Kalman filter data assimilation: Targeting observations and parameter estimation

Thomas Bellsky, Eric J. Kostelich, and Alex Mahalov

Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science **24**, 024406 (2014); doi: 10.1063/1.4871916
View online: http://dx.doi.org/10.1063/1.4871916
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/24/2?ver=pdfcov
Published by the AIP Publishing

Kalman filter data assimilation: Targeting observations and parameter estimation

Thomas Bellsky,[a] Eric J. Kostelich, and Alex Mahalov
School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona 85287, USA

(Received 31 October 2013; accepted 8 April 2014; published online 24 April 2014)

This paper studies the effect of targeted observations on state and parameter estimates determined with Kalman filter data assimilation (DA) techniques. We first provide an analytical result demonstrating that targeting observations within the Kalman filter for a linear model can significantly reduce state estimation error as opposed to fixed or randomly located observations. We next conduct observing system simulation experiments for a chaotic model of meteorological interest, where we demonstrate that the local ensemble transform Kalman filter (LETKF) with targeted observations based on largest ensemble variance is skillful in providing more accurate state estimates than the LETKF with randomly located observations. Additionally, we find that a hybrid ensemble Kalman filter parameter estimation method accurately updates model parameters within the targeted observation context to further improve state estimation. © *2014 AIP Publishing LLC*.
[http://dx.doi.org/10.1063/1.4871916]

For chaotic systems like the weather, an accurate forecast requires an accurate representation of the current state. Data assimilation (DA) is a methodology that re-initializes the current state of a system by combining observational data along with an estimated state determined by a forecast model. Within data assimilation schemes, the spatial locations of the observational data can have a significant effect on the accuracy of the analysis (re-initialized) state. In this work, we use a particular strategy for locating observations and show that this targeting strategy is successful in reducing state estimation error for a conceptual chaotic model, as compared to fixed or randomly located observations. An additional facet of weather modeling that we investigate is the estimation of model parameters. Parameter estimation is a technique used within modeling that often seeks to fit parameters with historical data or to characterize subgrid-scale effects. We show that our utilized observation targeting strategy within a particular parameter estimation scheme is successful in accurately estimating a model parameter for this chaotic model. To motivate our study of this chaotic model, we first establish a theorem which is used to justify the use of targeted observations for a linear data assimilation scheme.

I. INTRODUCTION

The weather is chaotic[1,2] where small errors in initial conditions can quickly result in forecasts diverging from the true state. In order to prevent forecast divergence, operational numerical weather prediction relies on the periodic use of atmospheric observations to update initial conditions. A common method for re-initializing initial conditions is the use of DA, which balances the uncertainty in a model forecast with the uncertainty in observable data to minimize an objective function. This minimizer is subsequently the initial condition for the next forecast.

This paper examines two DA topics of current interest:

1. strategically targeting observations to reduce state estimation and forecast error;[3–16]
2. estimating model parameters[17–22] within the targeting observation context to further reduce estimation error.

Targeted observations,[3,4] also known as adaptive observations, result from using some strategy to locate observations in order to improve state estimation and forecast accuracy within a DA scheme. In modern day weather forecasting, the locations of many observations are predetermined, such as fixed observing stations or fixed satellite orbits. Supplementing these fixed observations with additional strategically located observations can be very beneficial to improving weather forecasting, especially in data sparse regions or extreme weather situations.[5] Operational opportunities to spatially target observations include weather balloons and aircraft, which are used to forecast hurricanes.[6]

There are many existing methods for targeting observations, which often aim to locate observations where solutions of a dynamical system exhibit instability. Singular value based methods[4,7] spatially locate additional observations where singular vectors are largest corresponding to locations of highest instability.

Lorenz and Emanuel[8] used the Lorenz-96 model[23] to compare a variety of targeting strategies, including fixed locations, random locations, singular vectors, and breeding techniques. They found that ensemble strategies targeting areas of maximal background error performed optimally. Trevisan and Uboldi[5] further studied the Lorenz-96 model, where they targeted observations using ensemble breeding techniques. Breeding techniques[24] are often used to determine the shape and growth of instabilities of a system. A key step in the analysis of Trevisan and Uboldi is the use of breeding techniques to characterize the instabilities of the

[a]Electronic mail: bellskyt@asu.edu

data assimilation observing cycle, where they then confine their correction of the system state to this unstable subspace. This methodology can be sufficient in satisfying the linear observability condition, often stabilizing an otherwise unstable system state. These techniques have been carried out on a more realistic quasi-geostrophic model with similar success in stabilizing analysis system states.[9] Further research has determined a rigorous condition for when the assimilation of an observation will stabilize a system state,[10] where a mathematical description is provided showing how large an analysis correction must be in order to counteract any unstable growth.

Ensemble transform techniques[11] are used to determine where to target a real world additional observation by brute force modeling. In the case of choosing only one of many possible observational locations, these methods independently simulate an observation at each available observation location and perform a DA analysis for each observation. This results in a number of independent analysis state covariances, where the particular observation resulting in the smallest analysis state covariance is the location chosen. These methods are used in real world applications, for example with aircraft dropsondes, to determine the optimal spatial location for an additional observation.

Of note, many of the previously mentioned techniques have been used operationally or have been investigated in field campaigns resulting in varying degrees of success.[12–16] In these field campaigns, methods often resulted in improved forecasts, but it was also found that occasionally targeted observations led to deteriorated forecasts.[12,13]

The numerical targeting technique used in this manuscript is in a sense a simplification of ensemble transform methods, where the goal is to reduce subsequent analysis error covariances. We instead aim to reduce the DA analysis error by targeting locations where the current forecast ensemble variance is largest. Often large errors in state forecasts correspond to locations where the model background ensemble variance is largest; thus, locating observations where this background ensemble variance is largest can be useful in reducing state estimation error.

Testing this method on the chaotic Lorenz-96 model, we find this targeting method to be affective in reducing analysis error as compared to randomly located observations. We point out that locating observations where the forecast ensemble variance is largest may not always be successful, since for Ensemble Kalman filter (EnKF) methods the ensemble covariance is often an underestimate of the true state covariance.[25] Additionally, ensemble methods can introduce spurious long-distance correlations, which can further increase analysis error.

The main part of our numerical results investigates estimating parameters for the chaotic Lorenz-96 model, where we expand on previous literature by applying recently developed parameter estimation techniques[18–20] within the novel context of targeted observations. Model parameters are typically fixed quantities that encode physical information about a dynamical system which are often estimated within large weather and climate models. When accurately estimated, these parameters can provide an adequate description of subgrid-scale physics.[26] We show that using targeted observations for estimating parameters has notable success in improving the accuracy of parameter estimates and state estimates.

We additionally investigate the effect of the level of chaoticity within the model with regards to certain data assimilation protocols (number of observations, ensemble size, and localization radius size). We form conclusions on how these protocols affect state estimation and forecast error with respect to the model's chaoticity.

To motivate our numerical investigations for using targeted observations within a DA scheme, we first provide an analytical justification, culminating in Theorem 1. This theorem shows under certain assumptions that targeting observations within the Kalman filter for a linear system can significantly reduce estimation error. This result demonstrates, especially for non-stable systems, that the use of targeted observations can offer a more accurate state estimate as opposed to randomly located or fixed observations. This theorem also provides a justification that targeting observations at locations where the forecast covariance is largest can lead to the best reduction in analysis error, offering support for our numerical targeting strategy.

Section II describes the linear Kalman filter and Sec. III presents our analytical results on how targeting observations can reduce state estimation error for a linear system. Section IV discusses ensemble Kalman filter techniques and Section V details the Lorenz-96 model. Section VI characterizes our numerical methods, Sec. VII presents our numerical results, and we finish with a conclusion section. The Appendix details the analysis covariance used within our analytical results in Sec. III.

II. LINEAR KALMAN FILTER DATA ASSIMILATION

In this section, we describe the Kalman filter algorithm.[27] The Kalman filter is a recursive algorithm that updates an analysis state from a weighted average of a model prediction and observations of the true state. A successful Kalman filter update will result in the analysis state being a more accurate estimate of the true state than either the model prediction or the observations alone.

Let $z(t) \in \mathbb{R}^N$ be the true state of some dynamic phenomena and $A \in \mathbb{R}^{N\times N}$ be a discrete map that determines the state vector

$$z_j = Az_{j-1} + \eta_j. \tag{1}$$

Above, η_j is model process noise, which is assumed to be a Gaussian random variable, $\eta_j \sim N(0, Q_j)$, where Q_j is the covariance matrix for η_j. The Kalman filter algorithm assumes the true dynamics described in Eq. (1) is unknown. Instead, only the map A is known, where at time t_j the map A forecasts the background state estimate as

$$z_j^b = Az_{j-1}^a, \tag{2}$$

where z_{j-1}^a is the analysis state at the previous time step.

