CHIN. PHYS. LETT. Vol. 33, No. 10 (2016) 100501

Denoising Nonlinear Time Series Using Singular Spectrum Analysis and Fuzzy Entropy

Jian Jiang(江剑)[1], Hong-Bo Xie(谢洪波)[2**]

[1]*School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094*
[2]*ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Brisbane 4000, Australia*

(Received 9 June 2016)

We present a hybrid singular spectrum analysis (SSA) and fuzzy entropy method to filter noisy nonlinear time series. With this approach, SSA decomposes the noisy time series into its constituent components including both the deterministic behavior and noise, while fuzzy entropy automatically differentiates the optimal dominant components from the noise based on the complexity of each component. We demonstrate the effectiveness of the hybrid approach in reconstructing the Lorenz and Mackey–Glass attractors, as well as improving the multi-step prediction quality of these two series in noisy environments.

PACS: 05.45.Tp, 05.45.Ac, 95.75.Wx DOI: 10.1088/0256-307X/33/10/100501

Singular spectrum analysis (SSA) is an advanced technique for time series analysis and signal processing, incorporating the elements of classical time series analysis, linear algebra, multivariate statistics, multivariate geometry, dynamical systems and signal processing.[1–3] SSA aims to decompose a time series into the sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and 'structureless' noise.[2,3] Based on these components, SSA can be employed to denoise the time series and to improve the time series prediction accuracy. However, an effective approach to determine the dominant components or to identify the noise floor is still lacking. In existing SSA and its variant techniques, finding the maximal gap between the ranked eigenvalues or flat noise floor is often employed to group the components, which does not consider the individual component characteristics.[4] If the selected number of dominant components is large, the system dynamics will be overfitted with noise. On the other hand, if the number of components is low, the system dynamics is not well characterized and there will be information loss. Such simple criteria may give spurious outcomes in system identification or prediction, particularly when dealing with time series stemming from nonlinear processes.[4] Moreover, in many real systems, especially if the signal-to-noise ratio is not sufficiently large, SSA has a clear power-law form without a clear noise floor.[5] Fuzzy entropy (FuzzyEn) is a robust nonlinear measure that quantifies the degree of regularity or complexity in a time series, where low FuzzyEn often corresponds to repeatability or predictability and large FuzzyEn corresponds to irregularity and stochasticity within the time series.[6–8] In this Letter, we extend our previous studies on FuzzyEn and present a hybrid SSA-FuzzyEn denoising method, where SSA decomposes the noisy time series into its constituent components and the optimal dominant components are determined by using the FuzzyEn measure. To demonstrate the effectiveness of the proposed approach, we first apply it to reconstruct the Lorenz and Mackey–Glass attractors embedded in noise, and then to improve the multi-step prediction performance of these two benchmark time series.

Singular spectrum analysis consists of four steps, that is, embedding, singular value decomposition, grouping, and diagonal averaging.[2] Given a real-valued time series $T = \{x_1, x_2, \ldots, x_n\}$, where n is the number of samples, it is first mapped into a multi-dimensional state space as

$$X = \begin{bmatrix} X_1^{\mathrm{T}} \\ X_2^{\mathrm{T}} \\ \vdots \\ X_m^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \ldots & x_d \\ x_2 & x_3 & \ldots & x_{d+1} \\ \vdots & \vdots & \ldots & \vdots \\ x_m & x_{m+1} & \ldots & x_n \end{bmatrix}, \quad (1)$$

where d is the embedding dimension, and $m = n-d+1$ is the number of embedded delay-shift row vectors. The second step in SSA is a singular value composition (SVD) of the trajectory matrix X into a sum of rank-one bi-orthogonal elementary matrices

$$X = X_1 + X_2 + \ldots + X_d. \quad (2)$$

The elementary matrix X_i $(i = 1, 2, \ldots, d)$ is given by $X_i = \lambda_i U_i V_i^{\mathrm{T}}$, where λ_i, U_i and V_i are the ith singular value, left singular vector, and right singular vector of X, respectively. These first two steps comprise the decomposition stage of SSA and the next two steps are the reconstruction stage. In the grouping step, the task is to split the elementary rank-one matrices into several groups $L_1, L_2, \ldots, L_s$ and then sum the matrices within each group. Let $L = \{l_1, l_2, \ldots, l_p\}$ be a group of indices $l_1, l_2, \ldots, l_p$.

