Simple noise-reduction method based on nonlinear forecasting

James P. L. Tan*
Interdisciplinary Graduate School, Nanyang Technological University, Singapore
and Complexity Institute, Nanyang Technological University, Singapore

(Received 15 December 2016; revised manuscript received 18 February 2017; published 20 March 2017)

Nonparametric detrending or noise reduction methods are often employed to separate trends from noisy time series when no satisfactory models exist to fit the data. However, conventional noise reduction methods depend on subjective choices of smoothing parameters. Here we present a simple multivariate noise reduction method based on available nonlinear forecasting techniques. These are in turn based on state-space reconstruction for which a strong theoretical justification exists for their use in nonparametric forecasting. The noise reduction method presented here is conceptually similar to Schreiber's noise reduction method using state-space reconstruction. However, we show that Schreiber's method has a minor flaw that can be overcome with forecasting. Furthermore, our method contains a simple but nontrivial extension to multivariate time series. We apply the method to multivariate time series generated from the Van der Pol oscillator, the Lorenz equations, the Hindmarsh-Rose model of neuronal spiking activity, and to two other univariate real-world data sets. It is demonstrated that noise reduction heuristics can be objectively optimized with in-sample forecasting errors that correlate well with actual noise reduction errors.

DOI: 10.1103/PhysRevE.95.032218

I. INTRODUCTION

For time series obtained from real-world complex systems, it is often the case that one neither has nor knows an accurate mechanistic model to fit the data. Indeed, nonparametric models are becoming increasingly favored to capture the complexities and nuances that simplified mechanistic models cannot [1,2]. In the absence of any reliable mechanistic model, it becomes necessary to resort to nonparametric detrending methods to separate noise from deterministic trends. Semantically, such an endeavor may be known as noise reduction or detrending depending on what one wants to recover from the noisy time series. Regardless, there is generally no distinction between detrending and noise reduction methods since the intermediate goal of separating noise from trend is equivalent. Conventional nonparametric methods such as Loess smoothing and kernel smoothing are problematic due to the subjective choice of a time scale over which to smoothen data. Furthermore, it is unclear if recovered trends accurately represent any dynamics inherent in the time series. This ambiguity also afflicts a more recent and popular method known as empirical mode decomposition (EMD) which attempts to avoid the issue of having to subjectively choose an appropriate time scale [3].

Here we adopt an approach to the problem of nonparametric regression by obtaining the trend of a time series using in-sample forecasts. By casting the noise reduction problem as one of forecasting, we show that unambiguous trends can be objectively recovered from noisy time series. The intuition behind this endeavor is rather straightforward: A reliable forecast one time step ahead is a projection of reconstructed dynamics from available time series. Hence, a series of reliable forecasts represents a trend that captures essential dynamics inherent in the time series. We shall call such a trend a *dynamical trend*.

*tanp0100@e.ntu.edu.sg

II. DYNAMICAL TREND

Let a multivariate time series $\mathbf{Y}_t$ of dimension n be fully determined from its history of past states and noise terms,

$$\mathbf{Y}_t = f(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{t-1}, \ldots, \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots), \tag{1}$$

where $\boldsymbol{\epsilon}_t$ is a multivariate random variable of dimension l at time t with joint probability distribution parameterized by past states,

$$\boldsymbol{\epsilon}_t \sim F(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots). \tag{2}$$

Then the dynamical trend $\mathbf{Z}_t$ of $\mathbf{Y}_t$ is defined as

$$\mathbf{Z}_t = f_{\text{pred}}(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots, \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots), \tag{3}$$

such that the mean-squared error between $\mathbf{Z}_t$ and $\mathbf{Y}_t$,

$$\mathrm{E}[(\mathbf{Z}_t - \mathbf{Y}_t)^2], \tag{4}$$

is minimized. Hence, $\mathbf{Z}_t$ represents the best possible forecast of $\mathbf{Y}_t$ without future knowledge of the numbers that were sampled for $\boldsymbol{\epsilon}_t$ but with knowledge of the statistical distribution of $\boldsymbol{\epsilon}_t$.