Furthermore, the Kalman filter algorithm assumes that there exists a linear observation operator $H_j : \mathbb{R}^N \to \mathbb{R}^m$

from the state space to the observation space where $m \leq N$. At time t_j, the vector y_j consists of m observations of the true state z_j, where

$$y_j = H_j(z_j) + \epsilon_j \in \mathbb{R}^m. \tag{3}$$

Above, ϵ_j is observational noise, which is assumed to be a Gaussian random variable, $\epsilon_j \sim N(0, R_j)$, where R_j is the covariance matrix for ϵ_j. The Kalman filter formulates the background error covariance matrix P_j^b for the background state z_j^b as

$$P_j^b = AP_{j-1}^a A^T + Q_j, \tag{4}$$

where P_{j-1}^a is the analysis covariance matrix from the previous Kalman filter update step. These covariance matrices P_{j-1}^a and P_j^b, respectively, describe the uncertainty in the estimated analysis state and the estimated background state.

The Kalman filter update step is a weighted average between the model estimate background state z_j^b and the observations y_j which produces an analysis state z_j^a and its associated covariance matrix P_j^a. The most common formulation of the Kalman filter update step is

$$\underbrace{z_j^a}_{\text{analysis}} = \underbrace{z_j^b}_{\text{background}} + \underbrace{K_j}_{\text{Kalman gain}} \underbrace{\left(y_j - H_j z_j^b\right)}_{\text{innovation}}, \tag{5}$$

$$P_j^a = (I - K_j H_j) P_j^b, \tag{6}$$

where the Kalman gain matrix K_j in Eq. (6) is defined as

$$K_j = P_j^b H_j^T (H_j P_j^b H_j^T + R_j)^{-1}. \tag{7}$$

From Eq. (5), we see how the Kalman update step acts as a weighted average between the model estimate background state z_b and the observations y, based on uncertainties in both. We also see in Eq. (6) how the $K_j H_j$ term acts as a rank m correction from the background covariance to the analysis covariance, which can then result in a reduction in the error of subsequent state analyses.

By definition, a covariance matrix is positive semi-definite, so it might not be invertible. In order to ensure that the Kalman gain matrix K_j in Eq. (7) is well-defined, the linear Kalman filter formulation typically assumes observations are independent of each other, which implies that the covariance matrix R_j is positive definite and of full rank m. This assumption is often satisfied in real-world applications; for example, weather observations are typically independent of each other. Alternatively, the forecast covariance matrix P_j^b is often rank deficient in real-world applications; in fact, our numerical techniques produce low-rank approximations of the true forecast covariance using ensemble methods, further discussed in Sec. IV. But, even if P_j^b is rank deficient, the sum $H_j P_j^b H_j^T + R_j$ will be of full rank m when R_j is of full rank m, ensuring K_j is well-defined.

An equivalent formulation[28] of the analysis covariance in Eq. (6) is given by

$$P_j^a = (I + P_j^b H_j^T R_j^{-1} H_j)^{-1} P_j^b, \tag{8}$$

for the Kalman gain matrix K_j given by

$$K_j = P_j^a H_j^T R_j^{-1}. \tag{9}$$

In this paper, we use Eqs. (8) and (9) for the analysis covariance and the Kalman gain matrix, respectively.

The analysis state z_j^a from these two Kalman filter formulations is a unique, unbiased, minimum variance estimate of the true state z_j when the model and the observation operator are linear.[29] Unfortunately, such a unique, unbiased, minimum variance estimate does not necessarily exist for nonlinear models; thus, the linear Kalman filter is extended in some manner for nonlinear models. In Sec. IV, we describe one particular extension of the linear Kalman filter, the ensemble Kalman filter.

III. ANALYSIS: REDUCED ESTIMATION ERROR WITH TARGETED OBSERVATIONS

In this section, we provide an analytical result, summarized in Theorem 1, that shows targeted observations are useful in reducing errors in state estimation. In the following analysis, we assume a true dynamic state z is evolving according to Eq. (1). Initially, we assume the true state z is being approximated by $\hat{z}$ without any observations, where we define each forecast step in terms of the previous forecast step

$$\hat{z}_j = A\hat{z}_{j-1}.$$

We also define the error in the initial forecast state as

$$\delta_0 = \hat{z}_0 - z_0. \tag{10}$$

Over j time steps, if A is known and is used to forecast the estimated state $\hat{z}$, then the estimated state will have the following error:

$$\begin{aligned} \hat{z}_j - z_j &= A(\hat{z}_{j-1} - z_{j-1}) - \eta_j, \\ &= A^j(\hat{z}_0 - z_0) - \sum_{k=1}^{j} A^{j-k}\eta_k, \\ &= A^j\delta_0 - \sum_{k=1}^{j} A^{j-k}\eta_k. \end{aligned} \tag{11}$$

Without model process noise, Eq. (11) reduces to

$$\hat{z}_j - z_j = A^j\delta_0. \tag{12}$$

Thus, if A is stable, where the modulus of every eigenvalue is less than 1, we see from Eq. (11) that the estimated state will eventually converge to the true state, up to model noise. Otherwise, for A unstable, the estimated state diverges from the true state.

Next, we examine the error in the analysis estimate from the linear Kalman filter. We assume the observational noise ϵ_j is an independent and identically distributed random variable (i.i.d.). This assumption implies the covariance matrix R_j for the observational noise ϵ_j is only non-zero on the diagonal, where

$$\text{diag}(R_j) = \rho_j \in \mathbb{R}^N. \tag{13}$$

We further assume at each analysis step there is a single observation. Thus, the observation operator $H_j \in \mathbb{R}^{1\times N}$ is of the form in Eq. (3).

We first examine the initial forecast step, where

$$\begin{aligned} z_1^b &= A\hat{z}_0, \\ P_1^b &= AP_0A^T + Q_1. \end{aligned} \tag{14}$$

Above, P_0 is the covariance matrix for the initial error described in Eq. (10). The following equation determines the initial background state:

$$z_1^b = A\hat{z}_0 = Az_0 + A(\hat{z}_0 - z_0) = z_1 + A\delta_0 - \eta_1, \tag{15}$$

where Eqs. (1) and (10) are applied. Inserting Eq. (9) into the analysis step from Eq. (5), the initial Kalman filter update step is formulated as

$$z_1^a = z_1^b + P_1^a H_1^T R_1^{-1}\left(y_1 - H_1 z_1^b\right). \tag{16}$$

A key step to this analysis is to characterize the analysis covariance as the background covariance plus a rank one correction. Detailed in the Appendix, this allows us to express a general analysis covariance matrix P^a with one observations as

$$P^a = ((P^b)^{-1} + I_s)^{-1} = P^b - \frac{1}{\rho_s + \text{tr}(I_s P^b)} P^b I_s P^b.$$

As described in the Appendix, the matrix $I_s \in \mathbb{R}^{N\times N}$ is zero everywhere except for a single 1 on the diagonal at the (s, s) location corresponding to the one observed location. Also, ρ_s is the sth component of ρ at $t = 1$ as defined in (13) and tr$(\cdot)$ is the usual trace operator, summing the diagonal of matrix. Applying the above result to Eq. (16), we have

$$\begin{aligned} z_1^a &= \left(P_1^b - \frac{P_1^b I_s P_1^b}{\rho_s + \text{tr}\left(I_s P_1^b\right)}\right) H_1^T R_1^{-1}\left(H_1\left(z_1 - z_1^b\right) + H_1 H_1^T \epsilon_1\right) \\ &\quad + z_1 + A\delta_0 - \eta_1, \\ &= \left(P_1^b - \frac{P_1^b I_s P_1^b}{\rho_s + \text{tr}\left(I_s P_1^b\right)}\right) \rho_s^{-1} I_s\left(\eta_1 - A\delta_0 + H_1^T \epsilon_1\right) \\ &\quad + z_1 + A\delta_0 - \eta_1, \end{aligned} \tag{17}$$

where we also applied Eqs. (3) and (15). If we rearrange terms from Eq. (17) above, we have the difference between the first analysis estimate and the true state

$$\begin{aligned} z_1^a - z_1 &= \left(I - \rho_s^{-1} P_1^b I_s + \frac{\rho_s^{-1} P_1^b I_s P_1^b I_s}{\rho_s + \text{tr}\left(I_s P_1^b\right)}\right)(A\delta_0 - \eta_1) \\ &\quad + \left(\rho_s^{-1} P_1^b I_s - \frac{\rho_s^{-1} P_1^b I_s P_1^b I_s}{\rho_s + \text{tr}\left(I_s P_1^b\right)}\right) H_1^T \epsilon_1. \end{aligned} \tag{18}$$