**Corresponding author. Email: hongbo.xie@qut.edu.au
© 2016 Chinese Physical Society and IOP Publishing Ltd

Then matrix X_L corresponding to the group L is defined as $X_L = X_{l_1} + X_{l_2} + \ldots + X_{l_p}$. The split of the set of indices $\{1, 2, \ldots, d\}$ into disjoint subsets $L_1, L_2, \ldots, L_s$ corresponds to the representation

$$X = X_{L_1} + X_{L_2} + \ldots + X_{L_s}. \tag{3}$$

If the original series contains the signal and the noise, we then consider two groups of indices $L_1 = \{1, 2, \ldots, k\}$ and $L_2 = \{k+1, \ldots, d\}$, and associate the groups L_1 and L_2 with the signal and noise components, respectively.

The fourth and last step of SSA converts each regrouped matrix in Eq. (3) back to the initial time series of length n. Diagonal averaging is employed to generate reconstructed elements of each resultant matrix by averaging over positive-slope skew diagonal groupings. The new element has the same position as the corresponding elements in the original series. Consider matrix Y with elements y_{gh}, $1 \le g \le m$ and $1 \le h \le d$. Let $d^* = \min(m, d)$, $m^* = \max(m, d)$ and $n = m + d - 1$. Let $y^*_{ij} = y_{ij}$ if $m < d$ and $y^*_{ij} = y_{ji}$ otherwise. Diagonal averaging transfers matrix Y to a series $\{y_1, y_2, \ldots, y_n\}$ according to

$$y_t = \begin{cases} \frac{1}{t}\sum\limits_{q=1}^{t} y^*_{q,t-q+1}, & \text{for } 1 \le t < d^*, \\ \frac{1}{d^*}\sum\limits_{q=1}^{d^*} y^*_{q,t-q+1}, & \text{for } d^* \le t \le m^*, \\ \frac{1}{n-t+1}\sum\limits_{q=t-m^*+1}^{n-m^*+1} y^*_{q,t-q+1}, & \text{for } m^* < t \le n. \end{cases} \tag{4}$$

With Eq. (4), the original time series T can be recovered from the sum of a number of elementary series T_i. A pending issue in SSA is the estimation of optimal dominant components or equivalently, the noise floor in step three. However, a quantitative criterion to precisely differentiate dominant components from the noise is still lacking. In this study, we propose to split the set of indices $\{1, 2, \ldots, d\}$ based on the fuzzy entropy of X_i. We define the FuzzyEn of elementary matrix X_i, which is equivalent to the FuzzyEn of its recovered time series T_i ($\{1, 2, \ldots, d\}$) in step 4. The eth vector sequence of X_i in Eq. (1) is rewritten as

$$X^d_{i,e} = \{x^e_i, x^{e+1}_i, \ldots, x^{e+d-1}_i\},\ 1 \le e \le n - d + 1. \tag{5}$$

The distance D^d_{ef} between the vector $X^d_{i,e}$ and its neighbor $X^d_{i,f}$ is defined as

$$D^d_{ef} = D[X^d_{i,e}, X^d_{i,f}] = \max_{k\in[1,d]} |x^{e+k-1}_i - x^{f+k-1}_i|. \tag{6}$$

The degree of similarity between $X^d_{i,e}$ and $X^d_{i,f}$ is determined by a fuzzy membership function

$$SD^d_{ef} = \exp(-(D^d_{ef})^2/r). \tag{7}$$

Given the tolerance r, we average all the degrees of similarity between vector $X^d_{i,e}$ and its neighbor vectors $X^d_{i,f}$. The average degree of similarity is denoted as

