Correcting Observation Model Error
in Data Assimilation

Franz Hamilton*, Tyrus Berry† and Timothy Sauer‡

George Mason University, Fairfax, VA 22030, USA

*North Carolina State University, Department of Mathematics, Raleigh, 27695, USA

†George Mason University, Fairfax, VA 22030, USA

‡*Corresponding author address:* Timothy Sauer, George Mason University, Fairfax, VA 22030, USA

E-mail: tsauer@gmu.edu

ABSTRACT

Standard methods of data assimilation assume prior knowledge of a model that describes the system dynamics and an observation function that maps the model state to a predicted output. An accurate mapping from model state to observation space is crucial in filtering schemes when adjusting the estimate of the system state during the filter's analysis step. However, in many applications the true observation function may be unknown and the available observation model may have significant errors, resulting in a suboptimal state estimate. We propose a method for observation model error correction within the filtering framework. The procedure involves an alternating minimization algorithm used to iteratively update a given observation function to increase consistency with the model and prior observations, using ideas from attractor reconstruction. The method is demonstrated on the Lorenz 1963 and Lorenz 1996 models, and on a single-column radiative transfer model with multicloud parameterization.

1. Introduction

Data assimilation as a means of fusing mathematical models with observed data is a critical component of geophysical data analysis in general and numerical weather prediction in particular, and is steadily finding broader applications throughout nonlinear science. Standard applications of data assimilation algorithms require possession of the system equations of motion and observation modalities. In particular, the use of the Extended Kalman Filter (EKF) and Ensemble Kalman Filter (EnKF) (Houtekamer and Mitchell (1998); Burgers et al. (1998); Anderson (2001); Kalnay (2003); Rabier (2005); Hunt et al. (2004); Cummings (2005); Evensen (2007)) assume precise knowledge of the dynamical equations and the relationship between the system state and observables.

Some intriguing recent work has focused on investigating the effects of incomplete knowledge on this process, such as model error, missing equations and multiple sources of error in observations. In particular, the issue of observation errors, due to truncation, resolution differences, and instrument error, has received great attention (Dee (1995); Satterfield et al. (2017); Hodyss and Nichols (2015); Van Leeuwen (2015); Janjic et al. (2017); Berry and Sauer (2018)). In the case of unknown or incorrect observation models, there is interest in fixing these deficiencies. For example, a recent study Berry and Harlim (2017) discusses replacing an unknown observation function with a training set of observations and accompanying states.

In this article, an iterative approach to fixing observation model error is proposed which does not require training data, and can be applied as part of a sequential data assimilation implementation. The idea is based on an alternating minimization algorithm applied to the observation function. In the first step, a filter (eg. Kalman-type or variational filter) is applied to find the optimal state estimate based on the given observation model. In the second step, an observation model correction term is interpolated from the difference between the actual observations and the observation model applied to the state estimate produced by the filter; this interpolation is localized in the underlying phase space of the dynamical system. The model correction term is then applied to form a new observation model. The two steps are then repeated until convergence.

Fig. 2 shows an example application of the technique, to the Lorenz attractor with dynamical noise. The underlying model equations (the Lorenz equations) are assumed known. An initial guess is made for the observation function used in the filter, which is far from the function generating the observed data. Sequential filtering is applied iteratively, and the observation model correction is learned through the iteration. The RMSE of the filter decreased with iteration number, and after about a dozen iterations the minimum RMSE is approximately attained.

Several other examples illustrate the varying contexts in which the method can be applied. A critical hurdle for all filtering methods is the ability to scale up to large problems, which is typically achieved with a spatial localization. As a test case for spatiotemporal data we consider the Lorenz-96 system, in networks with 10 and 40 nodes. In the latter case, a spatial localization technique is developed which allows interpolation within each local region. Finally, we consider a more physically realistic example where observation model error can be especially detrimental to filtering, namely the case of radiative transfer models (RTM). To simulate severe observation model error, we assign the cloud fractions of a typical RTM to zero in the observation model. We then generate data using the full RTM (including the cloud fractions) and apply our method using the crippled observation function (with cloud fractions set to zero). The results show significant improvement in RMSE after three iterations of our observation model error correction algorithm.

The algorithm for correcting the observation model error is described in Section 2, along with its relation to alternating minimization methods in optimization theory, and details of its implementation in an ensemble filter. Sections 3 and 4 describe applications of the algorithm to Lorenz-63 and Lorenz-96 models, the latter to show how the method scales for spatiotemporal problems. The application to the radiative transfer model in shown in Section 5.

2. Filtering with an incorrect observation function

In the general filtering problem, we assume a system with n-dimensional state vector x and m-dimensional observation vector y defined by

$$\begin{aligned} x_k &= f(x_{k-1}) + w_{k-1} \\ y_k &= h(x_k) + v_k \end{aligned} \tag{1}$$

where w_{k-1} and v_k are white noise processes with covariance matrices Q and R, respectively. The function f represents the system dynamics and h is an observation function that maps the model state to a predicted output. The goal is to sequentially estimate the state of the system given some noisy observations. Below we will consider a specific filtering algorithm, however, at this point our approach can be formulated in terms of a generic filtering method.

a. The observation error correction algorithm

The effectiveness of standard filtering approaches is based on the assumption that the observation function h is perfectly known. The goal of this section is to address what happens when h is *not known*, and in its place an incorrect observation function g is used. In fact, observation model errors can have many sources, from truncation error due to downsampling high resolution state variables (also called representation error) to simple mismatch between the actual and available observation functions (often referred to as observation model error) Satterfield et al. (2017); Van Leeuwen (2015); Janjic et al. (2017). In this article we will take a very general outlook by considering h to be the true mapping from the fully resolved true state variables x_k into observed variables y_k, which is subject only to instrument error v_k. Meanwhile, g will denote a possibly incorrect mapping from state variables into observation variables which can be compared to the actual observations y_k. In such a situation, we can rewrite of the second part of Eq. 1 as

$$\begin{aligned} y_k &= h(x_k) + v_k \\ &= g(x_k) + b(x_k) + v_k \end{aligned} \tag{2}$$

where b is the error in our estimate resulting from use of the incorrect observation function. The term $b(\cdot)$ encapsulates all sources of error except for instrument noise which is the noise term v_k. We can write this error term as $b(x_k) = h(x_k) - g(x_k)$, or the difference between the true and incorrect observation functions at step k. Note that this error is dependent on the fully resolved state x_k.

Repairing observation model error was addressed recently Berry and Harlim (2017) by building a nonparametric estimate of the function b using a training set consisting of observations along with the corresponding true state. In the current article, we assume that the true state is *not* available. A novel approach will be proposed for empirically estimating the model error term b using only the observations y_k. We begin by describing our method generically for any filtering scheme. The general idea is to iteratively update the incorrect observation function g by obtaining successively improved estimates of the observation model error.