For this paper, we shall consider the noisy multivariate time series $\mathbf{Y}_t$ of the form

$$\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\epsilon}_t, \tag{5}$$

where $\boldsymbol{\epsilon}_t$ is a multivariate continuous random variable of joint probability density function $p(\boldsymbol{\epsilon})$ with mean $\mathbf{0}$, and $\mathbf{X}_t$ is a deterministic time series,

$$\mathbf{X}_t = f(\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots). \tag{6}$$

The mean-squared error is then

$$\int_{\Omega} (\mathbf{Z}_t - \mathbf{Y}_t)^2 \, p(\boldsymbol{\epsilon}_t) \, d\epsilon_t, \tag{7}$$

where the integral is over the sample space Ω of $\boldsymbol{\epsilon}_t$. Therefore, the mean-squared error is minimized when $\mathbf{Z}_t = \mathbf{X}_t$. In this case, the dynamical trend is simply the time series of the deterministic system. If the goal of noise reduction is to recover $\mathbf{X}_t$ from $\mathbf{Y}_t$, then forecasting ability is equivalent to

noise reduction performance in such a system. By associating noise reduction performance with forecasting ability, a noise reduction method can be made to be objective by optimizing its parameters based on the ability to forecast. In this paper, we will be concerned with recovering $\mathbf{X}_t$ from $\mathbf{Y}_t$, with $\boldsymbol{\epsilon}_t$ being a Gaussian white noise vector of variance σ^2. Therefore, the results contained in this paper are for additive observational noise in the absence of dynamical noise except in Sec. IIIA near the end of the paper, where we discuss multiplicative dynamical and observational noise including other limitations of the noise reduction method.

III. NOISE REDUCTION METHOD

Forecasts were conducted using a class of nonlinear forecasting techniques that derive from a method known as *state-space reconstruction* [4]. First introduced in a seminal paper by Packard *et al.* [5], and fleshed in mathematical rigor with Taken's theorem [6], state-space reconstruction allows for the reconstruction of a multidimensional state space from the lags of a single state variable. In this work, univariate time series were forecast using simplex projection [7]. For multivariate time series, multiview embedding (MVE) was used because embeddings from different combinations of variables and lags may not be equally useful in forecasting ability with the presence of noise and limited data [8–10]. Instead of relying on any particular state variable, MVE selects the best combinations of variables and lags from in-sample forecasts [9]. In essence, forecasting using state-space reconstruction means that each corrected point is obtained by forecasting using nearest neighbors in the reconstructed state spaces one time step before.

Utilizing state-space reconstruction for the purposes of noise reduction is not new and literature on such methods has existed since the mid-1990s [11]. Our noise reduction method is most conceptually similar to Schreiber's method [12]. In Schreiber's method, nearest neighbors in the reconstructed state space of a point to be corrected are averaged to produce the corrected point. This is not ideal because the noise terms that are supposed to be averaged over were involved in determining the nearest neighbors. Consider the case where nearest neighbors are identified from a noisy time series in a small neighborhood about the point to be corrected in the reconstructed state space. Then the corrected point is relatively unchanged from the original. This necessitates an increase of neighborhood size until a reasonable correction is available, which means the inclusion of nearest neighbours farther away from the original point. Such a problem can be mitigated by correcting the point using a forecast one time step ahead from nearest neighbors of the state one time step before. In this way, the noise terms to be averaged over would be independent of the terms used to identify the nearest neighbors.

In combination with these forecasting techniques, we make use of two heuristics inspired from previous literature that can be optimized with in-sample forecasting to improve noise reduction performance [11]:

(1) Under a time reversal, the time series also contains information on the dynamics of the system. Hence, forecasting performance may be improved if the ability to forecast backward is as good or even better than the ability to forecast forward. This leads to three possible noise reduction algorithms. The first is based on forward forecasting, the second is based on backward forecasting, while the third relies on a combination of both forward and backward forecasting where the forward forecast and the backward forecast are combined with a simple average. We call these three variants the forward algorithm, the backward algorithm, and the bidirectional algorithm.

(2) As pointed out by Schreiber [12], noise reduction from a first pass of the algorithm may not be optimal. The noise reduction algorithm may then be applied recursively on corrected time series to improve noise reduction performance. Thus, the number of times the noise reduction algorithm is run recursively, r, becomes a parameter to optimize.