A remaining task is to make sense of the sum of matrices in parentheses above, where

$$P_1^b I_s = \begin{pmatrix} 0 & \cdots & 0 & P_s^b & 0 & \cdots & 0 \end{pmatrix} \equiv P^{bs}, \tag{19}$$

for $P_s^b \in \mathbb{R}^{N\times 1}$ and $0 \in \mathbb{R}^{N\times 1}$. It follows that

$$\begin{aligned} \rho_s^{-1} P_1^b I_s - \frac{\rho_s^{-1} P_1^b I_s P_1^b I_s}{\rho_s + \text{tr}\left(I_s P_1^b\right)} &= \rho_s^{-1} P^{bs} - \frac{\rho_s^{-1} P_{ss}^b}{\rho_s + P_{ss}^b} P^{bs}, \\ &= \frac{1}{\rho_s + P_{ss}^b} P^{bs}, \end{aligned} \tag{20}$$

where P_{ss}^b is the (s, s) component of P_1^b.

Now, we make a further assumption, where we assume that P_1^b is diagonal. In order to satisfy this assumption, we need that A, P_0^a, and Q_1 are all diagonal to satisfy Eq. (4). Typically, in the algorithmic setting, these assumptions are not realistic; linear models are rarely diagonal, and typically there are non-zero covariances between model states and also model process noise. Additionally, this assumption would often hinder the ability of the filter, since the filter would only have the ability to update a state location with just observational information at that same location. Thus, this analytical result lies within the unique situation when this assumption holds.

With the above assumption, we plug Eq. (20) into Eq. (18) and have that the error in the analysis state is

$$\begin{aligned} z_1^a - z_1 &= \left(I - \frac{P_{ss}^b}{\rho_s + P_{ss}^b} I_s\right)(A\delta_0 - \eta_1) \\ &\quad + \frac{P_{ss}^b}{\rho_s + P_{ss}^b} I_s H_1^T \epsilon_1. \end{aligned} \tag{21}$$

Additionally, using the definitions of the estimated state covariance from Eq. (4) and the analysis from the Appendix, we have that the analysis covariance is updated as

$$P_1^a = P_1^b - \frac{(P_{ss}^b)^2}{\rho_s + P_{ss}^b} I_s. \tag{22}$$

Applying an induction argument, we advance the equations for the state error (21) and for the analysis covariance matrix error (22) to any time step t_n as described in the following theorem.

Theorem 1. *Given the initial guess $\hat{z}(t_0) = \hat{z}_0$ for a dynamic state $z(t_0) = z_0 \in \mathbb{R}^N$, define the initial error as*

$$\hat{z}_0 - z_0 = \delta_0,$$

where the covariance matrix P_0 for the error δ_0 is diagonal. Additionally, suppose the diagonal matrix $A \in \mathbb{R}^{N\times N}$ maps the true state subject to

$$z_j = Az_{j-1} + \eta_j,$$

where the covariance matrix Q_j for the model process noise η_j is diagonal, and A forecasts the estimated state as

$$z_j^b = Az_{j-1}^a. \tag{23}$$

Then applying the linear Kalman filter algorithm as described above, the following equations describe the error

in the predicted analysis state and the predicted analysis covariance matrix at time $t = t_n$:

$$z_n^a - z_n = \prod_{j=1}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right) A^n \delta_0 - \sum_{j=1}^{n}\left(\prod_{k=j}^{n}\left(I - \frac{P^b_{s_k s_k}}{\rho_{s_k} + P^b_{s_k s_k}} I_{s_k}\right)\right)\left(A^{n-j}\eta_j\right)\epsilon_n + \sum_{j=1}^{n-1}\left(\prod_{k=j+1}^{n}\left(I - \frac{P^b_{s_k s_k}}{\rho_{s_k} + P^b_{s_k s_k}} I_{s_k}\right)\right) \times \left(A^{n-j}\frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j} H_j^T \epsilon_j\right) \quad (24)$$

and

$$P_n^a = \prod_{j=1}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right) A^n P_0 (A^T)^n + \sum_{k=1}^{n}\left(\prod_{j=k}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right)\right) A^{n-k} Q_k (A^T)^{n-k}. \quad (25)$$

Remark 1. In the case of no model process noise nor observation noise, Eqs. (24) and (25) simplify to

$$z_n^a - z_n = \prod_{j=1}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right) A^n \delta_0 \quad (26)$$

and

$$P_n^a = \prod_{j=1}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right) A^n P_0 (A^T)^n. \quad (27)$$

In the case of no observational noise, the error in the analysis estimates from the linear Kalman filter will decrease for A stable. We can see this effect from the above statements, where the coefficient for I_{sj} (and I_{sk}) in Eqs. (24) and (26) will always be one when there is not any observational noise. So the difference of the identity matrix and the I_{sj} matrix will partially zero out A^n, thereby reducing analysis error. We also see that the update step will reduce a diagonal entry of P_n^a to zero in (27) when a previously unobserved location is observed.

When the observational noise and model process noise are reasonably small, the estimate from the Kalman filter algorithm will be more accurate than the estimate without observations from Eq. (11). Equations (24) and (26) again offer a mathematical description of this, where the coefficient for I_{sj} (and I_{sk}) in Eqs. (24) and (26) will be close to one; thus, the difference of the identity matrix and the I_{sj} matrix will be small, thereby significantly reducing analysis error.

This theorem especially describes how targeting the location where the forecast covariance is largest can be beneficial in reducing state error. If all observational locations exhibit equivalent error, then the location of largest forecast uncertainty corresponds to the largest value of $P^b_{s_j s_j}$ in Eqs. (24) and (26). Thus, targeting the location of largest forecast uncertainty will result in the smallest value of

$$I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j},$$

leading to the greatest reduction in analysis error.

The state estimate from the Kalman filter will accurately estimate the true solution even for A unstable when the linear observability condition is satisfied.[30] This condition is satisfied when the inherent instability of the system is nullified by enough appropriately located and accurate observations. Equation (24) allows us a mathematical description of how the state estimate can accurately estimate the true solution for A unstable, where the product,

$$\prod_{j=1}^{n}\left(I - \frac{P^b_{s_j s_j}}{\rho_{s_j} + P^b_{s_j s_j}} I_{s_j}\right),$$

will be small for reasonable forecast uncertainty and small observational noise when different locations are observed over different analysis steps.

We use this theorem to describe how targeted observations can significantly reduce error in state estimation, as compared to other methods of locating observations. To simplify the following argument, we assume the observational noise and model process noise are zero. First, assume a single observation is always at the same location at each time step. Then the difference term in parenthesis in Eq. (26) will be the identity except for one zero on the diagonal. Thus, the state estimate from the Kalman filter with one observation detailed in Eq. (26) will do better than the case of no observations described in Eq. (12), but only by reducing A^n by one dimension. Now, if at each time step, the filter observes a location that has not been previously observed, then we see from Eq. (26) that the estimated solution will decay to the true solution in only N time steps. Alternatively, if the observation is chosen at a random location at each time step, then after N time steps, there is $\frac{(N-1)!}{N^N}$ (which rapidly asymptotes to zero for increasing N) probability that Eq. (26) has decayed to zero. Thus, for this simplified linear example, a random strategy leads to more accurate state estimates than a fixed location strategy, but a particular targeting strategy produces the most accurate state estimates.

