

# Ensemble Kalman filtering without a model

Franz Hamilton,<sup>1</sup> Tyrus Berry,<sup>2</sup> and Timothy Sauer<sup>2,\*</sup>

<sup>1</sup>*North Carolina State University, Raleigh, NC 27695*

<sup>2</sup>*George Mason University, Fairfax, VA 22030*

Methods of data assimilation are established in physical sciences and engineering for the merging of observed data with dynamical models. When the model is nonlinear, methods such as the ensemble Kalman filter have been developed for this purpose. At the other end of the spectrum, when a model is not known, the delay coordinate method introduced by Takens has been used to reconstruct nonlinear dynamics. In this article, we merge these two important lines of research. A model-free filter is introduced based on the filtering equations of Kalman and the data-driven modeling of Takens. This procedure replaces the model with dynamics reconstructed from delay coordinates, while using the Kalman update formulation to reconcile new observations. We find that this combination of approaches results in comparable efficiency to parametric methods in identifying underlying dynamics, and may actually be superior in cases of model error.

PACS numbers: 05.45.Tp, 92.60.Aa, 05.45.Jn

## I. INTRODUCTION

Data assimilation plays an increasingly important role in nonlinear science, as a means of inferring unobserved model variables and constraining unknown parameters. Use of the Extended Kalman Filter (EKF) and Ensemble Kalman Filter (EnKF) is now standard in a wide range of geophysical problems [1–5] and several areas of physical and biological sciences where spatiotemporal dynamics is involved [6–9].

When a physically motivated model is available, parametric forecasting methods can be used in a variety of applications in noise filtering, prediction and control of systems. Nonlinear approaches to filtering [1, 2, 10, 11] allow forecasting models to use the model equations to develop close to optimal predictions. Even if some variables are not observable, they may be reconstructed, *provided that their model equations are known*.

In other cases, a model may not be available, or available models may be poor. In geophysical processes, basic principles may constrain a variable in terms of other driving variables in a way that is well understood, but the driving variables may be unmodeled or modeled with large error [12–15]. In numerical weather prediction codes, physics on the large scale is typically sparsely modeled [16, 17]. Some recent work has considered the case where only a partial model is known [18, 19].

When a physical model is completely unavailable, Takens' method of attractor reconstruction [20–23] has become the foundation of nonparametric time series prediction methods. Under suitable genericity hypotheses, the dynamical attractor may be represented by delay coordinate vectors built from the time series observations, and methods of prediction, control, and other applications from chaotic time series have been developed [24–26]. In particular, time series prediction algorithms locate the

current position in the delay coordinate representation and use analogues from previously recorded data to establish a predictive statistical model, which can be accomplished in several ways [27–40].

In this article we propose a “Kalman-Takens” method of data analysis which is able to filter a noisy time series without access to a model. We replace the model equations governing the evolution of the system, which are assumed to be known in standard data assimilation methods, with advancement of the dynamics nonparametrically using delay-coordinate vectors. We find that the Kalman-Takens algorithm as proposed in this article is able in many cases to filter noisy data at a rate comparable to parametric filtering methods that have full access to the exact model. In one sense, this is no surprise: The content of Takens’ theorem is that in the large data limit, the equations can be replaced by the data. But particularly surprising is the fact that by implementing the Kalman update, the nonparametric representation of the dynamics is able to handle substantial noise in the observed data. Moreover, in cases where the parametric filter is subject to model error, we will see that the Kalman-Takens approach may exhibit superior results.

Throughout this paper we will assume availability of a noisy set of historical data for which there is no model. The goal is to use this historical data set to predict future values using one of the many established nonparametric prediction methods. Our objective is to first filter these noisy observations using Kalman-Takens and reconstruct the underlying signal, which will allow for increased predictive capability. We will compare this novel approach to filtering, which does not use any model, to the ideal scenario which assumes that the perfect model is known, as well as to scenarios which include model error.

Nonparametric forecasting has received renewed interest recently [18, 19, 41, 42], however, all these methods have impractically large data requirements for intrinsically high-dimensional dynamics. Indeed, the approaches of [18, 19] are based on combining a partially known or imperfect model with nonparametric methods in order

---

\* tsauer@gmu.edu

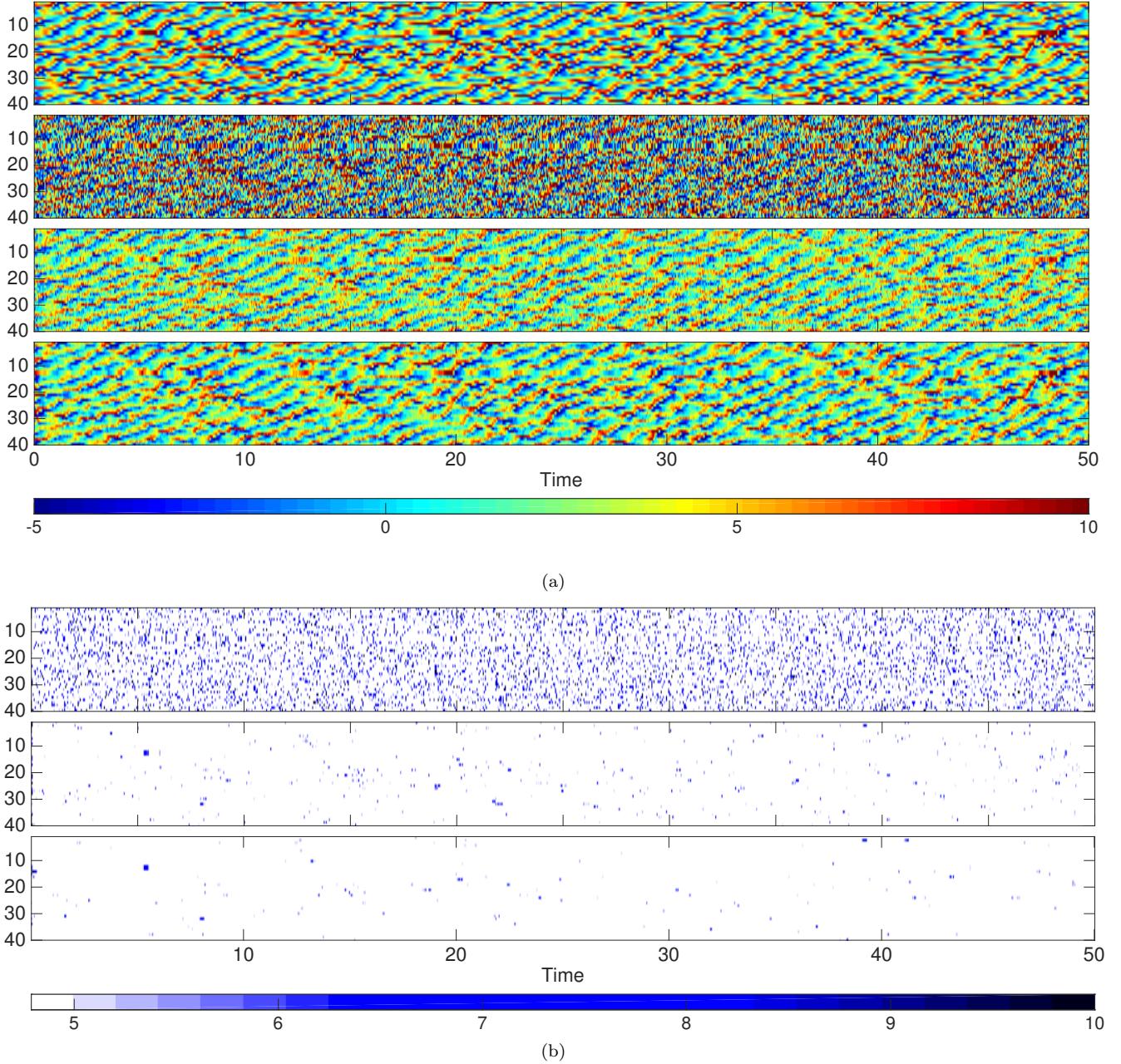


FIG. 1. Results of filtering the full 40-dimensional Lorenz-96 system. (a) Spatiotemporal plot of the system dynamics. The horizontal axis represents time, and the vertical axis represents the spatial position. From top to bottom, (1) the true system trajectory, (2) observations perturbed by 100% noise, (3) Kalman-Takens output and (4) parametric filter output. (b) Spatiotemporal plot of the resulting absolute errors. For readability, only absolute errors greater than 5 are shown. From top to bottom, error in (1) the observed signals, (2) Kalman-Takens output and (3) parametric filter output.

to overcome the high dimensionality which is present in many practical problems. While solving the forecasting problem requires representing the entire dynamical system, the filtering problem only requires the ability to perform a short term forecast (until the next observation becomes available).