$$B^d_r(e) = \frac{1}{n-d} \cdot \sum_{f=1, f\ne e}^{n-d+1} SD^d_{ef}, \tag{8}$$

and the probability that two vector sequences match is given by

$$B^d_r = \frac{1}{n-d+1} \cdot \sum_{e=1}^{n-d+1} B^d_r(e). \tag{9}$$

Similarly, we can construct $(d+1)$-dimensional embedding vectors $X^{d+1}_{i,e} = \{x^e_i, x^{e+1}_i, \ldots, x^{e+d}_i\}$ for each X_i, and define the degree of similarity $B^{d+1}_r(e)$ as well as the probability B^{d+1}_r in exactly the same way as $X^d_{i,e}$. Finally the estimated FuzzyEn of rank-one matrix X_i or the reconstructed component T_i (obtained in the limit of $n \to \infty$) is defined by

$$\text{FuzzyEn}(d, r, n) = -\ln(B^{d+1}_r/B^d_r). \tag{10}$$

Suggestions for selecting values of the parameters d and r in FuzzyEn estimation can be found in our previous studies.[6,7] It is well known that larger FuzzyEn values correspond to broader banded spectra, whilst lower FuzzyEn values correspond to more peaked spectra.[7] It is therefore natural to group those X_i with large FuzzyEn values into noise, and those with low FuzzyEn values into trend and oscillatory components. To demonstrate the usefulness of the proposed hybrid method, we first apply it to reconstruct two benchmark dynamic systems embedded in noise.

The Lorenz system is defined by the following three-state nonlinear ordinary differential equations[9]

$$\frac{dx}{dt} = \sigma(y - x), \tag{11}$$

$$\frac{dy}{dt} = -xz + rx - y, \tag{12}$$

$$\frac{dz}{dt} = xy - bz, \tag{13}$$

with parameter values $\sigma = 10$, $b = 8/3$, and $r = 28$.

The Mackey–Glass system is given by the delay differential equation[9]

$$\frac{dx}{dt} = -bx(t) + \frac{ax(t-\tau_0)}{1 + x^c(t-\tau_0)}, \tag{14}$$

where $a = 0.2$, $b = 0.1$, $c = 10$ and $\tau_0 = 17$.

We use the fourth-order Runge–Kutta numerical integration algorithm for the Lorenz and Glass–Mackey systems, respectively, to simulate the original noise-free time series. The integration step size for the Lorenz and Mackey–Glass equations are set to 0.015 and 0.25, respectively. The first 5000 points are discarded to eliminate initial transients.

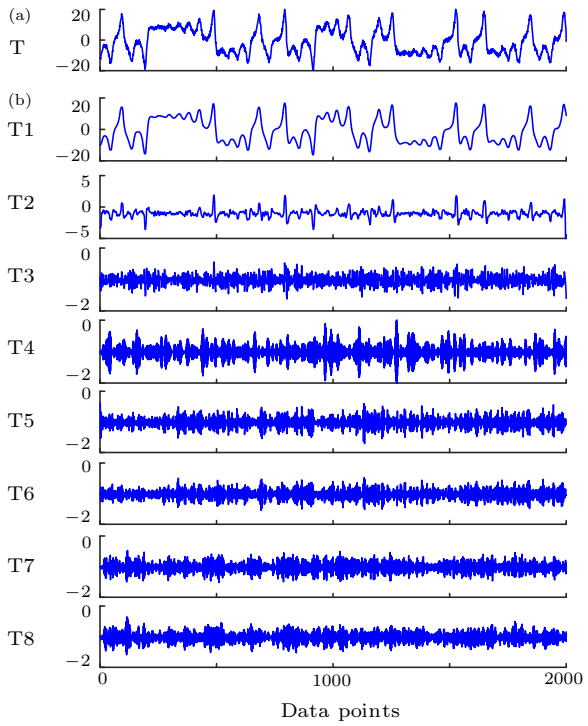

Fig. 1. The x component of noisy Lorenz series at 10% NL (a) and its eight reconstructed components using SSA (b).