Figure 1 numerically demonstrates the above argument from this theorem for small observational noise (normally distributed with mean zero and standard deviation 0.01) and small model process noise (normally distributed with mean zero and standard deviation 0.01). For this experiment, the matrix $A \in \mathbb{R}^{50\times 50}$ is unstable, with all eigenvalues equal to 1.001. The horizontal axis is the Kalman filter DA analysis time step and the vertical axis is the root mean square error (*RMS* error), an error quantity that we define in Eq. (31). This figure plots four different observational schemes: a scheme without any observations, a scheme with the same fixed observation at each time step, a scheme with an

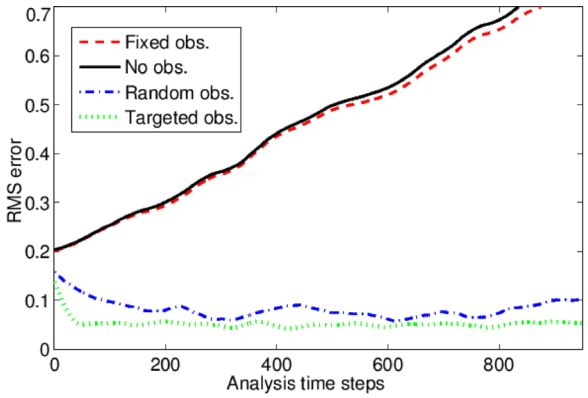

FIG. 1. The *RMS* error of four different observation schemes for *A* unstable: a fixed observation (dashed, red), no observation (solid, black), a randomly located observation (dashed dotted, blue), and a targeted observation (dotted, green). The horizontal axis is the analysis time step and the vertical axis is the *RMS* error. Observational noise and model process noise are both normally distributed with mean zero and standard deviation of 0.01.

observation located randomly in space at each time step, and finally a targeting scheme. Here, at the *j*th analysis step, the targeting scheme locates the observation at the component satisfying $max_{1\le i\le 50} P_j^b(i,i)$, which corresponds to the location where the forecast uncertainty is the highest. Since *A* is unstable, the error grows in time for the case of no observations. We find that a fixed observation scheme only does slightly better than no observations, with the error still increasing in time. The targeting scheme and the random scheme do well in reducing error where both result in a stable error bound, with the targeting scheme performing the best.

In the following numerical analysis, we examine a chaotic model where we instead apply ensemble Kalman filtering techniques. Still the above motivation is maintained: locating observations randomly often outperforms fixed observations in reducing state estimation error, and a targeting strategy is the most successful at reducing state estimation error. As can be extrapolated from Eq. (24), especially for a highly unstable or chaotic model, a good targeting strategy can be the difference between accurate state estimates and awful state estimates.

IV. ENSEMBLE KALMAN FILTER

Complex geophysical models are typically nonlinear, so various extensions of the linear Kalman have been developed.[31] Some techniques linearize a nonlinear model, which can be computationally expensive. Another difficulty of using Kalman filter techniques for large models is estimating a large background covariance matrix P^b, which can also be computationally expensive. To reduce this expense, 3D-Var and 4D-Var schemes replace P^b by a constant or slowly time-varying matrix representing typical forecast uncertainties.[32,33]

EnKF methods are useful, particularly when applied to meteorological systems,[34–36] in providing low-rank approximations of $P^b \in \mathbb{R}^{N\times N}$ from an ensemble of k forecasts where $k \ll N$. Traditional EnKF methods use the ensemble to update covariances at each step

$$P_{en}^b = \frac{1}{k-1} Z^b Z^{bT}, \tag{28}$$

where $Z^b \in \mathbb{R}^{N\times k}$ consists of k column vectors

$$Z_i^b = z_i^b - \bar{z}^b. \tag{29}$$

Above, z_i^b is one of k ensemble members, each a realization of the background state. Additionally, $\bar{z}^b$ is the mean of the background states.

Unfortunately, computational expense quickly increases for large ensembles, and a reasonably computable ensemble size k within an EnKF method may not be sufficiently large to provide an accurate approximation of the true background covariance over the entire spatial domain. One method to combat this drawback is spatial localization, which is used within the local ensemble transform Kalman filter (LETKF). Within the LETKF, the local analysis determines different linear combinations of ensemble members, so the combined global analysis explores a much larger dimensional space than the k ensembles alone.[29]

Spatial localization is a process that determines the analysis state independently at each model grid point using only observations within a prescribed distance of that model grid point. The prescribed distance, or radius, is indicated in this work as the localization radius l. Thus, for a localization radius of $l=5$, only observations within 5 spatial units are used to update a grid point at each analysis update.

The LETKF scheme allows for a multiplicative ensemble covariance inflation factor μ. Covariance inflation is an *ad hoc* procedure to avoid underestimating uncertainties,[37] a particular issue within ensemble Kalman filtering. For any constant μ, all entries in the background covariance matrix are inflated equally at each DA step to $P^b \to \mu P^b$, which will inflate uncertainties in the background covariance matrix.[29] This procedure is done to prevent P^b from collapsing to an overconfident, but incorrect state; something referred to as ensemble collapse. We take μ to be $\mu = 1.2$.

V. MODEL PROBLEM

The Lorenz-96 model is a system of ODEs that governs the time evolution of N periodic points

$$\begin{aligned} \frac{dX_i}{dt} &= (X_{i+1} - X_{i-2})X_{i-1} - X_i + F, \\ X_{i\pm N} &= X_i, \end{aligned} \tag{30}$$

where we fix $N=40$. Although the Lorenz-96 model is not a truncated version of any physical system, it is useful in simulating observed atmospheric characteristics on a latitude circle to examine questions of predictability in weather forecasting.[23] The nonlinear terms mimic advection and conserve total energy. The linear term dissipates the total energy and F is external forcing. For Eq. (30), we fix the forcing as a constant, so F is a global parameter. For $N=40$, where $F<0.895$, solutions decay to the steady state solution $X_1 = \cdots = X_N = F$; when $0.895 < F < 4$, solutions are periodic; and when $F>4$, solutions are chaotic.[8] We typically take $F=8$ in order to induce chaos. Also by varying F, we

examine both a chaotic and a non-chaotic Lorenz-96 in order to determine the effects of chaoticity with regards to different aspects of the LETKF.

We use the 4th order Runge-Kutta method to forecast solutions of Lorenz-96 with a fixed time step of $\Delta t = 0.05$, which corresponds to 6 atmospheric hours.[23] A DA analysis occurs every time step after a spin-up period, and each forecast step is simply one iteration of the Runge-Kutta scheme. Since there is not a closed form solution to this model, we instead define a reference solution that is integrated from a fixed initial condition. We perturb the reference solution to form ensembles, which are all spin-up for 360 time steps.

VI. METHODOLOGY

The spatial location of observations can affect forecast accuracy; thus, targeting the location of observations is one method to improve a meteorological forecast. As described in the introduction, Lorenz and Emanuel first used the Lorenz-96 model to study a selection of targeting strategies.[8] We offer a distinct targeting strategy applied to the Lorenz-96 model, where we target observations at the locations where the ensemble of model forecasted (background) solutions has the largest ensemble variance. By ensemble variance, we simply mean the pointwise variance computed over all ensemble members. Thus, we target observations at the locations where the ensemble of forecasted solutions has the largest ensemble variance, and we then use the ensemble of forecasted solutions and these targeted observations within the LETKF scheme to form an updated analysis solution for the Lorenz-96 model.

A main component of this paper is estimating model parameters using EnKF techniques. We use state augmentation methods,[18–20] where these methods augment an ensemble of background state vectors $\{z_i^b : i = 1, 2, \ldots, k\}$, with an ensemble of background parameters $\{p_i^b : i = 1, 2, \ldots, k\}$, where the augmented ensemble of background states $\{(z_i^b, p_i^b) : i = 1, 2, \ldots, k\}$ is updated with some DA technique resulting in an ensemble of augmented analysis states $\{(z_i^a, p_i^a) : i = 1, 2, \ldots, k\}$. These methods are classified into two types, one deemed simultaneous[21] and the other separate.[18–20,22] Simultaneous parameter estimation techniques form an update of both the state and the parameter from the ensemble of augmented analysis states $\{(z_i^a, p_i^a) : i = 1, 2, \ldots, k\}$, which is updated by some DA method from the ensemble of augmented background states $\{(z_i^b, p_i^b) : i = 1, 2, \ldots, k\}$. Separate parameter estimation techniques first use some DA method to update the ensemble of analysis states $\{z_i^a : i = 1, 2, \ldots, k\}$ from the ensemble of background states $\{z_i^b : i = 1, 2, \ldots, k\}$. They separately update the ensemble of analysis parameters $\{p_i^a : i = 1, 2, \ldots, k\}$ using some DA method on the ensemble of augmented background states $\{(z_i^b, p_i^b) : i = 1, 2, \ldots, k\}$ (where any update of the analysis state from the augmented analysis state is discarded). We utilize separate parameter estimation techniques, since they are more successful in reducing error.[20]

In Ref. 19, the same non-localized EnKF is used in both the update of the state and the update of the parameter. As in Ref. 20, we modify the separate technique of Koyama,[19] where we implement the LETKF to update the ensemble of analysis states $\{z_i^a : i = 1, 2, \ldots, k\}$ from the ensemble of background states $\{z_i^b : i = 1, 2, \ldots, k\}$. We subsequently use an EnKF without any localization to obtain the ensemble of analysis parameters $\{p_i^a : i = 1, 2, \ldots, k\}$ from the ensemble of augmented background states $\{(z_i^b, p_i^b) : i = 1, 2, \ldots, k\}$. In particular, a non-localized EnKF is used in the parameter estimation, since localization is not useful for estimating a global parameter;[19,20] we also provide evidence to support this statement with numerical results described in Figure 6. With this EnKF separate parameter estimation method, we update an ensemble of forcing parameters $\{F_i : i = 1, 2, \ldots, k\}$ for the Lorenz-96 model.