We demonstrate this fact in Figure 1 by filtering the Lorenz-96 spatiotemporal dynamical system without any knowledge of the underlying model. The Lorenz-96 sys-

tem represents a simplified weather system on a mid-latitude and is a common benchmark for spatiotemporal filtering methods due to its chaotic dynamics and moderate dimensionality. Notice that the Kalman-Takens filter does not equal the performance of the parametric filter, however, this should be expected since the parametric method is given full knowledge of the true model whereas the Kalman-Takens filter has only the short noisy training data set from which to learn the dynamics.

One key application of the Kalman-Takens filter is to datasets where one has no knowledge of the underlying dynamics. However, we will also show that the proposed approach has significant advantages when the true model is known, but imperfectly. In Sections III and IV we will show that the filter has only a small loss of performance compared to parametric methods which have the perfect model. In exchange for this small loss of performance compared to perfect knowledge, the Kalman-Takens method offers extreme robustness to model error. Since we make no model assumption at all, the results are free from the biases these strong assumptions introduce. In Section V we demonstrate the advantage of the new approach by showing that introducing a small amount of model error (by simply perturbing parameters) leads to the parametric filtering methods being outperformed by the Kalman-Takens filter. We will return to the Lorenz-96 example at the end of Section IV.

As a precursor to this article, we note that Kalman filtering was used in [43] to fit parameters of a global radial basis function model based at unstable fixed points in delay coordinates, which were in turn estimated by another method. In contrast, our approach is in the spirit of fully nonparametric statistics.

## II. KALMAN-TAKENS FILTER

The standard Kalman filtering context assumes a nonlinear system with  $n$ -dimensional state vector  $x$  and  $m$ -dimensional observation vector  $y$  defined by

$$\begin{aligned} x_{k+1} &= f(x_k, t_k) + w_k \\ y_k &= g(x_k, t_k) + v_k \end{aligned} \quad (1)$$

where  $f$  and  $g$  are known, and where  $w_k$  and  $v_k$  are white noise processes with covariance matrices  $Q$  and  $R$ , respectively. The ensemble Kalman filter (EnKF) approximates a nonlinear system by forming an ensemble, such as through the unscented transformation (see for example [44]). Here we initialize the filter with state vector  $x_0^+ = 0_{n \times 1}$  and covariance matrix  $P_0^+ = I_{n \times n}$ . At the  $k$ th step of the filter there is an estimate of the state  $x_{k-1}^+$  and the covariance matrix  $P_{k-1}^+$ . In the unscented version of EnKF, the singular value decomposition is used to find the symmetric positive definite square root  $S_{k-1}^+$  of the matrix  $P_{k-1}^+$ , allowing us to form an ensemble of  $E$  state vectors where the  $i^{th}$  ensemble member is denoted  $x_{i,k-1}^+$ .

The model  $f$  is applied to the ensemble, advancing it one time step, and then observed with function  $g$ . The mean of the resulting state ensemble gives the prior state estimate  $x_k^-$  and the mean of the observed ensemble is the predicted observation  $y_k^-$ . Denoting the covariance matrices  $P_k^-$  and  $P_k^y$  of the resulting state and observed ensemble, and the cross-covariance matrix  $P_k^{xy}$  between

the state and observed ensembles, we define

$$\begin{aligned} P_k^- &= \sum_{i=1}^E (x_{ik}^- - x_k^-)(x_{ik}^- - x_k^-)^T + Q \\ P_k^y &= \sum_{i=1}^E (y_{ik}^- - y_k^-)(y_{ik}^- - y_k^-)^T + R \\ P_k^{xy} &= \sum_{i=1}^E (x_{ik}^- - x_k^-)(y_{ik}^- - y_k^-)^T \end{aligned} \quad (2)$$

and use the equations

$$\begin{aligned} K_k &= P_k^{xy}(P_k^y)^{-1} \\ P_k^+ &= P_k^- - P_k^{xy}(P_k^y)^{-1}P_k^{yx} \\ x_k^+ &= x_k^- + K_k(y_k - y_k^-). \end{aligned} \quad (3)$$

to update the state and covariance estimates with the observation  $y_k$ . We refer to this throughout as the parametric filter, since a full set of model equations are assumed to be known. In some cases  $Q$  and  $R$  are not known; an algorithm was developed in [9] for adaptive estimation of  $Q$  and  $R$ . A brief description of this algorithm is included in the Appendix.

Contrary to (1), our assumption in this article is that neither the model  $f$  or observation function  $g$  are known, making outright implementation of the EnKF impossible. Instead, the filter described here requires no model while still leveraging the Kalman update described in (3). The idea is to replace the system evolution, traditionally done through application of  $f$ , with advancement of dynamics nonparametrically using delay-coordinate vectors. We describe this method with the assumption that a single variable is observable, say  $y$ , but the algorithm can be easily generalized to multiple system observables. In addition, we will assume in our examples that noise covariances  $Q$  and  $R$  are unknown and will be updated adaptively as in [9].

Given the observable  $y_k$ , consider the delay-coordinate vector  $x_k = [y_k, y_{k-1}, \dots, y_{k-d}]$  where  $d$  is the number of delays. This delay vector  $x_k$  represents the state of the system [20, 22]. At each step of the filter an ensemble of delay vectors is formed. However, the advancement of the ensemble one time step forward requires a nonparametric technique to serve as a local proxy  $\tilde{f}$  for the unknown model  $f$ .

The proxy is determined as follows. Given a delay coordinate vector  $x_k = [y_k, y_{k-1}, \dots, y_{k-d}]$ , we locate its  $N$  nearest neighbors  $[y_{k'}, y_{k'-1}, \dots, y_{k'-d}], [y_{k''}, y_{k''-1}, \dots, y_{k''-d}], \dots, [y_{k^N}, y_{k^N-1}, \dots, y_{k^N-d}]$  within the set of noisy training data, with respect to Euclidean distance. Once the neighbors are found, the known  $y_{k'+1}, y_{k''+1}, \dots, y_{k^N+1}$  values are used with a local model to predict  $y_{k+1}$ . In this article, we use a locally constant model which in its most basic form is an average of the nearest neighbors, namely

$$\tilde{f}(x_k) = \left[ \frac{y_{k'+1} + y_{k''+1} + \dots + y_{k^N+1}}{N}, y_k, \dots, y_{k-d+1} \right].$$

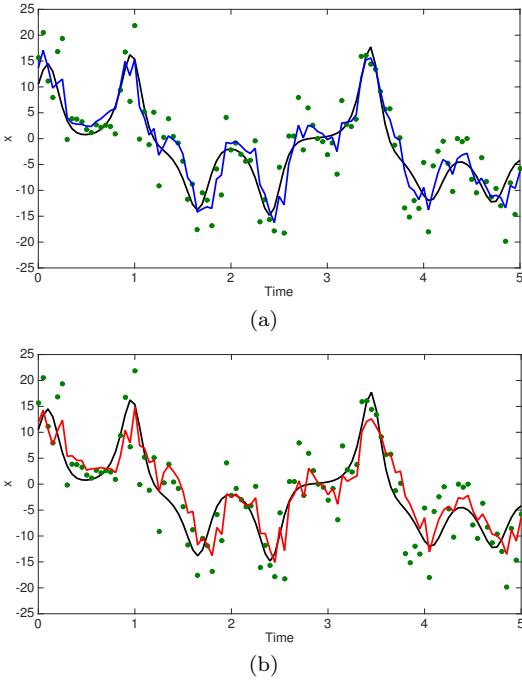


FIG. 2. Given noisy observations of the Lorenz-63  $x$  variable (green circles) sampled at rate  $h = 0.05$ , the goal is to reduce the signal error (RMSE = 4.76) using an ensemble Kalman filter where the true  $x$  trajectory (solid black curve) is desired. (a) If we have knowledge of the Lorenz-63 equations, the standard filtering methodology can be used whereby the noisy  $x$  data are assimilated to the model and using an EnKF a noise-reduced estimate of the  $x$  variable as well as estimates of the unobserved  $y$  and  $z$  variables are provided. We refer to this as the parametric filter (solid blue curve, RMSE = 2.87). (b) Kalman-Takens is able to significantly filter the noise in the observed signal (solid red curve, RMSE = 3.04) to a level comparable to the parametric filter.

