Nonlinear conditional model bias estimation for data assimilation

Article

Accepted Version

Otkin, J. A., Potthast, R. W. E. ORCID: https://orcid.org/0000-0001-6794-2500 and Lawless, A. S. ORCID: https://orcid.org/0000-0002-3016-6568 (2021) Nonlinear conditional model bias estimation for data assimilation. SIAM Journal on Applied Dynamical Systems, 20 (1). pp. 299-332. ISSN 1536-0040 doi: https://doi.org/10.1137/19M1294848 Available at https://centaur.reading.ac.uk/93076/

To link to this article DOI: http://dx.doi.org/10.1137/19M1294848

Publisher: Society for Industrial and Applied Mathematics

NONLINEAR CONDITIONAL MODEL BIAS ESTIMATION FOR DATA ASSIMILATION*

JASON A. OTKIN [†‡], ROLAND W. E. POTTHAST [‡§], AND AMOS S. LAWLESS [‡¶‖]

Abstract. In this study, we develop model bias estimators based on an asymptotic expansion of the model dynamics for small time scales and small perturbations in a model parameter, and then use the estimators to improve the performance of a data assimilation system. We employ the well-known Lorenz (1963) model so that we can study all aspects of the dynamical system and model bias estimators in a detailed way that would not be possible with a full physics numerical weather prediction model. In particular, we first work out the asymptotics of the Lorenz model for small changes in one of its parameters and then use statistics from cycled data assimilation experiments to demonstrate that the asymptotics accurately represent the behavior of the model and that the coefficients of the nonlinear asymptotical expansion can be reasonably estimated by solving a least squares minimization problem.

In data assimilation, the background error covariance matrix usually estimates the uncertainty of the model background, which is then used along with the observation error covariance matrix to produce an updated analysis. If the uncertainty of the model background is strongly influenced by time-dependent model biases, then the development of nonlinear bias estimators that also vary with time could improve the performance of the assimilation system and the accuracy of the updated analysis. We demonstrate this improvement through the combination of a constant background error covariance matrix with a dynamically-varying matrix computed using the model bias estimators. Numerical tests using the Lorenz (1963) model illustrate the feasibility of the approach and show that it leads to clear improvements in the analysis and forecast accuracy.

Key words. Variational data assimilation, asymptotic expansion, model error, parameter estimation

AMS subject classifications. 34A55, 65K10, 34E05

1. Introduction.

Partial differential equations are widely used in scientific and technological fields to simulate the evolution of natural phenomena. For initial boundary condition problems such as those that are commonly encountered in atmospheric science, an accurate prediction of the spatial and temporal characteristics of various weather and climate features depends not only on the ability of a numerical model to realistically simulate the physical processes controlling their evolution, but also on the ability of a data assimilation system to provide the forecast model an accurate estimate of the initial conditions. Atmospheric data assimilation systems typically combine information from a short-range model forecast, or ensemble of forecasts, with a set of observations gathered over a specified time period to produce an analysis of the current state of the dynamical system that then serves as the initial conditions for the next model forecast. Commonly used data assimilation methods include variational assimilation that determines the analysis through minimization of

*Submitted to the editors on 21 October 2019. Revised version submitted on 07 March 2020.

Funding: The lead author was partially supported by a University of Reading International Research Studentship. The third author was funded in part by the NERC National Centre for Earth Observation.

[†]Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin-Madison, Madison, WI, USA (jasono@ssec.wisc.edu, https://www.ssec.wisc.edu/~otkin/)

[‡]Department of Mathematics and Statistics, University of Reading, United Kingdom (r.w.e.potthast@reading.ac.uk, a.s.lawless@reading.ac.uk)

[§]German Meteorological Service - Deutscher Wetterdienst, Offenbach, Germany

[¶]Department of Meteorology, University of Reading, United Kingdom

[‖]National Centre for Earth Observations, University of Reading, United Kingdom

a cost function, ensemble methods that use an ensemble of forecasts to dynamically estimate the sample covariances between different state components when determining how new observations impact the ensemble analysis, and so-called hybrid methods that combine aspects of variational and ensemble data assimilation methods. A wide range of literature is known today introducing and studying different data assimilation methods, see for example [46, 36, 20, 3, 75, 62, 38, 52, 33, 10].

Regardless of which assimilation methodology is employed, generation of the best possible analysis state $x^{(a)}$ through combination of the model first guess or background state $x^{(b)}$ with the available observations requires knowledge of the observation error and the underlying uncertainty in the model background $x^{(b)}$. The observation error uncertainty is usually determined by the covariance matrix $R \in \mathbb{R}^{m\times m}$ of the observation vector $y \in \mathbb{R}^m$ in observation space $\mathbb{R}^m$, where $m \in \mathbb{N}$ denotes the number of observations. The uncertainty of the model background state $x^{(b)}$ is measured by the covariance matrix $B \in \mathbb{R}^{n\times n}$, where $n \in \mathbb{N}$ is the dimension of the model space $\mathbb{R}^n$ and $x^{(b)} \in \mathbb{R}^n$. Variational data assimilation methods calculate the analysis state $x^{(a)} \in \mathbb{R}^n$ by minimizing the functional

$$J(x) := ||x - x^{(b)}||^2_{B^{-1}} + ||y - H(x)||^2_{R^{-1}}, \quad x \in \mathbb{R}^n, y \in \mathbb{R}^m \tag{1.1}$$

where $H : \mathbb{R}^n \to \mathbb{R}^m$ is the forward observation operator that maps the model state x into the simulated observation $H(x) \in \mathbb{R}^m$. For linear observation operators, it is well-known (c.f. [52], Chapter 5) that the minimization of (1.1) is given by

$$x^{(a)} = x^{(b)} + BH^T(R + HBH^T)^{-1}(y - H(x^{(b)})). \tag{1.2}$$

Because the model background and observations are not perfect, accurate knowledge of the covariance matrices B and R is very important for data assimilation since they determine the weights that are applied to the model background and observations, respectively, when generating the updated analysis $x^{(a)}$. In addition, the matrix B spreads information spatially within a region surrounding the observation location and can also be used to add balance constraints between analysis variables based on physical principles [8, 9].

Despite its importance, an exact form for B cannot be determined for real-world applications because the true state of the dynamical system cannot be completely known due to a limited number of observations and the presence of errors in the observations that are available. For variational assimilation systems, the model background error covariances are often computed using the so-called National Meteorological Center (NMC) method that was first described by [58]. This method estimates B using differences between forecasts of different lengths valid at the same time. For example, forecast errors could be assessed by examining differences between 24 and 48 hour forecasts from model integrations initialized one day apart. These difference fields are usually obtained from a large collection of model forecasts covering time periods of a month or longer. As such, the NMC method generates a climatological estimate of B that may not properly represent the true model errors on any given day due to changes in the atmospheric conditions. Because of this, some operational weather forecasting centers have developed new methods to generate B. One approach is to use an ensemble of data assimilations (EDA) where an ensemble of reduced-resolution data assimilation cycles is performed in which the observations and model are perturbed in some manner. A theoretical analysis by [35] has shown that if the perturbations are drawn from the true distributions of observation and model errors, that the spread in the resultant EDA analyses about the unperturbed control analysis will be representative of the background error. This approach has the advantage of introducing some flow-dependency to the B matrix, thereby allowing it to better capture the errors of the day ([13, 35, 61]).

Ensemble data assimilation systems such as the ensemble Kalman filter (EnKF) (e.g. [20, 18, 28, 21, 29, 4, 80, 72, 30, 32]) on the other hand re-compute B for each assimilation step using an estimator based on output from an ensemble $x^{(b,\ell)}$ of forecasts valid at the current analysis time, where $\ell = 1, ..., L$, and L is the size of the ensemble. For most applications, the standard stochastic estimator

$$(1.3)\qquad B := \frac{1}{L-1}\sum_{\ell=1}^{L}(x^{(b,\ell)} - \bar{x}^{(b)})(x^{(b,\ell)} - \bar{x}^{(b)})^T, \qquad \bar{x}^{(b)} := \frac{1}{L}\sum_{\ell=1}^{L} x^{(b,\ell)},$$

is used to compute the first guess ensemble mean $\bar{x}^{(b)}$ and the background error covariance matrix B. The stochastic estimator includes the uncertainty of the previous model analysis propagated to the current analysis time. Because the forecast model is an approximation of the real dynamical system, the distribution of the first guess ensemble could be sub-optimal due to the impact of systematic errors on the ensemble mean and ensemble spread. Similar problems can arise when using the NMC method because in situations where the model error varies with time, the forecast differences used to compute the covariances in B will include the dynamically-varying model bias. This could result in incorrect statistical relationships between the model variables. Both of these outcomes are not desirable because the inclusion of systematic model errors when generating B can degrade the accuracy of the posterior analysis $x^{(a)}$ obtained during the data assimilation step.

Various studies have focused on improving methods to estimate the background error covariances used by modern data assimilation systems; however, accounting for model error is challenging because of the large size of geophysical models [16]. One approach is to add perturbations to a subset of the model variables, such as temperature, at the initialization time, whereas another technique adds random perturbations to specific parameters in the parameterization schemes used to simulate sub-grid scale processes during each model time step. The goal with both approaches is to increase the range of possible forecast solutions to realistically address the impact of systematic model errors and the underlying uncertainties in the parameterization schemes. Substantial research has been directed toward development of these methods, which have the potential to greatly improve the performance of assimilation systems [14, 79, 63, 25, 71, 11]. As a corollary to the above approaches, other studies have shown that the detrimental impact of systematic model errors in ensemble data assimilation systems can be reduced by using different parameterization schemes in each ensemble member [49, 22].

Another approach widely used in ensemble data assimilation systems to increase the ensemble spread is to apply additive or multiplicative covariance inflation during the assimilation step. Some amount of covariance inflation is often necessary because the rank deficiency of the system can lead to an underestimation of the ensemble variance and because systematic model errors can cause the model background $x^{(b)}$ to deviate greatly from reality. This in turn can lead to so-called filter divergence where the model analyses can no longer be pulled toward the observations during the data assimilation step [33]. In the case of additive covariance inflation, the impact of the unknown model error is treated by drawing random perturbations from some distribution and then adding them to either the model background $x^{(b)}$ or to the model analysis $x^{(a)}$. With multiplicative covariance inflation, the ensemble spread for selected model variables is multiplied by a real number to achieve the desired ensemble spread. Both methods have some adaptivity because observation-minus-background (OMB) statistics are used to estimate how much inflation is necessary. There is a very active community working on these approaches, see for example [26, 5, 6, 31, 44, 43, 51, 81].

Model error has often been ignored in variational data assimilation systems because it is difficult to quantify and has been viewed as having a minor impact compared to random errors in the initial conditions and systematic errors in the observations [15]. Unlike ensemble assimilation systems where the background error covariance matrix B is dynamically estimated for each assimilation cycle using the ensemble output, additional statistical or dynamical assumptions are generally necessary when creating these estimates for variational systems. Studies by [17, 84, 77, 73, 74] have shown that treating the model error as part of the state estimation problem substantially improves the accuracy of the state estimates. Theoretical model error frameworks were developed by [24, 54, 55, 56] based on the behavior of model errors in deterministic models. These frameworks were then used by [15] to derive evolution equations for the model error covariances and correlations that address errors due to parameterization schemes.

The desire to properly account for model error also underpins recent efforts to move from "strong-constraint" 4-dimensional variational systems that assume the forecast model is perfect to "weak-constraint" systems that include some estimate of the model error. This concept was introduced 50 years ago by [69], however, it was not implemented in a full-physics forecast model for several decades because of the lack of information with which to define and solve the problem and the computational burden associated with inverting the model error covariance matrix along with the other matrices already included in the strong-constraint formulation [53]. The basic premise behind the weak-constraint approach is that it is sufficient to only approximately satisfy the model equations because they are not exact anyway due to incomplete knowledge of the physical processes being modeled or the need to simplify the governing equations due to computational limitations. Despite the challenges associated with implementing weak-constraint systems, their use has generally led to more accurate model analyses and forecasts when compared to strong-constraint systems due to the higher number of degrees of freedom. As such, they are becoming more widely used in variational data assimilation systems [74, 45, 53]. A recent study by [34] has also shown that model errors can be accounted for in strong-constraint systems by allowing errors in both the model and the observations when considering the statistics of the innovation vector. They demonstrate that a more accurate estimate of the model state can be obtained when the combined model and observation error statistics are used instead of the standard observation-only error statistics.

In this paper, we seek to extend our understanding of how to identify and treat model bias in modern data assimilation systems. Key tasks of this research include: 1) studying the behavior of model errors in a nonlinear dynamical system, 2) developing nonlinear conditional model bias estimators using the observations and the model first guess and analysis states, and 3) employing these estimators during variational data assimilation experiments to assess their ability to improve the performance of the system. Numerical experiments are performed using the Lorenz-63 (L63) model [47], which is a well-known and popular study object within the data assimilation and dynamical systems communities.

We begin by carrying out an asymptotic analysis of the L63 model when one of its parameters, in this case, the normalized Rayleigh number ρ, varies with time. In the L63 model, the ρ parameter is usually set to a constant value; however, we allow it to vary with time in order to introduce a model bias. This is accomplished through use of a coupled version of the L63 model where a background or "hidden" system $S2$ is used to control how the $\rho = \rho(t)$ parameter changes with time in the "primary" system $S1$ that is used to represent the truth. Though we chose to focus on variations in the ρ parameter during this study, the approach works in the same way for the other L63 model parameters. We then develop a nonlinear model bias estimator method based on the initial ideas discussed in [57] where the bias estimator is formulated as a polynomial expansion of the model variables and the coefficients of this expansion are determined by solving a least-squares minimization problem. The ability of this method to dynamically estimate the model error contribution to the matrix B and to improve the resultant OMB statistics is demonstrated by carrying out an experiment where B is represented as the sum of static and dynamically-varying components. Finally, we demonstrate the feasibility and potential utility of the asymptotic expansion and nonlinear bias estimation method by running numerical experiments using a 3-dimensional variational (3DVAR) data assimilation system and a coupled version of the L63 model.

