \} \) with an ensemble of background parameter states \( \{ \mathbf{p}^b_j; j = 1, 2, \ldots, k' \} \) to produce an ensemble of augmented analysis states \( \{ (\mathbf{z}^a_i, \mathbf{p}^a_j); i = 1, 2, \ldots, k; j = 1, 2, \ldots, k' \} \). Note that we will use the same number of ensemble members for both state and parameter estimation (\( k = k' \)), although it is not essential to do so; in fact, for some situations it may be desirable to use more ensemble members for the state estimation than for the parameter estimation (Ruiz et al. 2013).

The parameter estimation methods we consider can be divided into two types, which we refer to as simultaneous and separate methods. In simultaneous parameter estimation the \( N \)-dimensional state and \( N_p \)-dimensional parameter background vectors are augmented to produce a \( N + N_p \) length vector \( \{ (\mathbf{z}^b_i, \mathbf{p}^b_j); j = 1, 2, \ldots, k \} \), and then an EnKF data assimilation scheme updates the state and parameter from \( (\mathbf{z}^a_i, \mathbf{p}^a_j) \) (Anderson 2001; Aksoy et al. 2006; Hacker and Angevine 2013). In separate parameter estimation, first the state is updated from the background state \( \{ \mathbf{z}^b_i; j = 1, 2, \ldots, k \} \), and then the parameters are updated in a separate step from the augmented state \( \{ (\mathbf{z}^a_i, \mathbf{p}^a_j); j = 1, 2, \ldots, k \} \) (Koyama and Watanabe 2010; Ruiz et al. 2013). This augmentation allows the DA method to update the parameter the same as it would any unobserved state variable, by using the background error covariance (Baek et al. 2006). We will utilize the separate method in this work; although this type is more computationally burdensome (as discussed below), it has been shown to more significantly reduce estimation error (Ruiz et al. 2013).

In the separate parameter estimation approach from Koyama and Watanabe (2010) and Ruiz et al. (2013), either the EnKF or the LETKF is used to update the state, and then an EnKF without localization is used to update a global parameter (\( N_p = 1 \)). Evidence is also provided that using localization when updating a global parameter offers little benefit in reducing analysis error.

Our parameter estimation approach is to use the LETKF, thus localization, to both produce an ensemble of analysis states \( \{ (\mathbf{z}^a_i); j = 1, 2, \ldots, k \} \) updated from \( \{ \mathbf{z}^b_i; j = 1, 2, \ldots, k \} \) and to produce an ensemble of parameter states \( \{ (\mathbf{p}^a_j); j = 1, 2, \ldots, k \} \) updated from \( \{ \mathbf{p}^b_j; j = 1, 2, \ldots, k \} \). Specifically, the update of this augmented background state produces an ensemble of analysis states each in \( \mathbb{R}^{N_a + N_p} \), where the final part of each analysis state within \( \mathbb{R}^{N_p} \) is the update of each ensemble...
member of parameter states \( \{ (p^j) : j = 1, 2, \ldots, k \} \), where each \( p^j \in \mathbb{R}^{N_p} \).

We use this methodology to estimate nonglobal parameters to determine how localization in the parameter update step will reduce parameter estimation error. Of note, our methodology is more computationally efficient than using an ETKF without localization, since each localized analysis update uses fewer observations, resulting in less computational complexity.

Recall the localization radius \( d \) for the LETKF. We represent the localization radius used in the analysis update as \( d = r_e \) and the localization radius used in the parameter update as \( d = r_p \). Thus, at the \( i \)th grid point, the LETKF update of \( z_i^d \) will only use observations within a radius of \( r_e \). Similarly, for the \( i \)th spatial location of the augmented state \( (z_i^b, p_i^b) \), only observations within a radius of \( r_p \) are used to update the \( i \)th spatial location of the augmented state.

One key investigation of our research is to determine the optimal value of the localization radius \( r_p \) for the parameter estimation step relative to the system dimension for different types of spatially and temporally varying parameters. Accordingly, we vary \( r_p \) in our experiments to determine the effect it has on parameter and forecast error. Of note, in this work we focus primarily on the effect of localization on the parameter estimation, hence, the localization radius \( r_e \) for the state estimation is held constant for similar experiments.