In state-space reconstruction, multivariate time series $\mathbf{B}_t$ in a multidimensional state space of dimension E (also called the embedding dimension) can be constructed from $E-1$ lags of a univariate time series Y_t, i.e.,

$$\mathbf{B}_t = (Y_t, Y_{t-1}, \ldots, Y_{t-E+1}). \tag{8}$$

In simplex projection, to obtain a forecast one time step ahead for a state vector $\mathbf{B}_0$, the $E+1$ nearest neighbors of $\mathbf{Y}'_0$ are identified and the forecast is computed from the corresponding vectors of these nearest neighbors one step ahead in time [7]. The computation is done by averaging with exponential weights according to the Euclidean distance to $\mathbf{B}_0$. Therefore, the forecast $\hat{\mathbf{B}}_1$ is given by

$$\hat{\mathbf{B}}_1 = \sum_{\text{nni}} \mathbf{B}_{nni+1} w_{\text{nni}}, \tag{9}$$

where nni (for nearest neighbor index) is the time index of one of the nearest neighbors of $\mathbf{B}_0$ and

$$w_{\text{nni}} = h \exp[-||\mathbf{B}_{\text{nni}} - \mathbf{B}_0||/\min(d)]. \tag{10}$$

Here h is a normalization constant for the weights and $\min(d)$ refers to the smallest distance between $\mathbf{B}_0$ and its nearest neighbors. If only the forecast for Y_1 is needed, then only the first coordinate of $\hat{\mathbf{B}}_1$ needs to be computed to obtain the forecast for Y_1.

In MVE, state-space reconstructions with embedding dimension E are done for all variable and lag combinations of a multivariate time series $\mathbf{Y}_t$ such that each combination consists of at least a variable of lag 0 [9]. The top k reconstructions for each coordinate of $\mathbf{Y}_t$ are then chosen based on in-sample leave-one-out cross-validation (LOOCV) forecasting performance using simplex projection. In this case, in-sample forecasting performance for different embeddings is ranked by correlation between forecasts and the noisy time series. To obtain a forecast for MVE, the nearest neighbor from each reconstruction in the top k reconstructions is identified and the vectors from these nearest neighbors one step ahead in time are averaged over to produce the forecast. Following Ye and Sugihara [9], we set $k = \sqrt{m}$, where m is the number of available variable and lag combinations.

A. Forward algorithm

To obtain a corrected time series from in-sample forecasts, MVE was used for multivariate time series, whereas simplex projection was used for univariate time series. Here it should

be noted that for a corrected time series $\hat{\mathbf{Y}}_t$, we also made use of $\mathbf{Y}_t$ such that

$$\hat{\mathbf{Y}}_t = \alpha \mathbf{Y}_t + (1-\alpha)\hat{\mathbf{Y}}_{t+}, \tag{11}$$

where $0 < \alpha < 1$ is a real number that we set at 0.5 for all noise reduction done in this paper, and $\hat{\mathbf{Y}}_{t+}$ indicates the time series obtained by in-sample forecasts using MVE or simplex projection one time step forward in time (+).

B. Backward algorithm

The backward algorithm is the same as the forward algorithm, except that time series is first flipped horizontally before noise reduction with the forward algorithm. The corrected time series is then flipped horizontally again to give the corrected time series by the forward algorithm. The forecast is

$$\hat{\mathbf{Y}}_t = \alpha \mathbf{Y}_t + (1-\alpha)\hat{\mathbf{Y}}_{t-}, \tag{12}$$

where $\hat{\mathbf{Y}}_{t-}$ indicates the forecast time series made by the backward forecasting of MVE or simplex projection.

C. Bidirectional algorithm

The bidirectional algorithm combines the forecast of the forward algorithm and the backward algorithm by a simple average. The forecast is

$$\hat{\mathbf{Y}}_t = \alpha \mathbf{Y}_t + (1-\alpha)\hat{\mathbf{Y}}_{t\pm}, \tag{13}$$

where $\hat{\mathbf{Y}}_{t\pm} = (\hat{\mathbf{Y}}_{t+} + \hat{\mathbf{Y}}_{t-})/2$ indicates a forecast made by the forward and backward forecasting of MVE or simplex projection.