We make an initial definition $g^{(0)} = g$. The filter is given the known system dynamics f, the initial incorrect observation function $g^{(0)}$, and the observations y, and provides an estimate of the state at each observation time k, which we denote $x_k^{(0)}$. This initial state estimate will be subject to large errors, due to the unaccounted-for observation model error. Using this imperfect state estimate, we calculate a noisy estimate $\hat{b}_k^{(0)}$ of the observation model error, corresponding to observation y_k where

$$\hat{b}_k^{(0)} = y_k - g\left(x_k^{(0)}\right). \tag{3}$$

Due to noise in the data as well as the imperfection of the state estimate, $\hat{b}_k^{(0)}$ will not accurately reflect the true observation model error, $b(x_k)$. To build a better estimate of $b(x_k)$, we use a standard method of nonparametric attractor reconstruction (Takens (1981); Packard et al. (1980); Sauer et al. (1991); Sauer (2004)) to interpolate the observation model error function, as follows. Given observation y_k, introduce the delay-coordinate vector $z_k = [y_k, y_{k-1}, \ldots, y_{k-d}]$, with d delays. The vector z_k is a representation of the system state Takens (1981); Sauer et al. (1991). The reconstruction is built by locating the N nearest neighbors $z_{k_1}, ..., z_{k_N}$ (with respect to Euclidean distance), where

$$z_{k_j} = [y_{k_j}, y_{k_j-1}, \ldots, y_{k_j-d}]$$

within the set of observations. The corresponding $\hat{b}_{k_1}^{(0)}, \hat{b}_{k_2}^{(0)}, \ldots, \hat{b}_{k_N}^{(0)}$ values are used to estimate $b(x_k)$ by the weighted average

$$b^{(0)}(x_k) = w_{k_1}\hat{b}_{k_1}^{(0)} + w_{k_2}\hat{b}_{k_2}^{(0)} + \ldots + w_{k_N}\hat{b}_{k_N}^{(0)}. \tag{4}$$

The weights may be chosen in many different ways (Hamilton et al. (2016, 2017)). To impose smoothness on the function $b^{(0)}$, we could use weights which decay exponentially in delay space distance. Namely, the weight for the j^{th} neighbor can be defined as

$$w_{k_j} = \frac{e^{-||z_{k_j} - z_k||/\sigma}}{\sum_{j=1}^{N} e^{-||z_{k_j} - z_k||/\sigma}}.$$

Here, $||z_{k_j} - z_k||$ is the distance of the j-th nearest neighbor, z_{k_j}, to the current delay-coordinate vector, z_k, and σ is the bandwidth which controls the weighting of the neighbors in the local model. Methods are available to tune the σ variable. In this work, we set it to half of the mean distance of the N nearest neighbors to give a smooth roll off of the weights with distance. This choice adapts to the local density of the data.

Note that Eq. (4) is still just an approximation of $b(x_k)$, although a more accurate estimate compared to Eq. (3). Our observation function can now be updated, namely

$$g^{(1)} = g + b^{(0)}.$$

This improved observation function is given to the filter, and the data are re-processed. An improved state estimate, $x_k^{(1)}$, at time k is obtained, a more accurate reconstruction, $b^{(1)}(x_k)$, of the observation model error is formed using Eqs. (3-4) and the observation function is again updated, $g^{(2)} = g + b^{(1)}$.

The method continues iteratively, each iteration an improved reconstruction of $b(x_k)$ is obtained resulting in a better estimate of the state on the next iteration. The method is summarized for steps $\ell = 0, 1, 2, \ldots$ as follows:

1. Initialize $g^{(0)} = g$, $\Delta g = \mathrm{Inf}$
2. While Δg is greater than threshold
   (a) For each observation y_k, use filter to estimate state $x_k^{(\ell)}$ given known f and observation function $g^{(\ell)}$
   (b) Calculate the noisy observation model error estimates $\hat{b}_k^{(\ell)} = y_k - g(x_k^{(\ell)})$
   (c) For each k, find the N-nearest neighbors of delay vector z_k and set

$$b^{(\ell)}(x_k) \quad = \quad w_{k_1}\hat{b}_{k_1}^{(\ell)} + w_{k_2}\hat{b}_{k_2}^{(\ell)} + \ldots + w_{k_N}\hat{b}_{k_N}^{(\ell)} \tag{5}$$

   (d) Update the observation function, $g^{(\ell+1)} = g + b^{(\ell)}$
   (e) Update $\Delta g = \frac{1}{T}\sum_{k=1}^{T} |\hat{b}_k^{(\ell)} - \hat{b}_k^{(\ell-1)}|$

In the absence of results on convergence for most nonlinear Kalman-type filters it is difficult to analyze the convergence of our method. At each step of the algorithm we estimate the local average of the observation model error from the previous estimates $\hat{b}_k^{(\ell)}$ and then add this estimate to the observation function. Notice that if the same state estimates $x_k^{(\ell+1)} = x_k^{(\ell)}$ were found in the next iteration of the Kalman filter, then the observation model error estimates would be unchanged. Informally, if the state estimates only change by a small amount and if g is continuous then the observation model error estimates should also only change by a relatively small amount. In the next section we will present an interpretation of the method as an alternating minimization approach for estimating the local observation model error parameters. Moreover, we will present numerical results demonstrating convergence for strongly nonlinear systems with extremely large error in the specification of the observation function.

b. Interpretation as alternating minimization algorithm

The method introduced above can be viewed as belonging to the family of projection algorithms in optimization theory called alternating minimization algorithms Wang et al. (2008); Niesen et al. (2009). Implicit to the above construction is the following nonparametric representation of the estimated global observation model error $b^{(\ell)}(x)$, which interpolates the errors at each x_k as

$$b^{(\ell)}(x_k) = \sum_{i=1}^{N} \hat{b}_{k_i}^{(\ell)} \frac{e^{-||z_{k_j} - z_k||/\sigma}}{\sum_{j=1}^{N} e^{-||z_{k_j} - z_k||/\sigma}} = \sum_{j=1}^{N} \hat{b}_{k_j}^{(\ell)} \frac{e^{-d(x,x_{k_j})/\sigma}}{\sum_{j=1}^{N} e^{-d(x,x_{k_j})/\sigma}},$$

where $\{x_{k_j}\}_{j=1}^{N}$ are the N nearest neighbors of the input x. Takens' theorem Takens (1981); Sauer et al. (1991) states that we can use the delay coordinate vectors z_{k_j} as a proxy for the unknown true states x_{k_j}. Using the Euclidean distance on the proxy vectors z_{k_j} implicitly changes the distance function in state space to a metric d, which is consistent since all metric are equivalent in Euclidean space, and this has really only affected the weights in the average. Notice that the finite set of parameters $\{\hat{b}_k^{(\ell)}\}$ determine the function $b^{(\ell)}(x)$. From (2) we assume that

$$y_k = g(x_k) + b(x_k) + \nu_k$$

where v_k is mean zero Gaussian noise with covariance matrix R. Thus, the likelihood of the estimated observation model error $b^{(\ell)}(x)$ can be estimated on the data set as

$$P\left(x_k^{(\ell)} \,|\, b^{(\ell)}\right) \propto \prod_{k=1}^{T} \exp\left(-\frac{1}{2}||y_k - g(x_k^{(\ell)}) - b^{(\ell)}(x_k^{(\ell)})||_R^2 - \frac{1}{2}||x_{k+1}^{(\ell)} - f(x_k^{(\ell)})||_Q^2\right) \tag{6}$$

where $||v||_R^2 = v^\top R^{-1} v$ is the norm induced by the covariance matrix R. Our goal is to maximize the probability simultaneously with respect to both the state estimate $x_k^{(\ell)}$ and the observation model error estimate $\hat{b}^{(\ell)}$, or equivalently, to minimize $-\log P$, the negative log likelihood.