We vary the chaoticity of the Lorenz-96 model and the number of observations, the ensemble size, and the localization radius to determine the effect that the chaoticity and these protocols have on DA analysis error. We fix the localization radius l anywhere from 1 to 20. The number of observations assimilated in the whole domain varies in separate experiments for the Lorenz-96 model at fixed values from 1 to 10. Additionally, we fix the number of ensembles within the EnKF analysis in separate experiments anywhere from 2 to 40. Of note, each set of experiments is always the mean of 50 model realizations.

We form synthetic observations from the corresponding reference solution by adding the product of Gaussian noise and a fixed standard deviation parameter. In the experiments described below, the standard deviation parameter is set so observations approximate the reference solution to an accuracy of four bits.

In these experiments, the DA schemes have no knowledge of the reference solution. To evaluate the performance of various LETKF schemes, we compare the DA analysis to the reference solution by examining the *RMS* error. For any DA scheme A, we define the corresponding *RMS* error for any time t as

$$RMSE_A(t) = \sqrt{\frac{\sum_{i=1}^{N} (z(x_i, t) - \bar{z}(x_i, t))^2}{N}}, \tag{31}$$

where z is the reference solution, $\bar{z}$ is the mean of the analysis ensemble, and N is the number of grid points. In each experiment, we will specify whether t is a DA analysis time (indicated by t^a) or a forecast time (indicated by t^f). We use the subscript T to denote the DA method using targeted observations, R for the DA method using randomly located observations, Tp for the DA method using targeted observations with parameter estimation, and Rp for the DA method using randomly located observations with parameter estimation; the subscript 0 indicates an analysis with no data assimilation.

We define the skill of any DA strategy A in terms of the relative decrease in the *RMS* error at some fixed time t as

$$\gamma_A(t) \equiv \frac{RMSE_0(t) - RMSE_A(t)}{RMSE_0(t)}. \tag{32}$$

Notice $\gamma = 1$ for a DA scheme that perfectly models the reference solution, and $\gamma = 0$ if the DA scheme provides no improvement over an analysis without any data assimilation. We say scheme A is more skillful than scheme B for fixed times if

$$\gamma_A(t) > \gamma_B(t). \tag{33}$$

In particular, we say a DA scheme with targeted observations is more skillful than one with random observations if $\gamma_T(t) > \gamma_R(t)$.

Our standard of a successful parameter estimation scheme using targeted observations is twofold: first, it must be more skillful than the parameter estimation scheme using random observations, $\gamma_{Tp}(t) > \gamma_{Rp}(t)$; second, it must outperform the targeting strategy without parameter estimation, $\gamma_{Tp}(t) > \gamma_T(t)$.

VII. RESULTS

A. Results: Skill of targeted DA schemes

We examine the targeting method for the chaotic Lorenz-96 system, (30). In Figure 2, we plot the *RMS* forecast errors for the targeting scheme, $RMSE_T(t)$, and for the random scheme, $RMSE_R(t)$. The forecasts are begun at the DA analysis time $t^a = 100$, where we consider forecasts of length $t^f = 28$ (corresponding to 7 days). We take $F = 8$ in Eq. (30) with 4 observations, 20 ensembles, and $l = 2$. Over time, the chaotic nature of Lorenz-96 results in the forecast diverging. We see that the targeting scheme results in a lower state estimation error than the random scheme at every forecast step, where the targeting scheme offers substantial improvement for longer forecast intervals. One explanation why the targeting scheme leads to a more accurate forecast over longer forecast times as opposed to the random observation scheme is that the targeting scheme is effectively eliminating most unstable modes,[9] resulting in a significant reduction in the growth rate of the forecast error.

Table I shows the skill of the targeting method, γ_T, and the skill of the random method, γ_R, for various protocols with $F = 8$. The left half of the table shows the accuracy at DA analysis times $t^a = 25$, 50, and 100. The right half shows the forecast accuracy begun at the DA analysis time $t^a = 100$ over forecasts of lengths $t^f = 8$, 12, and 28 (corresponding to 2, 3, and 7 days forecasts). For each of the columns underneath the number of observations in this table, the left value is the skill, Eq. (32), of the targeting scheme and the right value is the skill, Eq. (32), of the random scheme. Table I shows the targeting scheme offers a vast improvement over the random scheme where Eq. (33) is satisfied, particularly for limited observations. Although not contained in Table I, we find that similar results hold for many additional analysis and forecast times for a variety of forcings ranging from $F = 0.75$ to $F = 10$.

TABLE I. Skill: Eq. (32), 20 ensembles, $F = 8$.

		$\gamma_T(t^a)\ \gamma_R(t^a)$				$\gamma_T(t^f)\ \gamma_R(t^f)$			
l	t^a	2 obs.	4 obs.	8 obs.	t^f	2 obs.	4 obs.	8 obs.
2	25	0.46 0.27	0.83 0.63	0.94 0.91	8	0.50 0.24	0.74 0.63	0.75 0.72
2	50	0.65 0.36	0.90 0.76	0.94 0.92	12	0.42 0.15	0.72 0.51	0.74 0.67
2	100	0.75 0.47	0.93 0.87	0.95 0.93	28	0.23 0.08	0.46 0.29	0.49 0.41
4	25	0.40 0.24	0.86 0.61	0.94 0.87	8	0.59 0.18	0.74 0.68	0.75 0.64
4	50	0.64 0.26	0.91 0.80	0.95 0.92	12	0.51 0.09	0.70 0.58	0.73 0.52
4	100	0.72 0.37	0.93 0.88	0.96 0.94	28	0.30 0.06	0.43 0.33	0.45 0.32

Figure 3(a) shows the skill of the targeting method, γ_T, and of the random method, γ_R, up to analysis time $t^a = 100$. Figure 3(b) shows the skill of the targeting method, γ_T, and of the random method, γ_R, over a $t^f = 28$ forecast begun after $t^a = 100$ analysis steps. This figure corresponds to Table I for the cases when there are 4 observations, 20 ensembles, and a localization radius of $l = 2$.

Figure 4 illustrates how the DA analysis is updating the Lorenz-96 mean analysis state from the mean background state at the $t^a = 25$ analysis time step for the targeting observation methodology with 4 observations, 20 ensembles, and a localization radius of $l = 2$, corresponding to a case in Table I. Figure 4(a) plots the true Lorenz-96 state (solid, black), the mean model background state (dashed, green), the updated mean state (dashed-dotted, magenta), and observations (circles). This figure demonstrates how the DA analysis can improve the mean analysis state from the mean

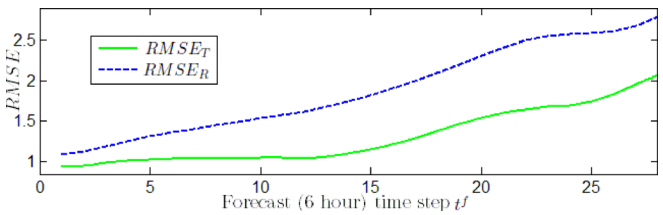

FIG. 2. The *RMS* error of the targeting scheme, $RMSE_T(t)$ (solid, green) and the *RMS* error of the random scheme, $RMSE_R(t)$ (dotted, blue). This is a $t^f = 28$ forecast after 100 DA analysis steps with 4 observations, 20 ensembles, and $l = 2$.