As alluded to before, another heuristic is to run the noise reduction algorithms recursively on corrected time series. Let $\hat{\mathbf{Y}}_t^{(r)}$ be the time series corrected by the bidirectional algorithm (for example) over r recursive iterations from the original coordinate time series $\mathbf{Y}_t$ such that

$$\hat{\mathbf{Y}}_t^{(r)} = \alpha \hat{\mathbf{Y}}_t^{(r-1)} + (1-\alpha)\hat{\mathbf{Y}}_{t\pm}^{(r-1)}, \tag{14}$$

where $\hat{\mathbf{Y}}_{t\pm}^{(r-1)}$ indicates the forecast made by the forward and backward forecasting of MVE or simplex projection using the time series $\hat{\mathbf{Y}}_t^{(r-1)}$. Hence, we define $\hat{\mathbf{Y}}_t^{(0)} = \mathbf{Y}_t$. The in-sample cross-validation forecasting error for a coordinate of $\hat{\mathbf{Y}}_t^{(r-1)}$ which we use as an estimate for the noise reduction performance of the corresponding coordinate in $\hat{\mathbf{Y}}_t^{(r)}$ is then the mean absolute error of $\hat{\mathbf{Y}}_{t\pm}^{(r-1)}$ measured against $\mathbf{Y}_t$ for that coordinate. The in-sample cross-validation errors for the other two algorithms were calculated in a similar way.

IV. RESULTS AND DISCUSSION

We test the noise reduction method on noisy time series sampled from the Van der Pol oscillator, the chaotic Lorenz equations, and the chaotic Hindmarsh-Rose model. Deterministic time series $\mathbf{X}_t$ sampled from these systems were combined with additive observational noise $\boldsymbol{\epsilon}_t$ to give $\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\epsilon}_t$, the noisy time series. The noise-reduced time series is obtained using in-sample forecasts one time step ahead. The errors from these forecasts (as calculated against the noisy time series) essentially constitute a performance measure (the mean absolute error, MAE) from LOOCV. This cross-validation error is used as an estimate of the potential noise reduction performance of the corrected time series obtained by the in-sample forecasts. This allows us to objectively identify the noise reduction parameters, i.e., which algorithm to run (forward, backward, or bidirectional) and how many times to run it recursively, based on the the lowest MAE. Ideally, the goals of noise reduction and forecasting are equivalent in these systems. However, we should not expect a perfect correlation between in-sample forecasting errors and actual noise reduction errors because in-sample forecasting errors are calculated against noisy time series, whereas actual noise reduction errors are calculated against the deterministic time series $\mathbf{X}_t$. A significant presence of noise also leads to complications such as an inaccurate reconstruction of state space which would significantly limit the ability to recover any meaningful trend in the noisy time series.

The results of the noise reduction method for several periods of the limit cycle from the Van der Pol oscillator can be seen in Figs. 1(a)–1(d). Here $\sigma^2 = 0.1$ and 800 data points were used. The in-sample forecasting errors for the x coordinate [Fig. 1(c)] from LOOCV correlates well with the actual noise reduction errors [Fig. 1(d)], i.e., the error between the noise-reduced time series and $\mathbf{X}_t$. In particular, the in-sample errors predict that making use of the bidirectional algorithm with five recursive iterations of the algorithm would be optimal, a result that was corroborated to a good degree by the actual noise reduction errors. The noise reduction of the noisy limit cycle for conventional noise reduction methods like Loess smoothing requires the subjective choice of the span. If $\mathbf{Y}_t$ contains only one oscillation of the limit cycle, then the behavior of the noise reduction algorithm presented here is conceptually similar to that of Loess smoothing in that cleaned data points are computed locally from nearest neighbors in time. This is the case because in a single oscillation, nearest neighbors in time are also nearest neighbors in space and cleaned data points in the algorithms are computed from nearest neighbors in reconstructed state spaces. However, if multiple oscillations are present, then, unlike Loess smoothing, a cleaned data point can also be computed across large differences in time. In this case the noise reduction method confers a higher performance over Loess smoothing whatever the span [13]. This is despite the fact that parameters from the noise reduction method were optimized objectively without knowledge of $\mathbf{X}_t$.

The results of the noise reduction method for the x coordinate of the chaotic Lorenz system can be seen in Figs. 1(e) and 1(f). Here $\sigma^2 = 20$ and 500 data points were used. It should be noted that with MVE, information from noisy time series belonging to the other two coordinates were also used in noise reduction of the noisy time series from the x coordinate. In-sample cross-validation errors also correlate well with actual noise reduction errors [Figs. 1(g) and 1(h)]. Optimized parameters from the in-sample cross-validation errors produces a noise-reduced time series that replicates the original deterministic time series remarkably well [Fig. 1(f)].