At the ℓ-th step of our approach, we first fix the observation model error estimate $b^{(\ell)}$ and use the nonlinear Kalman filter to approximate the best estimate of the state $x_k^{(\ell)}$ given the current estimate of the observation model error. The nonlinear Kalman filter is approximating the solution which maximizes (6) where $b^{(\ell)}$ is fixed. One could also apply a variational filtering method to achieve this maximization.

Next, we fix the estimate $x_k^{(\ell)}$ and estimate the parameters $\hat{b}_k^{(\ell+1)}$ to maximize (6). Since the second term in the exponential is independent of $\hat{b}_k^{(\ell+1)}$, the solution which maximizes (6) is simply the solution to the linear system of equations

$$y_k - g(x_k^{(\ell)}) = b^{(\ell)}(x_k^{(\ell)}) = \sum_{j=1}^{N} \hat{b}_{k_j}^{(\ell)} \frac{e^{-d(x,x_{k_j})/\sigma}}{\sum_{j=1}^{N} e^{-d(x,x_{k_j})/\sigma}}. \tag{7}$$

Instead of explicitly solving this system, in our implementation we simply used the approximate solution given by

$$\hat{b}_k^{(\ell)} = y_k - g(x_k^{(\ell)}) \tag{8}$$

since each point is its own nearest neighbor and $d_{k_1} = 0$ yields the largest weight in the summation. In Fig. 1 we show that the observation model error estimates (7) and (8) are very similar, but (8) is much faster to compute and is more numerically stable so we will use (8) in all the examples below.

c. Ensemble Kalman filtering with observation model error correction

In this section we assume a nonlinear system with n-dimensional state vector x and m-dimensional observation vector y defined by (1). The ensemble Kalman filter (EnKF) is a data assimilation algorithm designed for nonlinear systems, that forms an ensemble of states to handle the nonlinearity., One simple implementation is known as the unscented transformation (see Simon (2006); Julier et al. (2000, 2004), for example). The state estimate at step $k-1$ is denoted x^+_{k-1} and the covariance matrix is denoted P^+_{k-1}. The unscented version of the EnKF employs the singular value decomposition to calculate S^+_{k-1}, the symmetric positive definite square root of P^+_{k-1}. The singular directions form an ensemble of E new state vectors at step $k-1$, where $x^+_{i,k-1}$ identifies the i^{th} ensemble member .

On each step, the EnKF applies a forecast, predicting the state, followed by analysis, correcting the state prediction with benefit of the current observation. The model f advances the ensemble one time step, and then the observation function $g^{(\ell)}$ is applied:

$$\begin{aligned} x^-_{i,k} &= f\left(x^+_{i,k-1}\right) \\ y^-_{i,k} &= g^{(\ell)}\left(x^-_{i,k}\right). \end{aligned} \tag{9}$$

Notice that in the ideal filtering situation we would apply the true observation function h in (9). In this context of this article, we assume that we are only given an incorrect observation function g. In the initial iteration of the filter ($\ell = 0$) we simply use the best available observation function $g^{(0)} = g$, and in future iterations $(\ell > 0)$ we incorporate the ℓ-th observation model error estimate to form $g^{(\ell)} = g + \hat{b}^{(\ell)}$ as described above. Notice that each ensemble member has the same correction $\hat{b}^{(\ell)}$ applied since the correction is computed based on the neighbors in delay-embedded observation space, so the neighbors do not change based on the state estimate or iteration of the algorithm. We emphasize that the state estimate and observation model error estimates change at each iteration, but the indices of the neighbors, $k_1,...,k_N$ that are used to estimate the observation model error at time step k do not change (they are independent of ℓ).

The prior state estimate x_k^- is defined to be the mean of the state ensemble, and the predicted observation y_k^- is defined to be the mean of the observed ensemble. Define P_k^- and P_k^y to be the covariance matrices of the resulting state and observed ensembles, and let P_k^{xy} denote the cross-covariance matrix of the state and observed ensembles. More precisely, in the notation of Hamilton et al. (2017), we set

$$\begin{aligned}
P_k^- &= \frac{1}{E}\sum_{i=1}^{E}\left(x_{i,k}^- - x_k^-\right)\left(x_{i,k}^- - x_k^-\right)^T + Q \\
P_k^y &= \frac{1}{E}\sum_{i=1}^{E}\left(y_{i,k}^- - y_k^-\right)\left(y_{i,k}^- - y_k^-\right)^T + R \\
P_k^{xy} &= \frac{1}{E}\sum_{i=1}^{E}\left(x_{i,k}^- - x_k^-\right)\left(y_{i,k}^- - y_k^-\right)^T.
\end{aligned} \tag{10}$$

Then the Kalman update equations

$$\begin{aligned}
K_k &= P_k^{xy}(P_k^y)^{-1} \\
P_k^+ &= P_k^- - K_k P_k^{yx} \\
x_k^+ &= x_k^- + K_k\left(y_k - y_k^-\right).
\end{aligned} \tag{11}$$

are used to update the state x_k^+ and covariance estimates P_k^+ with the observation y_k. The covariance matrices Q and R are quantities that have to be known *a priori* or estimated from the data.

The method of Berry and Sauer (2013) will be used for the adaptive estimation of the covariance matrices Q and R. This is a key component in our method since the R covariance will be inflated by the adaptive filter to represent the error between the true observation function h and the observation function $g^{(\ell)}$ that we actually use in the filter. In other words, the adaptive filter is combining the covariance of the observation model error and the instrument noise into the R covariance matrix.

As we iterate the algorithm (as ℓ increases) we find that $g^{(\ell)}$ more closely approximates the true observation function h and the adaptive filter will find smaller values for R.

3. Assimilating Lorenz-63 with an incorrect observation model

In the results presented below, we assume noisy observations are available from a system of interest and we implement an ensemble Kalman filter (EnKF) for state estimation. The EnKF approximates a nonlinear system by forming an ensemble, such as through the unscented transformation (see for example Simon (2006)). Additionally, we use the method of Berry and Sauer (2013) for the adaptive estimation of the filter noise covariance matrices Q and R. The correct observation function h that maps the state to observation space is unknown, and in its place an incorrect function g is chosen for use by the EnKF. Throughout, we will compare our corrected filter with the standard filter (essentially, the $\ell = 0$ iteration) which assumes no correction.

As a feasibility test we consider the Lorenz-63 system Lorenz (1963)

$$\begin{aligned} \dot{x_1} &= \sigma(x_2 - x_1) \\ \dot{x_2} &= x_1(\rho - x_3) - x_2 \\ \dot{x_3} &= x_1 x_2 - \beta x_3 \end{aligned} \tag{12}$$

where $\sigma = 10$, $\rho = 28$, $\beta = 8/3$. We will assimilate 8000 noisy observations of the system, sampled at rate $dt = 0.1$, to which we add independent Gaussian observational noise, ν_k, with mean zero and covariance $R = 2I_{3\times 3}$. Our goal is to filter the observations

$$\vec{y} = h(\vec{x}) + \nu_k$$

(see Fig. 2, blue circles) and reconstruct the underlying state, $\vec{x}$, (Fig. 2, solid black lines). However, we assume that the true observation function h, given by

$$h(\vec{x}) = h\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} \sin(x_1) \\ x_2 - 6 \\ \cos(x_3) \end{bmatrix}$$

is unknown to us. Instead, the EnKF will use an incorrect observation function g, given by

$$g(\vec{x}) = g\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Using the incorrect mapping g, and with no estimate of the observation model error, the filter's reconstruction of the system state suffers substantially (Fig. 2(a)-(c), solid gray lines). We should note that even obtaining these poor estimates requires adaptive estimation of the system and observation noise covariance matrices Q and R used by the EnKF. The RMSE for reconstructing the three Lorenz-63 variables x_1, x_2 and x_3 using an EnKF with observation function g and no observation model error correction is 8.10, 6.77 and 22.33 respectively. This is not surprising, since without the correct observation function the analysis step of the EnKF, where the state and covariance estimates are updated, suffers due to the errors in mapping the predicted state to observation space.

