Denoising Nonlinear Time Series by Adaptive Filtering and Wavelet Shrinkage: A Comparison

Jianbo Gao, *Member, IEEE*, Hussain Sultan, Jing Hu, *Member, IEEE*, and Wen-Wen Tung

***Abstract*—Time series measured in real world is often nonlinear, even chaotic. To effectively extract desired information from measured time series, it is important to preprocess data to reduce noise. In this Letter, we propose an adaptive denoising algorithm. Using chaotic Lorenz data and calculating root-mean-square-error, Lyapunov exponent, and correlation dimension, we show that our adaptive algorithm more effectively reduces noise in the chaotic Lorenz system than wavelet denoising with three different thresholding choices. We further analyze an electroencephalogram (EEG) signal in sleep apnea and show that the adaptive algorithm again more effectively reduces the Electrocardiogram (ECG) and other types of noise contaminated in EEG than wavelet approaches.**

***Index Terms*—Adaptive denoising algorithm, EEG signal, Lorenz, wavelet.**

I. Introduction

TIME series analysis for parameter estimation, system identification, pattern classification, prediction, and so on, is an important exercise in many areas of science and engineering. When data is noisy, however, such tasks may not be effectively handled. Therefore, reducing noise in experimental time series is an important issue. With the rapid accumulation of complex data in health sciences, systems biology, nanosciences, information systems, and physical sciences, this issue has become increasingly important.

Noise reduction for time series may be classified as model-based and model-free. In the former case, if the model is a state-space model, then one may use extended Kalman filtering or particle filtering. As example applications, we refer to [1], [2]. In this letter, however, we shall focus on model-free noise-reduction techniques.

Three of the major model-free noise reduction techniques are linear lowpass filtering, chaos-based smoothing [3], [4], and wavelet thresholding [5], [6]. When a time series is nonlinear, especially chaotic, linear lowpass filtering is not a viable approach for reducing noise, since chaotic signals usually have a broad-band spectrum that overlaps with the spectrum of noise significantly. In fact, linear filtering even distorts clean chaotic signals severely [7]. Chaos-based approaches [3], while conceptually elegant, often are computationally expensive and may not be very effective either, especially when noise is strong [4]. It is thus no wonder that wavelet-based noise reduction techniques have found the broadest applications. However, it has not been systematically studied for chaotic signals whether wavelet-based noise reduction techniques are truly the best. In this Letter, we propose a nonlinear adaptive denoising algorithm, and show that our approach is more effective in reducing noise in nonlinear signals than a number of wavelet thresholding based noise reduction approaches. Note that the adaptive method was originally developed to determine the 11-year cycle of the sunspot numbers [9]. However, its capability for denoising nonlinear signals has not been examined.

The remainder of the letter is organized as follows. In Section II, we describe our adaptive denoising algorithm and three wavelet thresholding techniques. In Section III, we compare the effectiveness of our algorithm and wavelet based approaches in reducing noise from the chaotic Lorenz data and a measured EEG signal. In Section IV, we make a few concluding remarks.

II. Nonlinear Adaptive and Wavelet Shrinkage Denoising

A. Nonlinear Adaptive Denoising

Our nonlinear adaptive denoising algorithm works as follows. It first partitions a time series into segments (or windows) of length $2n+1$ points, where neighboring segments overlap by $n+1$ points. Thus, the time scale introduced by the algorithm is $n+1$ sample points. For each segment, we fit a best polynomial of order K. Note that $K=0$ and 1 correspond to piece-wise constant and linear fitting, respectively. Denote the fitted polynomial for the i-th and $(i+1)$-th segments by $y^{(i)}(l_1), y^{(i+1)}(l_2), l_1, l_2 = 1, \ldots, 2n+1$, respectively. Note the length of the last segment may be smaller than $2n+1$. We define the fitting for the overlapped region as

$$y^{(c)}(l) = w_1 y^{(i)}(l+n) + w_2 y^{(i+1)}(l), \quad l = 1, 2, \ldots, n+1 \quad (1)$$

where $w_1 = (1-(l-1)/(n)), w_2 = (l-1)/(n)$ can be written as $(1-d_j/n), j=1,2$, where d_j denotes the distances between the point and the centers of $y^{(i)}$ and $y^{(i+1)}$, respectively. This means the weights decrease linearly with the distance between the point and the center of the segment. Such a weighting ensures symmetry and effectively eliminates any jumps or discontinuities around the boundaries of neighboring segments. In fact, the scheme ensures that the fitting is smooth at the non-

Manuscript received September 21, 2009; revised November 18, 2009. First published December 04, 2009; current version published December 23, 2009. This work was supported in part by U.S. NSF Grants CMMI-0825311 and 0826119. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pina Marziliano.