Within the LETKF methodology, the computational time scales (at most) linearly to the size of the state space (Hunt et al. 2007). If the simultaneous method has a computational time of \( t_{sim} \), then all things being equal, the separate method has a computational time of \( t_{sep} \), where

\[
t_{sep} = \frac{2N + N_p}{N + N_p} t_{sim},
\]

recalling \( z \in \mathbb{R}^N \) and \( p \in \mathbb{R}^{N_p} \). Thus, for a global parameter \( (N_p = 1) \) the separate method takes less than twice as long as the simultaneous method. When the parameter space is the same as the state space \( (N_p = N) \), the separate method only takes 50% more computational time over the simultaneous method. Additionally, the computational time for the separate method is about 50% longer for a parameter that spatially varies over the whole state space \( (N_p = N) \) than if the parameter is global \( (N_p = 1) \).

c. Model problems

To investigate the efficacy of our LETKF parameter estimation techniques for determining nonglobal parameters, we will use a modified version of the well-known Lorenz-96 model (Lorenz 1996) with spatially varying forcing. The Lorenz-96 model is a system of ordinary differential equations (ODEs) that governs the time evolution of \( N \) periodic points,

\[
dX_i/dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F, \quad X_{i+N} = X_i,
\]

where we fix \( N = 40 \). Equation (15) was formulated to model certain aspects of the atmosphere, where the nonlinear terms model advection and conserve the total energy (Lorenz 1996). The linear terms dissipate the total energy and \( F \) represents external forcing, which strongly determines chaotic properties (Lorenz 1996). Within this model we include a spatially varying forcing term:

\[
F := F(i) = F_0 + \alpha \sin(2\pi\beta i/N), \quad i \pm N = i,
\]

where \( \alpha \) determines the amplitude of the forcing and \( \beta \) determines the wavelength. We choose the above form for the forcing in Eq. (16) for both simplicity and similarity to a Fourier sine series; we do not claim that it is representative of any specific atmospheric forcing. However, there are numerous examples in the literature of similar modifications to conceptual atmospheric models; Palmer (1999) modifies the Lorenz-63 by adding a sinusoidal forcing term, and both Orrell (2003) and Mitchell and Gottwald (2013) use modified versions of the Lorenz-96 model as testbeds for experiments in predictability and data assimilation.

To test our techniques on unobserved physics and implicit parameters in atmospheric climatology, we simulate such processes in two ways. First, we consider a fast–slow variant of the Lorenz-96 model consisting of \( N \) slow variables \( X_i \) that are coupled to \( M \times N \) fast variables \( Y_{j,i} \):

\[
\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F - \frac{hc}{b} \sum_{j=1}^{M} Y_{j,i},
\]

\[
X_{i+N} = X_i,
\]

\[
\frac{dY_{j,i}}{dt} = cb(Y_{j-1,i} - Y_{j+2,i})Y_{j+1,i} - cY_{j,i} + \frac{hc}{b} X_i,
\]

\[
Y_{j,N} = Y_{j,i}.
\]

This model is used to mimic the climatology of the atmosphere, where the fast components \( Y_{j,i} \) model weather properties of the atmosphere, which have an influence on the slowly evolving components \( X_i \), modeling properties of the climate. To study a sufficiently chaotic system (Lieb-Lappen and Danforth 2012), we
consider Eq. (17) with $N = 16$ slow components and $M = 8$ distinct fast components coupled to each slow component, as illustrated in Fig. 1. The parameter $h$ determines the level of coupling between the models ($h = 0$ leads to no coupling), $c$ determines the time-scale separation ($c = 1$ leads to variables on the same time scale), and $b$ scales the magnitude of the fast components. A common formulation is $h = 1$, $c = 10$, and $b = 10$, which induces leading-order coupling, where the fast variables $Y_{j\mu}$ are 10 times faster with a magnitude 10 times smaller than the slow variables $X_i$.

We also consider a model intermediate to the systems defined in Eqs. (15) and (17), which defines a stochastic forcing component to simulate the fast component in Eq. (17b). Whereas Eq. (17) exhibits bidirectional coupling between the subsystems in Eqs. (17a) and (17b) at different time scales, the stochastic forcing is unidirectional with only the slow component affected by the addition of a stochastic forcing term, simulating stochastic effects of subgrid-scale physics. We define each spatial component by

$$dX_i = dt[(X_{i+1} - X_{i-2})X_i + F(i)] + \sigma dW_i,$$  

where $F(i)$ is defined in Eq. (16) and corresponds to $F_1$ in Table 1. The term $W_i$ is standard Brownian motion (i.e., a standard Wiener process). For brevity, let $a[X(t)] = [X_{i+1}(t) - X_{i-2}(t)]X_i(t) + F(i)$ and $\beta(t) = \sigma$. Then the standard global Lipschitz condition necessary for strong convergence of the Euler–Maruyama algorithm,

$$|a[X(t)] - a[X(t')]| + |\beta(t) - \beta(t')| \leq C|X_i(t) - X_i(t')|,$$  

will be satisfied for some $C > 0$. That is, the approximate solution will converge in expectation to the true solution with rate $C\Delta t^{1/2}$ for sufficiently small time step $\Delta t$ (Higham et al. 2003). Observe that in terms of the stochastic variable $W_i$ we compute a discretized Brownian path with increments $W(\tau_j) - W(\tau_{j-1})$. A partition of the time interval $[0, T]$ into $\kappa \in \mathbb{N}$ subintervals gives a discrete value for the steps of $\Delta t$ that the set of points at which we evaluate the Brownian path is contained in the collection of points $\{t_i\}$ at which the Euler–Maruyama is evaluated.