A description of the coupled L63 modeling system and derivation of the model asymptotics are provided in Section 2. The utility of dynamically estimating the model background error covariance matrix B is discussed in Section 3, along with development of nonlinear conditional model error estimators. We then perform various numerical experiments using the L63 model in Section 4, first demonstrating the validity of the asymptotic expansion of the nonlinear model error estimators in Section 4.1. This is followed by a study of the optimality of the fixed and dynamic components of the B matrix used during data assimilation and then a study of the estimation of the nonlinear model error dynamics based on the first guess minus analysis statistics. Results from these sections will demonstrate the feasibility of using methods developed during this study to estimate nonlinear model errors without any prior knowledge or assumptions regarding the form of the model dynamics. Conclusions are presented in Section 5.

2. Estimating System Bias.

2.1. Coupled Lorenz 1963 Model. We want to use a relatively simple atmospheric model to assess the behavior of nonlinear model biases and to develop ways to take into account those biases in a way that is complex enough to represent nonlinear atmospheric processes while being simple enough to provide insight into the nonlinear behavior of the system. To accomplish this goal, we have chosen to employ the L63 model [47], which is widely used within the atmospheric data assimilation community (see for example [19, 76, 78, 59, 2, 15, 41, 27, 42, 67, 82, 83, 48, 23]) because it is less complex than a full physics numerical weather prediction model while maintaining strong nonlinearity representative of many atmospheric processes. The L63 model consists of a set of three coupled ordinary differential equations that provide a simplified description of dry convection. The model equations can be written as:

$$\tau \frac{dx_1}{dt} = \sigma(x_2 - x_1) \tag{2.1}$$

$$\tau \frac{dx_2}{dt} = \rho x_1 - x_2 - x_1 x_3 \tag{2.2}$$

$$\tau \frac{dx_3}{dt} = x_1 x_2 - \beta x_3 \tag{2.3}$$

where $x_1(t)$, $x_2(t)$, and $x_3(t)$ are the dependent variables, τ is a temporal scaling factor and σ, ρ, and β are the parameters of the model. For some parameter values, the system shows chaotic behavior because very small perturbations in the initial conditions can grow very rapidly into completely different solutions. The model was designed to simulate atmospheric dry cellular convection following the work of [68]. The model simulates the evolution of a forced dissipative hydrodynamic system that possesses non-periodic and unstable solutions. The x_1 variable measures the intensity of convective motion, the x_2 variable measures the temperature difference between the ascending and descending currents, and the x_3 variable measures the distortion of the vertical temperature profile from linearity. The model parameters represent the Prandtl number (σ), a normalized Rayleigh number (ρ), and a non-dimensional wave number (β). The critical Rayleigh number for the system is 24.74; however, ρ is typically set to the slightly supercritical value of 28 following the work of [47]. The σ and β parameters are set to 10 and 8/3, respectively. Together, the values for these three parameters sustain the chaotic nature of the model.

In this study, we investigate the sensitivity of the L63 model to perturbations in the ρ parameter and identify suitable predictors that can be used to estimate conditional biases in the state variables (x_1, x_2, x_3) due to these perturbations. We first generate a nature or "truth" simulation that tracks the evolution of the state variables over a certain period of time. The truth simulation is generated using a particular model for the behavior of ρ over time. Here, we choose to use a coupled version of the L63 model where each system $(S1, S2)$ is run at a different speed and one-way coupling occurs through the influence of $S2$ on the ρ parameter in $S1$, as is illustrated in Fig. 1. After some experimentation, we decided to set $\tau_{S1} = 1$ and $\tau_{S2} = 5$, which means that the hidden system $S2$ is integrated forward at one-fifth the speed of $S1$.

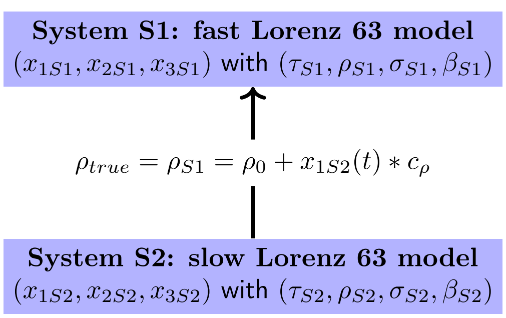

Fig. 1. *Coupled version of the Lorenz-63 model, with the fast system $S1$ dependent on the slow system $S2$. $S1$ is used to generate the nature simulation.*

The state location x_{S2} obtained from the hidden system is then scaled by a factor of $c_\rho = 0.2$, with the scaled value subsequently used to perturb ρ_0, such that

$$\rho_{S1} = \rho_0 + x_{1S2}(t) \cdot c_\rho, \quad t \in \mathbb{R}, \tag{2.4}$$

where $\rho_0 = \rho_{S2} = 28$, $x_{1S2}(t) \cdot c_\rho$ is the $\delta\rho$ perturbation obtained from $S2$, and ρ_{S1} is the resultant value used when integrating $S1$ during the next time step. The scaling of x_{1S2} by $c_p = 0.2$ means that $\delta\rho$ varies between approximately -4 and +4, which is reasonable because this represents departures up to 15% from ρ_0. The slowly varying autocorrelated $\delta\rho$ perturbations could be thought of as representing changes in the original L63 model equations due to the influence of the seasonal cycle on daily forecasts or the diurnal cycle on hourly forecasts. For example, parameters in cloud microphysics parameterization schemes are often assigned constant values even though some of them are known to vary, sometimes by up to several orders of magnitude, depending upon the stage of the cloud's life cycle. A similar approach was used by [83], where they attached an ocean slab model to the L63 model equations in order to represent the interaction between the slowly-varying ocean and the rapidly-changing atmosphere. Note that the parameters σ_{S1} and σ_{S2} were set to 10, and β_{S1} and β_{S2} were set to $8/3$, as is typically done in the L63 model.

II. After generating the "truth" simulation using $S1$ in which the ρ_{S1} parameter varied with time, observations were generated for each state variable $(x_{1S1}, x_{2S1}, x_{3S1})$ at each model time step and then used in cycled data assimilation experiments employing a 3DVAR data assimilation system. The truth simulation and data assimilation experiments were started with the same initial conditions $(x_{1S1}, x_{2S1}, x_{3S1}) = (2, 3, 11)$; however, in the absence of data assimilation, they will follow different trajectories thereafter due to differences in the ρ parameter. The L63 model is integrated to the next time step using a 4th order Runge-Kutta time integration scheme. Various tests were performed using different observation error magnitudes and time step lengths, as will be shown in Section 4. Figure 2 shows the trajectory of the model state variables and evolution of the ρ_{S1} parameter during the truth simulation.

The data assimilation experiments employed the typical L63 model equations, including $\rho = 28$; however, for these experiments, the equations represent an imperfect model because we know that ρ is not constant during the truth simulation. Let us assume that we know that ρ varies with time, but that we only know its mean value $(\bar{\rho} = 28)$ and not how it changes with time. The instantaneous difference between ρ in the data assimilation experiment and ρ in the truth simulation represents a model error; however, these differences correspond to conditional model biases when assessed over long time periods because $\delta\rho$ is a function of $S2$. Because errors in ρ directly impact the evolution of all three of the state variables in nonlinear ways, the instantaneous errors will potentially result in biases in the model state variables that are a nonlinear function of one or more predictors when assessed over long time periods. For example, a numerical weather prediction model may have the tendency to produce convection that is too strong during the day or too weak during the night, both of which will impact the sign and magnitude of the model biases in nonlinear ways during different parts of the diurnal cycle.

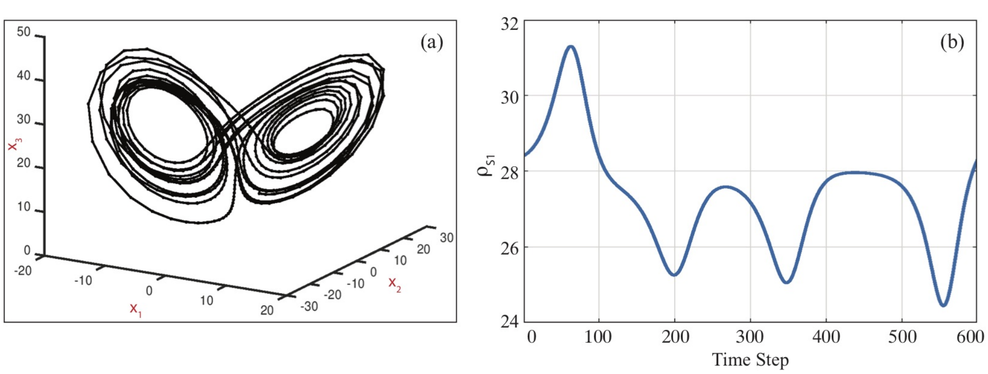

FIG. 2. *(a) Butterfly diagram showing the model trajectory during 600 time steps of the truth simulation using the coupled L63 system described in Section 2.1. (b) Time series showing the evolution of the ρ_{S1} parameter during the truth simulation, where ρ_{S1} for each model time step is set using (2.4).*

2.2. Asymptotics for Model Error of the Lorenz 1963 System. Here, we first evaluate how the model variables (x_1, x_2, x_3) change in dependence on the model parameter ρ. In particular, we aim to develop an asymptotic estimator for the error in (x_1, x_2, x_3) depending on ρ and time t. The asymptotic analysis is performed using a Taylor series expansion with an explicit integral form of the error term. This approach is necessary because some of the constants will be zero in the higher order terms; therefore, we need to take sufficiently many terms into account to get the correct higher order terms.

THEOREM 2.1. *The leading terms of the asymptotic analysis of the L63 system with respect to variations of $\rho = \rho_0 + \delta\rho$, where we use $t = t_0 + \delta t$ and $O(s)$ denotes a function bounded by $c|s|$ with some constant c in a neighborhood of $s = 0$, are given by*

$$x_1(\rho, t) - x_1(\rho_0, t) = \frac{1}{2}\sigma x_1(\rho_0, t_0) \cdot \delta\rho \cdot (\delta t)^2 + O(\delta\rho \cdot \delta t^3) \tag{2.5}$$

$$x_2(\rho, t) - x_2(\rho_0, t) = x_1(\rho_0, t_0) \cdot \delta\rho \cdot \delta t + O(\delta\rho \cdot \delta t^2) \tag{2.6}$$

$$x_3(\rho, t) - x_3(\rho_0, t) = x_1^2(\rho_0, t_0) \cdot \delta\rho \cdot (\delta t)^2 + O(\delta\rho \cdot \delta t^3) \tag{2.7}$$

Proof. We work out the proof in four steps, starting with some general setup and then considering the variables x_1, x_2, and x_3 in three steps.

Step 1. We begin by differentiating the equations (2.1) - (2.3) with respect to ρ using the product rule, where

$$x_1' = \frac{dx_1}{d\rho}, \quad x_2' = \frac{dx_2}{d\rho}, \quad x_3' = \frac{dx_3}{d\rho}$$

are the derivatives of the state variables with respect to ρ. Because the differentiation with respect to t and to ρ can be exchanged in the case of continuously differentiable functions, we obtain:

$$\frac{dx_1'}{dt} = \sigma x_2' - \sigma x_1' \tag{2.8}$$

$$\frac{dx_2'}{dt} = x_1'\rho + x_1 - x_2' - x_1' x_3 - x_1 x_3' \tag{2.9}$$

$$\frac{dx_3'}{dt} = x_1' x_2 + x_1 x_2' - \beta x_3'. \tag{2.10}$$

Note that all of the variables depend on time t and the parameter $\rho = \rho(t)$, and that the τ terms in equations (2.1) - (2.3) have been set to 1 to represent the original L63 model equations as described in [47].