J. Gao is with PMB Intelligence LLC, West Lafayette, IN 47996 USA (e-mail: jbgao@pmbintelligence.com).

H. Sultan is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville FL 32611 USA (email: hussainz@ufl.edu).

J. Hu is with Affymetrix, Inc., Santa Clara, CA 95051 USA (email: jing.hu@gmail.com).

W. Tung is with the Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, IN 47907 USA (email: wwtung@purdue.edu).

Digital Object Identifier 10.1109/LSP.2009.2037773

boundary points, and has at least the right- or left-derivative at the boundary points.

Note that the method contains two free parameters, K, the order of the polynomial, and $2n+1$, the segment length (or window size)—*by properly choosing K and making $2n+1$ small enough, the fitting can be very precise, since the difference between the data and the fitting (which may be called the residual data) can be exactly zero*. When analyzing experimental data, both parameters may be determined by requiring that the variance of the residual data no longer decreases significantly when K is further increased and/or $2n+1$ is further decreased. For an example illustrating this point, we refer to our work on determining the 11-year cycle of the sunspot numbers [9].

B. Wavelet Shrinkage Denoising

Assume that the observed data $x(i)$ is obtained with a sampling time τ and contains the true signal $s(i)$ and an additive noise $n(i)$,

$$x(i) = s(i) + n(i), \quad i = 1, \ldots, N. \tag{2}$$

Wavelet shrinkage denoising consists of three steps [5]:

1) wavelet transform of the observed data;
2) thresholding the resulting wavelet coefficients;
3) inverse wavelet transform to obtain an estimation of the signal.

More precisely, one uses wavelet multiresolution analysis and obtains [8]

$$x(i) = SA_J + SD_J + SD_{J-1} + \cdots + SD_1 \tag{3}$$

where J is the chosen maximal scale, SA_J is the approximate coefficients corresponding to time scale $2^J\tau$, and $SD_j, j = J, J-1, \ldots, 1$, are detail coefficients corresponding to time scales $2^{j-1}\tau$. Now step 2 may be expressed as

$$D(U, \lambda) = \operatorname{sgn}(U)\max(0, |U| - \lambda) \tag{4}$$

where U denotes wavelet detail coefficients, $\operatorname{sgn}(U)$ is 1 when $U > 0$ and -1 when $U < 0$ and λ is a threshold. Depending on how λ is chosen, there are many alternatives to implement step 2. One popular soft threshold is

$$\lambda = \hat{\sigma}\sqrt{2\ln N} \tag{5}$$

where $\hat{\sigma} = M/0.6745$ is a rough estimate of the signal variance, M is the median absolute deviation of detail coefficients at level J, and N is the total number of points.

To take into account that detail coefficients at different levels may have different variances, Han *et al.* [6] have examined two scale-dependent thresholding schemes. One is to replace $\hat{\sigma}$ in (5) by the variance σ_j of the detail coefficients at scale j,

$$\lambda_j^{(1)} = \sigma_j\sqrt{2\ln N}. \tag{6}$$

The other scheme consists of using (5) for detail coefficients at scale 1, and

$$\lambda_j^{(2)} = \sigma_j\sqrt{2\ln N}/\beta, \tag{7}$$

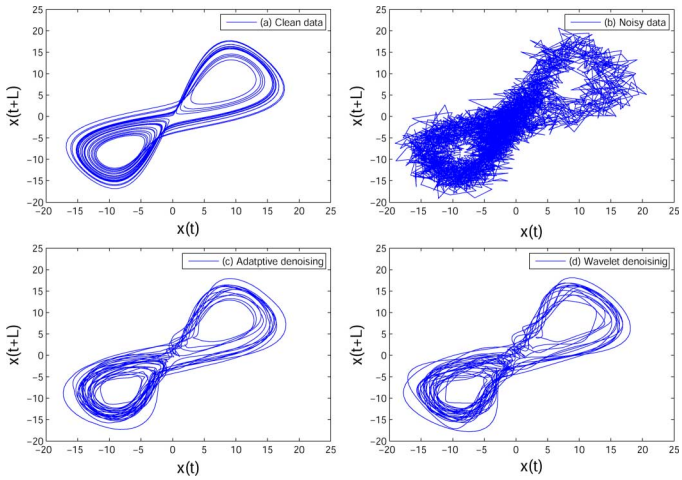

Fig. 1. Phase diagrams (with delay time $L = 12$) for (a) the clean Lorenz signal, (b) the noisy Lorenz signal, (c) the signal filtered by adaptive algorithm, and (d) the signal filtered by wavelet thresholding.