![Fig. 1. Illustration of the coupling topology for the Lorenz-96 fast–slow model for $N = 16$ and $M = 8$.](image-url)

**Table 1.** Experimental design for nonglobal forcing terms.

| No. | Experiment type                        | Forcing term                                      |
|-----|----------------------------------------|--------------------------------------------------|
| 1   | Spatially varying, $F = F(i)$          | $F_1(i) = F_0 + \alpha \sin(2\pi \beta i/N)$    |
| 2   | Spatially and temporally varying, $F = F(i, t)$ | $F_2(i, t) = F_0 + \nu t + \alpha \sin(2\pi \beta i/N)$ |
| 3a  | Missing physics, slow–fast deterministic, $F = F(i, j)$ | $F_{3a}(i, j) = F_1(i) - \frac{hc}{b} \sum_{j=1}^{N} Y_{j\mu}$ |
| 3b  | Missing physics, stochastic, $F - N(\mu, \sigma)$ | $F_{3b}(i) = F_1(i) + \sigma dW_i$ |
The quantity $W_t$ approximates the influence of coupling unobserved physics evolving on a fast time scale, and contained in spatially localized weather, on climatic changes modeled by the slow variables $X_t$. The magnitude of the influence due to the unobserved properties is determined by the deterministic parameter $\sigma$, which corresponds to the coupling magnitude $b$ in Eq. (17). In line with the fast–slow system in Eq. (17), we study a stochastically forced system of $N = 16$ (slow) variables where the evolution of each $X_t(t)$ is influenced by a predetermined number of neighbors, as in Eq. (15), along with an additional stochastic forcing element of unknown strength. This amounts to performing data assimilation on a noisy truth for which spatially localized physics affect the evolution of the system. It is worth being precise about the parameter $\sigma$: while we consider parameter estimation within the context of a stochastic model, the parameter $\sigma$ that we are estimating is itself deterministic [see DelSole and Yang (2010) and Yang and DelSole (2009) for more details on the pitfalls of estimating stochastic parameters].

Table 1 summarizes our three experimental designs.

d. Error and skill evaluation

To evaluate our localized parameter estimation methods when applied to each of the above model problems, we examine the root-mean-square (RMS) error of both the estimated state and the estimated parameter. We define the RMS error for the LETKF parameter estimation strategy as

$$\text{RMSE}(t) = \sqrt{\frac{\sum_{i=1}^{N} [Z(t) - \bar{Z(t)}]^2}{N}},$$  \hspace{1cm} (20)

where $Z$ is the referenced truth, $\bar{Z}$ is the mean of the analysis ensemble, and $N$ is the number of grid points. Here, $Z$ can refer to either the state or the parameter. Additionally, we denote the RMS error of the LETKF without localization, that is, the ETKF, as $\text{RMSE}_0$. With this, we define the skill of the LETKF parameter estimation strategy as

$$\gamma(t) = \frac{\text{RMSE}_0(t) - \text{RMSE}(t)}{\text{RMSE}_0(t)}. \hspace{1cm} (21)$$

Thus, Eq. (21) provides a nondimensionalized, normalized quantity with which to evaluate our localized methods for each model problem. When $\gamma = 1$, the localized scheme perfectly models the true solution (or true parameter), and when $\gamma = 0$, the localized scheme provides no improvement over an analysis by the ETKF. Note that while this skill measure is bounded above by $\gamma = 1$, no such lower bound exists. In the next section we will use this skill measure to quantify the improvements made by our nonglobal parameter estimation strategy in determining parameters that vary in both space and time.

3. Results

a. Spatially varying forcing

Examining the system in Eq. (15), we investigate how well the local parameter estimation methodology estimates a spatially nonuniform but temporally uniform parameter, that is $F = F(i)$. To calibrate the ensemble size we first perform 50 independent experiments where we use an ETKF with no localization or inflation, and estimated a global parameter $F_0 = 8$. Error bars show the standard deviation of the mean value over 50 independent realizations.