This prediction can be further refined by considering a weighted average of the nearest neighbors where the weights are assigned as a function of the neighbor's distance to the current delay-coordinate vector. Throughout the following examples, unless otherwise specified, 20 neighbors were used. This process is repeated for each member of the ensemble. Once the full ensemble is advanced forward, the remaining EnKF algorithm is then applied as previously described and our delay-coordinate vector is updated according to (3). We refer to this method as the *Kalman-Takens filter*.

### III. IMPROVED FORECASTING

As an introductory example we examine the Lorenz-63 system [45]

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z\end{aligned}\quad (4)$$

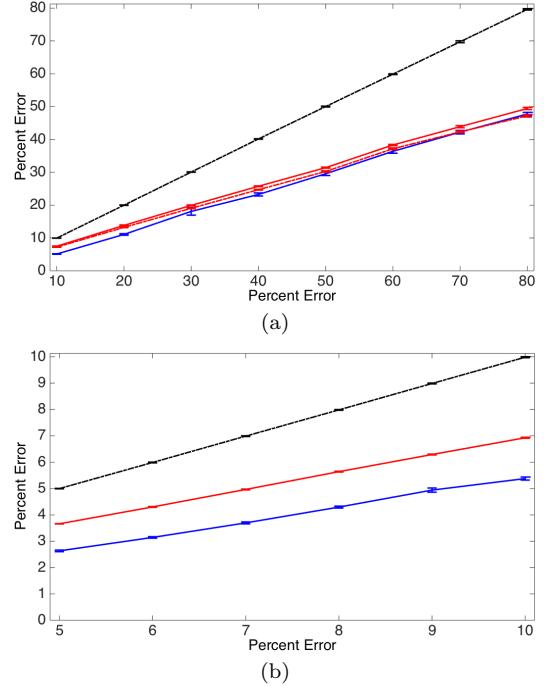


FIG. 3. Filter performance of both the parametric (solid blue curve) and nonparametric (solid red curves) filters when denoising the observed Lorenz-63  $x$  variable at increasing levels of (a) large and (b) low observation noise (signal error, dotted black curve). Time series sampled at rate  $h = 0.05$  and error bars denote standard error over 10 realizations.  $Q$  and  $R$  noise covariances are estimated adaptively. The nonparametric filter is able to significantly reduce the signal error to a level comparable to the parametric filter which has full use of the model. In some instances, the performance of the nonparametric filter improves by reprocessing the data (dotted red curve). At higher noise levels, the discrepancy between parametric and nonparametric filter become negligible.

where  $\sigma = 10$ ,  $\rho = 28$ ,  $\beta = 8/3$ . Assume we have a noisy set of training data observed from the  $x$  variable. Using this data, we want to develop a nonparametric forecasting method to predict future  $x$  values of the system. However, due to the presence of noise, outright application of a prediction method leads to inaccurate forecasts. If knowledge of (4) were available, the standard parametric filter could be used to assimilate the noisy  $x$  observations to the model, generate a denoised estimate of the  $x$  variable and simultaneously estimate the unobserved  $y$  and  $z$  variables.

Without knowledge of (4), we are limited to methods of nonparametric filtering. Using the Kalman-Takens method, we can effectively filter the noisy  $x$  observations, reconstructing the underlying signal, without any knowledge of the underlying system.

For this example we assume that 6000 noisy training points of the Lorenz-63  $x$  variable are available, sampled at rate  $h = 0.05$ . We build a delay coordinate vector  $[x(t), x(t-h), \dots, x(t-dh)]$  where  $d = 4$ . Given that we will be filtering and searching for neighbors within the same training set, we implement a 600 sample lock-

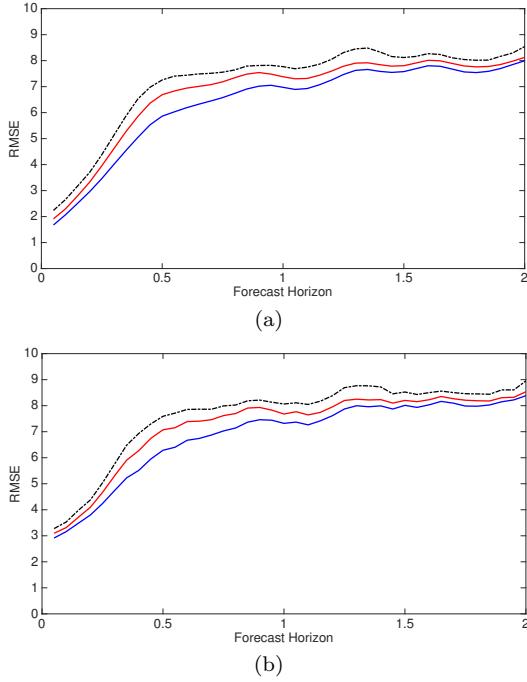


FIG. 4. Results of forecasting the Lorenz-63  $x$  variable when the training data are perturbed by 30% noise. Results averaged over 3000 realizations. Forecast error calculated with respect to (a) the non-noisy truth and (b) the noisy truth. When the training data are filtered by Kalman-Takens (solid red curve) we gain improved forecasting capability as compared to use of the unfiltered training data (dotted black curve). The forecasting results (using identical forecast algorithms) when the training data is filtered by the parametric filter (solid blue curve) is included for a point of reference.

out centered around the current delay-coordinate vector to prevent overfitting (effectively reducing the set from which to find neighbors to 5400 data points). The Kalman-Takens method can then be implemented, iteratively filtering the training data.

Figure 2 shows a comparison of the parametric and Kalman-Takens filter in reconstructing the Lorenz-63  $x$  time series given noisy observations of the variable. The observed time series is corrupted by 60% noise. The green circles denote the noisy observations and the black solid line denotes the true trajectory of the variable (signal RMSE = 4.76). In Figure 2(a), result of the parametric filter (solid blue curve) reconstruction are shown. The parametric filter, operating with full knowledge of the underlying dynamical equations, is able to significantly reduce the signal error (RMSE = 2.87). Figure 2(b) shows the result of the Kalman-Takens filter reconstruction (solid red curve). Without knowledge of any model equations, the Kalman-Takens method is able to significantly reduce the signal error to a level comparable with the parametric filter (RMSE = 3.04).

The results of reconstructing the observed Lorenz-63  $x$  variable over a range of noise levels are shown in Figure 3. Error bars denote standard error over 10 realizations. Estimation of the noise covariances  $Q$  and  $R$  for both

filter algorithms was done using the methodology of [9]. The Kalman-Takens filter (solid red curve) is able to substantially reduce the signal error (dotted black curve) to a level comparable to the parametric filter (solid blue curve). In some instances, the Kalman-Takens filter can be used to re-process the filtered data which can lead to a reduction in error (dotted red curve).