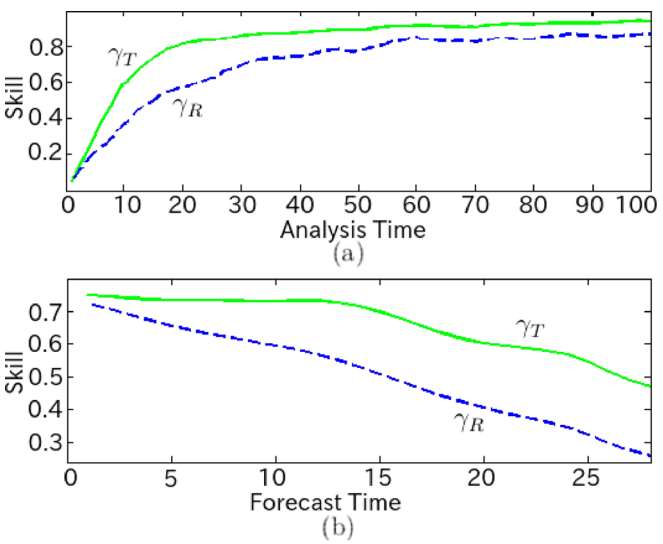

FIG. 3. The skill for the targeted observation scheme (solid, green) and the random observation scheme (dashed, blue) both for analysis time (a) and a subsequent forecast (b) for 4 observations, 20 ensembles, and a localization radius of $l = 2$.

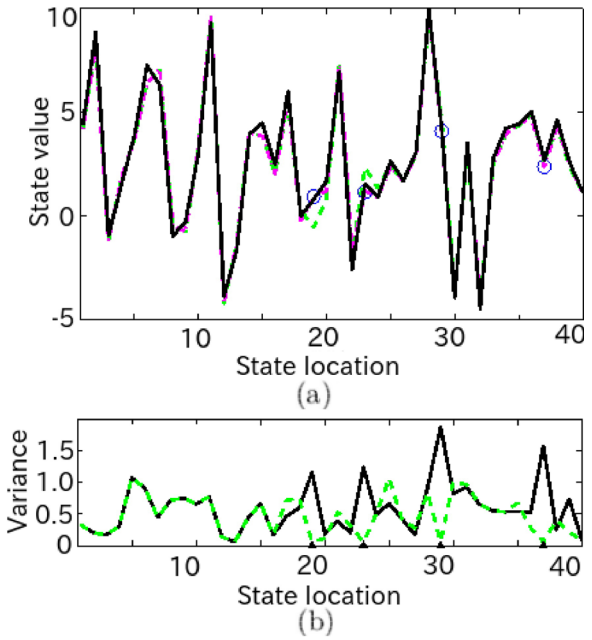

FIG. 4. Figure 4(a) plots the true Lorenz-96 state (solid, black), the mean background state (dashed, green), the mean analysis state (dashed-dotted, magenta), and four targeted observations (open circles, blue) at the 25th analysis step. Figure 4(b) plots the background ensemble variance (solid, black), the updated analysis ensemble variance (dashed, green), and indicates the four observational locations (triangles). This is for the targeted observation scheme with 4 observations, 20 ensembles, and a localization radius of $l = 2$.

background state, especially in the case of the observation located at 20.

Figure 4(b) plots the model background ensemble variance (solid, black), the updated analysis ensemble variance (dashed, green), and indicates the locations of the four targeted observations (triangles). This figure illustrates that the utilized targeting strategy locates observations at the locations where the background ensemble variance is largest. Figure 4(b) shows that the analysis ensemble variance is significantly reduced with respect to the background ensemble variance by the DA scheme at observed locations. The analysis ensemble variance is unchanged from the background ensemble variance far enough away from observations, since the LETKF DA scheme is localizing state updates only near observations.

B. Results: Parameter estimation

Next, we employ our parameter estimation methods and examine the skill of each scheme for a chaotic Lorenz-96. In the following results, we take the true forcing as $F = 8$, and we form an ensemble of perturbed forcings normally distributed about $\hat{F} = 6$, where the ensemble of perturbed forcings has a variance of 0.25. In Figure 5, we estimate the forcing with both the random observation scheme and the targeted observation scheme as a function of DA analysis step t^a. For this figure, there are 8 observations, a localization radius of $l = 4$, and 20 ensembles. We see from Figure 5 that the true forcing is more accurately estimated by the targeting observation scheme than by the random observation scheme.

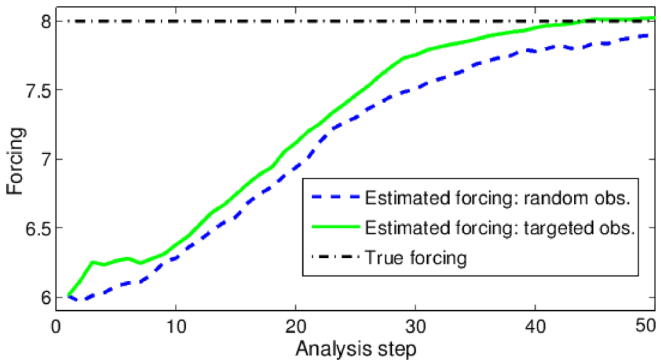

FIG. 5. The estimated forcing from the targeted observation scheme (solid, green), the estimated forcing from the random observation scheme (dashed, blue), and the true forcing (dashed dot, black) as a function of data assimilation analysis step. Here, there are 8 observations, a localization radius of $l = 4$, and 20 ensembles.

Table II lists the skills $\gamma_{Tp}(t)$, $\gamma_{Rp}(t)$, and $\gamma_T(t)$ at DA analysis times $t^a = 25$, 50, 100 for various protocols. For each of the columns underneath the number of observations in this table, the left value is the skill, Eq. (32), of the parameter estimation scheme with targeted observations, the middle value is the skill, Eq. (32), of the parameter estimation scheme with random observations, and the right value is skill, Eq. (32), of the targeted scheme with no parameter estimation. Table II shows the parameter estimation scheme with targeted observations is skillful in significantly satisfying both $\gamma_{Tp} > \gamma_{Rp}$ and $\gamma_{Tp} > \gamma_T$. Results analogous to this are obtained for every protocol and every DA analysis time up to $t^a = 100$.

We also perform a number of similar experiments on Lorenz-96 with small values for both the true and perturbed forcings, so the system will have little to no chaoticity. In every case, the state is resolved exceptionally well where the targeting scheme substantially satisfies both $\gamma_{Tp} > \gamma_{Rp}$ and $\gamma_{Tp} > \gamma_T$.

Next, we investigate the effect of localization within the parameter estimation scheme. Figure 6(a) shows the *RMS* error of the estimated state and Figure 6(b) shows the *RMS* error of the estimated forcing parameter F for the Lorenz-96 model. In both figures, the horizontal axis varies the localization radius for the DA update step of the parameter, the vertical axis varies the localization radius for the DA update step of the state, and the *RMS* error is plotted after 25 analysis steps where there are 4 observations and 10 ensembles. In Figure 6(a), a smaller localization radius in the state update step can lead to a significantly smaller error in the state

TABLE II. Skill (32): Parameter estimation, 20 ensembles, $F = 8$, $\hat{F} = 6$.

		$\gamma_{Tp}(t^a)\ \gamma_{Rp}(t^a)\ \gamma_T(t^a)$		
l	t^a	2 obs.	4 obs.	8 obs.
2	25	0.365 0.237 0.340	0.834 0.585 0.790	0.906 0.835 0.876
2	50	0.492 0.297 0.362	0.908 0.682 0.792	0.942 0.917 0.883
2	100	0.528 0.375 0.287	0.930 0.850 0.771	0.949 0.924 0.876
4	25	0.380 0.139 0.374	0.790 0.502 0.763	0.902 0.848 0.885
4	50	0.441 0.235 0.409	0.908 0.673 0.804	0.946 0.917 0.899
4	100	0.486 0.301 0.352	0.920 0.829 0.826	0.954 0.931 0.896