FIG. 2. RMS errors in the state estimate as a function of ensemble size $k$. We have used an ETKF with no localization or inflation, and estimated a global parameter $F_0 = 8$. Error bars show the standard deviation of the mean value over 50 independent realizations.

Next, we calibrate the state estimation inflation and localization parameters $\delta$ and $r_z$ for our experiments on the uncoupled Lorenz-96 system in Eq. (15) when $N = 40$. We perform similar experiments as above for $F_0 = 8$, and vary both $\delta$ and $r_z$. The results for both state and parameter estimation are shown in the contour plots in Figs. 3a and 3b, respectively. Based on these plots we use a global covariance inflation factor of $\delta = 1.05$ and state localization radius of $r_z = 8$. Furthermore, in all experiments for the uncoupled Lorenz-96 system, we take 20 random observations of the state with Gaussian noise of variance 0.05, roughly 5% of the climatic variance.
As a first example of how the parameter estimation performs, Fig. 4 shows examples of the quality of the parameter estimation for a parameter localization radius $r_p = 5, 15,$ and 20 (the latter equivalent to estimating the parameters using the ETKF). This figure gives the final estimate for the spatially varying forcing $F(i)$ where forcing amplitude $a = 2$, wavelength $b = 1$, and $F_0 = 8$ after 1500 analysis steps, as averaged over 50 realizations. The ETKF ($r_p = 20$) approximately recovers the true forcing mean of $F_0 = 8$, however, being a global method, it cannot recover the local spatial variation in $F$. When $r_p = 15$ the filter is unable to reproduce either the correct mean or spatial variation in $F$ due to spurious long-range correlations between sites.

As Fig. 5 shows, this filter divergence for $r_p = 15$ occurs for a range of values of the forcing amplitude $a$ and will be discussed further. We have checked that the filter divergence is due to finite ensemble size effects; it does not occur for this value of $r_p$ when using the larger ensemble size of $k = 50$. Finally, we remark that while it appears in Fig. 4 that the average analysis estimate of $F(i)$ is biased toward a lower mean value $F_0$ for $r_p = 15$, the standard deviation of forcing values across realizations is in fact of the same order of magnitude as the estimates themselves.

We now make this comparison of results for different localization radii $r_p$ clearer by showing the skill improvement factor $g$ for all possible localization radii $r_p$. Figure 5 shows how well the localized method performs at estimating the state and nonlocal forcing parameters as a function of the forcing amplitude $a$ and parameter ensemble localization radius $r_p$. Contours in Fig. 5a show skill improvement for estimates of the state $z$, while the contours in Fig. 5b show the skill improvement for estimates of the parameter $F(i)$ in Eq. (16). The vertical plots to the right of each contour plot show this RMS error of the ETKF as a function of $a$. As the variation in the forcing increases with increasing amplitude $a$, progressively smaller local regions must be used in order to obtain accurate estimates of the parameters $F(i)$, these being a substantial improvement over the global ETKF method. Furthermore, the increase in errors from increasing the localization radius $r_p$ is nonmonotonic. We see a band of localization radii centered around $r_p = 14$ where the filter diverges, and errors are much greater than they are for smaller or larger values of $r_p$. This band

![Fig. 3. Contour plots showing RMS errors as a function of localization radius and multiplicative inflation factor for (a) state estimation and (b) parameter estimation. We have used an ETKF with $k = 20$ ensemble members and the results are averaged over 50 independent realizations.](image)

![Fig. 4. Final estimates of $F(i)$ after 1500 analysis cycles for $r_p = 5$ (circles), 15 (triangles), and 20 (squares). We have set $a = 2, \beta = 1$ and $F_0 = 8$, and averaged over 50 realizations. The true forcing $F(i)$ is visually indistinguishable from the curve with $r_p = 5$.](image)
increases in width and shifts to smaller $r_p$ as $\alpha$ increases. The effect is more pronounced for the parameter estimation than for the state estimation.

We remark that this filter divergence is a finite ensemble size effect; increasing the ensemble size to $k = 50$ greatly reduces the number of divergences; however, the general trends with $\alpha$ and $r_p$ are preserved. The range of localization windows over which parameters can be well estimated grows smaller as the parameter amplitude $\alpha$ increases. This is intuitive, as the variation in forcing $F(i)$ becomes greater as $\alpha$ increases, meaning that only the information coming from observations in increasingly local regions will be useful. If we instead hold the forcing amplitude constant at $\alpha = 2$ and vary $\beta$ we observe similar trends (not shown) as in Fig. 5, where the skill passes through a local minimum and then a local maximum as $r_p$ increases from 1 up to 20 but little variation in this trend as $\beta$ varies.