Using our proposed method, the EnKF state estimate can be improved by iteratively building an approximation of the observation model error, essentially augmenting our observation function. In building our reconstruction of the observation model error, we use $d = 2$ delays and $N = 100$ nearest neighbors. After $M = 20$ iterations of our method, we are able to obtain and accurate estimate of the Lorenz-63 state (Fig. 2(a)-(c), solid red lines). The resulting error in our estimates is significantly smaller (RMSE of 2.11, 1.77 and 2.91 for x, y and z respectively) compared to filtering without an observation model error correction.

Fig. 2(d) shows the error in our estimation of x (solid black line), y (dashed black line) and z (dotted black line) as a function of number of iterations of our algorithm. We note that $\ell = 0$ corresponds to running the EnKF without any observation model error. At each iteration, we obtain a better reconstruction of the observation model error which helps improve our estimate of the state in the next iteration. At a certain point, our reconstruction of the observation model error and system state converges, a period indicated by the plateau in our RMSE plot.

4. Spatiotemporal observation model error correction

To show the method can work in a spatially extended system, we consider the system introduced by Lorenz (1996), which represents a ring of K nodes coupled by the equations

$$\dot{x}_i = (ax_{i+1} - x_{i-2})x_{i-1} - x_i + F \tag{13}$$

with parameter settings $a = 1$ and $F = 8$. The Lorenz-96 system exhibits higher dimensional complex behavior, that can be adjusted by changing the number of nodes and the forcing parameter F. In this example, we generate 10000 observations, corrupted by mean-zero Gaussian noise with variance equal to 2, from each node in the ring. Denoting $\mathbf{x} = [x_1, x_2, \ldots, x_K]$, the true observation function h for this system is defined as $h(\mathbf{x}) = C\mathbf{x}$, where

$$C = \begin{bmatrix} c_1 & c_2 & 0 & \cdots & \cdots & \cdots & \cdots & c_3 \\ c_3 & c_1 & c_2 & 0 & & & & \vdots \\ 0 & c_3 & c_1 & c_2 & \ddots & & & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & & & \ddots & c_3 & c_1 & c_2 & 0 \\ \vdots & & & & 0 & c_3 & c_1 & c_2 \\ c_2 & \cdots & \cdots & \cdots & \cdots & 0 & c_3 & c_1 \end{bmatrix},$$

$c_1 = 1, c_2 = 1.2, c_3 = 1.1$. In effect, our observations at each node in the ring is a linear combination of the current node and its two spatial neighbors. The true observation map h is assumed unknown to us, and in its place we use the incorrect function

$$g(\mathbf{x}) = \mathbf{I}_{K\times K}\mathbf{x}$$

where $I_{K\times K}$ is the $K \times K$ identity matrix.

We first consider a $K = 10$ dimensional Lorenz-96 ring. Fig. 3 shows the results of reconstructing the 10 dimensional Lorenz-96 state. Fig. 3(a) shows a representative reconstruction of the x_2 state (similar results are obtained for each node of the ring). Given the noisy observations (blue circles), the EnKF without observation model error correction (solid gray line) is unable to estimate the true trajectory (solid black line), resulting in an RMSE of 5.83. Accounting for the observation model error ($M = 15$ iterations, $d = 2$ delays and $N = 100$ neighbors), we are able to improve our reconstruction of the x_2 trajectory (solid red line, RMSE = 2.37). Similarly as in the Lorenz-63 example, we see in Fig. 3(b) that as the number of iterations of our observation model error correction method increases we eventually converge to a stable RMSE.

We next consider a $K = 40$ dimensional ring. Fig. 4 shows the spatiotemporal plots of the system. The top plot shows the true system dynamics and the second plot our noisy observations of the system. Similarly to the 10 dimensional ring, the filtering without observation model error correction is unable to provide an accurate reconstruction of the system state (third plot). The high dimensionality of the system can make finding accurate nearest neighbors for observation model error reconstruction difficult. We implement a spatial localization technique when finding neighbors, whereby for each node we look for neighbors in a delay-coordinate space consisting of its delays and the delays of its two spatial neighbors. While our method can be successfully implemented in this high dimensional example without localization, results are improved through use of the localization technique. The bottom plot of Fig. 4 shows the resulting filter estimate with observation model error correction. Again, we see that there is a substantial improvement in the state reconstruction and we are able to obtain a more accurate representation of the true system dynamics.

5. Correcting error in cloudy satellite-like observations without training data

The presence of clouds is a significant issue in assimilation of satellite observations. Clouds can introduce significant observation model error into the results of radiative transfer models (RTM). As previously mentioned, a recently developed method Berry and Harlim (2017) is able to learn a probabilistic observation model error correction using training data consisting of pairs of the true state and the corresponding observations. Of course, requiring knowledge of the true state in the training data is a significant restriction, and while methods such as reanalysis or local large-scale data gathering are possible, it would be extremely advantageous to remove this requirement. The innovation of the method introduced here is that we do not require knowledge of the true state in the training data. Instead, we use an iterative approach to learn local observation model error corrections based on delay reconstruction in observation space. In this section we will apply our method to an RTM and show that the observation model error can be iteratively learned without the training data.

The model Khouider et al. (2010) presented here represents a single column of atmosphere with three temperature variables $\theta_1, \theta_2, \theta_{eb}$ and a vertically averaged water vapor variable q. The RTM also contains a stochastic multicloud parameterization with three variables f_c, f_d, and f_s which represent fractions of congestus, deep, and stratiform clouds respectively. The three temperature variables are extrapolated to yield the temperature as a continuous function of the height, and then a simplified RTM can be used to integrate over this vertical profile to determine the radiation at various frequencies (see Berry and Harlim Berry and Harlim (2017) for details). We follow Liou Liou (2002) to incorporate information from the cloud fractions into the RTM in order to produce synthetic 'true' observations at 16 different frequencies. Each frequency has a different height profile which is integrated against the vertical temperature profile. The presence of the different types of clouds influences these height profiles to simulate the cloud 'blocking' radiation from below it. We first show that the EnKF is capable of recovering most of the state variables from the observations when the correct observation model is specified (meaning the RTM includes the cloud fraction information from the model). In Fig. 5 we show the true state (grey) along with the estimates produced using the correct observation model (black).

Next, we assume that the cloud fractions are unknown or that their effect on the RTM is poorly understood, and we attempt to assimilate the true observations using an RTM where the cloud fractions are held constant at zero (note that the cloud fractions are still present and evolving in the model used by the filter, but they are not included in the RTM used for the observation function of the filter). We should note that this assimilation is impossible without artificially inflating the observation covariance matrix R by a factor of 100. The results of assimilating are shown in Fig. 5

(red, dotted). Finally we apply the iterative observation model error correction (3 iterations) and the results are shown in Fig. 5 (blue, dashed). Similar to the results of Berry and Harlim Berry and Harlim (2017) the water vapor variable, q is difficult to reconstruct in the presence of observation model error, however the cloud and temperature variables are significantly improved.