Next we show how Kalman-Takens filtering of the noisy dataset can enhance forecasting. We utilize a simple nonparametric prediction method similar to the local reconstruction of  $\tilde{f}$  above. More sophisticated methods exist, but we want to avoid complicating the comparison of the filtering methods. Assume the current time is  $k$ , and we want to forecast the variable  $y$  ahead  $j$  time units into the future. Using the delay coordinate vector  $x_k = [y_k, y_{k-1}, \dots, y_{k-d}]$ , the vector's  $N$  nearest neighbors  $[y_{k'}, y_{k'-1}, \dots, y_{k'-d}]$ ,  $[y_{k''}, y_{k''-1}, \dots, y_{k''-d}], \dots, [y_{k^N}, y_{k^N-1}, \dots, y_{k^N-d}]$  within the set of noisy training data are located. Then, the known  $y_{k'+j}, y_{k''+j}, \dots, y_{k^N+j}$  are averaged to predict  $y_{k+j}$ .

Figure 4 shows the error in predicting two time units of the Lorenz-63  $x$  variable when the training set is influenced by 30% noise. Results are averaged over 3000 realizations. Prediction error was calculated with respect to the non-noisy truth in Figure 4(a) and the noisy truth in Figure 4(b). When using the noisy training data without any filtering (dotted black curve) for the nonparametric forecast, the prediction error is highest as would be expected. We observe a significant improvement in prediction error when the training data are filtered first by the Kalman-Takens method (solid red curve). This improvement in predictive capabilities when using the Kalman-Takens method compares favorably to the improvement gained when filtering with the full set of model equations (solid blue curve).

We emphasize that our purpose is to demonstrate the utility of the Kalman-Takens filter, not to compare different methods of forecasting. As such, all the forecast methods in this paper are model-free methods; even when the true model is used to filter the historical data, in order to focus on the filtering comparison between the Kalman-Takens and parametric approaches. In practice, if the correct model were available, the noisy initial conditions could be filtered and the model advanced in time to calculate the forecasted values.

#### IV. SPATIOTEMPORAL DYNAMICS

Given the success of the Kalman-Takens method in filtering noisy observations in the low dimensional Lorenz-63 system, allowing for improved forecasting ability, we now examine the filter-forecast problem in a higher dimensional system. Consider a coupled ring of  $N$  Lorenz-96 nodes [46]

$$\dot{x}_i = (ax_{i+1} - x_{i-2})x_{i-1} - x_i + F \quad (5)$$

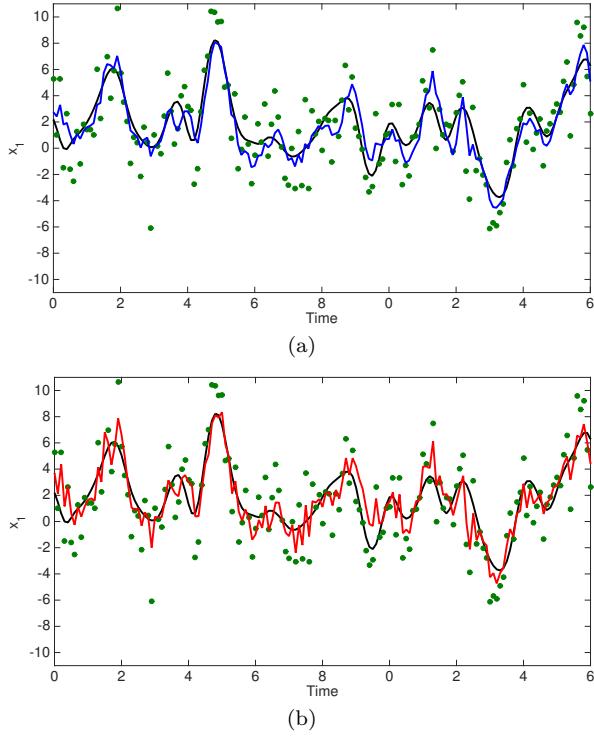


FIG. 5. The filtering problem in a 40 dimensional Lorenz-96 system where noisy observations from 3 of the network nodes are available. Observations are perturbed by 60% noise. The goal is to filter the noisy observations of the  $x_1$  node (black circles) and reduce the signal error (RMSE = 2.16). Black solid curve denotes true trajectory of variable. (a) The parametric filter (solid blue curve) with full knowledge of the system is able to filter the signal to RMSE = 1.17. This requires full estimation of the 40 dimensional state. (b) The nonparametric Kalman-Takens filter (solid red curve), with no knowledge of the underlying physical system and utilizing only delay coordinates of the observables, is able to filter the signal to RMSE = 1.36.

where  $a = 1$  and  $F = 8$ . The Lorenz-96 system is a convenient demonstrative example since it allows for a range of higher dimensional complex behavior by adjusting the number of nodes in the system. We first consider a 40 dimensional Lorenz-96 system in which three nodes are observable, namely  $x_1, x_2$  and  $x_{40}$ . We increase the number of observables in this example since filtering a 40 dimensional Lorenz-96 with one observation, and the full set of model equations, is itself a difficult task. Therefore we introduce the assumption that in this higher dimensional system we are afforded additional observations. Even though we have additional observations, our goal is only to filter and predict the  $x_1$  variable.

We assume that 10000 noisy training data points from each of the three nodes, sampled at rate  $h = 0.05$ , are available. Since three nodes are observable in this example, our delay-coordinate vector takes the form  $[x_1(t), \dots, x_1(t - dh), x_2(t), \dots, x_2(t - dh), x_{40}(t), \dots, x_{40}(t - dh)]$ , where we set  $d = 3$ . Figure 5 shows the results of filtering the  $x_1$  node in this system

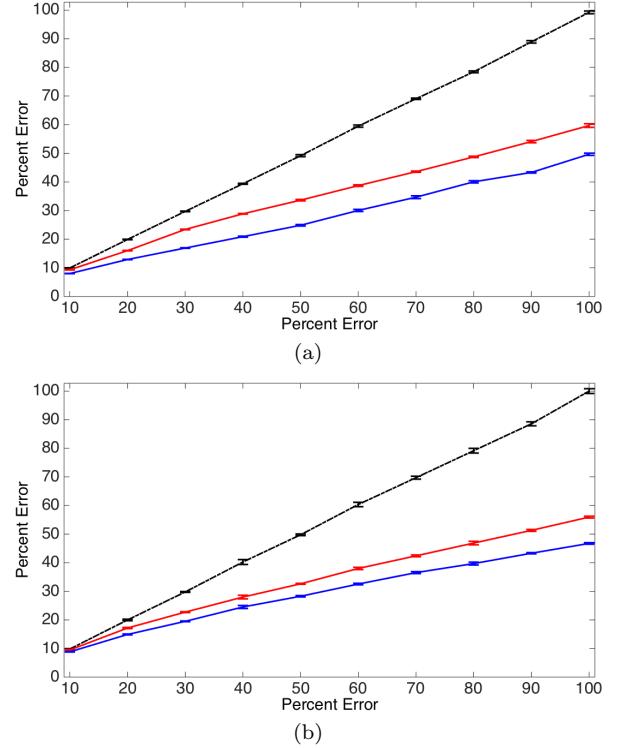


FIG. 6. Filter performance of the parametric (solid blue curve) and Kalman-Takens filter (solid red curve) when denoising the observed Lorenz-96  $x_1$  variable at increasing levels of observation noise (signal error, dotted black curve). The variable was sampled at rate  $h = 0.05$  and the error bars denote standard error over 10 realizations.  $Q$  and  $R$  noise covariances were tuned offline for optimal performance of the filters. Results presented for both a (a) 5 dimensional Lorenz-96 system with only  $x_1$  observable and a (b) 40 dimensional Lorenz-96 system where  $x_1, x_2$  and  $x_{40}$  are observable. We assume additional nodes are observable in the 40 dimensional case due to the high dimensionality of the system which inhibits both the parametric and Kalman-Takens filter. In both systems, the Kalman-Takens filter is able to significantly reduce the signal error to a level comparable to the parametric filter that has full knowledge of the system.

when the observations are perturbed by 60% noise (green circles). With knowledge of the full 40 dimensional system a parametric filter can be implemented, solid blue curve in Figure 5(a), to reduce the signal error (RMSE = 1.17). With no model available the Kalman-Takens method can be utilized, solid red curve in Figure 5(b), and even in this high dimensional example is able to reduce signal error to a level comparable to the parametric filter (RMSE = 1.36).