where $\beta = \sqrt{J}$ for $j = J$ and $\ln(j+1)$ for $j = 2, \ldots, J-1$. Below, we shall simply denote the three wavelet thresholding algorithms, corresponding to (5)–(7), by wavelet-1, wavelet-2, wavelet-3, respectively.

III. Performance Evaluation

In this section, we evaluate the effectiveness of our adaptive denoising algorithm by studying two nonlinear time series: the chaotic Lorenz data and EEG corrupted by ECG.

A. Lorenz System

Following Han *et al.* [6], we study the Lorenz system

$$\begin{aligned} dx/dt &= -\sigma(x - y) \\ dy/dt &= -xz + \gamma x - y \\ dz/dt &= xy - bz \end{aligned} \tag{8}$$

with $\sigma = 10, \gamma = 28, b = 8/3$. The system is solved using a 4th order Runge-Kutte method. The x-component is recorded with a sampling time of 0.01. To facilitate computation of invariant measures including Lyapunov exponent and fractal dimension, we have simulated a time series of length 6×10^4 points. Gaussian white noise of zero mean is added to the data such that the SNR of the resulting data is 13.89 dB. The phase diagrams of the clean and noisy data are shown in Fig. 1(a) and (b), respectively.

We have denoised the noisy Lorenz data by the three wavelet thresholding techniques using the same $J = 4$ as in [6]. The phase diagram obtained after using "wavelet-3" is shown in Fig. 1(d) (for ease of presentation, the corresponding curves in Figs. 2–4 are also based on wavelet-3). For ease of comparison, we also have calculated the root-mean-square-error (RMSE) defined by

$$\text{RMSE} = \sqrt{\frac{1}{2N}\sum_{i=1}^{N}(s(i) - \hat{s}(i))^2} \tag{9}$$

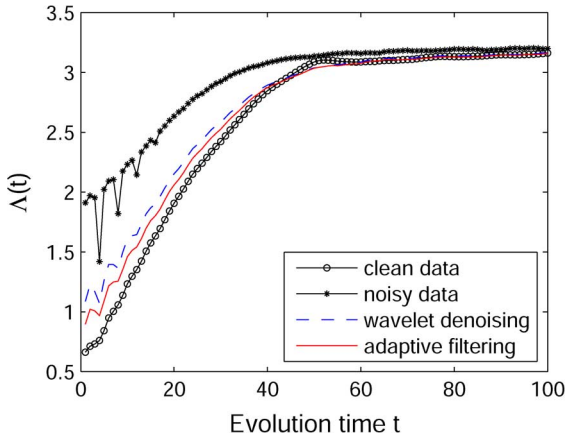

Fig. 2. $\Lambda(t)$ curves for the clean, noisy, and filtered Lorenz data. Here, $m = 5, L = 4$. In estimating the LE, a common time interval of 10–30 δt, where δt is the sampling time, is used for all the four curves. The estimated LE does not change much when different m, L are used.

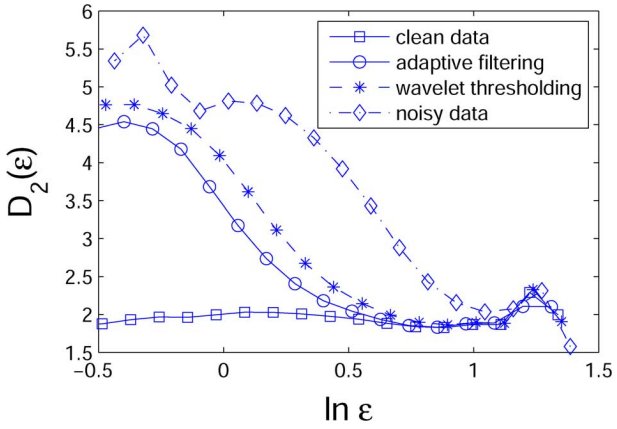

Fig. 3. Correlation dimension for the clean, noisy, and filtered Lorenz data with $m = 5, L = 4$. Again, the results remain almost the same when different m, L are used.

where $s(i)$ and $\hat{s}(i)$ are the clean and denoised data, respectively. By experimenting with more than ten mother wavelets, we have found that the smallest RMSE obtained using "wavelet-1" and "wavelet-2" is slightly smaller than that reported by Han *et al.* [6], while the RMSE with 'wavelet-3" is consistent with their results. See Table I.