To assess the sensitivity of the L63 model equations to variations in ρ at times t close to some initial time, t_0, we begin by looking at the scenario where the initial values for (x_1, x_2, x_3) are prescribed and identical for all ρ under consideration, such that at $t = t_0$:

$$x_1(\rho, t_0) = x_{1,0} \tag{2.11}$$

$$x_2(\rho, t_0) = x_{2,0} \tag{2.12}$$

$$x_3(\rho, t_0) = x_{3,0}. \tag{2.13}$$

This is an initial value problem where the derivatives of each equation with respect to ρ, (x_1', x_2', x_3'), are equal to zero at $t = t_0$, i.e.

$$x_1'(\rho, t_0) = 0, \quad x_2'(\rho, t_0) = 0, \quad x_3'(\rho, t_0) = 0. \tag{2.14}$$

After inserting these initial values into (2.8) - (2.10), we obtain:

$$\frac{dx_1'}{dt}(\rho, t_0) = 0 \tag{2.15}$$

$$\frac{dx_2'}{dt}(\rho, t_0) = x_1(\rho, t_0) \tag{2.16}$$

$$\frac{dx_3'}{dt}(\rho, t_0) = 0 \tag{2.17}$$

Step 2. Equation (2.9) reveals that the time rate of change of the sensitivity of x_2 with respect to ρ (i.e., x_2') is a function of its location along the x_1-axis. We now carry out an asymptotic analysis by an expansion of the functions with respect to variations in time $t = t_0 + \delta t$ and the parameter $\rho = \rho_0 + \delta\rho$. To assess the sensitivity of x_2 with respect to small variations in ρ, we employ (2.14) and (2.16) as follows. We estimate

$$\begin{aligned}
x_2(\rho, t) - x_2(\rho_0, t) &= \int_{\rho_0}^{\rho} x_2'(\tilde\rho, t)\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \Big(\underbrace{x_2'(\tilde\rho, t_0)}_{=0} + \int_{t_0}^{t} \frac{dx_2'(\tilde\rho, \tilde t)}{d\tilde t}\, d\tilde t \Big)\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \int_{t_0}^{t} \frac{dx_2'(\tilde\rho, \tilde t)}{d\tilde t}\, d\tilde t\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \int_{t_0}^{t} \Big(\underbrace{\frac{dx_2'(\tilde\rho, \tilde t)}{d\tilde t}|_{t_0}}_{=x_1(\tilde\rho, t_0)} + \int_{t_0}^{\tilde t} \frac{d^2 x_2'(\tilde\rho, s)}{ds^2} ds \Big)\, d\tilde t\, d\tilde\rho. \qquad (2.18)
\end{aligned}$$

We estimate both terms in (2.18) separately. For the first term T_1, by (2.11) we obtain

$$\begin{aligned}
T_1 &= \int_{\rho_0}^{\rho} \int_{t_0}^{t} x_1(\tilde\rho, t_0)\, d\tilde t\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \int_{t_0}^{t} x_1(\rho_0, t_0)\, d\tilde t\, d\tilde\rho \\
&= x_1(\rho_0, t_0) \cdot \delta\rho \cdot \delta t, \qquad (2.19)
\end{aligned}$$

where $x_1(\tilde\rho, t_0)$ is replaced by $x_1(\rho_0, t_0)$ because the derivative of x_1 with respect to ρ is zero at t_0 following (2.14). The $\delta\rho$ and δt terms are obtained by solving the definite integrals, with $\delta\rho$ denoting the interval $[\rho_0, \rho]$ and δt denoting the interval $[t_0, t]$. The second term is estimated in a similar way by

$$\begin{aligned}
T_2 &= \int_{\rho_0}^{\rho} \int_{t_0}^{t} \int_{t_0}^{\tilde t} \frac{d^2 x_2'(\tilde\rho, s)}{ds^2}\, ds\, d\tilde t\, d\tilde\rho \\
&= O(\delta\rho \cdot \delta t^2). \qquad (2.20)
\end{aligned}$$

Combining the estimates (2.19) and (2.20) then leads to

$$x_2(\rho, t) - x_2(\rho_0, t) = x_1(\rho_0, t_0) \cdot \delta\rho \cdot \delta t + O(\delta\rho \cdot \delta t^2). \tag{2.21}$$

This proves equation (2.6) in Theorem 2.1.

Step 3. To obtain an estimate for $x_1(\rho, t)$, we proceed as in equation (2.18) and, for a twice continuously differentiable function $x_1(\rho, t)$, estimate

$$x_1(\rho, t) = x_1(\rho_0, t) + \int_{\rho_0}^{\rho} x_1'(\tilde{\rho}, t)\, d\tilde{\rho} \tag{2.22}$$

$$= x_1(\rho_0, t) + \int_{\rho_0}^{\rho} \Big(x_1'(\rho_0, t) + \int_{\rho_0}^{\tilde{\rho}} x_1''(\tilde{\tilde{\rho}}, t)\, d\tilde{\tilde{\rho}} \Big)\, d\tilde{\rho} \tag{2.23}$$

We note that by taking the derivative of (2.8) with respect to time and inserting (2.16) into the resultant equation, we obtain

$$\begin{aligned} \frac{d^2 x_1'(\rho, t_0)}{dt^2} &= \sigma \frac{dx_2'(\rho, t_0)}{dt} - \sigma \frac{dx_1'(\rho, t_0)}{dt} \\ &= \sigma x_1(\rho, t_0) \end{aligned} \tag{2.24}$$

and thus, the derivative of (2.24) with respect to time gives

$$\frac{d^2 x_1''(\rho, t_0)}{dt^2} = \sigma x_1'(\rho, t_0) = 0. \tag{2.25}$$

Performing a third order expansion around t_0 then leads to an estimate for $x_1''(\rho, t)$:

$$x_1''(\rho, t) = O(\delta t^3). \tag{2.26}$$

After inserting (2.26) into (2.23) and then solving the definite integrals, we obtain:

$$x_1(\rho, t) = x_1(\rho_0, t) + x_1'(\rho_0, t) \cdot \delta\rho + O(\delta\rho^2 \cdot \delta t^3) \tag{2.27}$$

To estimate $x_1'(\rho, t)$, with the help of (2.14) and (2.15), we derive:

$$\begin{aligned} x_1'(\rho, t) &= \underbrace{x_1'(\rho, t_0)}_{=0} + \int_{t_0}^{t} \frac{dx_1'(\rho, \tilde{t})}{d\tilde{t}}\, d\tilde{t} \\ &= \int_{t_0}^{t} \Big(\underbrace{\frac{dx_1'(\rho, \tilde{t})}{d\tilde{t}}\Big|_{t_0}}_{=0} + \int_{t_0}^{\tilde{t}} \frac{d^2 x_1'(\rho, s)}{ds^2}\, ds \Big)\, d\tilde{t}. \end{aligned} \tag{2.28}$$

The second derivative of $x_1'(\rho, t)$ with respect to time t can be estimated by differentiating (2.8) with respect to t, and using (2.9) and (2.16), which yields:

$$\begin{aligned} \frac{d^2 x_1'(\rho, t)}{dt^2} &= \frac{d}{dt}\Big(\frac{dx_1'(\rho, t)}{dt} \Big) \\ &= \frac{d}{dt}\Big(\sigma x_2'(\rho, t) - \sigma x_1'(\rho, t) \Big) \\ &= \sigma \frac{dx_2'}{dt}(\rho, t) - \sigma \frac{dx_1'}{dt}(\rho, t) \\ &= \sigma \frac{dx_2'}{dt}(\rho, t_0) - \sigma \underbrace{\frac{dx_1'}{dt}(\rho, t_0)}_{=0} + O(\delta t) \\ &= \sigma x_1(\rho, t_0) \ + \ O(\delta t). \end{aligned} \tag{2.29}$$

We insert this into (2.28) to conclude with

$$\begin{aligned}
x_1'(\rho,t) &= \sigma x_1(\rho,t_0)\cdot\int_{t_0}^{t}\int_{t_0}^{\tilde t} 1\; ds\; d\tilde t \;+\; O(\delta t^3)\\
&= \sigma x_1(\rho,t_0)\cdot\int_{t_0}^{t}(\tilde t - t_0)\; d\tilde t \;+\; O(\delta t^3)\\
&= \sigma x_1(\rho,t_0)\cdot\frac{1}{2}(\delta t)^2 \;+\; O(\delta t^3).\\
&= \sigma x_1(\rho_0,t_0)\cdot\frac{1}{2}(\delta t)^2 \;+\; O(\delta t^3). \qquad (2.30)
\end{aligned}$$

Finally, we insert (2.30) into (2.27) with the help of (2.14) to obtain (2.5) in Theorem 2.1.

Step 4. In our final step, we estimate the behavior of $x_3(\rho,t)$. We note that similarly to $x_1'(\rho,t)$ given by (2.30) as in (2.18) we obtain:

$$\begin{aligned}
x_2'(\rho,t) &= \underbrace{x_2'(\rho,t_0)}_{=0} \;+\; \int_{t_0}^{t}\frac{dx_2'(\rho,\tilde t)}{d\tilde t}\; d\tilde t\\
&= x_1(\rho,t_0)\cdot\delta t + O(\delta t^2). \qquad (2.31)
\end{aligned}$$

Also, based on (2.17) we calculate

$$\begin{aligned}
x_3'(\rho,t) &= \underbrace{x_3'(\rho,t_0)}_{=0} \;+\; \int_{t_0}^{t}\frac{dx_3'(\rho,\tilde t)}{d\tilde t}\; d\tilde t\\
&= \int_{t_0}^{t}\Big(\underbrace{\frac{dx_3'(\rho,\tilde t)}{d\tilde t}|_{t_0}}_{=0} \;+\; \int_{t_0}^{\tilde t}\frac{d^2x_3'(\rho,s)}{ds^2}\; ds\Big)\; d\tilde t\\
&= O(\delta t^2). \qquad (2.32)
\end{aligned}$$

Now, we follow the above lines to estimate

$$\begin{aligned}
x_3(\rho,t) - x_3(\rho_0,t) &= \int_{\rho_0}^{\rho} x_3'(\tilde\rho,t)\; d\tilde\rho\\
&= \int_{\rho_0}^{\rho}\Big(\underbrace{x_3'(\tilde\rho,t_0)}_{=0} \;+\; \int_{t_0}^{t}\frac{dx_3'(\tilde\rho,\tilde t)}{d\tilde t}\; d\tilde t\Big)\; d\tilde\rho \qquad (2.33)
\end{aligned}$$

Here, to obtain a sharper estimate than (2.32) and to evaluate the constant explicitly, we insert (2.10) into (2.33), which yields:

$$x_3(\rho,t) - x_3(\rho_0,t) = \int_{\rho_0}^{\rho}\int_{t_0}^{t}\Big(x_1'(\tilde\rho,\tilde t)x_2(\tilde\rho,\tilde t) + x_1(\tilde\rho,\tilde t)x_2'(\tilde\rho,\tilde t) - \beta x_3'(\tilde\rho,\tilde t)\Big)\; d\tilde t\; d\tilde\rho \qquad (2.34)$$

Because $(x_1', x_2', x_3') = 0$ at t_0, we need to estimate the leading order term by its temporal change at t_0 as given in (2.15) - (2.17). We insert the asymptotics for $x_1'(\rho,t)$, $x_2'(\rho,t)$, and $x_3'(\rho,t)$ given by (2.30), (2.31), and (2.32) into (2.34) to estimate:

$$\begin{aligned}
x_3(\rho,t) - x_3(\rho_0,t) &= \int_{\rho_0}^{\rho}\int_{t_0}^{t}\Big(x_1^2(\rho,t_0)\delta t + O(\delta t^2)\Big)\; d\tilde t\; d\tilde\rho\\
&= x_1^2(\rho,t_0)\cdot\delta t^2\cdot\delta\rho + O(\delta\rho\cdot\delta t^3), \qquad (2.35)
\end{aligned}$$

where $x_1^2(\rho, t_0)\delta t$ is the leading order term, and all other terms have been absorbed into the order $O(\delta t^2)$ term. Thus, we have derived (2.7) in Theorem 2.1 and the proof is complete. $\square$

Remark. In Step 3 of the proof, we could have performed the estimate slightly differently. Using an approach similar to Steps 2 and 4, we obtain:

$$
\begin{aligned}
x_1(\rho,t) - x_1(\rho_0,t) &= \int_{\rho_0}^{\rho} x_1'(\tilde\rho,t)\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \Big(\underbrace{x_1'(\tilde\rho,t_0)}_{=0} + \int_{t_0}^{t} \frac{dx_1'(\tilde\rho,\tilde t)}{d\tilde t}\, d\tilde t \Big)\, d\tilde\rho \\
&= \int_{\rho_0}^{\rho} \int_{t_0}^{t} \Big(\underbrace{\frac{dx_1'(\tilde\rho,\tilde t)}{d\tilde t}\Big|_{t_0}}_{=0} + \int_{t_0}^{\tilde t} \frac{d^2 x_1'(\tilde\rho,s)}{ds}\, ds \Big)\, d\tilde t\, d\tilde\rho \qquad (2.36)
\end{aligned}
$$

and then proceed as in (2.29) and (2.30) to obtain (2.5) as above. $\square$

3. Improving Data Assimilation using Bias Estimators. Being able to accurately estimate errors in the model background $x^{(b)}$ is important for any practical implementation of a data assimilation algorithm. In this section, we first discuss the model error and model bias terminology and then study a simple Bayesian example to illustrate the importance of correctly estimating the model background error covariance matrix B. We then develop a generalized model error estimation method that is subsequently applied to the L63 model discussed in Section 2.2 to demonstrate the feasibility of dynamically estimating the model errors using nonlinear estimators based on the model variables. In Section 3.4, we show how the bias correction coefficient vector obtained through solving a least squares minimization problem can be used to estimate the unknown parameter using the analysis increments from the data assimilation system.

3.1. Nonlinear Model Bias and Error Terminology. In this section, we sharpen the terminology for model error, model bias, and conditional model bias, and compare the concepts. For a particular location, the model error is the instantaneous difference between the background state $x^{(b)}$ and the true state $x^{(true)}$ of the system. Model bias is then defined as the $x^{(b)} - x^{(true)}$ differences averaged over some period of time or region:

$$b_b := \mathbb{E}\{x^{(b)} - x^{(true)}\}, \qquad (3.1)$$

where the bias is computed separately for different model quantities such as temperature, humidity, or cloud water path. If we then assume that the analysis state $x^{(a)}$ obtained during each assimilation cycle is the best estimate of the true system state, we can use the resultant $x^{(b)} - x^{(a)}$ differences as an approximation to the true model bias, with appropriate summation over particular regions or periods of time:

$$b_{b-a} := \mathbb{E}\{x^{(b)} - x^{(a)}\}. \qquad (3.2)$$

The conditional model bias can then be defined as the mean deviation of the dependent variable from the true system state when the bias is a function of some other parameter or variable p referred to as the predictor. The conditional model bias can be estimated using:

$$b_{b-a}(p) = \mathbb{E}\{x^{(b)}(p) - x^{(a)}(p)\}. \qquad (3.3)$$

For this study, we are interested in the situation where the bias predictor is a component of the model state.

If $b_{b-a}(p)$ varies in a nonlinear manner, then this behavior represents a nonlinear conditional bias and we will need to use nonlinear bias correction methods to remove the bias from the model variables. In this case, let us assume that the function $b_{b-a}(p)$ can be written as a superposition

$$b_{b-a}(p) = \sum_{\xi=1}^{N} \psi_\xi(p)\alpha_\xi \tag{3.4}$$

of nonlinear basis functions ψ_ξ with N unknown coefficients α_ξ. The solution of (3.4) can be understood as a generalized bias estimation equation because it structures the set of differences according to the predictor p and searches for a functional estimation of its behavior. We can then employ nonlinear bias correction methods such as that described in [57] to determine the bias correction coefficients based on a set of $b_{b-a}(p)$ differences. To do this effectively, we will need to obtain a large sample of differences covering a diverse range of system states.