Figure 6 shows the results of filtering the  $x_1$  node at various noise levels. Figure 6(a) shows the results in a 5 dimensional Lorenz-96 network where only  $x_1$  is observable and Figure 6(b) shows the results in a 40 dimensional network where  $x_1, x_2$  and  $x_{40}$  are observable. In this example,  $Q$  and  $R$  noise covariances were tuned offline for

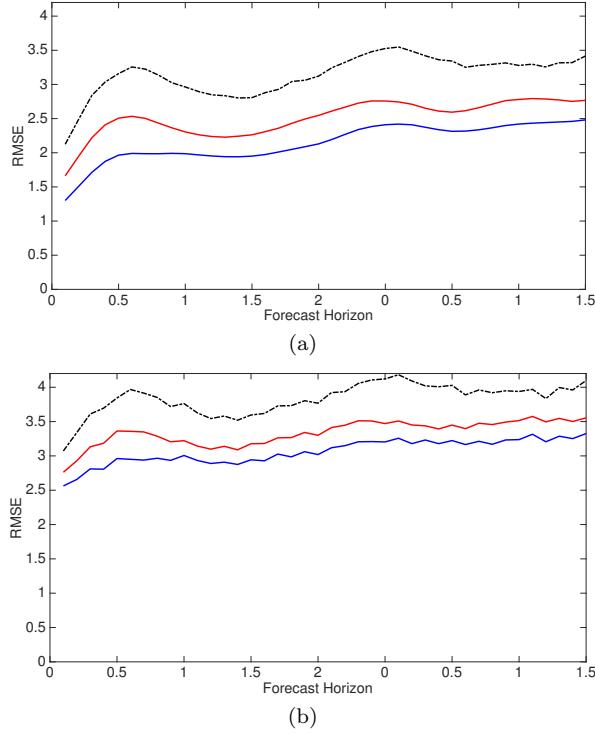


FIG. 7. Results of forecasting the  $x_1$  node in a 5 dimensional Lorenz-96 system when the training data are affected by 60% noise. Results averaged over 2500 realizations. Forecast error calculated with respect to (a) the non-noisy truth and (b) the noisy truth. Even in this higher dimensional system, we see that use of Kalman-Takens to filter the training data results in improved forecasting capability (solid red curve) as compared to forecasting without filtering (dotted black curve). The forecast based on Kalman-Takens initial conditions compares well to the same Takens based forecast method using initial conditions from the parametric filter (solid blue curve).

each filter to ensure optimal performance. In both network sizes, the Kalman-Takens method (solid red curve) is able to significantly reduce the signal error (dotted black curve) to a level comparable with the parametric filter (solid blue curve) which operates with knowledge of the full system equations.

Figure 7 shows the results of forecasting  $x_1$  in a 5 dimensional Lorenz-96 network, when the training data is corrupted by 60% noise, with respect to the non-noisy truth in Figure 7(a) and the noisy truth in Figure 7(b). Prediction with the unfiltered training data (dotted black curve) results in large errors. Once again, application of the Kalman-Takens method results in predictions with smaller error (solid red curve) and is comparable if a parametric filter is used (solid blue curve). The success of the Kalman-Takens filter in these higher dimensional systems is encouraging for its use in real-world physical systems where the system is complex, high dimensional and a model is most likely unknown.

We emphasize that we do not expect to obtain long

term forecasts of intrinsically high-dimensional dynamics from a nonparametric technique (due to the curse of dimensionality). However, we now return to the example in Figure 1 which shows that filtering such systems without a model is sometimes possible. Indeed, having shown that we can filter a single spatial location, we can easily filter the entire spatiotemporal system one location at a time. At each spatial node,  $x_i$ , of the 40-dimensional Lorenz-96 network, we build the short-term forecast model used by the Kalman-Takens filter from  $x_i$  along with its two spatial neighbors  $x_{i+1}$  and  $x_{i-1}$ .

The full spatiotemporal filtering results are shown in Fig. 1 of Section I. The Kalman-Takens filter is applied to the full 40-dimensional Lorenz-96 system when the observations are corrupted by 100% noise, i.e. additive Gaussian noise with mean zero and variance equal to the variance of the signal. Figure 1(a) shows the spatiotemporal plot of the system dynamics along with the noisy signal and the filter outputs. Note that the Kalman-Takens filter output (third panel from the top) is able to filter a significant amount of noise and reconstruct the underlying system dynamics at level comparable to the parametric filter (fourth from the top). Figure 1(b) shows the spatiotemporal plot of absolute errors that are greater than 5 for the (top) noisy signal, (middle) Kalman-Takens output, and (bottom) parametric filter output.

The key to this application of the Kalman-Takens filter is that the dynamical system is spatially localized on short time scales. This allows us to learn short-term forecasting models for each spatial location using only its two spatial neighbors. These short-term forecasting models appear to be low dimensional enough that they can be learned from small training data sets with Takens based methods. Of course, not all dynamical systems will have this localization property, in particular systems with an infinite speed of propagation of information may be problematic.

## V. MODEL ERROR

So far, we have been working under the assumption that while a set of noisy data exists from a system we have no knowledge of the dynamical equations. To that extent, we have shown the power of the Kalman-Takens method in filtering the noisy observations allowing for better predictions. Next, we ask an even more provocative question: Are there circumstances in which a system with noisy observations can be predicted better without a model than with a model?

This may not be possible if a perfect model is available. However, in any typical filtering problem, the question of model error becomes relevant: How accurately do the dynamics of the model describe those of the observed system? Moreover, if the model is slightly imperfect, how is the filtering affected? Of course, model error is not relevant to the Kalman-Takens method as it relies solely

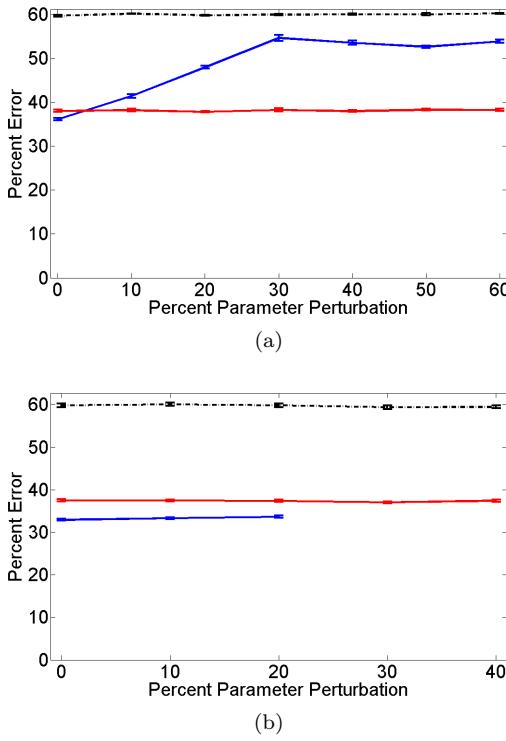


FIG. 8. Under situations of model error, the performance of the parametric filter can degrade rapidly. In our examples, this model error was simulated by perturbing the system parameters. Even under this relatively mild form of model error, it is evident that the parametric filter can have difficulty in filtering a noisy observation. Error bars denote standard error over 10 realizations. (a) Results for filtering the Lorenz-63  $x$  variable when the observations are perturbed by 60% noise (dotted black curve indicates signal error). The performance of the parametric filter (solid blue curve) under slight model error becomes much worse than the Kalman-Takens filter (solid red curve) which is completely robust to any form of model error. (b) Results for filtering the  $x_1$  node in a 40 dimensional Lorenz-96 system when observations are perturbed by 60% noise. Only three data points shown for the parametric filter because starting at 30% parameter perturbation the filter diverges, once again showing the crippling effect of model error on the performance of the parametric filter.

on the observed data to nonparametrically advance the dynamics.