In Table 1 we summarize the RMSE of each variable averaged over 4500 discrete filter steps (15.6 model time units with $dt = .0035$) for each filter, the observation noise variance was set at 0.5% of the variance of each observed variable. The observation model error correction is able to improve the estimation of all of the cloud fraction variables f_c, f_d, and f_s along with two of the temperature variables. The estimation of θ_2 was only slightly degraded. The estimation of q was more significantly degraded by the observation model error correction, probably because q does not enter into the observation function as directly as the other variables. These results compare favorably with Berry and Harlim Berry and Harlim (2017) who also found that the q variable was difficult to reconstruct in the presence of this observation model error, even using training data that included the true state.

Since our approach here does not depend on perfect training data, we also found that our results were more robust to observation noise than the results of Berry and Harlim Berry and Harlim (2017). In that approach, this was a significant issue since it was assumed that the observation noise was small in order to be able to recover the true model error from the training data. As a result, the results were only robust up to observation noise levels of about 1% of the variance of the observations.

In Fig. 6 we show the robustness of the observation model error correction proposed here to increasing levels of observation noise. We find that the iterative observation model error correction is robust at noise levels over 10% of the variance of the observations. At extremely low noise levels, such as levels near 0.1%, the method of Berry and Harlim (2017) has performance comparable to the true observation function, so when perfect full state training data is available and observation noise is small the methods have roughly equivalent behavior.

6. Discussion

Accurate linear and nonlinear filtering depends on thorough knowledge of model dynamics and the function connecting states to observations. The method proposed here uses an alternating minimization approach to iteratively correct observation model error, assuming knowledge of the correct dynamical model. This approach was shown to succeed in temporal and spatiotemporal examples as well as a cloud model.

Although the iteration converges to eliminate observation model error in a wide variety of examples, there is no proof of global convergence of the method. This is typical for alternating minimization methods. A better understanding of the basin of convergence would be helpful, and the object of further study.

The increasing diversity of measurement devices used in meteorological data assimilation is subject to a wide variety of separate errors. It is possible that more refined versions of the method can be designed to target particular subsets of the total observation error. The proof of concept carried out in this article show the potential for a relatively simple iterative solution to the problem, that can result in significant improvement in total RMSE.

We envision additional applications in other science and engineering areas, including hydrology, physical and biological experiments. A particular problem of interest in physiology is the common usage of intracellular neural models to assimilate extracellular measurements from single electrodes and electrode arrays. The observation function that connects such measurements to the model is not well understood by first principles and may vary by preparation. An automated way to solve this issue would potentially be a significant advance in data assimilation for neuroscience problems.

Acknowledgments. This research was partially supported by grants RTG/DMS-1246991 and DMS-1723175 from the National Science Foundation.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, **129**, 2884–2903.

Berry, T., and J. Harlim, 2017: Correcting biased observation model error in data assimilation. *Monthly Weather Review*, **145**, 2833–2853.

Berry, T., and T. Sauer, 2013: Adaptive ensemble Kalman filtering of nonlinear systems. *Tellus A*, **65**, 20 331.

Berry, T., and T. Sauer, 2018: Correlation between system and observation errors in data assimilation. *In revision, Monthly Weather Review*.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, **126**, 1719–1724.

Cummings, J. A., 2005: Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3583–3604, doi:10.1256/qj.05.105.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, **123 (4)**, 1128–1145.

Evensen, G., 2007: *Data assimilation: The Ensemble Kalman Filter*. Springer: Heidelberg.

Hamilton, F., T. Berry, and T. Sauer, 2016: Ensemble Kalman filtering without a model. *Phys. Rev. X*, **6**, 011 021, doi:10.1103/PhysRevX.6.011021, URL https://link.aps.org/doi/10.1103/PhysRevX.6.011021.

Hamilton, F., T. Berry, and T. Sauer, 2017: Kalman-Takens filtering in the presence of dynamical noise. *European Physical Journal Special Topics*, **226**, 3239–3250.

Hodyss, D., and N. Nichols, 2015: The error of representation: basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, **67 (1)**, 24 822, doi:10.3402/tellusa.v67.24822.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811.

Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus A*, **56 (4)**, 273–277.

Janjic, T., and Coauthors, 2017: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3130.

Julier, S., J. Uhlmann, and H. Durrant-Whyte, 2000: A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Control*, **45**, 477–482.

Julier, S., J. Uhlmann, and H. Durrant-Whyte, 2004: Unscented filtering and nonlinear estimation. *Proc. IEEE*, **92**, 401–422.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability*. Cambridge Univ. Press.

Khouider, B., J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8 (1)**, 187–216.

Liou, K.-N., 2002: *An introduction to atmospheric radiation*, Vol. 84. Academic press.

Lorenz, E., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proceedings: Seminar on predictability*, AMS, Reading, UK, Vol. 1, 1—18.

Niesen, U., D. Shah, and G. W. Wornell, 2009: Adaptive alternating minimization algorithms. *IEEE Transactions on Information Theory*, **55 (3)**, 1423–1429.

Packard, N., J. Crutchfield, D. Farmer, and R. Shaw, 1980: Geometry from a time series. *Phys. Rev. Lett.*, **45**, 712–715.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3215–3233.

Satterfield, E., D. Hodyss, D. D. Kuhl, and C. H. Bishop, 2017: Investigating the use of ensemble variance to predict observation error of representation. *Monthly Weather Review*, **145 (2)**, 653–667, doi:10.1175/MWR-D-16-0299.1.

Sauer, T., 2004: Reconstruction of shared nonlinear dynamics in a network. *Phys. Rev. Lett.*, **93**, 198 701–4.

Sauer, T., J. Yorke, and M. Casdagli, 1991: Embedology. *J. Stat. Phys.*, **65**, 579–616.

Simon, D., 2006: *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. John Wiley and Sons.

Takens, F., 1981: Detecting strange attractors in turbulence. *Lecture Notes in Math. Springer-Verlag: Berlin*, **898**.

Van Leeuwen, P. J., 2015: Representation errors and retrievals in linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **141 (690)**, 1612–1623.

Wang, Y., J. Yang, W. Yin, and Y. Zhang, 2008: A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, **1 (3)**, 248–25.

LIST OF TABLES

Percent Error (RMSE)	θ_1	θ_2	θ_{eb}	q	f_c	f_d	f_s
True Observation Function	2.8	1.6	6.2	10.6	8.1	3.1	8.2
Wrong Observation Function	30.3	9.1	51.0	62.8	44.2	76.2	93.1
Model Error Correction	11.8	12.0	31.5	103.9	15.6	25.8	45.4

TABLE 1. Root mean squared error of cloud model variables averaged over 4500 filter steps. Estimation of the cloud fraction variables is significantly improved by the observation model error correction.