We have also denoised the data by our adaptive algorithm using $K = 3, 2n + 1 = 33$. The phase diagram for the cleaned data is shown in Fig. 1(c). We observe that it is cleaner than Fig. 1(d). The resulting RMSE, shown also in Table I, is slightly less than 0.34, indicating that it is indeed better than the three wavelet thresholding methods.

To assess whether a noisy chaotic signal has been properly denoised, it is important to check if the invariant measures for the clean chaotic signal, including the Lyapunov exponent (LE), the Kolmogorov entropy, and the fractal dimension [10], can be properly estimated from the denoised signal. Noting that the Kolmogorov entropy is equal to the summation of all positive LEs, while the Lorenz system only has one positive LE, we thus only need to compute LE and the fractal dimension from the denoised signal.

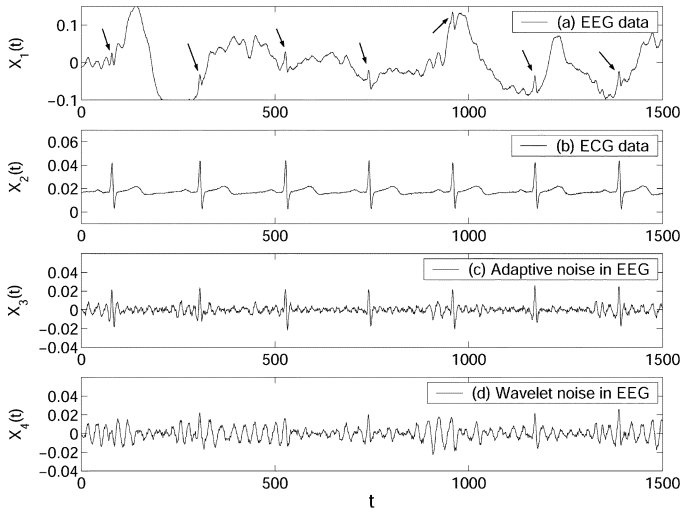

Fig. 4. (a), (b) EEG and ECG signals measured with a sampling frequency of 250 Hz and a unit of mV; (c), (d) noise extracted from EEG by using adaptive algorithm and wavelet thresholding. The arrows in (a) highlight the ECG components contained in EEG signal.

TABLE I
RMSE AND THE ESTIMATED LARGEST POSITIVE LYAPUNOV EXPONENT FOR THE NOISY LORENZ SIGNALS DENOISED BY THREE WAVELET THRESHOLDING METHODS AND OUR ADAPTIVE FILTERING ALGORITHM

Algorithm	RMSE	Lyapunov exponent
Wavelet-1 (Eq. (5))	0.4062	0.80 ± 0.03
Wavelet-2 (Eq. (6))	0.4061	0.80 ± 0.03
Wavelet-3 (Eq. (7))	0.3903	0.81 ± 0.03
Adaptive algorithm	0.3396	0.89 ± 0.02

The LE, often denoted by λ_1, is a dynamical quantity. It characterizes the exponential growth of an infinitesimal line segment ϵ_0, i.e., $\epsilon_t \sim \epsilon_0 e^{\lambda_1 t}, \epsilon_0 \to 0$. While the classical Wolf *et al.*'s algorithm [11] does a good job in estimating LE from clean chaotic signals, estimation of LE from noisy and filtered data is a much harder task. Here, we employ the algorithm of Gao and Zheng [8], [12], which has found numerous applications, including characterization of noise-induced chaos [13], [14] and development of a new multiscale complexity measure, the scale-dependent Lyapunov exponent [8], [15], [16]. The algorithm involves constructing vectors of the form $X_i = [x(i), x(i + L), \ldots, x(i + (m - 1)L)]$, where m and L are the the embedding dimension and the delay time, respectively, then monitors how $\ln \epsilon_t$ grows with time. Concretely, one (arbitrarily) chooses a set of shells defined by