Finally, in Fig. 6, for $\alpha = 2$ and $\beta = 1$, we show how the skill depends on the number of observations taken. Figure 6a shows contours of skill for state estimates as a function of localization radius $r_p$ and number of observations, while Fig. 6b shows the same for parameter estimates. As expected, when fewer observations are available, it is necessary to use a smaller parameter localization radius $r_p$, as fewer observations, on average, will be spread farther apart in this experiment. Thus, a larger localization radius will mean that individual sites are informed to a greater degree by distant observations, thereby corrupting the analysis. As in Fig. 5 there is a band of localization radii for which the filter diverges; however, we remark that there is a slight

![Figure 5](image-url)  
**Fig. 5.** Contours of skill improvement $\gamma$ for (a) state estimation and (b) parameter estimation as a function of localization radius $r_p$ and parameter amplitude $\alpha$. The insets to the right of each contour plot show the RMS errors for the ETKF, RMSE$_0$, used for the normalization of $\gamma$. The dashed line shows $\gamma = 0$.

![Figure 6](image-url)  
**Fig. 6.** As in Fig. 5, but as a function of both parameter localization radius $r_p$ and number of observations, for $\alpha = 2$ and $\beta = 1$. 
difference between the state and parameter estimation problems. Whereas for the state estimation problem the divergence is worst around $r_p = 16$, for the parameter estimation the divergence is centered around $r_p = 14$. This could be due to the fact that the observations used are direct measurements of the state, whereas they only implicitly contain information about the parameters.

b. Spatially and temporally varying forcing

Next, again for the system in Eq. (15), we investigate how well the LETKF estimates a parameter that varies both spatially and temporally, as described in Table 1 for the second experiment where $F = F(i, t)$. Accordingly, we vary Eq. (16) so that the mean value $F_0$ increases at a constant rate of 1 forcing unit per five units of system time by setting $\gamma = 0.2$ (see experiment 2 in Table 1), and set $\alpha = 2, \beta = 1$.

Figure 7 plots the effect of the localization radius on parameter estimation error, where each figure results from the average of 100 similar experiments. Figure 7a shows time series estimates of $F_1 = F(1, t)$ for various localization radii. As $F_0$ is continually increasing, all estimates of the forcing $F(i, t)$ eventually diverge from the true value, regardless of the parameter localization $r_p$. Therefore, in Fig. 7b we also show the average time until this divergence occurs as a function of $r_p$, where the time until divergence is defined as the time when the parameter RMS errors first exceed the threshold value of 3. The inset to Fig. 7b shows the average RMS error for $r_p \approx 5$, showing a clear minimum when $r_p = 2$. Time series estimates of $F_1$ for $i \neq 1$ look qualitatively similar to that in Fig. 7a for $F_1$ over the time period shown. As for the case of the spatially varying parameter, we find that it is necessary to use a small localization window to accurately estimate the parameter over long times. For $r_p \leq 5$ the analysis estimate of $F(i, t)$ tracks the true parameter value very well, but diverges from the truth for larger localization windows after increasingly short times. Increasing the size of the ensemble increases the time until filter divergence for all values of $r_p$; however, we still observe the same trend shown here that filtering with smaller $r_p$ increases the length of fidelity of the experiment.

c. Missing physics problem

All experiments so far have been performed in the perfect model scenario, where the truth is produced by a run of the same model as used in the LETKF forecasting step. This is, of course, overly optimistic—in an operational situation one is forced to use inadequate models that cannot hope to accurately represent the full atmospheric dynamics, either due to coarse resolution or missing physics (Carrassi and Vannitsem 2011). Now, we use the uncoupled Lorenz-96 system in Eq. (15) to model the true state, where we define the true state to be a solution to the coupled fast–slow Lorenz-96 system in Eq. (17a). The model will be missing physics from the additional coupled fast term in Eq. (17a), and observations of the truth will include the effects of this additional fast term. Thus, when the parameter estimation scheme updates the forcing parameter, it will actually be estimating the forcing parameter plus the additional physics from this fast term.

We again use the local parameter estimation methodology to estimate the spatially varying forcing parameter described in Table 1 under experiment 3a where

$$F_{3a}(i,j) = F_1(i) - \frac{hc}{b} \sum_{j=1}^{M} Y_{ij},$$
with \( \alpha = 2 \) and \( \beta = 1 \). To balance low RMS errors and filter stability for the coupled Lorenz-96 system with \( N = 16 \), we have calibrated the ensemble size as \( k = 20 \), the multiplicative inflation factor as \( \rho = 1.05 \), and the localization radius for the state analysis is fixed as \( r_x = 3 \). This calibration has been done in a similar manner as described in section 3a and in Figs. 2 and 3 for the uncoupled Lorenz-96 with \( N = 40 \).