Last, results of the noise reduction method for the chaotic Hindmarsh-Rose model can be seen in Figs. 1(i) and 1(h). Here $\sigma^2 = 0.3$ and 500 data points were used. The Hindmarsh-Rose model is a model of neuronal spiking activity in the brain

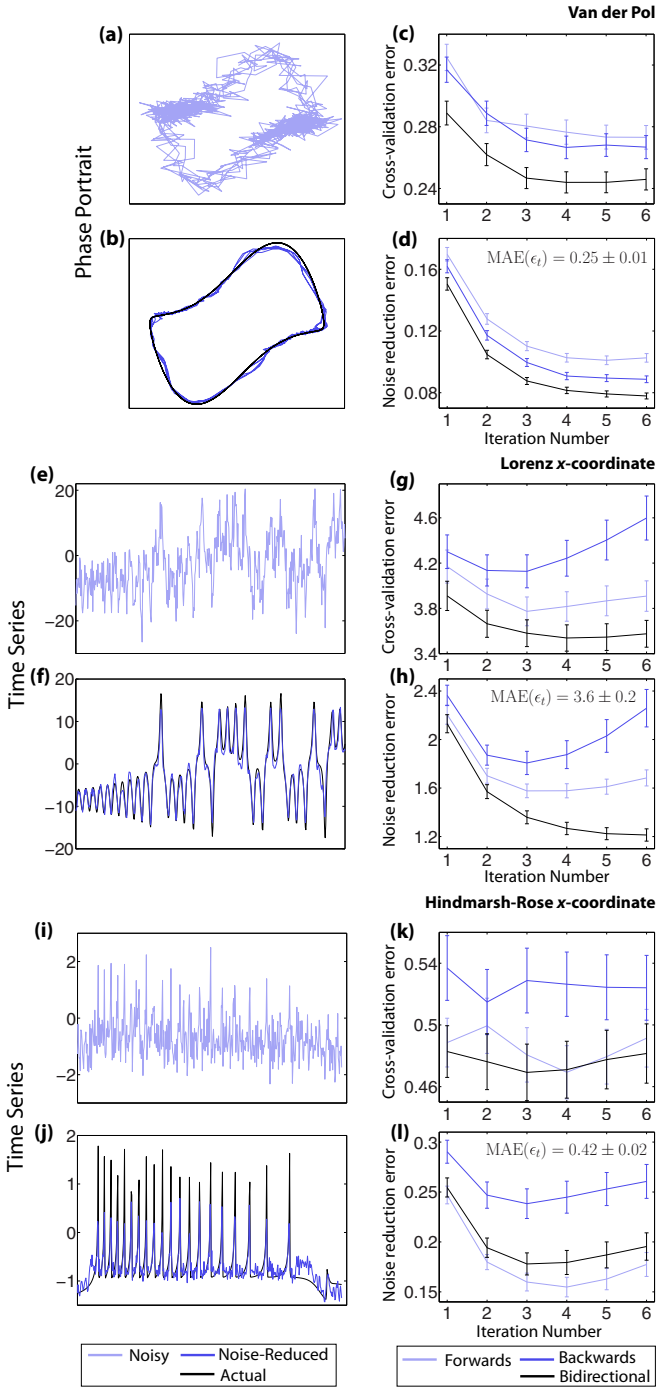

FIG. 1. Performance of the noise reduction method. Error bars are estimates of standard errors. (a) The phase portrait of the noisy Van der Pol oscillator. (b) The phase portrait of the noise-reduced Van der Pol oscillator in red, including the deterministic Van der Pol oscillator in black. The corrected time series were calculated with the bidirectional algorithm for five and four recursive iterations for the x and y coordinates, respectively. These optimized parameters (for the x coordinate) were determined with (c), the in-sample cross-validation forecasting error for the x coordinate versus the recursive iteration number r. An in-sample forecasting error for a time series cleaned $r-1$ times is associated with the potential noise reduction performance of the time series cleaned r times. Therefore, a data point at r is the in-sample forecasting error of a time series that has been cleaned $r-1$ times. (d) The actual noise reduction errors (MAE) of the corrected time series as calculated from $\mathbf{X}_t$. Also indicated on the plot is MAE(ϵ_t), the MAE of the noisy time series as calculated from $\mathbf{X}_t$. [(e)–(h)] Same as (a) to (d) but for the x coordinate of the Lorenz system. Corrected time series were calculated with the bidirectional algorithm with four recursive iterations. Time series are shown for (**E** and **F**) instead of phase portraits but it should be noted that noise reduction was conducted concurrently for all three variables of the Lorenz system. [(i)–(l)] Same as (e) to (h) but for the x coordinate of the Hindmarsh-Rose model. Noise-reduced time series were calculated with the bidirectional algorithm with three recursive iterations.