LIST OF FIGURES

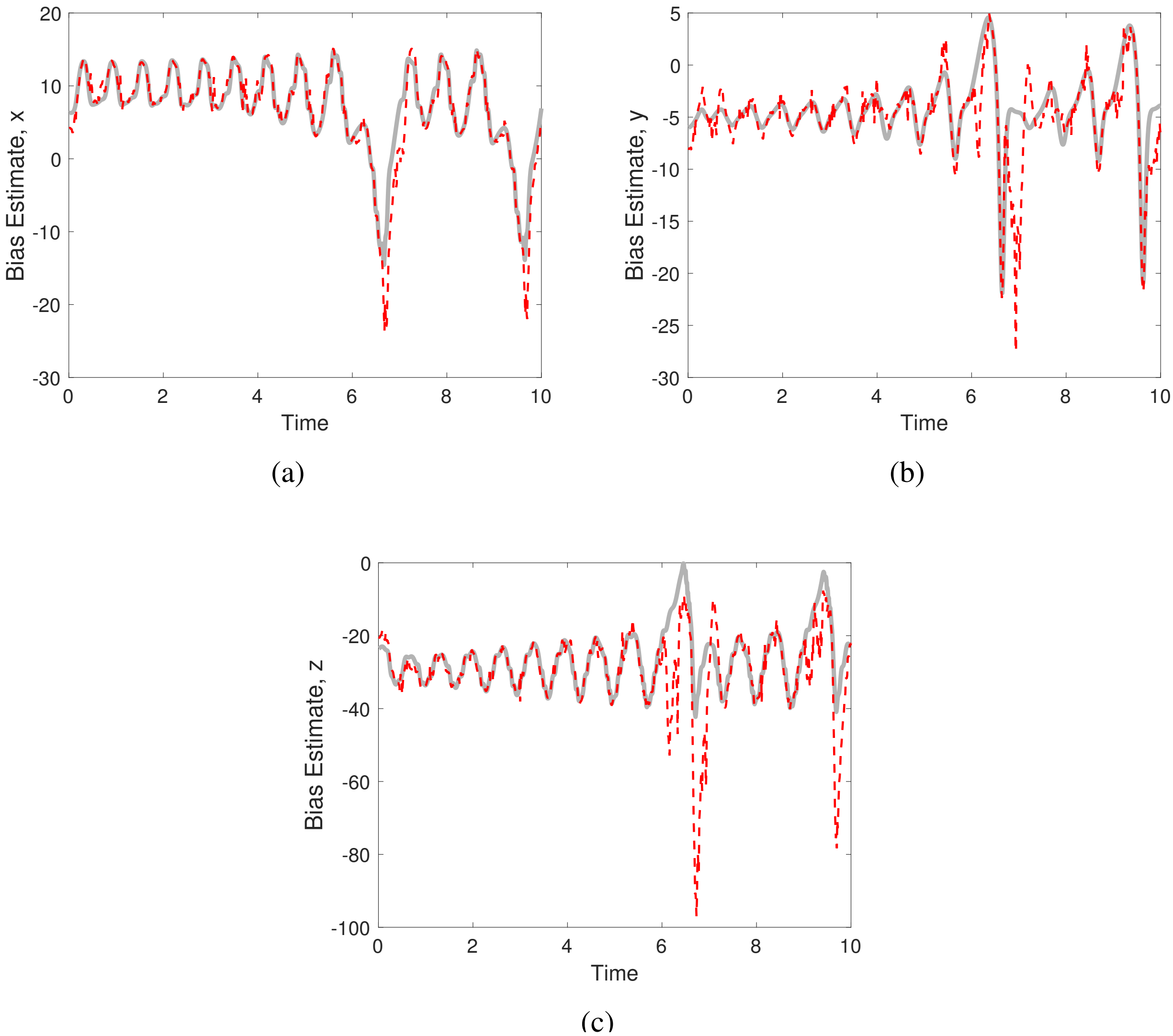


FIG. 1. Comparison of the observation model error correction which solves (7) (red, dashed) to the correction given by (8) (grey, solid) which is used in all the examples below. Observation errors are shown from the Lorenz-63 example described below (see Fig. 2).

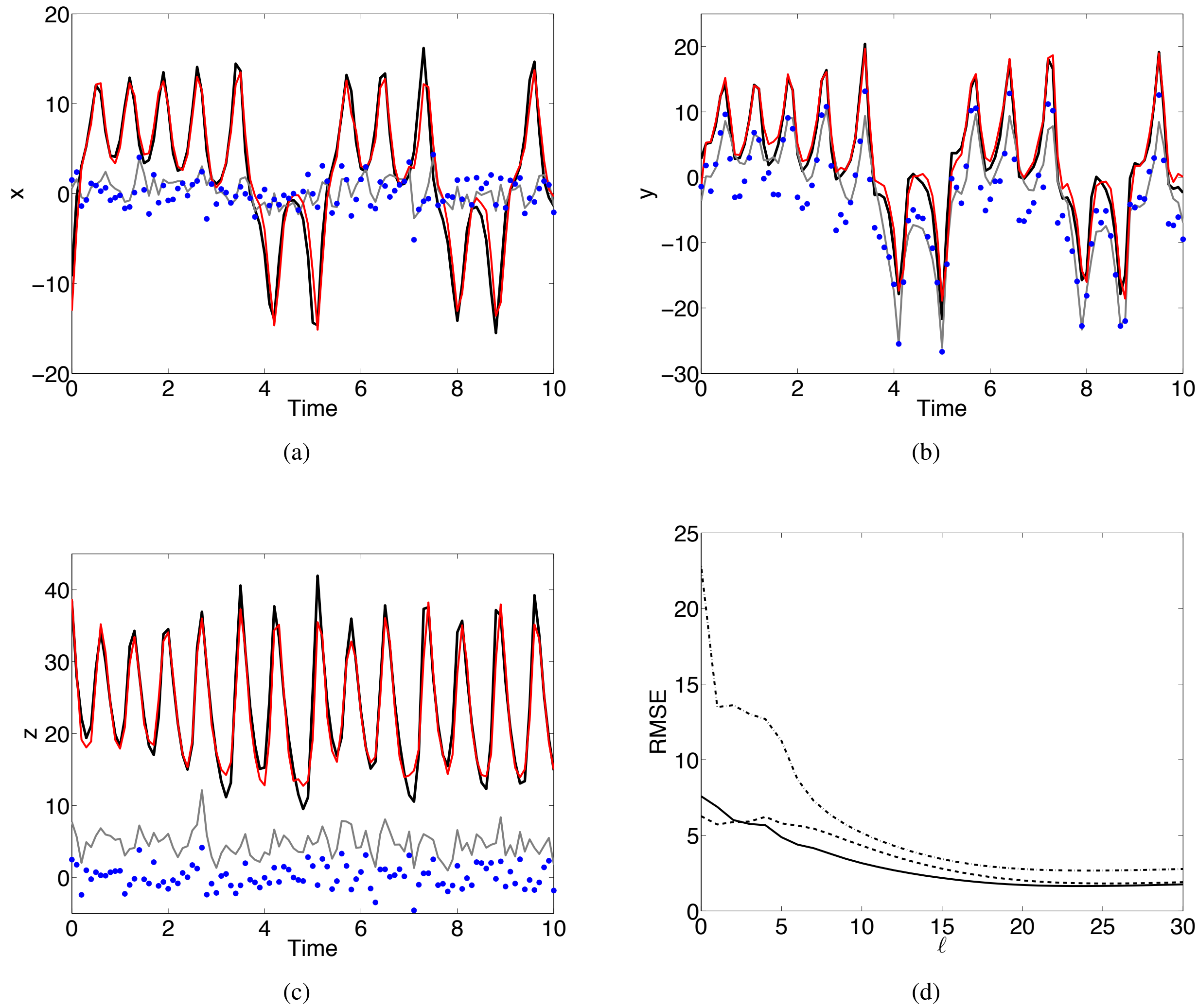
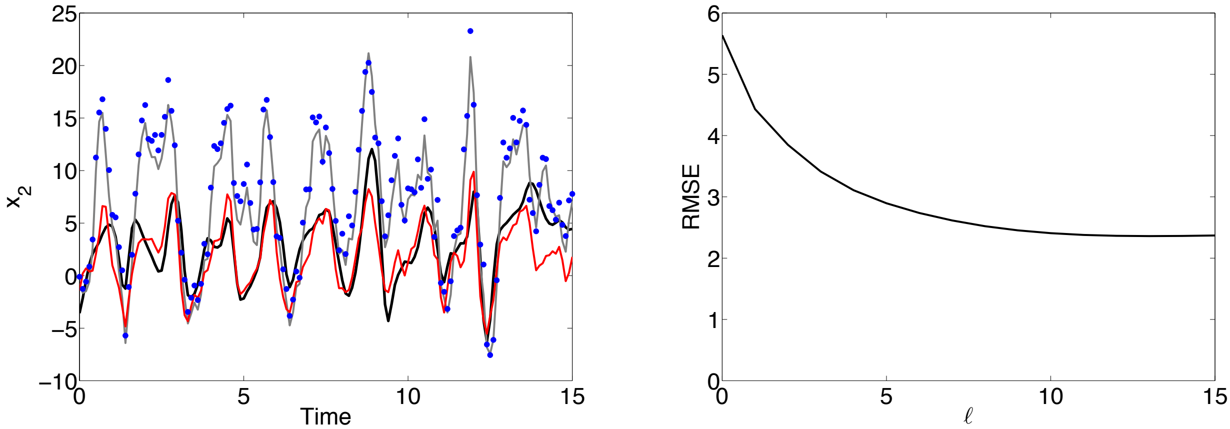