It should also be noted that the estimation of the coefficients α_ξ in (3.4) using $x^{(b)} - x^{(a)}$ differences accumulated over multiple assimilation cycles subsequently leads to the capability to predict the instantaneous model error when those coefficients are applied to the current state during an individual assimilation cycle. This demonstrates that conditional model bias estimation and model error estimation are strongly related and show significant overlap. As discussed in Section 3.3, the forecast error in general can be represented as a combination of state estimation error associated with the propagation of errors in the prior analysis to the current time and a second component that represents the true model error arising from the use of an imperfect model. The instantaneous model errors can therefore be viewed as conditional model biases because their characteristics likely depend on the state of the system.

The conditional model error estimators can be used for various purposes, including a) model bias correction where the model background is corrected prior to its use in the data assimilation system, b) model uncertainty estimation where the model error estimates are used to improve the background error covariance matrix B, and c) model development efforts where the error statistics are used to improve the accuracy of the numerical model. In this paper, we focus on application b) because we seek to employ knowledge regarding the behavior of the model errors to improve estimates of the model background uncertainty.

3.2. Study of a Simple Bayesian Example. A Bayesian data assimilation step employs Bayes formula

$$p^{(a)}(x) = cp^{(b)}(x)p(y|x), \;\; x \in \mathbb{R}^n \tag{3.5}$$

for estimating the posterior probability distribution $p^{(a)}(x)$ based on the prior probability distribution $p^{(b)}(x)$ and the observation error distribution $p(y|x)$. The prior distribution is usually assumed to be Gaussian in data assimilation systems, such that:

$$p^{(b)}(x) := \tilde{c}e^{-\frac{1}{2}(x-x^{(b)})^T B^{-1}(x-x^{(b)})}, \;\; x \in \mathbb{R}^n, \tag{3.6}$$

where $\tilde{c}$ is a constant and the background error covariance matrix B is estimated climatologically in classical variational assimilation systems or based on an ensemble of model states in an EnKF.

Here, we discuss and demonstrate the role of the correct estimate of B on the quality of the analysis mean and analysis distribution. For an EnKF system, the ensemble spread is used to estimate B, however, this estimate only contains part of the error when a numerical model is used because it does not include the difference between the model and the true state of the system. Variational data assimilation systems, such as 3DVAR, are also unable to consider these differences because B is chosen as fixed for a particular time period due to the way in which it is constructed. This means that the model bias and how it changes with time is not taken into account by either assimilation methodology, which can substantially degrade their performance. For the remainder of this work, we restrict our attention to 3DVAR because that is what we used during the numerical experiments discussed in Section 4. We note however that similar arguments apply for ensemble and hybrid data assimilation systems.

As a starting point, we derive the error representation explicitly for a one-dimensional Gaussian case with observation operator $H = I$. In one dimension, the best estimate of the current state (or analysis) during an assimilation step is given by:

$$x^{(a)} = x^{(b)} + \frac{q}{r+q}(y - x^{(b)}), \tag{3.7}$$

where y is the observation, r is the observation error uncertainty, $x^{(b)}$ is the first guess or background, and q represents the estimated variance of the error in the variable x. Now, let us assume that q_0 is the true background error variance that includes model error, such that the correct analysis $x_0^{(a)}$ is represented as:

$$x_0^{(a)} = x^{(b)} + \frac{q_0}{r+q_0}(y - x^{(b)}) \tag{3.8}$$

The error between the analysis based on some uncertainty or variance q and the correct uncertainty or variance q_0 is then given by:

$$\begin{aligned} |x^{(a)} - x_0^{(a)}| &= \left| \frac{q}{r+q} - \frac{q_0}{r+q_0} \right| \cdot |y - x^{(b)}| \\ &= \left| \frac{r(q-q_0)}{(r+q)\cdot(r+q_0)} \right| \cdot |y - x^{(b)}|. \end{aligned} \tag{3.9}$$

This result shows that the analysis error for each assimilation cycle is proportional to the observation departure $|y - x^{(b)}|$ and to the accuracy of the background error variance estimate $|q - q_0|$. Thus, development of new methods that can be used to generate a more accurate estimate of q will directly improve the quality of the analysis and performance of the assimilation system.

3.3. Dynamical Error and Bias Estimators. In this section, we develop a generalized method to diagnose model biases using the model variables. First, let us assume that the the forecast error $x_k^{(b)} - x_k^{(true)}$ at a given time k can be represented as the difference between the dynamical states that are obtained when the prior analysis $x_{k-1}^{(a)}$ is propagated by an imperfect model M and the true prior state $x_{k-1}^{(true)}$ is propagated by the perfect model M^{true}:

$$x_k^{(b)} - x_k^{(true)} = M(x_{k-1}^{(a)}) - M^{true}(x_{k-1}^{true}) \tag{3.10}$$

The forecast error can then be decomposed into one part that is due to the propagation of the uncertainty error associated with the prior analysis state $M(x_{k-1}^{(a)}) - M(x_{k-1}^{true})$, and a second part that represents the true model error, $E = M(x_{k-1}^{true}) - M^{true}(x_{k-1}^{true})$, during the propagation from the prior time:

$$x_k^{(b)} - x_k^{(true)} = \Big(M(x_{k-1}^{(a)}) - M(x_{k-1}^{true})\Big) + \Big(M(x_{k-1}^{true}) - M^{true}(x_{k-1}^{true})\Big). \tag{3.11}$$

Taking the variance on both sides of (3.11), and using

$$q^{state} := \mathrm{Var}(M(x_{k-1}^{(a)}) - M(x_{k-1}^{true})) \tag{3.12}$$

and

$$q^{model} = \mathrm{Var}(M(x_{k-1}^{true}) - M^{true}(x_{k-1}^{true})), \tag{3.13}$$

we obtain the total variance of the forecast error:

$$q^{total} := \mathrm{Var}\Big(x_k^{(b)} - x_k^{(true)}\Big) \tag{3.14}$$

$$= q^{state} + q^{model} + 2 \cdot \mathrm{Cov}\Big(M(x_{k-1}^{(a)}) - M(x_{k-1}^{true}),\ \ M(x_{k-1}^{true}) - M^{true}(x_{k-1}^{true})\Big). \tag{3.15}$$

It is a standard approach in data assimilation to assume that the initial condition uncertainty and true model error are uncorrelated [50], which means that the covariance term on the righthand side of (3.15) will equal zero and therefore the total variance of the forecast error can be given by

$$q^{total} = q^{state} + q^{model}, \tag{3.16}$$

where q^{state} reflects the influence of the variance of the estimate of the prior analysis propagated to the current analysis time using the model equations, and q^{model} is the variance in the model error E due to the use of an imperfect numerical model.

If some error estimators such as those shown in Theorem 2.1 are available, we can employ (3.16) to estimate q^{total} and then use it to improve the estimate of the analysis during a given data assimilation step. Though we typically will not know q^{state} in a complex real-world system, the development of a method that can be used to estimate the time-varying model error E, and thus the variance q^{model}, allows us to employ a lower fixed q^{state} in our approach. This outcome is better than having to use a larger fixed q^{state}, which would otherwise be the case, because that would lead to an overestimate of the total error variance. In general, it will not be possible to carry out a full assessment of the model error due to incomplete knowledge of the governing equations; however, Theorem 2.1 shows that the model error asymptotically depends on the model variables, here in particular, $x_1(\rho_0, t_0)$. We can therefore employ nonlinear model error estimators to diagnose such dependencies as follows.

II. We begin with a general example where we study the estimation of a error that depends on the model state x and time t. We model the dependence on the states using basis functions $\varphi_\ell(x)$, $x \in \mathbb{R}^n$, with $\ell = 1, ..., N_\ell$. The dependence on time is modeled using basis functions ψ_k, $k = 1, ..., N_k$. Let us assume an ansatz of the form

$$E_j(x,t) = \sum_{\ell=1}^{N_\ell} \sum_{k=1}^{N_k} \beta_{\ell,k}^{(j)}\, \varphi_\ell(x)\psi_k(t), \ \ x \in \mathbb{R}^n,\ t \in \mathbb{R}, \tag{3.17}$$

for the model error E_j. For illustrative purposes, the functions $\psi_k(t)$ could be represented by $\sin(t)$ and $\cos(t)$ or by higher order trigonometric functions, whereas the functions $\varphi_\ell(x)$ could be represented by the polynomial terms in Theorem 2.1. In this situation, the terms would correspond to $\varphi_\ell(x) = x_1^{\xi_1} x_2^{\xi_2} x_3^{\xi_3}$, with ξ_1, ξ_2, ξ_3 counted by $\ell = 1, ..., N_\ell$, $\psi_1(t) \equiv 1$, and $\psi_k(t) = 0$ for $k > 1$. The coefficients $\beta_{\ell,k}^{(j)}$ are the unknown coefficients linking the true dynamics with the numerical model.

If we then observe the model error $E_j(x,t)$ for a selection of states $(x[\eta], t[\eta])$, $\eta = 1, ..., N_\eta$ such that the linear independence of φ_ℓ on $x[\eta]$ is satisfied and a set $t[\eta] \in [0,T]$ such that the linear independence of ψ_k is satisfied on this set, we know that the linear system

$$E_j(x[\eta]) = \sum_{\ell=1}^{N_\ell} \sum_{k=1}^{N_k} \beta_{\ell,k}^{(j)} \, \varphi_\ell(x[\eta]) \, \psi_k(t[\eta]), \tag{3.18}$$

$\eta = 1, ..., N_\eta$, has at most one solution for each $j = 1, ..., n$. It may be overdetermined if $N_\eta > N_\ell \cdot N_k$, and if the data is inconsistent would have no exact solution. In that case, we can use least squares methods to calculate approximate solutions.

Let us also discuss the case of non-uniqueness for the calculation of the bias correction coefficients. This situation can arise if two or more variables in the dynamical system under consideration are correlated. For example, the x_1 and x_2 variables in the L63 system display strong correlations in parts of the trajectory. Though the non-unique solution will not affect the quality of the bias estimate for the time interval used to calculate the coefficients, it could potentially lead to large errors if these coefficients are used outside of the training period. Thus, we note that: 1) for time-local estimation of model biases, the consequences of non-uniqueness should be small, and 2) when the bias estimation tool is employed for longer time periods or for forecasting, it is important to have training periods that include conditions representative of the full climatology of the dynamical model.

III. Here, we illustrate the utility of the generalized framework developed in the previous section by applying it to the L63 model. First, let us assume that the true evolution of a hypothetical dynamical system, represented by M^{true}, depends on a particular parameter that varies with time, but that limitations in our understanding of the physical system means that it is assigned a constant value in the imperfect numerical model M used to represent the true dynamical system. An example is the dependence of the parameter ρ in the coupled L63 model described in Section 2.1, for which we have worked out the behavior of the model error for small time intervals δt and small changes $\delta\rho$ of ρ in Section 2.2. For this particular system, we observe the dependence of the error

$$E(\delta\rho) := \|x[\rho] - x[\rho_0]\|^2 \tag{3.19}$$

on the model state $x = (x_1, x_2, x_3)$ in Theorem 2.1, where ρ_0 is the true value at a given time t_0 in M^{true} and ρ is the constant value used by the imperfect model M. This dependence leads to the error estimate for the coupled L63 system:

$$E(\delta\rho) = x_1^2(\rho_0, t_0) \cdot \delta\rho^2 \cdot \delta t^2 + O(\delta\rho^2 \cdot \delta t^4), \tag{3.20}$$

where we added the squares of (2.5), (2.6), and (2.7), and then absorbed the higher order terms into the $O(\delta\rho^2 \cdot \delta t^4)$ term. It can be seen in (3.20) that the leading error term is proportional to x_1^2, which means that the expected model error is largest when the system state is located near the tips of the butterfly wings.

For this work, we use the analysis $x^{(a)}$ from each assimilation step as an approximation of the true state $x^{(true)}$ because the true state is unknown in a real-world system. Note that this approximation means that we will be unable to recover the full model error; however, because $x^{(a)}$ will be pulled toward the observations, we will still be able to estimate part of the model error under the assumption that the observations have small errors. The current model error E_j of the component x_j of the state $x \in \mathbb{R}^n$ is approximated by:

$$E_j := |x_j^{(a)} - x_j^{(b)}|, \tag{3.21}$$

where $j = 1, 2, 3$ corresponds to the three variables in the L63 system. Let us assume that knowledge of those parts of the system leading to model error at a specific time is such that after some manipulation the model error can be rewritten in the form of a triple sum:

$$E_j = \sum_{\xi_1,\xi_2,\xi_3=0}^{N_{coef}} \alpha^{(j)}_{\xi_1,\xi_2,\xi_3} x_1^{\xi_1} x_2^{\xi_2} x_3^{\xi_3}, \tag{3.22}$$

with coefficients $\alpha^{(j)}_{\xi_1,\xi_2,\xi_3}$, $\xi_1, \xi_2, \xi_3 = 0, ..., N_{coef}$, where N_{coef} is the total number of coefficients determined by the maximum order of the polynomial and the number of model variables under consideration. For the L63 system containing three variables, $N_{coef} = 10$ for a 2nd order polynomial. The model error can be expressed as in (3.22) if we know that a hidden model exists but that we do not know the dependence of the true system because we cannot derive the asymptotics of the model equations. The ansatz (3.22) assumes some polynomial dependence of this relationship on the model variables $x \in \mathbb{R}^n$, as we have shown to be the case for the coupled L63 system. We also assume that the model errors do not have a temporal dependence such that the basis functions $\psi_k(t)$ in (3.17) can be set to 1.