and is capable of chaotic behavior [14,15]. The deterministic time series which we used consisted of a chaotic burst of spikes [Fig. 1(j)]. From the in-sample cross-validation errors [Fig. 1(k)], there is difficulty in evaluating the performance of the forward algorithm and the bidirectional algorithm. Furthermore, the cross-validation errors do not correlate as well as the other two systems. These problems are, in this case, presumably due to the considerable noise involved since these problems alleviated with a smaller amount of observational noise [13]. The parameters determined with the in-sample forecasting errors are the bidirectional algorithm with three recursive counts [Fig. 1(k)]. These parameters are suboptimal according to the actual noise reduction errors [Fig. 1(l)]. Nonetheless, even with suboptimal parameters, the noise-reduced time series still manages to resolve the spiking peaks rather well [Fig. 1(j)]. Noise reduction errors and cross-validation errors for the other coordinates of the three systems analyzed also show that optimal or near-optimal parameters can be identified from the cross-validation errors [13].

Figure 2 shows the noise-reduced time series on the noisy Lorenz x coordinate [Fig. 1(e)] using Schreiber's method. The embedding dimension used for Schreiber's method was the same as that using the noise reduction method presented in this paper. The optimal recursive counts and neighborhood size for Schreiber's method were determined from the actual noise reduction errors. The actual noise reduction error using the noise-reduced method presented here is 1.27 ± 0.05 compared to an MAE of 2.05 ± 0.08 using Schreiber's method. Moreover, Schreiber's method appears to be unable to correct for rare and extreme observational errors (Fig. 2) in contrast with our method [Fig. 1(f)].

An obvious application of being able to reduce the noise in a time series satisfactorily is to use the noise-reduced time series for the purposes of forecasting. By reducing the uncertainty in a training data set or library used to make forecasts, out-of-sample forecasts should be improved since there is less error in reconstructed state spaces [11]. In a similar vein, out-of-sample forecasts may also be used to determine the extent of in-sample noise reduction. We made out-of-sample forecasts for the three systems analyzed in addition to two real-world data sets. The first real-world data set is the prevaccination measles incidence rate from the state of New York which is at least partly chaotic due to the chaotic incidence rate of measles in New York City [7,16,17]. The second real-world data set is the global surface temperature record [18]. The noisy time series from Fig. 1 and noisy time series for the other respective coordinates in the three systems were used

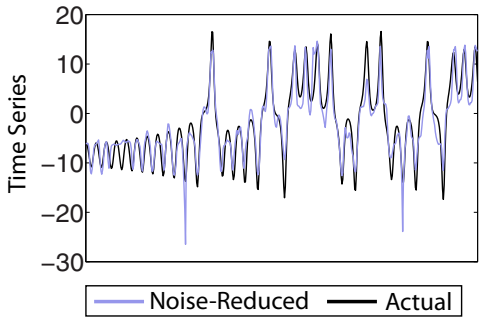

FIG. 2. Noise-reduced time series of the Lorenz x coordinate using Schreiber's method. Noisy time series from the Lorenz x coordinate [Fig. 1(e)] were noise reduced with Schreiber's method [12]. The resulting noise reduced time series (shown in red) and the actual deterministic time series (shown in black) are plotted in this figure.

as libraries. Out-of-sample forecasts one time step ahead with MVE (for the three multivariate systems) and simplex projection (for the measles data set and the temperature data set) using the noisy time series were contrasted against those using the noise-reduced time series. Noise-reduced time series were obtained by objectively optimizing the heuristics based on the in-sample cross-validation errors (Fig. 1 and Supplemental Material [13]). In all systems, forecasts with noise-reduced time series produce less error than the noisy time series except the y coordinate of the Hindmarsh-Rose model, which had a marginally higher error than the noisy time series (Fig. 3). This deviation of performance from the other systems and coordinates is notwithstanding the fact that using the dynamical trend as a library produces a better forecast [13] and the fact that the noise reduction method did reduce the error of the time series as measured against $\mathbf{X}_t$ (Fig. S5). Therefore, a likely explanation for the poorer performance of the noise-reduced time series is that the noise reduction method had smoothened over certain sections of essential dynamics in the noisy time series. Nonetheless, this marginal decrease in forecasting performance in the y coordinate should be measured against the more significant increase in forecasting performance in the x and z coordinates of the Hindmarsh-Rose model, of which the x coordinate, the membrane potential, is the primary variable of interest in the model. These improved forecasts using the noise-reduced time series further demonstrates the ability of the noise reduction method to recover dynamics from noisy time series.