$$\epsilon_k \le \|X_i - X_j\| \le \epsilon_k + \Delta\epsilon_k, \quad k = 1, 2, 3, \ldots \tag{10}$$

where X_i, X_j are reconstructed vectors, ϵ_k (the radius of the shell) and $\Delta\epsilon_k$ (the width of the shell) are arbitrarily chosen small distances ($\Delta\epsilon_k$ is not necessarily a constant). Then one monitors the evolution of all pairs of points (X_i, X_j) (where $|i - j|$ is greater than certain chosen window size) within a shell and computes

$$\ln \epsilon_t - \ln \epsilon_0 = \Lambda(t) = \left\langle \ln \left(\frac{\|X_{i+t} - X_{j+t}\|}{\|X_i - X_j\|} \right) \right\rangle. \tag{11}$$

For truly chaotic signals, $\ln \epsilon_t$ (and thus $\Lambda(t)$) is linear for a range of t, with the slope estimates the LE. The $\ln \epsilon_t$ curve for the clean Lorenz data is shown in Fig. 2, with the estimated LE being about 0.90, the same as the known value in the literature [8]. When such a procedure is applied to the noisy data, the LE is only about 0.5. In fact, the $\ln \epsilon_t$ shown in Fig. 2 is not very linear, suggesting that noise is too high for the LE to be objectively determined. Filtering using wavelet and adaptive methods recovers the linearity of the $\ln \epsilon_t$ curves. Since the curve using adaptive filtering is closer to that of the clean curve than that using wavelet thresholding, we conclude that adaptive filtering is indeed better. In fact, LE for the filtered data using adaptive filtering and wavelet thresholding is 0.89 and 0.80, respectively. See Table I.

We now discuss fractal dimension of the Lorenz signal. It is a geometrical quantity characterizing the minimal number of variables that are needed to fully describe the dynamics of a motion. It is often estimated by the correlation dimension D_2 using the Grassberger-Procaccia's algorithm.

$$C(\epsilon) \sim \frac{1}{N'^2} \sum_{i,j=1}^{N'} \theta(\epsilon - \|X_i - X_j\|) \sim \epsilon^{D_2} \qquad (12)$$

where $C(\epsilon)$ is called the correlation integral, θ is the Heaviside step function, X_i and X_j are reconstructed vectors, $N' = N-(m-1)L$ is the number of points in the reconstructed phase space, and ϵ is a prescribed small distance. Fig. 3 shows the variation of $D_2(\epsilon)$ vs. $\ln \epsilon$ for the clean, noisy, and filtered Lorenz data. We observe that for the clean data, there is a plateau with a value consistent with the correlation dimension of 2.02; for the noisy data, a plateau is hardly detectable; the plateau is recovered a little after denoising by wavelet thresholding; it is recovered more after adaptive denoising. Therefore, adaptive filtering is more effective than wavelet thresholding.

B. EEG

As an example application, we consider denoising nonstationary EEG in sleep apnea. The EEG analyzed here is downloaded from MIT-BIH polysomnography database, which is downloadable at http://www.physionet.org/physiobank/database/slpdb/. It was originally recorded for the purpose of determining sleep stages. A short segment of the data is shown in Fig. 4(a). Also shown in Fig. 4(b) is the simultaneously measured ECG data. Comparison between Fig. 4(a) and (b) makes it clear that the EEG signal contains a component of ECG. In order to reliably determine the sleep stages from the EEG data, this ECG component has to be properly removed. Note that this ECG component is typically time-varying—the ECG component can be stronger or weaker depending on one's posture of sleep, breathing, etc. Trying to remove it by subtracting from EEG data a signal of the form $\alpha \cdot \mathrm{ECG}$, where $0 < \alpha < 1$ is certain constant, does not work well [17].

We have tried to remove the ECG component from EEG using adaptive (with $K = 2, 2n + 1 = 25$) and wavelet denoising (with $J = 4$). The resulting noise signals are shown in Figs. 4(c) and (d), respectively. Evidently, the ECG component obtained by adaptive filtering resembles more the original ECG signal than that by wavelet thresholding. We have further analyzed the data filtered by our adaptive algorithm as well as by wavelet denoising using recurrence time statistics [18]–[21], and found that the sleep stages can be much more accurately determined from the data filtered by adaptive algorithm. Due to space limitations, the details of this study will be published elsewhere. From these brief discussions, we can conclude that adaptive filtering is more effective in removing the ECG component from the EEG signals in sleep apnea.