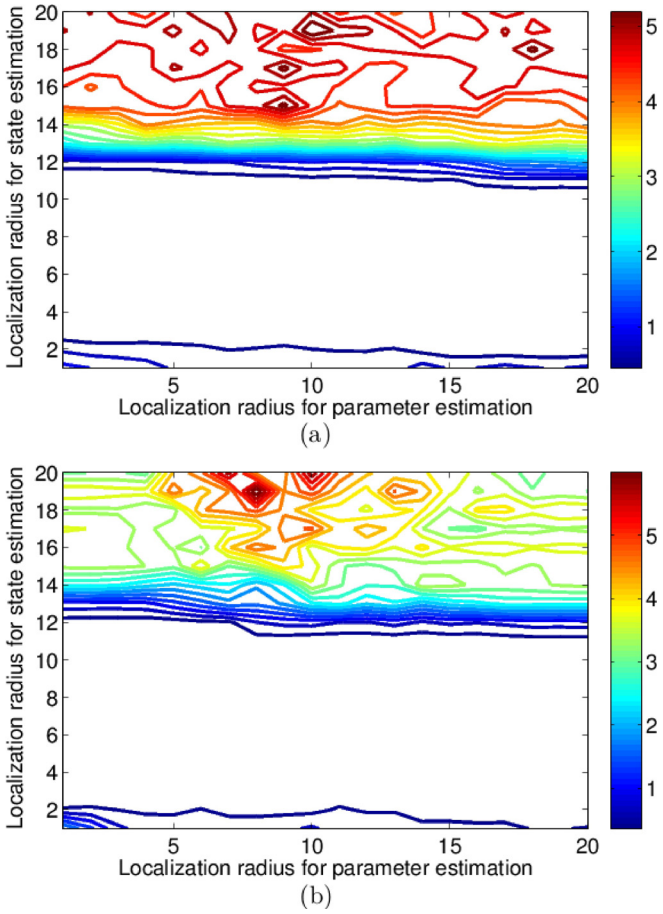

FIG. 6. Figure 6(a) plots the *RMS* error of the state estimate after 25 analysis steps using parameter estimation on the forcing F. Figure 6(b) plots the *RMS* error of the estimated forcing parameter F after 25 analysis steps using parameter estimation on this forcing. In each, the horizontal axis varies the localization radius l from 1 to 20 for the DA update step of the parameter and the vertical axis varies the localization radius l from 1 to 20 for the update step of the state. Additionally, there are 10 ensembles and 4 observations at each update step.

estimate; however, the localization radius for the parameter update step contributes little to the error in the state estimate. Figure 6(b) is similar to Figure 6(a), where a smaller localization radius in the state update step can lead to a smaller error in the forcing estimate, but the localization radius for the parameter update step does not significantly affect the error in the forcing estimate. Thus, for this particular global forcing parameter, we find that localizing in the DA update step of the parameter estimate offers little to no improvement in the accuracy of the state nor the accuracy of the estimated parameter. This supports previous findings that localization in the parameter estimation scheme is not useful for estimating a global parameter.[19,20]

Of note in both contours in Figure 6, if the localization radius in the state update is too large, the DA scheme loses any benefit from localization. Thus, the filter resolves the state and forcing poorly since there is not a sufficient number of ensembles to combat the chaotic growth of the Lorenz-96 system. If 40 ensembles are utilized instead of the 10 ensembles in Figures 6(a) and 6(b), there would be roughly the same RMS error no matter the value of the localization radius in the state estimate.

C. Results: Effects of chaoticity on state error within targeting scheme

Next, we contrast how well the targeting strategy estimates the state for both a chaotic and non-chaotic Lorenz-96 when the number of total observations, the ensemble size, and localization radius l are varied. In the following results, we vary the forcing F from 1 to 8. Each of the contours in Figure 7 plots the RMS error of the targeting scheme, $RMSE_T$, after 25 analysis steps.

Figure 7(a) plots a contour of the $RMSE_T$ as a function of the number of observations, varied from 1 to 8, and the forcing, where there are always 10 ensembles and a localization radius $l = 4$. We find that increasing the number of

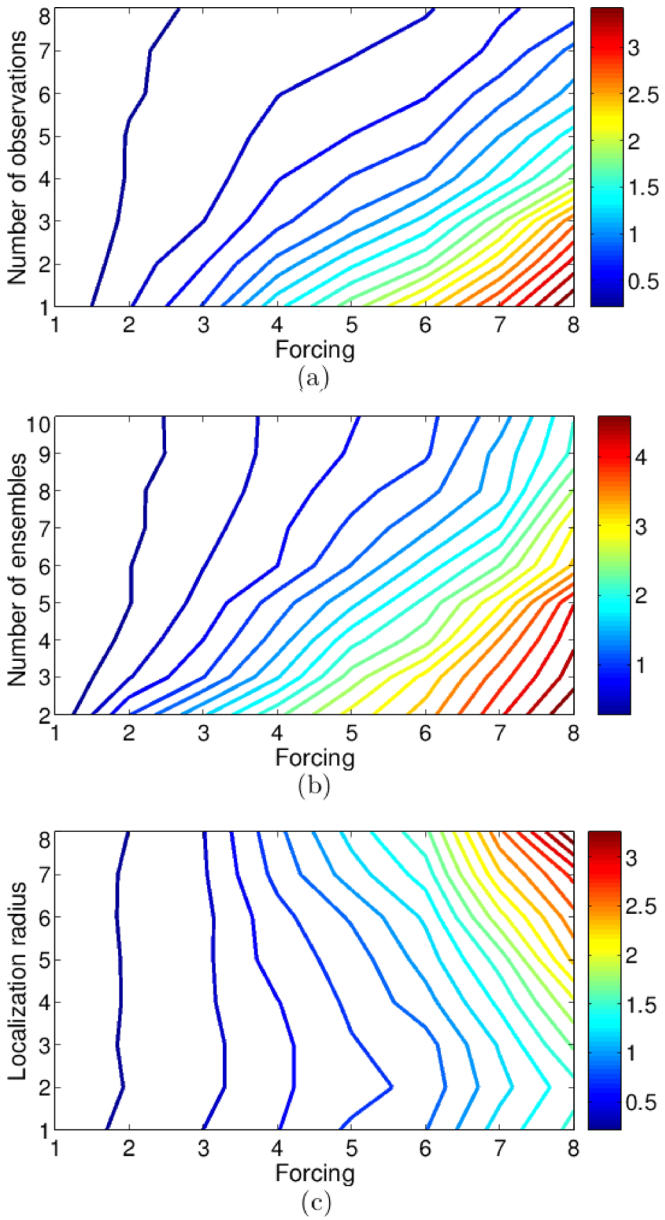

FIG. 7. All three figures above plot contours of the *RMS* error of the targeting scheme, $RMSE_T$. For each figure, the horizontal axis varies the forcing F from 1 to 8. In (a), there are 10 ensembles, a localization radius of $l = 4$, and the vertical axis varies the number of observations from 1 to 8. In (b), there are 4 observations, a localization radius of $l = 4$, and the vertical axis varies the number of ensembles from 2 to 10. In (c), there are 4 observations, 10 ensembles, and the vertical axis varies the localization radius l from 1 to 8.

observations will always substantially decrease $RMSE_T$, whether Lorenz-96 is chaotic or not. Thus, extra observations will always substantially improve the state estimate.

Figure 7(b) contains a contour of the $RMSE_T$ as a function of the number of ensembles, varied from 2 to 10, and the forcing, where there are always 4 observations and a localization radius $l=4$. For the Lorenz-96 model with a forcing inducing chaos, a larger ensemble size significantly reduces the $RMSE_T$. This is expected, since for a more chaotic flow, more ensembles are needed to capture all dimensions. Additionally, we find that increasing the number of ensembles for smaller forcings not inducing chaos ($F<2$) has a lesser effect on the $RMSE_T$.

Finally, Figure 7(c) plots a contour of the $RMSE_T$ as a function of the localization radius l, varied from 1 to 8, and the forcing, where there are always 4 observations and 10 ensembles. In this figure, we see that the localization radius has a lesser effect on the $RMSE_T$ for small forcings. For larger forcings inducing chaos, too large of a localization radius results in a significant increase in error. There is in fact a sweet spot for the localization radius, where for $l=2$, the $RMSE_T$ is the least.

VIII. CONCLUSION

This paper has investigated how targeting observations for both state estimation and parameter estimation can lead to reductions in error. First, we have provided an analytical argument investigating the effect of targeting observations for the Kalman filter applied to a linear system. The result in Theorem 1 demonstrates, especially for an unstable linear system, that spatially locating observations can play a significant role in improving state estimation accuracy.

Our numerical results indicate that using the LETKF with observations targeted at the locations of greatest ensemble variance is skillful at accurately estimating and forecasting solutions to Lorenz-96, (30). Additionally, the EnKF separate parameter estimation method is skillful at reducing analysis error in the novel context of targeted observations for the Lorenz-96 model. Finally, we have examined how the chaoticity of the model and various LETKF filter protocols affect state estimation. We have determined that an increase in observations will significantly reduce state estimation error for both a chaotic and non-chaotic Lorenz-96, an increase in ensemble size significantly reduces state estimation error for a chaotic Lorenz-96 but has a lesser effect on a non-chaotic Lorenz-96, and that a smaller localization radius leads to the smallest state estimation error with a sweet spot at $l=2$ for a chaotic Lorenz-96 model.