While we have shown that the two heuristics introduced here can be optimized with in-sample cross-validation errors, it is conceivable that other parameters, such as the embedding dimension (which we had set at 3 for this study), number of variable and lag combinations to use (for MVE), and α, may also be optimized with the in-sample errors. The optimization of these parameters and other potential ones identified by Ye and Sugihara in MVE [9] may provide room for greater improvement to the noise reduction performance of the noise reduction method. We refrain from exploring any of these other parameters in detail so as not to depart from the intention of this work as a concise presentation on a simple and multivariate nonparametric noise reduction technique.

A. Dynamical noise, autocorrelated noise, and other limitations

Consider the case where the multivariate time series $\mathbf{X}_t$ of dimension n is affected by multiplicative dynamical noise

$$\mathbf{X}_t = f(\mathbf{X}_{t-1},\mathbf{X}_{t-2},\dots) + g(\mathbf{X}_{t-1},\mathbf{X}_{t-2},\dots)\boldsymbol{\mu}_t, \quad (15)$$

where $\boldsymbol{\mu}_t$ are independent multivariate error terms of dimension n with identical multivariate probability distributions for all t. They also satisfy $\mathrm{E}(\boldsymbol{\mu}_t) = \mathbf{0}$. The observation of $\mathbf{X}_t$ gives $\mathbf{Y}_t$ which is affected by observational error

$$\mathbf{Y}_t = \mathbf{X}_t + h(\mathbf{X}_{t-1},\mathbf{X}_{t-2},\dots)\boldsymbol{\epsilon}_t \quad (16)$$

where $\boldsymbol{\epsilon}_t$ are independent multivariate error terms of dimension n with identical multivariate probability distributions for all t. They also satisfy $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$. Then the dynamical trend $\mathbf{Z}_t$ of $\mathbf{Y}_t$ is

$$\mathbf{Z}_t = \mathrm{E}(\mathbf{X}_t|\mathbf{X}_{t-1} = \mathbf{x}_{t-1},\mathbf{X}_{t-2} = \mathbf{x}_{t-2},\dots), \quad (17)$$

$$= f(\mathbf{x}_{t-1},\mathbf{x}_{t-2},\dots). \quad (18)$$

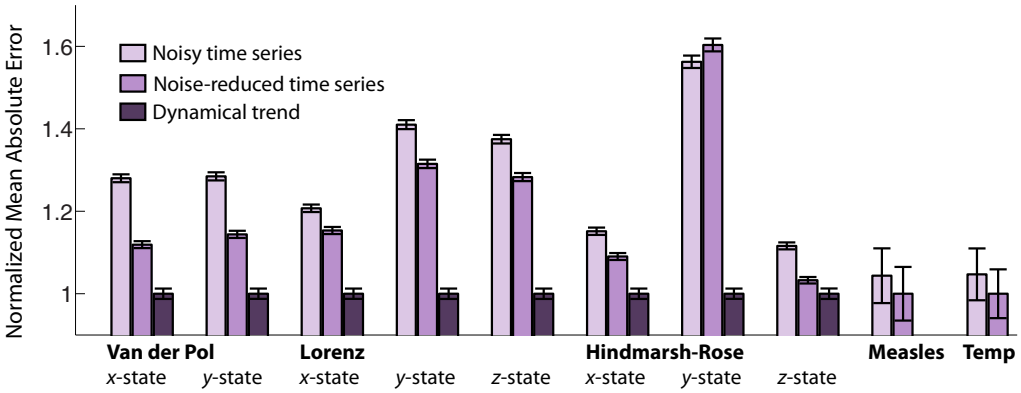

FIG. 3. Out-of-sample forecast performance of noise-reduced time series. The out-of-sample forecast performance using the original noisy time series and noise-reduced time series as libraries for forecasting. Noise-reduced time series were objectively corrected with the noise reduction method. Forecast performance is measured by the normalized MAE for the various systems. Forecast MAEs (blue and red bars) are normalized with the MAE of the dynamical trend (black bars) which is calculated against the noisy out-of-sample time series. For the measles data set and global surface temperature data set (Temp), MAEs are normalized against the MAE from the noise-reduced time series instead due to an unknown dynamical trend. Error bars are estimates of standard errors.