We vary the localization radius for the parameter estimation; recall we have fixed the dimension of the slow variable as \( \mathbf{X} \in \mathbb{R}^{16} \), so a localization radius of \( r_p = 8 \) will observe the whole space at every analysis step, thus removing any actual localization and reducing the scheme to the ETKF. We again take randomly located observations of the state at 50% of the state space. Hence, there are eight observations of \( \mathbf{X} \) at each time step, where each observation has Gaussian noise of variance 0.05.

Figure 8 demonstrates how well the nonglobal forcing parameter and the state may be estimated as a function of the time-scale separation factor \( c \) and the localization radius \( r_p \), where we fix the coupling parameter \( h = 1 \) and \( b = 10 \). As in Figs. 5 and 6, Fig. 8a shows errors in the state estimation, and Fig. 8b shows errors in the parameter estimation. This figure shows skill improvement \( \gamma \) contours for estimates of the forcing parameter \( F(i,j) \), described in Table 1, and the resulting state estimates. In every scenario in Fig. 8, \( \gamma > 0 \), thus the localization strategy offered by the LETKF always outperforms the global ETKF strategy. When estimating the forcing parameter for small time-scale separations, there exists a sweet spot for the localization radius \( r_p \) where we find that the skill improvement in the estimated forcing is highest for \( r_p \in [3, 6] \). As seen in both contours, the error is smallest and the skill is highest when the fast and slow components are on the same time scale. As the time-scale separation becomes larger, the contribution from the fast component to the slow component acts increasingly like noise, making the state and the forcing parameter more difficult to predict. Additionally, we see that a smaller localization radius is more optimal as the time-scale separation \( c \) increases. This is natural for an increased time-scale separation, since the contribution of the fast components will behave similarly to noise and the filter will determine decreased slow component structure from observations located farther away.

In Fig. 9, we fix \( b = 10 \) and \( c = 10 \) to induce the coupled Lorenz-96 system as explained in section 2c, where the fast terms are an order of magnitude smaller and an order of magnitude faster than the slow components. We vary the coupling parameter \( h \) and parameter ensemble localization radius \( r_p \) to determine the optimal localization radius for various levels of coupling. For \( h = 0 \), the contribution from the fast component will always be zero, so in this case the model and the truth are both just the uncoupled Lorenz-96 system as in sections 3a and 3b. As \( h \) increases, the contribution from the fast components will increase. Figure 9a shows the skill improvement contour for estimates of the state \( \mathbf{X} \) and Fig. 9b shows the skill improvement contour for estimates of the forcing parameter

\[
F_{3a}(i,j) = F_{i}(i) - \frac{hc}{b} \sum_{j=1}^{M} Y_{ij}.
\]

As before, the localization scheme always shows improvement over the ETKF. We find that the skill is highest in the estimate of the forcing parameter near \( r_p = 5 \). Additionally, for \( r_p = 5 \), the state estimate has the highest skill when \( h = 0.2 \). In general, skill is higher for
smaller localization radii versus larger localization radii, but we see a sweet spot for smaller levels of couplings where the skill is highest approximately for \( r_p \in [3, 6] \). Again, as expected, the error is smallest when the coupling from the fast component is smallest. As the coupling increases, the contribution from the fast component increases, thus resulting in less accuracy in the estimate of the state and the estimate of the forcing parameter.

d. Stochastic model

As described in section 2, we approximate the contribution of the fast component of a fast–slow system by generating a “noisy” truth using Eq. (18), where now we are estimating the forcing described in Table 1 under experiment 3b as \( F_{3b} = F_1(i) + \sigma dW_t \). The strength of the noise is determined by the nonstochastic parameter \( \sigma \) in Eq. (18), and is analogous to the coupling parameter, \( h \), in Eq. (17). We hold the forcing parameters constant at \( \alpha = 2 \) and \( \beta = 1 \) and vary the localization radius and magnitude of the stochastic component \( \sigma \). Thus, as with the fast–slow system in Eq. (17), when we update the parameters we are simulating the incorporation of unobserved, small-spatial-scale physics.

Figure 10a shows the skill improvement in the estimated state while Fig. 10b shows the skill improvement for estimates of the forcing parameter \( F(i) \) as the localization radius is increased. Overall, for a fixed \( \sigma \), a smaller localization radius yields more accurate estimates, due to the additive effects of the stochastic component in each \( X_i \) in Eq. (18). But as before, we see a sweet spot in the localization radius \( r_p \), where a localization radius spanning about half of the phase space, in particular \( r_p \in [2, 5] \), is most effective at increasing the skill in the estimation of \( F(i) \) for a fixed value of \( \sigma \). In this figure, the skill remains relatively constant for a wide range of localization windows as well. Except for fairly high noise levels, the LETKF performs much better than the global ETKF, until larger localization radii above \( r_p = 6 \), where the accuracy of the localization scheme decreases markedly.