FIG. 2. Results of filtering noisy Lorenz-63 (a) x_1 (b) x_2 and (c) x_3 time series when true observation function, h, is unknown and $R = 2I_{3\times 3}$. Notice the large difference between the true observations $h(\vec{x}_k) + \nu_k$ (blue circles) to the true state variables (solid black curve). We compare the EnKF estimate using the wrong observation function, g, without observation model error correction (solid gray lines) and the EnKF estimate with correction (solid red lines) shown. (d) Plot of RMSE vs. iteration of the observation model error correction method, where $\ell = 0$ corresponds to the standard EnKF without correction. RMSE for x (solid black line), y (dashed black line) and z (dotted black line) shown. After a sufficient number of iterations, the observation model error estimates converge as does the RMSE of the state estimate.



(a) (b)

FIG. 3. Results of filtering a noisy 10 dimensional Lorenz- 96 ring when the true observation function is unknown. (a) Representative results demonstrated by the x_2 node. We filter the noisy observation (blue circles) in an attempt to reconstruct the underling state (solid black line). Without observation model error correction, the EnKF estimate (solid gray line) is unable to track the true state (RMSE = 5.83). With observation model error correction (solid red line), our estimate of the state improves substantially (RMSE = 2.37). (b) Average RMSE of Lorenz-96 ring as a function of iteration shown. Similarly to the previous example, after a sufficient number of iterations our method converges to an estimate of the observation model error and system state, demonstrated by the convergence of the RMSE.

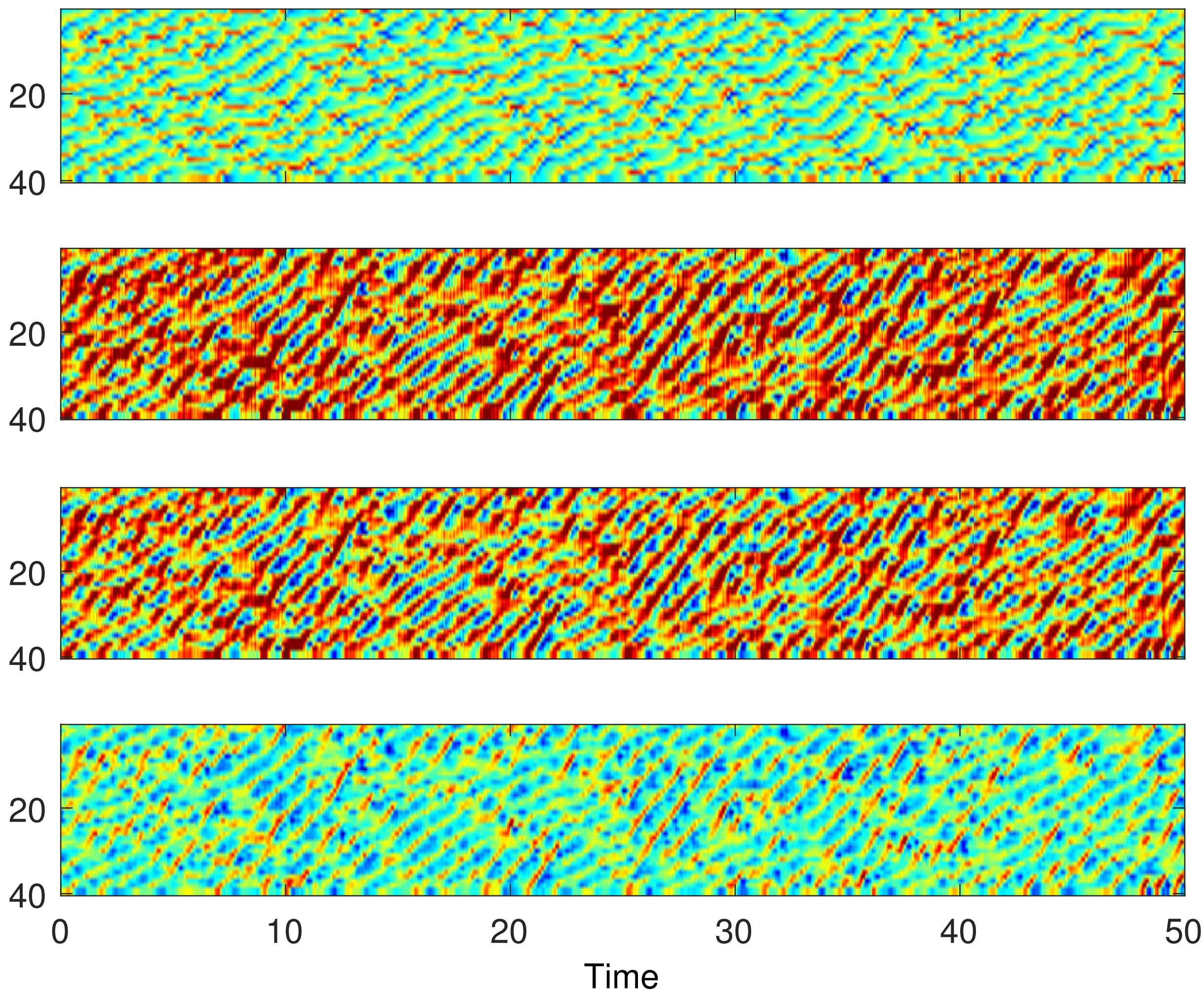


FIG. 4. Results of filtering a noisy 40 dimensional Lorenz-96 system. True spatiotemporal dynamics (top), noisy observations of the system (second plot), estimate without observation model error correction (third plot) and estimate with observation model error correction (bottom plot) shown. Without correction, we obtain a poor estimate of the system dynamics (average RMSE = 5.12). With correction, our estimate is improved (average RMSE = 2.50).