We examined this question of model error when filtering the  $x$  variable from Lorenz-63 and the  $x_1$  node from a 40 dimensional Lorenz-96 system, each corrupted by 60% noise. A mild form of model error was simulated in the corresponding parametric filters by perturbing the model parameters. Specifically  $\sigma, \rho, \beta$  in the Lorenz-63 filter and the  $a$  parameter in the Lorenz-96 filter were perturbed from their correct values.

Figure 8 shows the filtering results at increasing levels of model error in Lorenz-63 in Figure 8(a) and Lorenz-96

in Figure 8(b). As the model error increases, the performance of the parametric filter (solid blue curve) breaks down and is unable to reduce the signal error (dotted black curve). Of note, in Figure 8(b) there are only three data points plotted for the parametric filter due to filter divergence at higher levels of model error. The Kalman-Takens filter (solid red curve) is robust to model error and at higher levels of model error outperforms the parametric filter.

Of course, in a situation of incorrect parameters, parameter-fitting methods could be used to correct the parameters and reduce the error. For example, state-augmentation methods based on Kalman filtering [47] could be used. However, realistic model error often manifests itself as missing variables in the model, a mismatch between the governing physics of the system and those described by the model or a completely unknown model. The consideration of incorrect parameter values is an extremely mild form of model error since the underlying equations used to generate the data match those of the model used by the parametric filter. Our purpose in this example is to compare the performance of the two filtering approaches in the mildest of model error situations – a more realistic comparison would cause the parametric approach to fail even more substantially.

In many complex modeling problems, empirical observation leads to ad hoc identification of model error phenomena. One example of this is the El Niño phenomenon, an irregular variation in sea surface temperatures (SST) in the eastern Pacific ocean. The El Niño phenomenon is commonly described by various univariate indices [48], which are time series that describe the SST field in a particular region. El Niño was shown in [49] to be more accurately forecast by a data-driven model than by reduced models used in operational forecasts at the time. While the method of [49] forecasts the entire SST field, comparable forecast skill for the El Niño index was obtained in [41] using only the historical time series of the one dimensional index.

The diffusion forecast of [41] uses training data to represent an unknown stochastic dynamical system on a custom basis of orthogonal functions,  $\varphi_j(x_i)$  which are defined on the attractor  $\mathcal{M}$  of the dynamical system described by the data  $\{x_i\} \subset \mathcal{M} \subset \mathbb{R}^n$ . By representing an initial probability density in this basis as  $p(x, 0) = \sum_j c_j(0)\varphi_j(x_i)$ , the diffusion forecast evolves the density using a Markov process on the coefficients  $c_j(t + \tau) = \sum_k A_{jk}c_k(\tau)$ . The key result of [41] is that  $A_{jk} = \frac{1}{N} \sum_i \varphi_k(x_i)\varphi_j(x_{i+1})$  is an unbiased estimate of the forward operator  $e^{\tau\mathcal{L}}$  for the unknown dynamical system, where  $\tau$  is the time step from  $x_i$  to  $x_{i+1}$  and  $\mathcal{L}$  is the generator of the stochastic dynamical system. We can then obtain the initial probability density from the noisy observations using an iterative Bayesian filter introduced in [19, 42]. These initial densities are projected into the basis  $\{\varphi_j\}$ , the coefficients are evolved forward in time, and the diffusion forecast is the mean of the reconstructed forecast density.

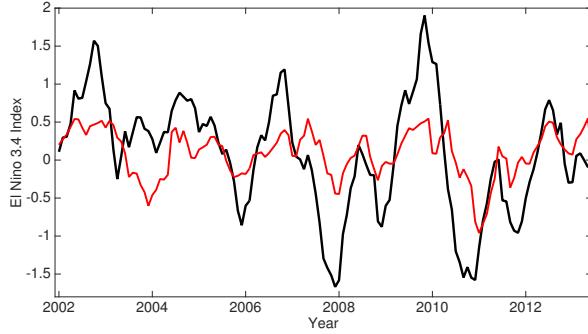
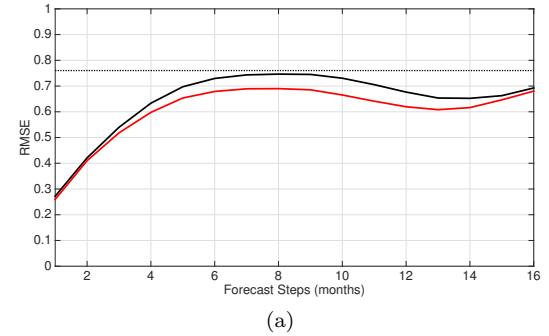


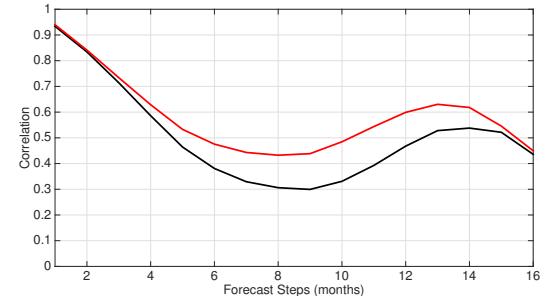
FIG. 9. The Kalman-Takens method was used to filter a set of historical data from the El Niño 3.4 index. The resulting 14-month lead nonparametric diffusion forecast using the filtered historical data is shown (solid red curve). The forecasted trajectory captures the general pattern of the true trajectory (solid black curve).

The difficulty in applying data-driven methods such as [41, 49] to empirically identified model error phenomena such as El Niño is that it is extremely difficult for the modeler to completely isolate the model error phenomena. A trivial example which illustrates this difficulty is that the El Niño indices contain seasonal variations which are not part of the model error (these are empirically removed by computing the monthly anomalies). Of course, even if the seasonality could be perfectly removed, many other aspects of the observed El Niño index are likely to be artifacts of the projection from a high-dimensional chaotic dynamical system to a 1-dimensional time series [15]. As a result, we expect that the observed data contains a dominant dynamical component which is possible to probabilistically forecast, and this dominant component is the target that the modeler is hoping to explain. However, the observed data also contains unpredictable ‘noise’ components which result from the projection of the high-dimensional dynamics, and are nonlinearly intertwined with the dominant dynamical component that we are interested in. This common scenario is the perfect context for application of the Kalman-Takens filter to remove the unpredictable ‘noise’ component of the time series. This will allow a method such as the diffusion forecast to be trained on a cleaner time series which is more focused on the desired feature.

To demonstrate this we examined the filter-forecast problem in the context of the El Niño 3.4 index which was also used in [41]. We used historical data from January 1950 to December 1999 to build our nonparametric forecast. Prior to building the diffusion forecast model, we first filter the data using the Kalman-Takens method with  $d = 9$ , 5 neighbors and  $Q$  and  $R$  noise covariance matrices tuned for optimal performance. Figure 9 compares the 14-month diffusion forecast using the filtered training data (solid red curve) to the true value of the El Niño index at the corresponding time (solid black curve). We note that our predictions follow the same general pat-



(a)



(b)

FIG. 10. Results of the diffusion forecast when trained on the unfiltered historical data (solid black curve) and Kalman-Takens filtered data (solid red curve). (a) Both forecasts compare favorably to the climatological error (dotted grey curve). However, processing the historical data with the Kalman-Takens filter results in improved forecast error. (b) Similarly, forecast correlation is improved by filtering the data.

tern of the true trajectory. However, our real interest is in determining if use of the Kalman-Takens filter results in any improvement in forecasting capability. In Figure 10(a) we show the resulting forecast error and in Figure 10(b) the forecast correlation of the diffusion forecast when the historical data is unfiltered (solid black curve) and filtered using Kalman-Takens (solid red curve). We observe that denoising the time series using the Kalman-Takens filter improved the isolation of the predictable component of the El Niño 3.4 index as shown by the improved forecast skill of the diffusion forecast when trained on the denoised data.