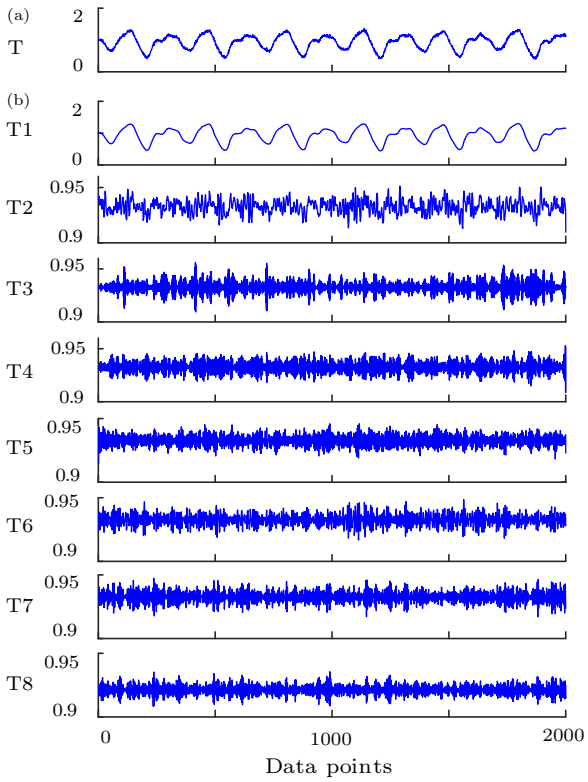

Fig. 2. The x component of noisy Mackey–Glass series at 10% NL (a) and eight reconstructed components using SSA (b).

Noise is then superimposed onto the time series through the addition of independent and identically distributed (i.i.d.) Gaussian white noise at various noise levels (NLs). The embedding dimension for both systems is set to eight.[2] Figures 1 and 2 illustrate the SSA decomposition for the noisy Lorenz and Mackey–Glass series at 10% NL. The top panels of both figures show the original noisy data, while the lower panels illustrate eight reconstructed components using SSA. The first component, showing the main trend, is similar to the original series. A simple Fourier transform can be applied to each component to indicate different frequency-band distributions. Figure 3 shows the FuzzyEn of clean Lorenz and Mackey–Glass series, as well as these series superimposed with noise from 2% to 10% NL in steps of 2%. The tolerance is set to 0.2 and the vector sequence dimension 2 used in the FuzzyEn calculations are suggested in Refs. [6,7]. Previous studies have indicated that the FuzzyEn value for both Gaussian and uniform distribution noise is more than 1.25 under such parameter settings.[6,7] The FuzzyEn of both panels from left to right correspond in order from leading to subordinate singular values. For each sub-panel the FuzzyEn first increases, corresponding to the dominant components of the dynamic systems. In particular, the first component has almost the same FuzzyEn value as the clean series. However, most of the subsequent FuzzyEns are over one, corresponding to the noise. On the other hand, the maximal FuzzyEn of each component obtained from the clean Lorenz and Mackey–Glass series is around 0.25, much lower than the corresponding noisy components. This is consistent with our previous studies showing that deterministic time series without contamination have low complexity with low FuzzyEn values.[6,7] We therefore select those components with FuzzyEn less than 1 while in the ascending phase of each FuzzyEn distribution curve as the dominant components. Based on this criterion, the first two Lorenz components and the first three Mackey–Glass components are selected as the dominant components.

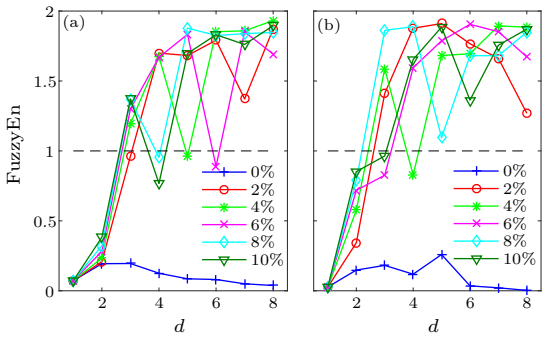

Fig. 3. Fuzzy entropy distribution of Lorenz (a) and Mackey–Glass (b) series with NLs from 0% to 10% in steps of 2%.

We can use these dominant components obtained from SSA FuzzyEn to reconstruct their attractors. The left panels of Figs. 4 and 5 show the attractor for clean data (red), along with the attractor of noisy time series (blue) at 10% NL. The right panels of both

figures show the attractor reconstructed from the sum of optimal dominant components (blue), along with the attractor of clean data (red). The envelope of red and blue curves in both right panels is well matched, which demonstrates that the dynamics of Lorenz and Mackey–Glass systems can be accurately recovered from noisy series using SSA FuzzyEn.