Next, given a sequence of states $x[\eta]$ and their corresponding model errors $E_j[\eta]$ for $\eta = 1, ..., N_{states}$ over some period of time, the above estimate leads to a linear system of equations:

$$A\alpha^{(j)} = q \tag{3.23}$$

for the $N_{coef} \times 1$ coefficient vector $\alpha^{(j)} = (\alpha^{(j)}_{0,0,0}, \alpha^{(j)}_{1,0,0}, \alpha^{(j)}_{0,1,0}, \alpha^{(j)}_{0,0,1}, \alpha^{(j)}_{1,1,0}, ...)^T$, where the sub-indices correspond to the polynomial order for the predictors (x_1, x_2, x_3) and the superscript denotes the model variable x_j. For example, the zeroth order coefficient for the x_1 variable is denoted as $\alpha^{(1)}_{0,0,0}$, whereas the second order coefficient for the $x_1 \cdot x_2$ mixed term is denoted as $\alpha^{(1)}_{1,1,0}$. Then, A is an $N_{states} \times N_{coef}$ matrix containing the N_{coef} polynomial terms for each observation:

$$A = A^{(j)} := \Big(x_1^{\xi_1}[\eta] x_2^{\xi_2}[\eta] x_3^{\xi_3}[\eta] \Big)_{\eta=1,\ldots,N_{states};\ \xi_1,\xi_2,\xi_3=0,\ldots,N_{coef}} \tag{3.24}$$

where η counts the rows and ξ_1, ξ_2, ξ_3 are subsequently ordered as column indices consistent with the ordering of the components of α, and

$$q = q^{(j)} := \Big(E_j[\eta] \Big)_{\eta=1,\ldots,N_{states}} \tag{3.25}$$

is the $N_{states} \times 1$ vector containing the model errors, with row index η. Finally, we can find the coefficients α that best fit the system of equations by solving the quadratic minimization problem, which leads to:

$$\alpha = (A^T A)^{-1} A^T q. \tag{3.26}$$

3.4. Parameter Estimation. We begin this section by noting that the asymptotics for the coupled L63 model shown in Theorem 2.1 reveal that the error, E_j, for each model variable $j = 1, 2, 3$ is proportional to the size of the hidden parameter $\delta\rho$, which means that the diagnosed conditional model bias should also be proportional to this parameter. In practice, however, this is not an easy relationship to capture because their proportionality depends in a very dynamic way on the current state of a modeling system characterized by chaotic behavior. Thus, without explicit knowledge of the model variables and the relationship between them and $\delta\rho$, it is impossible to draw conclusions about the size of $\delta\rho$.

However, based on the nonlinear model error estimators given by (2.5) - (2.7), we expect that the coefficient vector α in (3.22) will also be proportional to the size of the model bias. This vector depends on the average size of the analysis increment $x^{(a)} - x^{(b)}$ during a sequence of data assimilation steps rather than on the model state. The explicit dependence, unknown in general, is part of the estimation of the coefficients. Thus, we obtain a tool that can be used to dynamically diagnose the average size of the unknown parameter $\delta\rho$ by computing the mean of the coefficient vector α for each model variable $x_j = 1, 2, 3$. This leads to the following estimates for $\delta\rho$:

(3.27)
$$\delta\rho^{(1)}_{diag}(t) \approx c_1 \alpha^{(1)}_{1,0,0}(t) \quad \text{or} \quad \delta\rho^{(2)}_{diag}(t) \approx c_2 \alpha^{(2)}_{1,0,0}(t) \quad \text{or} \quad \delta\rho^{(3)}_{diag}(t) \approx c_3 \alpha^{(3)}_{2,0,0}(t)$$

where $c_1 = 2/\sigma(\delta t)^2$, $c_2 = 1/\delta t$, and $c_3 = 1/(\delta t)^2$, and we now need to carry out the bias estimation over time intervals $[t - \Delta t, t + \Delta t]$ with some $\Delta t > 0$ for which $\delta\rho$ can be considered a constant.

Many prior studies have performed parameter estimation within data assimilation systems, primarily through use of an augmented state vector and based on statistical assumptions about the distribution of the model parameter ([7, 1, 40, 37, 12, 60, 66, 65, 70, 64, 39]). These studies have generally shown that reasonably accurate parameter estimates can be obtained if the data assimilation statistics are used to estimate a single model parameter. Unlike these previous studies, however, our approach uses the asymptotics of the model dynamics to provide a functional form for the relationship between the unknown model parameter and the estimated model error when accumulated over a sequence of assimilation cycles. We will demonstrate in Section 4.4 that this simple diagnostic tool provides a reasonable approach to parameter estimation for the dynamical system under consideration.

4. Numerical Results using the L63 Model. The purpose of this section is to use the L63 model to perform numerical experiments that demonstrate the validity of the model error identification and correction methods developed in the previous sections and their use within a data assimilation system. We begin by showing in Section 4.1 that the error asymptotics developed in Theorem 2.1 accurately represent the behavior of the L63 model and that they are able to capture the rapid evolution of the model error in each of the state variables. We then demonstrate in Section 4.2 that the model error asymptotics can be used to improve the model background error covariance matrix B through inclusion of a dynamic component that captures the current model errors. It is then shown in Section 4.3 that the coefficients of the nonlinear asymptotical expansion can be reasonably estimated by solving a regularized least squares minimization problem without explicit a priori knowledge of the error behavior. This is accomplished through use of a polynomial expansion of the model variables. Finally, we show in Section 4.4 that the ρ parameter can be reconstructed using the bias correction coefficient vector. Moreover, it is shown that it is possible to reconstruct this parameter using the analysis increments that are readily available in all data assimilation systems.

4.1. Analysis of the Asymptotic Error Estimators for the L63 Model. In this section, we assess the ability of the asymptotics derived in Theorem 2.1 to accurately capture the rapid evolution of model errors in the coupled L63 system during a cycled data assimilation experiment covering $N_t = 600$ assimilation cycles with an assimilation frequency $\delta t_{assim} = 0.06$. Though the true ρ parameter in the coupled L63 system varies with time following (2.4), it was set to a constant value ($\rho = 28$) during the data assimilation experiment to represent a dynamic and unknown model bias. Output from the truth simulation employing the time-varying ρ parameter was used to generate observations with zero measurement error ($\epsilon = 0$) for (x_1, x_2, x_3), which were then assimilated using a 3DVAR system. The analysis $x^{(a)}$ during a given assimilation cycle was determined using:

$$x^{(a)} = x^{(b)} + BH^T(R + HBH^T)^{-1}(y - H(x^{(b)})), \tag{4.1}$$

where $H = I$, the observation error covariance matrix R was given the form of the identity matrix scaled by the factor r,

$$R = r \cdot I, \tag{4.2}$$

and the background error covariance matrix B was given the form:

$$B = \begin{pmatrix} (x_1^{(b)} - x_1^{(true)})^2 & 0 & 0 \\ 0 & (x_2^{(b)} - x_2^{(true)})^2 & 0 \\ 0 & 0 & (x_3^{(b)} - x_3^{(true)})^2 \end{pmatrix}, \tag{4.3}$$

with $x^{(b)}$ being the background state, $x^{(true)}$ being the true dynamical state obtained from the truth simulation, and the diagonal elements of B containing the model error variances. We chose to use a diagonal matrix here because it is a reasonable place to start and, as is shown in this section, still has a positive impact on the assimilation performance. Given the strong correlations between errors in the x_1 and x_2 variables (see Fig. 3), it is possible that including the off-diagonal elements would have led to even better results; however, their inclusion in the B matrix is left for future work. Note that even though this is a perfect observation experiment, we chose to set the scaling factor r to a small non-zero value so that we could use the data assimilation system rather than directly inserting the observations into the model. This approach maintains consistency with the other experiments presented in this section and is a reasonable approach because we generally would not know that the observations are perfect in a real data assimilation system and therefore would likely still assume that the observation errors come from a Gaussian distribution.

Figure 3 shows the evolution of the true ρ parameter and the model errors $x_1^{(b)} - x_1^{(true)}$, $x_2^{(b)} - x_2^{(true)}$, and $x_3^{(b)} - x_3^{(true)}$ during the assimilation experiment. The true error for each model variable is shown in blue, whereas the model errors estimated using the asymptotic error estimators in (2.5) - (2.7) are depicted by the red dashed lines. For the asymptotic model error estimates, $x_1(\rho_0, t_0)$ is taken to be its instantaneous value at each assimilation time. Inspection of the error time series (Figs. 3a-c) reveals that the asymptotic error estimators are able to accurately capture the magnitude of the true errors in the model background, as well as their rapid changes with time, when all other errors in the system are eliminated. The model errors display more rapid variations than the ρ parameter (Fig. 3d) because the time step used by the coupled model is five times faster than that used in the hidden model $S2$ to perturb ρ. The true ρ parameter oscillates in a quasi-periodic manner for an extended period of time either below or above $\rho = 28$, with occasional transitions between values less than or greater than this threshold as the hidden model driving the changes in ρ_{true} propagates from one wing of the butterfly to the other (see Fig. 2a). These quasi-periodic oscillations could be thought of as representing biases associated with the diurnal or seasonal cycles in atmospheric models.

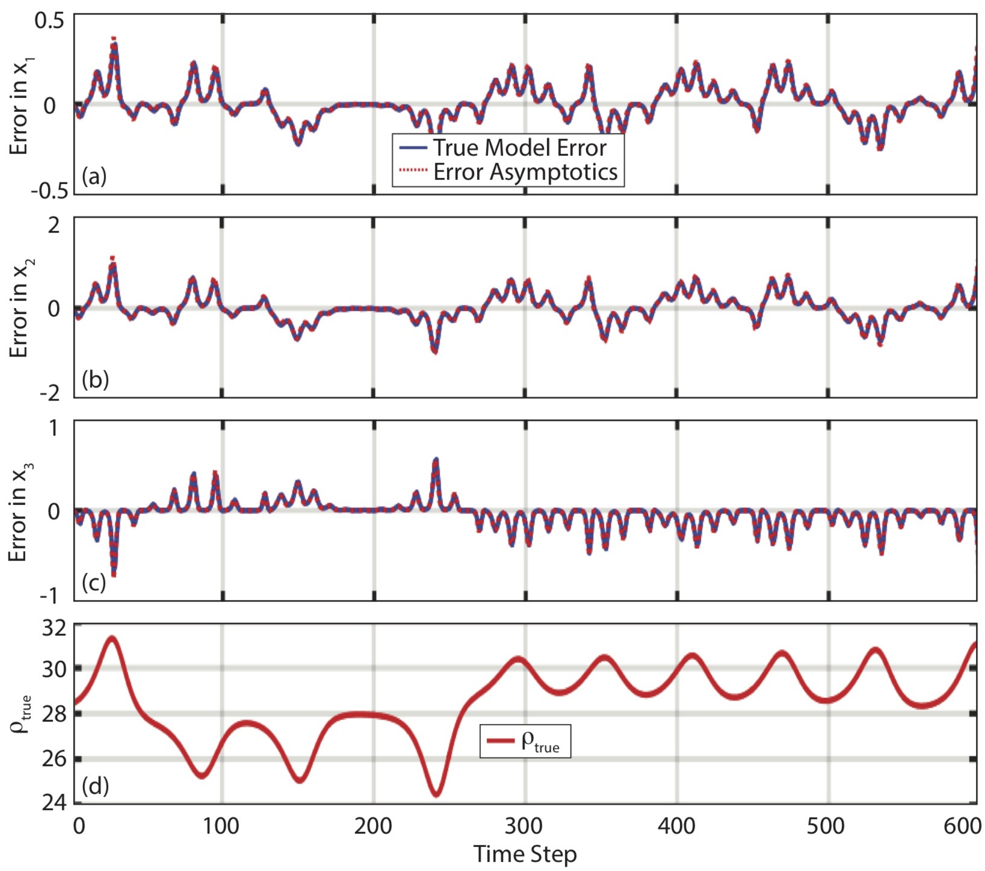

Fig. 3. *Time series showing the evolution of the true model error (blue lines) and asymptotic error estimations (red dashed lines) for the (a) x_1, (b) x_2, and (c) x_3 model state variables and for the (d) ρ_{true} parameter (red line) for an experiment lasting $N_t = 600$ assimilation cycles with $\delta t_{assim} = 0.06$ and the measurement error ϵ set to zero.*

4.2. Using Bias Estimators to Improve Assimilation Performance. The development of methods to accurately estimate the model background error covariance matrix B is important for all data assimilation algorithms. In this section, we demonstrate that the assimilation quality, as measured using OMB statistics, can be improved through inclusion of appropriate model error estimators during the data assimilation step. We also examine the optimality of using either a fixed or dynamically varying B matrix and assess the influence of the observation error on these estimates.

For this exercise, we performed cycled 3DVAR data assimilation experiments using two versions of the L63 model where we chose to use a constant $\delta\rho = 1$ in the truth simulation or where we allowed $\delta\rho$ to vary with time based on the influence of the hidden system $S2$ described in Section 2.1. The first version is used to represent a situation where a given parameter that does not vary in the real world is assigned the wrong constant value in the numerical model. Here, we assume that we know the asymptotics describing the sensitivity of the model to small perturbations in ρ, but that we do not know the correct scaling factor c for $\delta\rho$. In other words, we know the true value of $\delta\rho$ only up to a constant $c \in \mathbb{R}$, which includes the case of a constant but unknown $\delta\rho$. For brevity, this section only includes results for the scenario in which $\delta\rho$ is allowed to vary with time. Note that even though the errors in the asymptotic estimates will be larger in this situation because the maximum size of $\delta\rho$ is larger, the conclusions regarding the importance of using the dynamically varying B matrix are the same for the experiments using the constant and time-varying $\delta\rho$ perturbations.