FIG. 9. As in Fig. 8, but as a function of localization radius \( r_p \) and coupling parameter \( h \).

FIG. 10. As in Fig. 8, but for the estimated forcing parameter for the stochastic truth defined in Eq. (18), measured as a function of the localization radius \( r_p \) and the magnitude of the noise \( \sigma \) in the stochastic truth.
In Fig. 10 as the strength of the noise increases, the ability to predict the forcing parameter decreases. This is similar to how increasing the coupling parameter \( h \) in the fast–slow system in Eq. (17) leads to more error as seen in Fig. 9. In both experiments \( \gamma > 0 \) for localization radii \( r_p < 8 \), showing that the localization technique provides improved estimates compared to the nonlocal procedure.

4. Discussion

In this article we have presented a methodology for estimating nonglobal parameters in numerical models from observational data using the local ensemble transform Kalman filtering (LETKF) technique. The efficacy of this method is compared to an existing global parameter estimation (ETKF) technique by applying both methods to a low-dimensional conceptual atmospheric system exhibiting chaotic dynamics. Using this model we have examined the performance of the local parameter estimation method for three different types of nonglobal parameters, namely spatially varying parameters, spatially and temporally varying parameters, as well as parameters representing unobserved physics. For the third set of experiments we have considered two different types of unobserved physics, provided first by the dynamics of a fully unobserved fast subsystem with coupling to the slow observed system; and second by a stochastic process with spatial and temporal correlations.

In most cases the local method is significantly more skillful than the global method at estimating both the parameter and the state, with a general feature being that the global method will accurately estimate the parameter mean but not spatial deviations from this mean. The localization radius \( r_p \) has been found to be a significant parameter; while there are a wide range of values for which the LETKF method produces improvements over the ETKF method, the optimal localization radius depends upon spatial characteristics of the parameter being estimated. In particular, smaller localization radii are required as the spatial variation of the parameter is increased, and as the number of available state observations is decreased. We have found that often there exists a sweet spot for the localization radii, larger than one and less than half the size of the state space, where we find the most skill in parameter estimation. We have obtained similar results for both model error experiments; in both cases, as the unobserved physics in the nature run becomes noisier—either by directly increasing the time-scale separation between the slow and the fast systems or the stochasticity—it is required to make the parameter localization radius smaller to compensate.

Since this local parameter estimation method performs better than the global method for a low-dimensional conceptual model, it is of interest to investigate whether the localized method will offer improvement over the nonlocalized method for a more realistic, higher-dimensional model. In the future, we plan to investigate this by applying these methods in order to estimate parameters within an intermediate-complexity global circulation model (GCM) such as the simplified parameterization, primitive equation dynamics model (SPEEDY; Molteni 2003) or the Weather Research and Forecasting Model (WRF; Done et al. 2004). In such studies the shape of the localization region will become a consideration, as we expect that flow-dependent localization structures will be more desirable than cube-shaped ones. There may, in fact, be situations where it could be desirable for the localization radius \( r_p \) to vary in space and/or time in response to the local dynamics of the system. For example, in three dimensions it may become important to use a flow-dependent localization scheme (Bishop and Hodyss 2007) to respect the stratified nature of the (relatively flat) atmosphere, as well as the dynamics of the prevailing flow around orography. Computationally efficient, adaptive diagnostic tools such as bred vectors (Toth and Kalnay 1993) or singular vectors (Buizza and Palmer 1995), as well as the ensemble dimension (Patil et al. 2001), may be useful in such situations.

Acknowledgments. This work is supported by the National Science Foundation under the Grants DMS-0940363, DMS-0940314, and DMS-0940271. This work was performed in part using computational facilities at the College of William and Mary that were provided with the assistance of the National Science Foundation, the Virginia Port Authority, and Virginia’s Commonwealth Technology Research Fund. LM is grateful for the computational resources provided by the Vermont Advanced Computing Core, which is supported by NASA (NNX08A096G), and the Vermont Complex Systems Center. A portion of the work presented in this paper was carried out while JB was a postdoctoral fellow at the Institute for Mathematics and its Applications during the IMA’s annual program on Scientific and Engineering Applications of Algebraic Topology.

REFERENCES

Aksu, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Mon. Wea. Rev., 134, 2951–2970, doi:10.1175/MWR3224.1.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev., 129, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Wea. Rev., 129, 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DFFCBE>2.0.CO;2.