For a noise reduction algorithm optimized on forecasting ability, the objective is then to recover $\mathbf{Z}_t = f(\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots)$, the drift component of the stochastic process. Hence, for a satisfactory forecasting algorithm, the recovered time series $\hat{\mathbf{Z}}_t$ may be used to supplement current methods in filtering the drift and diffusion components of a stochastic process [19–22]. We apply the algorithm on synthetic time-series data from an Ornstein-Uhlenbeck process of the form $dx = -x dt + 0.5 dW$ affected by white additive Gaussian noise of standard deviation 0.2. The resulting MAE between $\hat{Z}_t$ and Z_t is 0.0682 ± 0.0016 compared to an MAE of 0.1636 ± 0.0040 between X_t and Z_t.

From Eqs. (15), (16), and (17), it follows that if an autocorrelation exists for $\boldsymbol{\epsilon}_t$ or $\boldsymbol{\mu}_t$, then one cannot expect to recover Eq. (18) from $\mathbf{Y}_t$ or, in the absence of dynamical noise, $\mathbf{X}_t$ from $\mathbf{Y}_t$. While the probability distributions of the noise terms can be quite general in that we only require the independence of the noise terms and that the noise terms have identical distributions with $\mathbf{0}$ mean, the algorithm is not expected to perform well with multimodal distributions which can lead to greater uncertainty in the identification of nearest neighbors.

In conclusion, a multivariate nonparametric noise reduction method based on state-space reconstruction was demonstrated, yielding results that compare favorably to current methods. It was also shown that cross-validation errors from in-sample forecasting correlate well with actual noise reduction errors, implying that the parameters of the method can be objectively optimized with the cross-validation errors.

ACKNOWLEDGMENTS

The author thanks Chew Lock Yue and Christopher Monterola for comments on the manuscript. The author also thanks Hao Ye and Eric Kostelich for short discussions on their work.

[1] H. Ye, R. Beamish, S. Glaser, S. Grant, C. Hsieh, L. Richards, J. Schnute, and G. Sugihara, Proc. Natl. Acad. Sci. U.S.A. **112**, E1569 (2015).
[2] D. DeAngelis and S. Yurek, Proc. Natl. Acad. Sci. U.S.A. **112**, 3856 (2015).
[3] Z. Wu, N. Huang, S. Long, and C. Peng, Proc. Natl. Acad. Sci. U.S.A. **104**, 14889 (2007).
[4] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987).
[5] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[6] F. Takens, in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, Berlin, 1981) pp. 366–381.
[7] G. Sugihara and R. May, Nature **344**, 734 (1990).
[8] T. Sauer, J. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[9] H. Ye and G. Sugihara, Science **353**, 922 (2016).
[10] M. Casdagli, S. Eubank, J. Farmer, and J. Gibson, Physica D **51**, 52 (1991).
[11] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).
[12] T. Schreiber, Phys. Rev. E **47**, 2401 (1993).
[13] See Supplemental Material http://link.aps.org/supplemental/10.1103/PhysRevE.95.032218 for more information.
[14] J. Hindmarsh and R. Rose, Proc. R. Soc. Lond. B **221**, 87 (1984).
[15] X. Wang, Physica D **62**, 263 (1993).
[16] B. Dalziel, O. Bjørnstad, W. van Panhuis, D. Burke, C. Metcalf, and B. Grenfell, PLoS Comput. Biol. **12**, e1004655 (2016).
[17] W. Schaffer, L. Olsen, G. Truty, S. Fulmer, and D. Graser, *From Chemical to Biological Organization* (Springer-Verlag, Berlin, 1981).
[18] C. Morice, J. Kennedy, N. Rayner, and P. Jones, J. Geophys. Res. **117**, D08101 (2012).
[19] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase, Phys. Rev. Lett. **97**, 090603 (2006).
[20] P. G. Lind, M. Haase, F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich, Phys. Rev. E **81**, 041125 (2010).
[21] B. Lehle, Phys. Rev. E **83**, 021113 (2011).
[22] T. Scholz, F. Raischel, V. Lopes, B. Lehle, M. Wächter, J. Peinke, and P. Lind, Phys. Lett. A **381**, 194 (2017).