To assess the sensitivity to the matrix B, we initially performed an experiment where a constant covariance matrix of the form $B = b \cdot I \in \mathbb{R}^{3\times3}$ was used during each assimilation cycle, where b is used to scale the identity matrix. We then searched for the constant b that produced the smallest OMB errors averaged over $N_t = 600$ assimilation cycles. Finally, we repeated the search using a dynamical B matrix, which as in (3.16), is the sum of a constant matrix as in (3.12) and a dynamical part as given by the term (3.13) that is computed using the model error estimators described in Theorem 2.1. The form of $B = B_k$ at time t_k, with the index $k = 1, 2, ..., N_t$ of analysis steps, is chosen as:

$$B_k = b \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} error_{1,k}^2 & 0 & 0 \\ 0 & error_{2,k}^2 & 0 \\ 0 & 0 & error_{3,k}^2 \end{pmatrix}, \tag{4.4}$$

where the diagonal elements in the second part of (4.4) are defined as:

$$error_{1,k} = c \cdot 0.5 \cdot \sigma \cdot x_1(\rho_0, t_k) \cdot (\delta t)^2 \cdot \delta\rho_k \tag{4.5}$$

$$error_{2,k} = c \cdot x_1(\rho_0, t_k) \cdot \delta t \cdot \delta\rho_k \tag{4.6}$$

$$error_{3,k} = c \cdot x_1^2(\rho_0, t_k) \cdot (\delta t)^2 \cdot \delta\rho_k \tag{4.7}$$

Equations (4.5) - (4.7) correspond to the model first guess errors for x_1, x_2, and x_3, respectively, for each assimilation time t_k. The numerical experiments evaluated in this section were carried out using $c = 1$.

Two examples illustrating the relationship between the size of b and the average model first guess errors when using either the constant or dynamic estimates for B during the assimilation experiments are shown in Fig. 4. The first example (Fig. 4a) has relatively frequent assimilation cycles ($\delta t_{assim} = 0.02$) and small random observation errors ($\epsilon = 0.2$), whereas the observation errors are larger ($\epsilon = 0.5$) and the observations are assimilated less frequently ($\delta t_{assim} = 0.04$) during the second example (Fig. 4b). Random errors added to each observation were drawn from a Gaussian distribution scaled by the value of ϵ chosen for each case.

In both examples, the behavior of the relationship shown in Fig. 4 is well-known in the field of inverse problems where a regularization that is too small increases the influence of the observation errors and a regularization that is too large will not be able to fully exploit the new information provided by the observations. The optimal

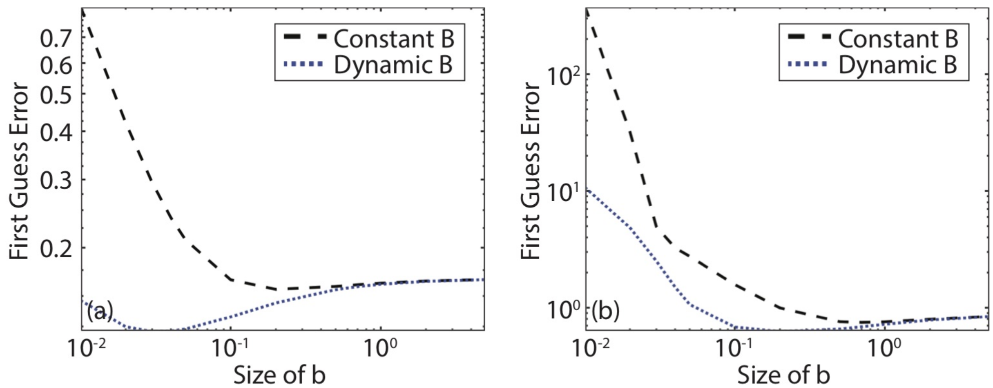

FIG. 4. *Scan of the average model first guess errors plotted as a function of the size of b when the background error covariance matrix B is a multiple of the identity matrix ($B = b \cdot I$) (black dashed line) or when it is obtained using the dynamic B estimator presented in (4.4) (blue dotted line). Panels (a) and (b) show results from experiments using assimilation update intervals δt_{assim} and random observation errors ϵ set to ($\delta t_{assim} = 0.02, \epsilon = 0.2$) and ($\delta t_{assim} = 0.04, \epsilon = 0.5$), respectively. The first guess error statistics were computed using output from 600 time steps.*

B, which varies depending upon the observation and model errors present during a given assimilation cycle, will lead to the smallest first guess errors. Of importance for this discussion is that the smallest first guess errors for both examples occur when the dynamic B matrix is used. It is also evident that the optimal size of b decreases when the dynamical error estimators are used to scale B because they are better able to capture the actual errors in the model background during each assimilation cycle. Together, these examples demonstrate that it is highly desirable to employ dynamical estimators of the model first guess error in data assimilation algorithms.

4.3. Numerical Estimation of the Bias Estimator Polynomial Coefficients. In this section, we investigate the determination of the model bias estimator coefficients α using output from cycled 3DVAR experiments employing different assimilation intervals and observation error magnitudes. For these experiments, we employ the dynamical background error covariance matrix B shown in (4.4) during each data assimilation cycle, with the dynamic model errors for (x_1, x_2, x_3) computed using the asymptotic error estimators in (4.5) - (4.7) with the scaling factor c set to 1. Sensitivity tests revealed that the model error coefficients were stable over a broad range of values for the scaling factor b; therefore, for convenience, it was set to 0.1 during the experiments discussed in this section. This behavior and the chosen value for b are consistent with the results shown in Fig. 4.

Experimentation also revealed that the matrix A used to determine the bias correction coefficients α in (3.26) is ill-posed with singular values smaller than 10^{-4} and a condition number larger than 10^4. Therefore, to improve its conditioning, Tikhonov regularization was used by replacing the least squares estimator $A^\dagger = (A^T A)^{-1} A^T$ in (3.26) with the Tikhonov inverse:

$$Q := (\alpha_{reg} I + A^T A)^{-1} A^T \tag{4.8}$$

where α_{reg} is the Tikhonov regularization parameter. Sensitivity tests showed that setting α_{reg} to a small value (10^{-5}) provided the most accurate results. This means that the bias correction coefficients for a given model variable can be determined using:

$$\alpha = (\alpha_{reg} I + A^T A)^{-1} A^T q \tag{4.9}$$

Table 1 shows results computed using truth-minus-background statistics accumulated over $N_t = 600$ assimilation cycles for two experiments, including one where perfect observations ($\epsilon = 0$) were assimilated at $\delta t_{assim} = 0.01$ time intervals (left columns) and a second experiment where random errors were added to the observations ($\epsilon = 0.01$) and the assimilation interval was increased to $\delta t_{assim} = 0.02$. The scaling factor r for the observation error covariance matrix in (4.2) was set to 10^{-5} and 10^{-4}, respectively, for each of these experiments, with $\delta\rho$ for a given time step obtained from the hidden system $S2$ described in Section 2.1. The coefficients of the polynomial expansion of the model bias are computed separately for each model variable (x_1, x_2, x_3). Here, we have used all polynomial terms up to the 2nd order when computing the dynamic B matrix in (4.4) because of the presence of the x_1^2 term in the asymptotics shown in (3.20). To ease interpretation of the results, we have included $\delta\rho$ and the constant 0.5, σ, δt, and (δt^2) terms as they appear in (4.5), (4.6), and (4.7) such that the estimation outcomes shown in Table 1 should be either 0 or 1 depending upon whether or not a given term is in the polynomial expansion. This means that the reconstructed bias correction coefficient $\alpha_{recon}(1, 0, 0)$ should equal one for x_1 and x_2, $\alpha_{recon}(2, 0, 0)$ should equal one for x_3, and all of the other α_{recon} values should be zero.

Inspection of Table 1 shows that the maximum error for each state variable (x_1, x_2, x_3) is 8% (e.g., $\alpha_{recon} = 0.92$) for the experiment in which perfect observations were assimilated, and that the errors for most of the remaining α_{recon} terms are very small. This demonstrates that the bias correction coefficients can be accurately estimated in this situation such that the only remaining sources of error are likely associated with numerical discretization errors or the exclusion of higher order polynomial terms from the asymptotical expansion (e.g., higher than the 2nd order). The error in each α_{recon} term increases during the second experiment where measurement errors were added to the observations prior to their assimilation. Even so, the results show that the method is still able to identify the dominant terms and that it is possible to obtain reasonable estimates for the bias correction coefficients in the presence of observation error. Finally, other experiments were performed where the size of the observation error and the length of the assimilation cycling interval were varied, with all of the experiments showing similar effects to those demonstrated in Table 1 if reasonable observation errors and cycling intervals were used.

4.4. Reconstruction of the ρ Parameter. In this section, we explore the effectiveness of using the bias correction coefficient vector α to reconstruct the ρ parameter within the data assimilation system. The truth simulation for this particular exercise was performed using the coupled L63 model described in Section 2.1. A cycled data assimilation experiment covering $N_t = 600$ assimilation cycles with $\delta t_{assim} = 0.04$ was then performed using observations from the truth simulation. Given that the true state of a real-world system is unknown, here we choose to use the analysis-minus-background difference as a proxy for the model error q in (3.26) because the model background $x^{(b)}$ and model analysis $x^{(a)}$ are both readily available from data assimilation systems.

Because ρ varies with time in the coupled L63 system used to perform the truth simulation, it is not advantageous to use assimilation statistics accumulated over a long time period to estimate the value of this parameter for a specific assimilation

	Exp 1			Exp 2		
	for x_1	for x_2	for x_3	for x_1	for x_2	for x_3
$\alpha_{recon}(0,0,0)$	4.94E-02	-5.30E-03	-4.06E-03	1.16E-01	-7.26E-03	-1.28E-01
$\alpha_{recon}(1,0,0)$	0.92	1.06	2.37E-03	0.76	1.16	-5.95E-01
$\alpha_{recon}(2,0,0)$	-2.57E-04	4.08E-05	1.03	9.79E-03	1.52E-03	0.94
$\alpha_{recon}(0,1,0)$	2.38E-02	-4.92E-02	1.31E-02	2.65E-02	-1.13E-01	-2.45E-01
$\alpha_{recon}(0,2,0)$	3.13E-04	-2.35E-05	-3.94E-03	2.32E-03	-1.84E-04	5.25E-02
$\alpha_{recon}(0,0,1)$	-1.11E-02	6.09E-04	-4.77E-02	1.88E-04	5.97E-03	-1.30E-01
$\alpha_{recon}(0,0,2)$	2.13E-04	-1.85E-05	2.86E-03	-8.31E-04	-3.04E-04	3.10E-03
$\alpha_{recon}(1,1,0)$	-1.88E-04	-3.84E-06	-4.02E-02	-9.54E-03	-9.32E-04	-2.04E-01
$\alpha_{recon}(1,0,1)$	3.54E-03	-4.07E-04	4.04E-04	5.71E-03	-2.03E-03	2.42E-03
$\alpha_{recon}(0,1,1)$	-2.21E-03	-4.42E-05	-1.08E-04	-4.80E-03	6.95E-05	3.10E-02

TABLE 1

Reconstructed bias correction coefficients (α_{recon}) for each model variable (x_1, x_2, x_3) determined using (3.23) and truth-minus-background statistics accumulated over 600 assimilation cycles for two experiments employing different observation errors and assimilation update intervals. The 0th to 2nd order terms are shown in each row. Columns 2-4 and 5-7 show the results for experiments employing ($\delta t_{assim} = 0.01; \epsilon = 0$) and ($\delta t_{assim} = 0.02; \epsilon = 0.01$), respectively. The Tikhonov regularization parameter a_{reg} was set to 10^{-5} for both experiments.

cycle. Instead, we compute the coefficient vector α using output from 10 consecutive assimilation cycles rather than from the full assimilation period. This length was chosen as a balance between the desire to acquire a large enough sample to robustly estimate $\delta\rho$ and the need to use a short enough time period to ensure that the instantaneous $\delta\rho$ values during a given time interval do not deviate strongly from the mean $\delta\rho$ over that interval. To ease comparison to the reconstructed mean $\delta\rho$, the average of the individual $\delta\rho$ estimates obtained using the simple diagnostic tools shown in (3.27) are used to represent the true mean $\delta\rho$ over each time period. Together, these choices are consistent with the constraints that would be encountered in a real-world data assimilation system.

Figure 5 shows the evolution of the instantaneous model errors $x_1^{(b)} - x_1^{(a)}$, $x_2^{(b)} - x_2^{(a)}$, and $x_3^{(b)} - x_3^{(a)}$, along with the actual and reconstructed values for $\delta\rho$ for three experiments employing different observation errors. The images on the left show the true error for each model variable in blue, whereas the dashed red lines show the model errors estimated using the asymptotic error estimators in Theorem 2.1. For the images on the right, the black and blue lines denote the true instantaneous and true mean $\delta\rho$ values, respectively, whereas the red lines depict the corresponding mean $\delta\rho$ estimates reconstructed using the α vector. Results are shown for three experiments assimilating observations with measurement errors $\epsilon = \{0, 0.02, \text{and } 0.04\}$ and scaling factors $r = \{0.0004, 0.0004, \text{and } 0.0016\}$ for the observation error covariance matrices.

Inspection of the time series in Fig. 5 reveals that the mean $\delta\rho$ values reconstructed from the coefficient vector α accurately capture the magnitude and evolution of the true $\delta\rho$ for the case where the assimilated observations have zero measurement error (Fig. 5b). The asymptotic error estimators also do an excellent job representing the true model errors during this experiment (Fig. 5a). As the observation error increases, however, the model error time series become more noisy (Fig. 5c, e) and the accuracy of the $\delta\rho$ reconstruction decreases due to the increased noise (Fig. 5d, f). The errors in the $\delta\rho$ reconstruction are largest for time periods when the true $\delta\rho$ reaches a local minimum or maximum because the rapid variation with time during those situations makes it more difficult to properly reconstruct $\delta\rho$. Regardless, these results show that it is possible to use the coefficient vector α to obtain useful information about the trajectory of $\delta\rho$ during the truth simulation. Because the true state was not used during this exercise, these results also demonstrate that reasonable parameter and model bias estimates can be obtained using differences between the model analysis and background states. This is important because whereas the true state of a real-world system is generally unknown, the model analysis and background states are both readily available from data assimilation systems.