## VI. DISCUSSION

The blending of the Takens embedding method with Kalman filtering is designed to exploit the complementary strengths of the two methodologies. Under favorable conditions, delay coordinate embedding can replace the underlying evolution equations with no loss in accuracy, and Kalman update provides a maximum likelihood estimate of the reconstructed state in the presence of observational noise. While the Kalman update equations were proposed with the assumption that a known set of

dynamical evolution equations were available, we show here that they can be used to effectively reduce observational noise and increase predictability when using nonparametric representations, such as those derived from delay coordinates.

Perhaps more surprisingly, our results show that the Kalman-Takens filter may outperform the standard, parametric Kalman filter when model error is present. The sensitivity of filter output to model error is difficult to quantify; we show that in certain cases, even when the model error is relatively small, the nonparametric approach may be a welcome alternative.

The Kalman-Takens filter shares the limitations of all Kalman filtering to the extent that it is limited to Gaussian noise assumptions. For more general, multimodal noise, a more complex type of data assimilation may be necessary. In this article, we restricted the computational examples to the EnKF form, but the fundamental strategy is not restricted to any particular version of Kalman filtering.

We expect this idea to be applicable to a wide range of nonlinear systems that are measured with observational noise. For forecasting, the data requirements are similar to other applications of Takens' theorem, in that enough observations must be available to reconstruct the underlying dynamical attractor to a sufficient extent. However, even if the system exhibits high-dimensional dynamics, we show that filtering may still be possible if some spatial localization can be done. Intuitively, this means that on short time scales, local information is sufficient to determine short-term dynamics. This may be used productively in geophysical applications where considerable pre-processing is needed, as a kind of *nonparametric reanalysis*. Such a reanalysis may reduce bias from incorrect priors such as insufficient models or poorly constrained parameters.

**Acknowledgments.** We thank the referees for several helpful suggestions. The research was partially supported by grants CMMI-1300007, DMS-1216568, DMS-1250936 and RTG/DMS-1246991 from the National Science Foundation.

- 
- [1] E. Kalnay, *Atmospheric modeling, data assimilation, and predictability* (Cambridge Univ. Press, 2003).
  - [2] Geir Evensen, *Data assimilation: The Ensemble Kalman Filter* (Springer: Heidelberg, 2009).
  - [3] Florence Rabier, “Overview of global data assimilation developments in numerical weather-prediction centres,” *Quarterly Journal of the Royal Meteorological Society* **131**, 3215–3233 (2005).
  - [4] B. R. Hunt, E. Kalnay, E. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin, “Four-dimensional ensemble kalman filtering,” *Tellus A* **56**, 273–277 (2004).
  - [5] James A. Cummings, “Operational multivariate ocean data assimilation,” *Quarterly Journal of the Royal Meteorological Society* **131**, 3583–3604 (2005).
  - [6] Kyo Yoshida, Junzo Yamaguchi, and Yukio Kaneda, “Regeneration of small eddies by data assimilation in turbulence,” *Phys. Rev. Lett.* **94**, 014501 (2005).
  - [7] K. Law and A. Stuart, “Evaluating data assimilation algorithms,” *Mon. Wea. Rev.* **140**, 3757–3782 (2012).
  - [8] S.J. Schiff, *Neural control engineering* (MIT Press, 2012).
  - [9] T. Berry and T. Sauer, “Adaptive ensemble kalman filtering of nonlinear systems,” *Tellus A* **65**, 20331 (2013).
  - [10] S. Julier, J. Uhlmann, and H. Durrant-Whyte, “A new method for the nonlinear transformation of means and covariances in filters and estimators,” *IEEE Trans. Automat. Control* **45**, 477–482 (2000).
  - [11] S. Julier, J. Uhlmann, and H. Durrant-Whyte, “Unscented filtering and nonlinear estimation,” *Proc. IEEE* **92**, 401–422 (2004).
  - [12] Rolf H Reichle and Randal D Koster, “Global assimilation of satellite surface soil moisture retrievals into the nasa catchment land surface model,” *Geophysical Research Letters* **32** (2005).
  - [13] H Hersbach, A Stoffelen, and S De Haan, “An improved c-band scatterometer ocean geophysical model function: Cmod5,” *Journal of Geophysical Research: Oceans* (1978–2012) **112** (2007).
  - [14] HM Arnold, IM Moroz, and TN Palmer, “Stochastic parametrizations and model uncertainty in the lorenz’96 system,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20110479 (2013).
  - [15] Tyrus Berry and John Harlim, “Linear theory for filtering nonlinear multiscale systems with model error,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **470**, 20140168 (2014).
  - [16] Edward N Lorenz and Kerry A Emanuel, “Optimal sites for supplementary weather observations: Simulation with a small model,” *J. Atmos. Sci.* **55**, 399–414 (1998).
  - [17] Tim N Palmer, “A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models,” *Quart. J. Roy. Met. Soc.* **127**, 279–304 (2001).
  - [18] F. Hamilton, T. Berry, and T. Sauer, “Predicting chaotic time series with a partial model,” *Physical Review E* **92**, 010902 (2015).
  - [19] Tyrus Berry and John Harlim, “Semiparametric forecasting and filtering: correcting low-dimensional model error in parametric models,” Submitted to *Journal of Computational Physics*, available at <http://arxiv.org/abs/1502.07766> (2015).
  - [20] F. Takens, “Detecting strange attractors in turbulence,” *Lecture Notes in Math.* Springer-Verlag: Berlin **898** (1981).
  - [21] N. Packard, J. Crutchfield, D. Farmer, and R. Shaw, “Geometry from a time series,” *Phys. Rev. Lett.* **45**, 712–715 (1980).
  - [22] T. Sauer, J.A. Yorke, and M. Casdagli, “Embedology,” *J. Stat. Phys.* **65**, 579–616 (1991).
  - [23] T. Sauer, “Reconstruction of shared nonlinear dynamics in a network,” *Phys. Rev. Lett.* **93**, 198701–4 (2004).