ACKNOWLEDGMENTS

All authors graciously acknowledge the support of the National Science Foundation under Grant No. DMS-0940314. A.M. acknowledges partial support from AFOSR FA9550-11-1-0220.

APPENDIX: UPDATE OF ANALYSIS COVARIANCE

In the immediate analysis below, the Kalman filter algorithm is always at the jth step; thus, we drop the subscript j. We also assume there is only one observation. We manipulate the analysis covariance matrix in Eq. (8) as

$$P^a = (I + P^b H^T R^{-1} H)^{-1} P^b = ((P^b)^{-1} + \rho_s^{-1} I_s)^{-1}, \quad \text{(A1)}$$

where

$$\rho_s^{-1} I_s = H^T R^{-1} H, \quad \text{(A2)}$$

and ρ_s is the sth entry of the vector ρ as defined in Eq. (13). The subscript s indicates there is one observation at the sth position, where $I_s \in \mathbb{R}^{N\times N}$ is the zero matrix except for a 1 at the (s, s) location.

The following lemma of Miller[38] gives a succinct formula for the inversion of the sum of a matrix and a rank one perturbation.

Lemma. If G and $G+H$ are invertible and H is rank one, then

$$(G+H)^{-1} = G^{-1} - \frac{1}{1+\mathrm{tr}(HG^{-1})} G^{-1} H G^{-1}. \quad \text{(A3)}$$

Above, tr(·) is the trace operator, which sums the diagonal elements of a matrix. This lemma can be applied to the analysis covariance in Eq. (A1) above, where $G=(P^b)^{-1}$ and $H=I_s$, and it follows that

$$P^a = \left((P^b)^{-1} + I_s\right)^{-1} = P^b - \frac{1}{\rho_s + \mathrm{tr}(I_s P^b)} P^b I_s P^b. \quad \text{(A4)}$$

Equation (A4) shows that one observation acts as a rank one correction within the update from P^a to P^b for the linear Kalman filter. Recall, we use this formula in Sec. III to analytically show that targeting observations for a linear Kalman filter can significantly reduce state estimation error as compared to using fixed or randomly located observations.

[1]E. Kalnay, *Atmospheric Modeling, Data Assimilation, and Predictability* (Cambridge University Press, Cambridge, 2002).
[2]E. Lorenz, "Deterministic non-periodic flow," J. Atmos. Sci. **20**, 130–141 (1963).
[3]C. Snyder, "Summary of an informal workshop on adaptive observations and FASTEX," Bull. Am. Meteor. Soc. **77**, 953–961 (1996).
[4]T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza, "Singular vectors, metrics, and adaptive observations," J. Atmos. Sci. **55**, 633–653 (1998).
[5]A. Trevisan and F. Uboldi, "Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle system," J. Atmos. Sci. **61**, 103–113 (2004).
[6]S. Aberson, S. Majumdar, C. Reynolds, and B. Etherton, "An observing system experiment for tropical cyclone targeting techniques using the Global Forecast System," Mon. Weather Rev. **139**, 895–907 (2011).
[7]R. Buizza and T. N. Palmer, "The singular-vector structure of the atmospheric global circulation," J. Atmos. Sci. **52**, 1434–1456 (1995).
[8]E. Lorenz and K. Emanuel, "Optimal sites for supplementary weather observations: Simulation with a small model," J. Atmos. Sci. **55**, 399–414 (1998).
[9]A. Carrassi, A. Trevisan, and F. Uboldi, "Adaptive observations and assimilation in the unstable subspace by breeding on the data assimilation system," Tellus **59A**, 101–113 (2007).
[10]A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi, "Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system," Chaos **18**, 023112 (2008).
[11]C. H. Bishop and Z. Toth, "Ensemble transformation and adaptive observations," J. Atmos. Sci. **56**, 1748–1765 (1999).
[12]I. Szunyogh, Z. Toth, K. A. Emanuel, C. Bishop, J. Woolen, T. Marchok, R. Morss, and C. Snyder, "Ensemble based targeting experiments during

FASTEX: The impact of dropsonde data from the Lear jet," Q. J. R. Meteorolog. Soc. **125**, 3189–3218 (1999).
[13]R. H. Langland, R. Gelaro, G. D. Rohaly, and M. A. Shapiro, "Targeted observations in FASTEX: Adjoint based targeting procedures and data impact experiments in IOP17 and IOP18," Q. J. R. Meteorolog. Soc. **125**, 3241–3270 (1999).
[14]I. Szunyogh, Z. Toth, A. V. Zimin, S. J. Majumdar, and A. Persson, "Propagation of the effect of targeted observations: The 2000 Winter Storm Reconnaissance Program," Mon. Weather Rev. **130**, 1144–1165 (2002).
[15]N. Fourrié, D. Marchal, F. Rabier, B. Chapnik, and G. Desroziers, "Impact study of the 2003 North Atlantic THORPEX regional campaign," Q. J. R. Meteorolog. Soc. **132**, 275–295 (2006).
[16]T. M. Hamill, F. Yang, C. Cardinali, and S. J. Majumdar, "Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions," Mon. Weather Rev. **141**, 2058–2065 (2013).
[17]S.-J. Baek, B. Hunt, E. Kalnay, E. Ott, and I. Szunyogh, "Local ensemble Kalman filtering in the presence of model bias," Tellus A **58**, 293–306 (2006).
[18]T. Bellsky, J. Berwald, and L. Mitchell, "Non-global parameter estimation using local ensemble Kalman filtering," Mon. Weather Rev. (published online).
[19]H. Koyama and M. Watanabe, "Reducing forecast errors due to model imperfections using ensemble Kalman filtering," Mon. Weather Rev. **138**, 3316–3332 (2010).
[20]J. Ruiz, M. Pulido, and T. Miyoshi, "Estimating model parameters with ensemble-based data assimilation: A review," J. Meteorol. Soc. Jpn. **91**, 79–99 (2013).
[21]J. L. Anderson, "An ensemble adjustment Kalman filter for data assimilation," Mon. Weather Rev. **129**, 2884–2903 (2001).
[22]T. DelSole and X. Yang, "State and parameter estimation in stochastic dynamic models," Physica D **239**, 1781–1788 (2010).
[23]E. Lorenz, *in Proceedings of the Seminar on Predictability*, ECWF Seminar, edited by T. Palmer (ECMWF, Reading, UK, 1996), Vol. I, pp. 1–18.
[24]Z. Toth and E. Kalnay, "Ensemble forecasting at NMC: The generation of perturbations," Bull. Am. Meteorol. Soc. **74**, 2317–2330 (1993).
[25]K. J. H. Law and A. M. Stuart, "Evaluating data assimilation algorithms," Mon. Weather Rev. **140**, 3757–3782 (2012).
[26]T. N. Palmer, "A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models," Q. J. R. Meteorolog. Soc. **127**, 279–304 (2001).
[27]R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME J. Basic Eng. **82**, 35–45 (1960).
[28]D. E. Catlin, "Estimation, control, and the discrete Kalman filter," Appl. Math. Sci. **71**, 137–140 (1989).
[29]B. Hunt, E. Kostelich, and I. Szunyogh, "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter," Physica D **230**, 112–126 (2007).
[30]R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," J. Basic Eng. **83**, 95–108 (1961).
[31]M. Ghil and P. Malanotte-Rizzoli, "Data assimilation in meteorology and oceanography," Adv. Geophys. **33**, 141–266 (1991).
[32]M. Corazza, E. Kalnay, and S.-C. Yang, "An implementation of the local ensemble Kalman filter in a quasigeostrophic model and comparison with 3D-Var," Nonlinear Proc. Phys. **14**, 89–101 (2007).
[33]S.-C. Yang, M. Corazza, A. Carrassi, E. Kalnay, and T. Miyoshi, "Comparison of local ensemble transform Kalman filter, 3DVAR, and 4DVAR in a quasigeostrophic model," Mon. Weather Rev. **137**, 693–709 (2009).
[34]G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics," J. Geophys. Res. **99**, 10143–10162, doi:10.1029/94JC00572 (1994).
[35]P. Houtekamer and H. Mitchell, "Data assimilation using an ensemble Kalman filter technique," Mon. Weather Rev. **126**, 796–811 (1998).
[36]T. M. Hamill, J. S. Whitaker, and C. Snyder, "Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter," Mon. Weather Rev. **129**, 2776–2790 (2001).
[37]J. L. Anderson and S. L. Anderson, "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," Mon. Weather Rev. **127**, 2741–2758 (1999).
[38]K. S. Miller, "On the inverse of the sum of matrices," Math. Mag. **54**, 67–72 (1981).