Higham, D. J., X. Mao, and A. M. Stuart, 2003: Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal., 40, 1041–1063, doi:10.1137/ S0036142901389530.

Houtekamer, P., and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUEKE>2.0.CO;2.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for stiopatemporal chaos: A local ensemble transform Kalman filter. Physica D, 230, 112–126, doi:10.1016/j.physd.2006.11.008.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng., 82, 35–45, doi:10.1115/1.3662552.

Kalnay, E., 2002: Atmospheric Modeling, Data Assimilation, and Predictability. Cambridge University Press, 364 pp.

Kang, J. S., E. Kalnay, J. Liu, and I. Fung, 2011: “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation. J. Geophys. Res., 116, D09110, doi:10.1029/2010JD014673.

Koyama, H., and M. Watanabe, 2010: Reducing forecast errors due to model imperfections using ensemble Kalman filtering. Mon. Wea. Rev., 138, 3316–3323, doi:10.1175/2010MWR3067.1.

Lieb-Lappen, R. M., and C. M. Danforth, 2012: Aggressive shadowing of a low-dimensional model of atmospheric dynamics. Physica D, 241, 637–648, doi:10.1016/j.physd.2011.12.001.

Lorenz, E. N., 1996: Predictability: A problem partly solved. Proc. Seminar on Predictability, Vol. I. Reading, United Kingdom, ECMWF, 1–19.

Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 2001: A mathematical framework for stochastic climate models. Commun. Pure Appl. Math., 54, 891–974, doi:10.1002/cpa.1014.

Mitchell, L., and G. A. Gottwald, 2013: Controlling model error of underdamped forecast models in sparse observational networks using a variance-limiting Kalman filter. Quart. J. Roy. Meteor. Soc., 139, 212–225, doi:10.1002/qj.1968.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. Climate Dyn., 20, 175–191, doi:10.1007/s00382-002-0268-2.

Orrell, D., 2003: Model error and predictability over different timescales in the Lorenz 96 systems. J. Atmos. Sci., 60, 2219–2228, doi:10.1175/1520-0469(2003)060<2219:MEAPOD>2.0.CO;2.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. Tellus, 56A, 415–428, doi:10.1111/j.1600-0870.2004.00076.x.

Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. J. Climate, 12, 575–591, doi:10.1175/1520-0442(1999)012<0575:ANPDPO>2.0.CO;2.

——, 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. Quart. J. Roy. Meteor. Soc., 127, 279–304, doi:10.1256/qj.00.72.

Patil, D., B. R. Hunt, E. Kalnay, J. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. Phys. Rev. Lett., 86, 5878–5881, doi:10.1103/PhysRevLett.86.5878.

Peacock, S., 2012: Projected twenty-first-century changes in temperature, precipitation, and snow cover over North America in CCSM4. J. Climate, 25, 4405–4429, doi:10.1175/JCLI-D-11-00214.1.
Randall, D., M. Khairoutdinov, A. Arakawa, and W. W. Grabowski, 2003: Breaking the cloud parameterization deadlock. Bull. Amer. Meteor. Soc., 84, 1547–1564, doi:10.1175/BAMS-84-11-1547.

Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. Quart. J. Roy. Meteor. Soc., 133, 129–146, doi:10.1002/qj.23.

Ruiz, J. J., M. Pulido, and T. Miyoshi, 2013: Estimating model parameters with ensemble-based data assimilation: A review. J. Meteor. Soc. Japan, 91, 79–99, doi:10.2151/jmsj.2013-201.

Solonen, A., P. Ollinaho, M. Laine, H. Haario, J. Tamminen, and H. Järvinen, 2012: Efficient MCMC for climate model parameter estimation: Parallel adaptive chains and early rejection. Bayesian Anal., 7, 715–736, doi:10.1214/12-BA724.

Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature, 433, 403–406, doi:10.1038/nature03301.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. Bull. Amer. Meteor. Soc., 74, 2317–2330, doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Urban, N. M., and K. Keller, 2010: Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: A Bayesian fusion of century-scale observations with a simple model. Tellus, 62A, 737–750, doi:10.1111/j.1600-0870.2010.00471.x.

Xu, K. M., and D. A. Randall, 2001: Updraft and downdraft statistics of simulated tropical and midlatitude cumulus convection. J. Atmos. Sci., 58, 1630–1649, doi:10.1175/1520-0469(2001)058<1630:UDSOTC>2.0.CO;2.

Yang, X., and T. DelSole, 2009: Using the ensemble Kalman filter to estimate multiplicative model parameters. Tellus, 61A, 601–609, doi:10.1111/j.1600-0870.2009.00407.x.