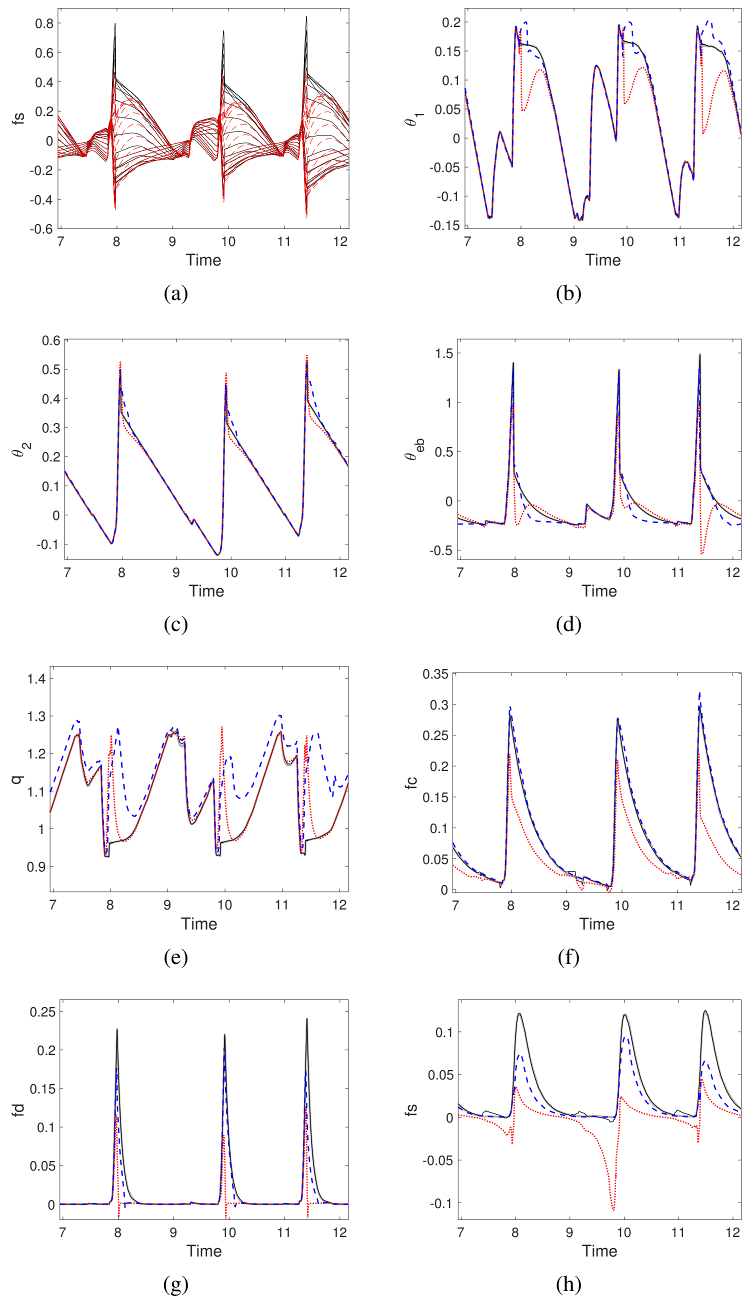


FIG. 5. (a) True observations (red, dashed) incorporating cloud information are compared to the incorrect observation function (black, solid) which sets all the cloud fractions to zero in the RTM. (b-h) True state (gray, thick curve) compared to the result of filtering with the true observation function (black), the wrong observation function using only inflation of the observation covariance matrix (red, dashed) and the wrong observation function with iterative observation model error correction (blue, dashed).

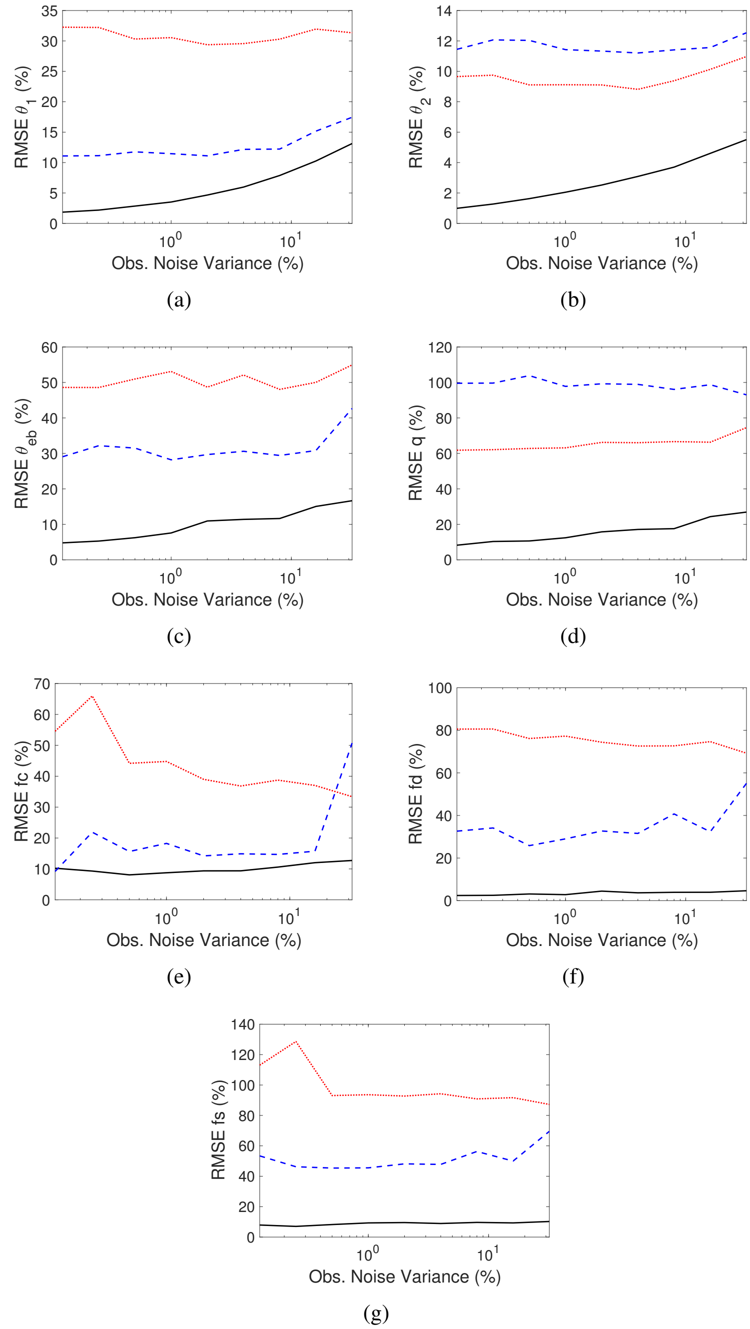


FIG. 6. Robustness of filter estimates. RMSE as a percentage of the standard deviation of each variable is shown as a function of observation noise percentage (noise variance is the given percentage of the the observation variance for each observed variable). The filter using the true observation function (black, solid curve) is compared to the result of filtering with the wrong observation function using only inflation of the observation covariance matrix (red, dashed) and the wrong observation function with iterative observation model error correction (blue, dashed).