IV. Concluding Remarks

In this Letter, we have proposed an adaptive denoising algorithm. Using chaotic Lorenz data and calculating RMSE, LE, and correlation dimension, we have shown that the adaptive algorithm more effectively reduces noise in the chaotic Lorenz system than wavelet denoising with three different thresholding choices. We have further analyzed an EEG signal in sleep apnea. Again, the adaptive algorithm more effectively reduces the ECG component contaminated in EEG than wavelet approaches.

References

[1] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Tran. Biomed. Eng.*, vol. 54, pp. 2172–2185, 2007.

[2] T. Chen, J. Morris, and E. Martin, "Dynamic data rectification using particle filters," *Comput. Chem. Eng.*, vol. 32, pp. 451–462, 2008.

[3] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] S. P. Lalley and A. B. Nobel, "Denoising deterministic time series," *Dynam. PDE*, vol. 3, pp. 259–279, 2006.

[5] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[6] M. Han, Y. H. Liu, J. H. Xi, and W. Guo, "Noise smoothing for nonlinear time series using wavelet soft threshold," *IEEE Signal Process. Lett.*, vol. 14, pp. 62–65, 2007.

[7] R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M. A. Rubio, "Dimension increase in filtered chaotic signals," *Phys. Rev. Lett.*, vol. 60, pp. 979–982, 1988.

[8] J. B. Gao, Y. H. Cao, W.-W. Tung, and J. Hu, *Multiscale Analysis of Complex Time Series—Integration of Chaos and Random Fractal Theory, and Beyond*. New York: Wiley Interscience, 2007.

[9] J. Hu, J. B. Gao, and X. S. Wang, "Multifractal analysis of sunspot time series: The effects of the 11-year cycle and Fourier truncation," *J. Statist. Mech.*, 2009/02/P02066.

[10] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Phys. Rev. Lett.*, vol. 50, pp. 346–349, 1983.

[11] A. Wolf, A. , J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Phys. D*, vol. 16, p. 285, 1985.

[12] J. B. Gao and Z. M. Zheng, "Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series," *Phys. Rev. E*, vol. 49, pp. 3807–3814, 1994.

[13] J. B. Gao, S. K. Hwang, and J. M. Liu, "When can noise induce chaos?," *Phys. Rev. Lett.*, vol. 82, pp. 1132–1135, 1999.

[14] J. B. Gao, W. W. Tung, and N. Rao, "Noise induced Hopf bifurcation-like sequence to chaos in the Lorenz equations," *Phys. Rev. Lett.*, vol. 89, p. 254101, 2002.

[15] J. B. Gao, J. Hu, W.-W. Tung, and Y. H. Cao, "Distinguishing chaos from noise by scale-dependent Lyapunov exponent," *Phys. Rev. E*, vol. 74, p. 066204, 2006.

[16] J. Hu, J. B. Gao, and W. W. Tung, "Characterizing heart rate variability by scale-dependent Lyapunov exponent," *Chaos*, vol. 19, p. 028506, 2009.

[17] J. M. Lee, D. J. Kim, I.-Y. Kim, K.-S. Park, and S. I. Kim, "Detrended fuctuation analysis of EEG in sleep apnea using MIT = BIH polysomnography data," *Comput. Biol. Medi.*, vol. 32, p. 3747, 2002.

[18] J. B. Gao, "Recurrence time statistics for chaotic systems and their applications," *Phys. Rev. Lett.*, vol. 83, p. 3178, 1999.

[19] J. B. Gao and H. Q. Cai, "On the structures and quantification of recurrence plots," *Phys. Lett. A*, vol. 270, pp. 75–87, 2000.

[20] J. B. Gao, "Detecting nonstationarity and state transitions in a time series," *Phys. Rev. E*, vol. 63, pp. 066202-1–066202-8, 2001.

[21] J. B. Gao, Y. Cao, L. Gu, J. G. Harris, and J. C. Principe, "Detection of weak transitions in signal dynamics using recurrence time statistics," *Phys. Lett. A*, vol. 317, pp. 64–72, 2003.