5. Conclusions. In this study, we have examined the behavior of dynamic model errors and their influence on the quality of the model analysis and first guess during cycled data assimilation experiments using the L63 model and a 3DVAR data assimilation system. We showed that conditional model biases due to errors in the specification of a model parameter can be represented as a polynomial function that can be estimated using the model background-minus-truth or background-minus-analysis statistics for the realistic situation where the modeling system consists of polynomial forcing terms. We have also suggested a regularized least squares regression method to estimate the model biases and then described how these model error estimators could be used in the data assimilation system to improve the accuracy of the model analysis and first guess.

We have carried out all derivations, estimations, and numerical experiments using the well-known L63 model to demonstrate the validity and feasibility of the ideas developed during this study. The L63 model allows us to study all parts of the system, bias estimators, and tools in a detailed way that would not be possible if we had used a full physics numerical model while still being able to represent the chaotic nonlinear characteristics of the real atmosphere. The results showed that the asymptotics are indeed a valid method to estimate an important part of the model first guess error, and that their use in data assimilation has the potential to improve the accuracy of the model background and analysis. We showed that model error estimators computed using the difference between the model background and analysis, which are readily available from all assimilation systems, are an effective way to estimate model error. In this framework, the model analysis serves as an approximation of the true state, which is unknown in a real-world system. Reasonable results can be achieved even when relatively large errors are present in the observations if Tikhonov regularization is employed during the estimation of the polynomial model error coefficients. Finally, we also show that the polynomial model bias coefficient vector can be used to reconstruct $\delta\rho$ during the assimilation experiments.

In the current work, we have restricted our attention to a small-scale system containing three state variables. Real-world NWP models and data assimilation systems have much deeper complexity and their dimensions are much larger than the system used here. Thus, future work is necessary to investigate the validity of the above ideas in high-dimensional models and to determine if the methods developed during this study can improve the representation of the background error covariance matrix B used by such systems. For the experiments presented in this paper, all of the state variables were observed during each assimilation cycle, which of course is not possible in a real data assimilation system. It will be important to evaluate the utility of the method when the observation uncertainty is higher or the measurements do not observe the full state of the model. It is reasonable to expect that it will be more difficult to estimate the model errors in such situations. It is also possible that the size of the initial condition uncertainty relative to the model error could impact

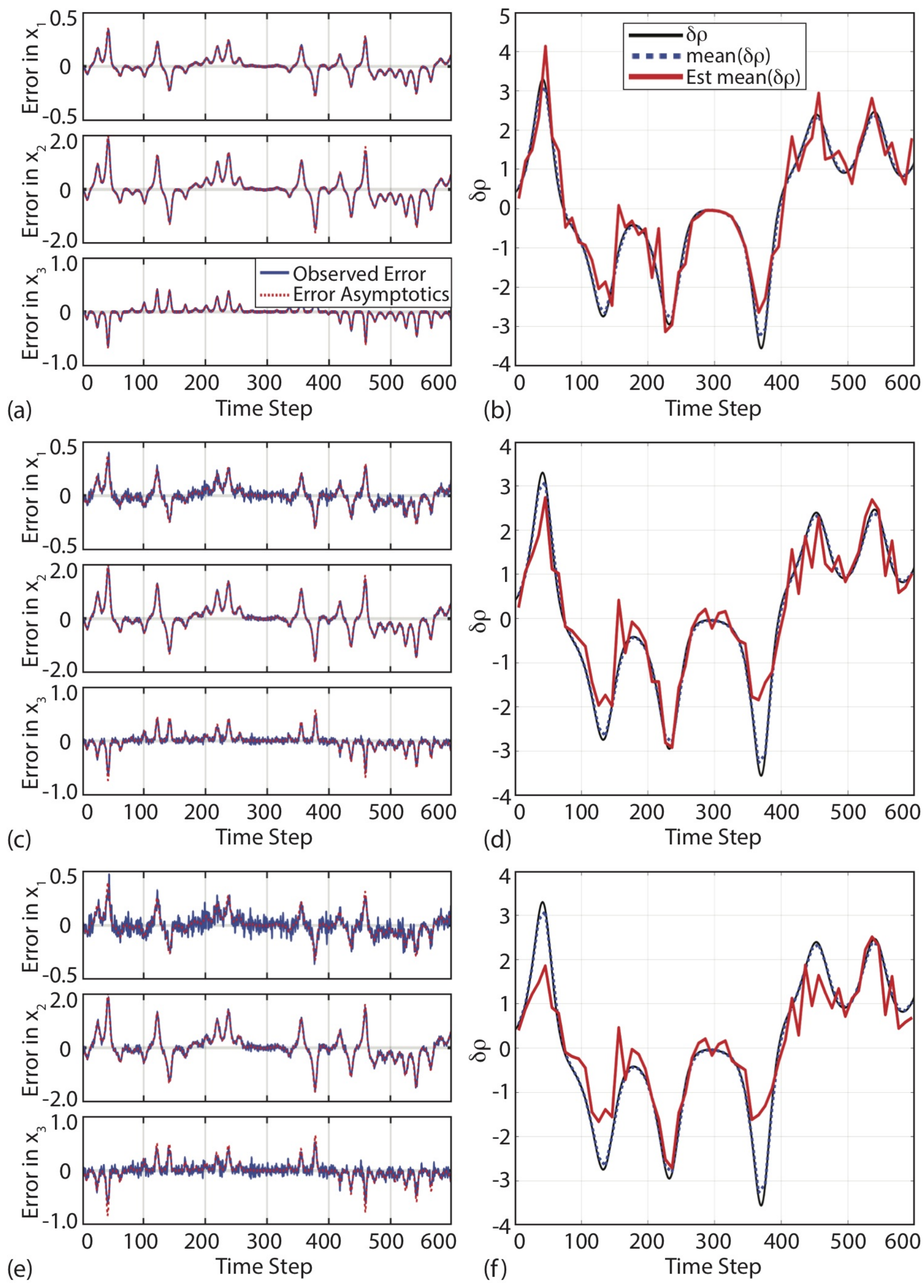

Fig. 5. *(a) Time series showing the evolution of the model error given by the first guess minus analysis (blue line) for $x_1, x_2,$ and x_3, and their estimation computed using the error asymptotics (dashed red line). Here, $\delta t = 0.04$ and $\epsilon = 0$. (b) Time series showing the evolution of the true $\delta\rho$ (black line). The mean $\delta\rho$ parameter computed over intervals of 10 assimilation cycles is shown by the dashed blue line, with the corresponding dynamic estimation computed using the mean bias correction coefficients shown by the red lines. (c-d) Same as (a-b), except for the case where the assimilation experiment was performed using $\delta t = 0.04$ and $\epsilon = 0.02$. (e-f) Same as (a-b), except for the case where the assimilation experiment was carried out using $\delta t = 0.04$ and $\epsilon = 0.04$.*

the performance of this method. For example, the model error contribution to the forecast uncertainty will typically increase relative to the initial condition uncertainty over longer time periods. This would suggest that the model error estimation method may be especially useful for longer assimilation windows or when the observations are assimilated less frequently. A final point to consider is that we already knew which model parameter was incorrectly specified in the L63 model during the data assimilation experiments, which made it possible for us to target its reconstruction using the bias correction coefficient vector. Though this knowledge made the problem easier to solve, it is still consistent with many real-world situations where it is known a priori that a certain parameter varies with time but has been assigned a constant value in the NWP model due to computational constraints or incomplete knowledge on how to predict its evolution. With this knowledge, it should be possible to use the general polynomial expansion of the model variables method developed in Section 4.3 to determine if there are relationships between any of the polynomial terms and a chosen parameter and then use that information to reconstruct the value of the parameter.

The dynamic B method developed during this study could be interpreted as providing dynamic additive covariance inflation capturing systematic model errors that are not represented by the static B used by variational systems nor by the dynamic B used by hybrid and EnKF assimilation methods. Inclusion of the dynamic model bias estimates in the B matrix could therefore make it possible to reduce the amount of covariance inflation that is used during the data assimilation step in EnKF systems. This is potentially advantageous because the dynamic B is computed based on the current conditions rather than using random perturbations drawn from a climatology as is typically done with additive covariance inflation methods. It may also provide a complementary approach to weak-constraint 4DVAR where instead of providing the model an additional degree of freedom through introduction of a model error forcing term, we instead enhance the quality of the B matrix through inclusion of the model bias estimates before it is used by the assimilation algorithm. More detailed investigations of these and other topics are left for future work.

Acknowledgments. The lead author was partially supported by a University of Reading International Research Studentship. The third author was funded in part by the NERC National Centre for Earth Observation. We thank the reviewers for their valuable comments that helped improve the quality of the manuscript.

REFERENCES

[1] A. Aksoy, F. Zhang, and J. W. Nielsen-Gammon, *Ensemble-based simultaneous state and parameter estimation with MM5*, Geophysical Research Letters, 33 (2006), https://doi.org/10.1029/2006GL026186.

[2] J. T. Ambadan and Y. Tang, *Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems*, Journal of the Atmospheric Sciences, 66 (2009), pp. 261–285, https://doi.org/10.1175/2008JAS2681.1.

[3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Dover Books on Electrical Engineering Series, Dover Publications, Incorporated, 2012, http://books.google.de/books?id=iYMqLQp49UMC.

[4] J. L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Monthly Weather Review, 129 (2001), pp. 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

[5] J. L. Anderson, *An adaptive covariance inflation error correction algorithm for ensemble filters*, Tellus A, 59 (2007), pp. 210–224, https://doi.org/{10.1111/j.1600-0870.2006.00216.x}.

[6] J. L. Anderson, *Spatially and temporally varying adaptive covariance inflation for ensemble*

filters, Tellus A, 61 (2009), pp. 72–83, https://doi.org/10.1111/j.1600-0870.2008.0361.x.

[7] J. Annan, *Parameter estimation using chaotic time series*, Tellus A: Dynamic Meteorology and Oceanography, 57 (2005), pp. 709–714, https://doi.org/10.3402/tellusa.v57i5.14735.

[8] R. N. Bannister, *A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances*, Quarterly Journal of the Royal Meteorological Society, 134 (2008), pp. 1951–1970, https://doi.org/10.1002/qj.339.

[9] R. N. Bannister, *A review of forecast error covariance statistics in atmospheric variational data assimilation. ii: Modelling the forecast error covariance statistics*, Quarterly Journal of the Royal Meteorological Society, 134 (2008), pp. 1971–1996, https://doi.org/10.1002/qj.340.

[10] R. N. Bannister, *A review of operational methods of variational and ensemble-variational data assimilation*, Quarterly Journal of the Royal Meteorological Society, 143 (2017), pp. 607–633, https://doi.org/10.1002/qj.2982.

[11] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. d. l. Cámara, H. M. Christensen, M. Colangeli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P. D. Williams, and J.-I. Yano, *Stochastic parameterization: Toward a new view of weather and climate models*, Bulletin of the American Meteorological Society, 98 (2017), pp. 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

[12] M. Bocquet, *Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var*, Quarterly Journal of the Royal Meteorological Society, 138 (2012), pp. 664–681, https://doi.org/10.1002/qj.961.

[13] M. Buehner, P. Gauthier, and Z. Liu, *Evaluation of new estimates of background- and observation-error covariances for variational assimilation*, Quarterly Journal of the Royal Meteorological Society, 131 (2005), pp. 3373–3383, https://doi.org/10.1256/qj.05.101.

[14] R. Buizza, M. Miller, and T. N. Palmer, *Stochastic representation of model uncertainties in the ECMWF ensemble prediction system*, Quarterly Journal of the Royal Meteorological Society, 125 (1999), pp. 2887–2908, https://doi.org/10.1002/qj.49712556006.

[15] A. Carrassi and S. Vannitsem, *Accounting for model error in variational data assimilation: A deterministic formulation*, Monthly Weather Review, 138 (2010), pp. 3369–3386, https://doi.org/10.1175/2010MWR3192.1.

[16] D. P. Dee, *On-line estimation of error covariance parameters for atmospheric data assimilation*, Monthly Weather Review, 123 (1995), pp. 1128–1145, https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.

[17] J. C. Derber, *A variational continuous assimilation technique*, Monthly Weather Review, 117 (1989), pp. 2437–2446, https://doi.org/10.1175/1520-0493(1989)117<2437:AVCAT>2.0.CO;2.

[18] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, Journal of Geophysical Research: Oceans, 99 (1994), pp. 10143–10162, https://doi.org/10.1029/94JC00572.

[19] G. Evensen, *Advanced data assimilation for strongly nonlinear dynamics*, Monthly Weather Review, 125 (1997), pp. 1342–1354, https://doi.org/10.1175/1520-0493(1997)125<1342:ADAFSN>2.0.CO;2.

[20] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Earth and Environmental Science, Springer, 2009, http://books.google.de/books?id=2_zaTb_O1AkC.

[21] G. Evensen and P. J. van Leeuwen, *An ensemble Kalman smoother for nonlinear dynamics*, Monthly Weather Review, 128 (2000), pp. 1852–1867, https://doi.org/doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

[22] T. Fujita, D. J. Stensrud, and D. C. Dowell, *Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties*, Monthly Weather Review, 135 (2007), pp. 1846–1868, https://doi.org/10.1175/MWR3391.1.

[23] M. Goodliff, J. Amezcua, and P. J. V. Leeuwen, *Comparing hybrid data assimilation methods on the Lorenz 1963 model with increasing non-linearity*, Tellus A: Dynamic Meteorology and Oceanography, 67 (2015), p. 26928, https://doi.org/10.3402/tellusa.v67.26928.