- [24] Edward Ott, Tim Sauer, and James A Yorke, *Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems* (Wiley New York, 1994).
- [25] H Abarbanel, *Analysis of observed chaotic data* (Springer-Verlag, New York, 1996).
- [26] Holger Kantz and Thomas Schreiber, *Nonlinear time series analysis* (Cambridge University Press, 2004).
- [27] J Farmer and J Sidorowich, “Predicting chaotic time series,” Phys. Rev. Lett. **59**, 845–848 (1987).
- [28] Martin Casdagli, “Nonlinear prediction of chaotic time series,” Physica D: Nonlinear Phenomena **35**, 335–356 (1989).
- [29] George Sugihara and Robert M. May, “Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series,” Nature **344**, 734–741 (1990).
- [30] Leonard A Smith, “Identification and prediction of low dimensional dynamics,” Physica D: Nonlinear Phenomena **58**, 50–76 (1992).
- [31] J Jimenez, JA Moreno, and GJ Ruggeri, “Forecasting on chaotic time series: A local optimal linear-reconstruction method,” Phys. Rev. A **45**, 3553 (1992).
- [32] T Sauer, “Time series prediction by using delay coordinate embedding,” in *Time Series Prediction: Forecasting the Future and Understanding the Past* (Addison Wesley, 1994) pp. 175–193.
- [33] George Sugihara, “Nonlinear forecasting for the classification of natural time series,” Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences **348**, 477–495 (1994).
- [34] Christian G Schroer, Tim Sauer, Edward Ott, and James A Yorke, “Predicting chaos most of the time from embeddings with self-intersections,” Phys. Rev. Lett. **80**, 1410 (1998).
- [35] D Kugiumtzis, OC Lingjærde, and N Christoffersen, “Regularized local linear prediction of chaotic time series,” Physica D: Nonlinear Phenomena **112**, 344–360 (1998).
- [36] G Yuan, M Lozier, L Pratt, C Jones, and K Helfrich, “Estimating the predictability of an oceanic time series using linear and nonlinear methods,” J. Geophys. Res. **109**, C08002 (2004).
- [37] Chih-Hao Hsieh, Sarah M Glaser, Andrew J Lucas, and George Sugihara, “Distinguishing random environmental fluctuations from ecological catastrophes for the north pacific ocean,” Nature **435**, 336–340 (2005).
- [38] Christopher C Strelcoff and Alfred W Hübner, “Medium-term prediction of chaos,” Phys. Rev. Lett. **96**, 044101 (2006).
- [39] S Regonda, B Rajagopalan, U Lall, M Clark, and Y-I Moon, “Local polynomial method for ensemble forecast of time series,” Nonlin. Proc. in Geophys. **12**, 397–406 (2005).
- [40] Björn Schelter, Matthias Winterhalder, and Jens Timmer, *Handbook of time series analysis: recent theoretical developments and applications* (John Wiley and Sons, 2006).
- [41] Tyrus Berry, Dimitrios Giannakis, and John Harlim, “Nonparametric forecasting of low-dimensional dynamical systems,” Phys. Rev. E **91**, 032915 (2015).
- [42] Tyrus Berry and John Harlim, “Forecasting turbulent modes with nonparametric diffusion models,” Submitted to Physica D, available at <http://arxiv.org/abs/1501.06848> (2015).
- [43] David M Walker, “Parameter estimation using kalman filters with constraints,” International Journal of Bifurcation and Chaos **16**, 1067–1078 (2006).
- [44] D.. Simon, *Optimal State Estimation: Kalman,  $H_\infty$ , and Nonlinear Approaches* (John Wiley and Sons, 2006).
- [45] E Lorenz, “Deterministic nonperiodic flow,” J. Atmos. Sci. **20**, 130–141 (1963).
- [46] Edward N Lorenz, “Predictability: A problem partly solved,” in *Proceedings: Seminar on predictability*, Vol. 1 (AMS, Reading, UK, 1996) pp. 1–18.
- [47] A. Sitz, U. Schwarz, J. Kurths, and H. Voss, “Estimation of parameters and unobserved components for nonlinear systems from noisy time series.” Physical Review E **66**, 16210 (2002).
- [48] NOAA, “Sst and other climate timeseries,” (2014), [http://www.esrl.noaa.gov/psd/gcos\\_wgsp/Timeseries/](http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/).
- [49] M. Chekroun, D. Kondrashow, and M. Ghil, “Predicting stochastic systems by noise sampling and application to the el nino - southern oscillation,” Proc. Nat. Acad. Sci. **108**, 11766–11771 (2011).

## Appendix: Adaptive updating of covariance matrices.

Since we cannot assume that the noise covariance matrices  $Q$  and  $R$  in (1) are known, we used a recently-developed method for the adaptive fitting of these matrices [9] as part of the filtering algorithm. The method uses the innovations  $\epsilon_k \equiv y_k - \bar{y}_k^-$  in observation space from (2) to update the estimates  $Q_k$  and  $R_k$  of the covariances  $Q$  and  $R$ , respectively, at step  $k$  of the filter.

First, we construct linearizations of the dynamics and the observation function that are reconstructed from the ensembles used by the EnKF. Let  $x_{i,k-1}^+ \sim \mathcal{N}(x_{k-1}^+, P_{k-1}^+)$  be the analysis ensemble at step  $k-1$ , where the index  $1 \leq i \leq E$  indicates the ensemble member, and let  $x_{i,k}^- = \mathcal{F}(x_{i,k-1}^+)$  be the forecast ensemble which results from moving from  $t_{k-1}$  to  $t_k$  using the Takens embedding with initial condition  $x_{i,k-1}^+$ . Define  $x_k^- = \frac{1}{E} \sum_{i=1}^E x_{ik}^-$ , the matrix of analysis perturbations

$$X_{k-1}^+ = [(x_{1,k-1}^+ - x_{k-1}^+)^T, \dots, (x_{E,k-1}^+ - x_{k-1}^+)^T]$$

and the matrix of forecast perturbations

$$X_k^- = [(x_{1k}^- - x_k^-)^T, \dots, (x_{Ek}^- - x_k^-)^T].$$

Then an approximation for the one-step dynamics is given by the matrix  $F_k = X_k^- (X_{k-1}^+)^{\dagger}$ , where  $\dagger$  denotes the matrix pseudo-inverse. Similarly, let  $\tilde{x}_{i,k}^- \sim \mathcal{N}(x_k^-, P_k^- + Q)$  be the inflated forecast ensemble and let  $z_{i,k}^- = h(\tilde{x}_{i,k}^-)$  be the projection of this ensemble into the observation space. Then we can define  $H_k = Z_k^- (\tilde{X}_k^-)^{\dagger}$ , which we think of as a local linearization of the observation function  $h$ , where

$$\tilde{X}_k^- = [(\tilde{x}_{1k}^- - x_k^-)^T, \dots, (\tilde{x}_{Ek}^- - x_k^-)^T]$$

$$Z_k^- = [(z_{1k}^- - z_k^-)^T, \dots, (z_{Ek}^- - z_k^-)^T]$$

are the matrix of inflated forecast perturbations and the matrix of observed forecast perturbations, respectively, and where  $z_k^- = \frac{1}{E} \sum_{i=1}^E z_{ik}^-$ .

After computing  $F_k$  and  $H_k$ , we use them to update covariances  $Q$  and  $R$  as follows. We produce empirical estimates  $Q_{k-1}^e$  and  $R_{k-1}^e$  of  $Q$  and  $R$  based on the innovations at time  $k$  and  $k-1$  using the formulas

$$\begin{aligned} P_{k-1}^e &= F_{k-1}^{-1} H_k^{-1} \epsilon_k \epsilon_{k-1}^T H_{k-1}^{-T} + K_{k-1} \epsilon_{k-1} \epsilon_{k-1}^T H_{k-1}^{-T} \\ Q_{k-1}^e &= P_{k-1}^e - F_{k-2} P_{k-2}^a F_{k-2}^T \\ R_{k-1}^e &= \epsilon_{k-1} \epsilon_{k-1}^T - H_{k-1} P_{k-1}^f H_{k-1}^T. \end{aligned} \quad (6)$$

It was shown in [9] that  $P_{k-1}^e$  is an empirical estimate of the background covariance at time index  $k-1$ . Notice that this procedure requires us to save the linearizations  $F_{k-2}, F_{k-1}, H_{k-1}, H_k$ , innovations  $\epsilon_{k-1}, \epsilon_k$ , and the analysis  $P_{k-2}^a$  and Kalman gain matrix,  $K_{k-1}$ , from the  $k-1$  and  $k-2$  steps of the EnKF.

The estimates  $Q_{k-1}^e$  and  $R_{k-1}^e$  are low-rank, noisy estimates of the covariance matrices  $Q$  and  $R$  that will make the posterior estimate statistics from the filter consistent with the empirical statistics in the sense of (6). In order

to form stable full-rank estimates of  $Q$  and  $R$  we assimilate these estimates using an exponential moving average with window  $\tau$ :

$$\begin{aligned} Q_k &= Q_{k-1} + (Q_{k-1}^e - Q_{k-1})/\tau \\ R_k &= R_{k-1} + (R_{k-1}^e - R_{k-1})/\tau. \end{aligned} \quad (7)$$

We interpret the moving average in (7) as a moving average filter that stabilizes the noisy empirical estimates  $Q_k$  and  $R_k$ . The stochastic nature of the estimate of  $Q_k$  can lead to excursions that fail to be symmetric and/or positive definite, leading to instability in the EnKF. While the matrix  $Q_k$  is not changed, the matrix used in the  $k$ -th step of the filter is a modified version of  $Q_k$  which is forced to be symmetric and positive definite by taking  $\tilde{Q}_k = (Q_k + Q_k^T)/2$  and then taking the max of the eigenvalues of  $\tilde{Q}_k$  with zero. Again, we emphasize that  $\tilde{Q}_k$  is only used in the  $k$ -th filter step and no ad-hoc corrections are made to the matrix  $Q_k$  which eventually stabilizes at a symmetric and positive definite matrix naturally via the moving average in (7). These ad-hoc corrections are only needed during the transient period of the estimation of  $Q$ , in particular when the trivial initialization  $Q_1 = 0$  is used.