[24] A. Griffith and N. Nichols, *Adjoint methods in data assimilation for estimating model error*, Flow, Turbulence and Combustion, 65 (2000), pp. 469–488, https://doi.org/10.1023/A:1011454109203.

[25] S. Ha, J. Berner, and C. Snyder, *A comparison of model error representations in mesoscale ensemble data assimilation*, Monthly Weather Review, 143 (2015), pp. 3893–3911, https:

//doi.org/10.1175/MWR-D-14-00395.1.

[26] T. M. Hamill and J. S. Whitaker, *Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches*, Monthly Weather Review, 133 (2005), pp. 3132–3147, https://doi.org/10.1175/MWR3020.1.

[27] D. Hodyss, *Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation*, Monthly Weather Review, 139 (2011), pp. 3571–3588, https://doi.org/10.1175/2011MWR3558.1.

[28] P. L. Houtekamer and H. L. Mitchell, *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Review, 126 (1998), pp. 796–811.

[29] P. L. Houtekamer and H. L. Mitchell, *A sequential ensemble Kalman filter for atmospheric data assimilation*, Monthly Weather Review, 129 (2001), pp. 123–137, https://doi.org/doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

[30] P. L. Houtekamer and H. L. Mitchell, *Ensemble Kalman filtering*, Quarterly Journal of the Royal Meteorological Society, 131 (2005), pp. 3269–3289, https://doi.org/10.1256/qj.05.135.

[31] P. L. Houtekamer, H. L. Mitchell, and X. Deng, *Model error representation in an operational ensemble Kalman filter*, Monthly Weather Review, 137 (2009), pp. 2126–2143, https://doi.org/10.1175/2008MWR2737.1.

[32] P. L. Houtekamer, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, *Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations*, Monthly Weather Review, 133 (2005), pp. 604–620, https://doi.org/doi:10.1175/MWR-2864.1.

[33] P. L. Houtekamer and F. Zhang, *Review of the ensemble Kalman filter for atmospheric data assimilation*, Monthly Weather Review, 144 (2016), pp. 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

[34] K. E. Howes, A. M. Fowler, and A. S. Lawless, *Accounting for model error in strong-constraint 4D-Var data assimilation*, Quarterly Journal of the Royal Meteorological Society, 143 (2017), pp. 1227–1240, https://doi.org/10.1002/qj.2996.

[35] L. Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, *Ensemble of data assimilations at ECMWF*, ECMWF Technical Memoranda, (2010), p. 45, https://doi.org/10.21957/obke4k60, https://www.ecmwf.int/node/10125.

[36] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, 2003, http://books.google.de/books?id=Uqc7zC7NULMC.

[37] J.-S. Kang, E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, *"variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation*, Journal of Geophysical Research: Atmospheres, 116 (2011), https://doi.org/10.1029/2010JD014673.

[38] D. T. Kleist, D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, *Introduction of the GSI into the ncep global data assimilation system*, Weather and Forecasting, 24 (2009), pp. 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

[39] S. Kotsuki, K. Terasaki, H. Yashiro, H. Tomita, M. Satoh, and T. Miyoshi, *Online model parameter estimation with ensemble data assimilation in the real global atmosphere: A case with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) and the global satellite mapping of precipitation data*, Journal of Geophysical Research: Atmospheres, 123 (2018), pp. 7375–7392, https://doi.org/10.1029/2017JD028092.

[40] H. Koyama and M. Watanabe, *Reducing forecast errors due to model imperfections using ensemble Kalman filtering*, Monthly Weather Review, 138 (2010), pp. 3316–3332, https://doi.org/10.1175/2010MWR3067.1.

[41] J. Lei and P. Bickel, *A moment matching ensemble filter for nonlinear non-Gaussian data assimilation*, Monthly Weather Review, 139 (2011), pp. 3964–3973, https://doi.org/10.1175/2011MWR3553.1.

[42] L. Lei, D. R. Stauffer, S. E. Haupt, and G. S. Young, *A hybrid nudging-ensemble Kalman filter approach to data assimilation. part i: application in the Lorenz system*, Tellus A: Dynamic Meteorology and Oceanography, 64 (2012), p. 18484, https://doi.org/10.3402/tellusa.v64i0.18484.

[43] H. Li, E. Kalnay, and T. Miyoshi, *Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter*, Quarterly Journal of the Royal Meteorological Society, 135 (2009), pp. 523–533, https://doi.org/10.1002/qj.371.

[44] H. Li, E. Kalnay, T. Miyoshi, and C. M. Danforth, *Accounting for model errors in ensemble data assimilation*, Monthly Weather Review, 137 (2009), pp. 3407–3419, https://doi.org/10.1175/2009MWR2766.1.

[45] M. Lindskog, D. Dee, Y. Trémolet, E. Andersson, G. Radnóti, and M. Fisher, *A weak-constraint four-dimensional variational analysis system in the stratosphere*, Quar-

terly Journal of the Royal Meteorological Society, 135 (2009), pp. 695–706, https://doi.org/10.1002/qj.392.

[46] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, *The Met. Office global three-dimensional variational data assimilation scheme*, Quarterly Journal of the Royal Meteorological Society, 126 (2000), pp. 2991–3012, https://doi.org/10.1002/qj.49712657002.

[47] E. N. Lorenz, *Deterministic nonperiodic flow*, Journal of the Atmospheric Sciences, 20 (1963), pp. 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[48] C. Marzban, *Variance-based sensitivity analysis: An illustration on the Lorenz'63 model*, Monthly Weather Review, 141 (2013), pp. 4069–4079, https://doi.org/10.1175/MWR-D-13-00032.1.

[49] Z. Meng and F. Zhang, *Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. part ii: Imperfect model experiments*, Monthly Weather Review, 135 (2007), pp. 1403–1423, https://doi.org/10.1175/MWR3352.1.

[50] L. Mitchell and A. Carrassi, *Accounting for model error due to unresolved scales within ensemble Kalman filtering*, Quarterly Journal of the Royal Meteorological Society, 141 (2015), pp. 1417–1428, https://doi.org/10.1002/qj.2451.

[51] T. Miyoshi, *The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter*, Monthly Weather Review, 139 (2011), pp. 1519–1535, https://doi.org/10.1175/2010MWR3570.1.

[52] G. Nakamura and R. Potthast, *Inverse Modeling*, 2053-2563, IOP Publishing, 2015, https://doi.org/10.1088/978-0-7503-1218-9.

[53] H. Ngodock, M. Carrier, S. Smith, and I. Souopgui, *Weak and strong constraints variational data assimilation with the NCOM-4DVAR in the Agulhas region using the representer method*, Monthly Weather Review, 145 (2017), pp. 1755–1764, https://doi.org/10.1175/MWR-D-16-0264.1.

[54] C. Nicolis, *Dynamics of model error: Some generic features*, Journal of the Atmospheric Sciences, 60 (2003), pp. 2208–2218, https://doi.org/10.1175/1520-0469(2003)060<2208:DOMESG>2.0.CO;2.

[55] C. Nicolis, *Dynamics of model error: The role of unresolved scales revisited*, Journal of the Atmospheric Sciences, 61 (2004), pp. 1740–1753, https://doi.org/10.1175/1520-0469(2004)061<1740:DOMETR>2.0.CO;2.

[56] C. Nicolis, R. A. P. Perdigao, and S. Vannitsem, *Dynamics of prediction errors under the combined effect of initial condition and model errors*, Journal of the Atmospheric Sciences, 66 (2009), pp. 766–778, https://doi.org/10.1175/2008JAS2781.1.

[57] J. A. Otkin, R. Potthast, and A. S. Lawless, *Nonlinear bias correction for satellite data assimilation using Taylor series polynomials*, Monthly Weather Review, 146 (2018), pp. 263–285, https://doi.org/10.1175/MWR-D-17-0171.1.

[58] D. F. Parrish and J. C. Derber, *The National Meteorological Center's spectral statistical-interpolation analysis system*, Monthly Weather Review, 120 (1992), pp. 1747–1763, https://doi.org/http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

[59] Z. Pu and J. Hacker, *Ensemble-based Kalman filters in strongly nonlinear dynamics*, Advances in Atmospheric Sciences, 26 (2009), pp. 373–380, https://doi.org/10.1007/s00376-009-0373-9.

[60] M. Pulido, S. Polavarapu, T. G. Shepherd, and J. Thuburn, *Estimation of optimal gravity wave parameters for climate models using data assimilation*, Quarterly Journal of the Royal Meteorological Society, 138 (2012), pp. 298–309, https://doi.org/10.1002/qj.932.

[61] L. Raynaud, L. Berre, and G. Desroziers, *An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system*, Quarterly Journal of the Royal Meteorological Society, 137 (2011), pp. 607–619, https://doi.org/10.1002/qj.795.

[62] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, 2015, https://doi.org/10.1017/CBO9781107706804.

[63] G. S. Romine, C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, *Representing forecast error in a convection-permitting ensemble system*, Monthly Weather Review, 142 (2014), pp. 4519–4541, https://doi.org/10.1175/MWR-D-14-00100.1.

[64] J. Ruiz and M. Pulido, *Parameter estimation using ensemble-based data assimilation in the presence of model error*, Monthly Weather Review, 143 (2015), pp. 1568–1582, https://doi.org/10.1175/MWR-D-14-00017.1.

[65] J. J. RUIZ, M. PULIDO, and T. MIYOSHI, *Estimating model parameters with ensemble-based data assimilation: A review*, Journal of the Meteorological Society of Japan. Ser. II,

91 (2013), pp. 79–99, https://doi.org/10.2151/jmsj.2013-201.

[66] J. J. RUIZ, M. PULIDO, and T. MIYOSHI, *Estimating model parameters with ensemble-based data assimilation: Parameter covariance treatment*, Journal of the Meteorological Society of Japan. Ser. II, 91 (2013), pp. 453–469, https://doi.org/10.2151/jmsj.2013-403.

[67] P. Sakov, D. S. Oliver, and L. Bertino, *An iterative enkf for strongly nonlinear systems*, Monthly Weather Review, 140 (2012), pp. 1988–2004, https://doi.org/10.1175/MWR-D-11-00176.1.

[68] B. Saltzman, *Finite amplitude free convection as an initial value problem—i*, Journal of the Atmospheric Sciences, 19 (1962), pp. 329–341, https://doi.org/10.1175/1520-0469(1962)019<0329:FAFCAA>2.0.CO;2.

[69] Y. SASAKI, *Some basic formalisms in numerical variational analysis*, Monthly Weather Review, 98 (1970), pp. 875–883, https://doi.org/10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2.

[70] S. Schirber, D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson, *Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world*, Journal of Advances in Modeling Earth Systems, 5 (2013), pp. 58–70, https://doi.org/10.1029/2012MS000167.

[71] G. Shutts, *A stochastic convective backscatter scheme for use in ensemble prediction systems*, Quarterly Journal of the Royal Meteorological Society, 141 (2015), pp. 2602–2616, https://doi.org/10.1002/qj.2547.

[72] C. Snyder and F. Zhang, *Assimilation of simulated doppler radar observations with an ensemble Kalman filter*, Monthly Weather Review, 131 (2003), pp. 1663–1677, https://doi.org/10.1175//2555.1.

[73] Y. Tr'emolet, *Accounting for an imperfect model in 4D-Var*, Quarterly Journal of the Royal Meteorological Society, 132 (2006), pp. 2483–2504, https://doi.org/10.1256/qj.05.224.

[74] Y. Trémolet, *Model-error estimation in 4D-Var*, Quarterly Journal of the Royal Meteorological Society, 133 (2007), pp. 1267–1280, https://doi.org/10.1002/qj.94.

[75] P. J. van Leeuwen, Y. Cheng, and S. Reich, *Nonlinear Data Assimilation*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Springer, 2015, https://doi.org/10.1007/978-3-319-18347-3.

[76] M. Verlaan and A. W. Heemink, *Nonlinearity in data assimilation applications: A practical method for analysis*, Monthly Weather Review, 129 (2001), pp. 1578–1589, https://doi.org/10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2.

[77] P. A. VIDARD, A. PIACENTINI, and F.-X. LE DIMET, *Variational data analysis with control of the forecast bias*, Tellus A, 56 (2004), pp. 177–188, https://doi.org/10.1111/j.1600-0870.2004.00057.x.

[78] T. Vukicevic and D. Posselt, *Analysis of the impact of model nonlinearities in inverse problem solving*, Journal of the Atmospheric Sciences, 65 (2008), pp. 2803–2823, https://doi.org/10.1175/2008JAS2534.1.

[79] A. Weisheimer, S. Corti, T. Palmer, and F. Vitart, *Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372 (2014), p. 20130290, https://doi.org/10.1098/rsta.2013.0290.

[80] J. S. Whitaker and T. M. Hamill, *Ensemble data assimilation without perturbed observations*, Monthly Weather Review, 130 (2002), pp. 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

[81] J. S. Whitaker and T. M. Hamill, *Evaluating methods to account for system errors in ensemble data assimilation*, Monthly Weather Review, 140 (2012), pp. 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

[82] S.-C. Yang, E. Kalnay, and B. Hunt, *Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model*, Monthly Weather Review, 140 (2012), pp. 2628–2646, https://doi.org/10.1175/MWR-D-11-00313.1.

[83] S. Zhang, Z. Liu, A. Rosati, and T. Delworth, *A study of enhancive parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model*, Tellus A: Dynamic Meteorology and Oceanography, 64 (2012), p. 10963, https://doi.org/10.3402/tellusa.v64i0.10963.

[84] D. Zupanski, *A general weak constraint applicable to operational 4DVAR data assimilation systems*, Monthly Weather Review, 125 (1997), pp. 2274–2292, https://doi.org/10.1175/1520-0493(1997)125<2274:AGWCAT>2.0.CO;2.