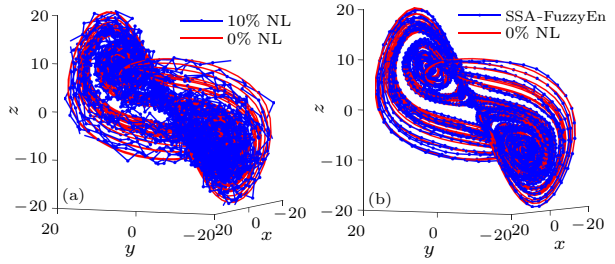

Fig. 4. The 3D reconstructed attractors for clean Lorenz series (red) and noisy Lorenz series with 10% NL (blue) (a), and 3D reconstructed attractors for clean Lorenz series and the sum of first two components obtained from SSA FuzzyEn (blue) with 10% Gaussian noise (b).

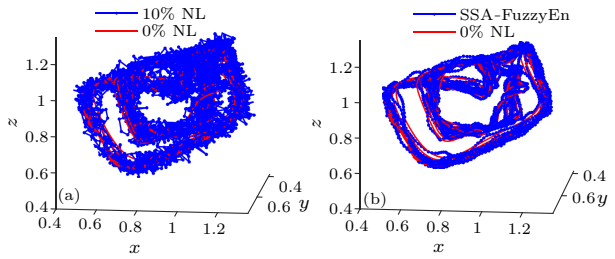

Fig. 5. The 3D reconstructed attractors for clean Mackey–Glass series (red) and Mackey–Glass series with 10% NL (blue) (a), and 3D reconstructed attractors for clean Mackey–Glass series and the sum of first three components obtained from SSA FuzzyEn (blue) with 10% Gaussian noise (b).

We can also employ the SSA-FuzzyEn approach to make a multi-step prediction of noisy time series. Nonlinear prediction (NLP) using local approximation is a popular method in chaotic time series prediction.[10] The dynamical system generating the time series can be interpreted in the form of a nonlinear function F_τ, i.e., $X_{i+\tau} = F_\tau X_i$, where X_i and $X_{i+\tau}$ are the current state space and the future state space, respectively. To predict $X_{i+\tau}$ from the current state X_i, a function $\hat{F}_\tau$ that maps a d-dimensional state space to a single scalar variable of the system is employed as

$$x_{i+\tau} = \hat{F}_\tau X_i. \tag{15}$$

In practice, the least-square criterion is often used to determine the local $\hat{F}$ for the first-order approximation, and the predicted value of $x_{i+\tau}$ is obtained from Eq. (15). To make predictions of more than one step ahead, earlier predictions are used in constructing a new local approximation function $\hat{F}$[10]. In the development of the NLP approach, it is assumed that the prediction accuracy is determined by the approximation quality, which, in turn, depends on the volume of the available data. Therefore, the method is effective for a time series consisting of an infinite sequence of noise-free observations.[11] However, in real life, experimental observations are inevitably corrupted by various sources of noise. Due to the noise, the probability that false neighbors appear and true neighbors are rejected increases. The prediction accuracy is naturally limited by the erroneous choice of phase space neighbors, even when increasing the number of observations. Moreover, many previous studies have indicated that the iterative local regression model outperformed the direct prediction.[12] In the former scheme, the predicted values are typically iterated to train the new model for obtaining further future values in multi-step ahead prediction. Hence, any prediction error at a given time will propagate and accumulate in later predictions. It is therefore expected that an effective denoising method can improve the nonlinear time series prediction quality. Here we use the sum of the dominant components to construct the embedded vector as the input of the NLP model, 1500 data points are used to train the model, and 500 data points are used to test the prediction performance. Meanwhile, the support vector machine (SVM) has been reported to be superior to the existing NLP model in prediction of nonlinear time series without contamination.[13,14] The prediction quality of the proposed approach is compared with existing NLP and SVM models in terms of the mean squared error (MSE),

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2, \tag{16}$$

and the correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2]\cdot[\sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}})^2]}}, \tag{17}$$

where $\hat{x}_i$ is the predicted value of x_i, and $\bar{x}$ and $\bar{\hat{x}}$ are the mean of original and predicted series, respectively.

Figures 6 and 7 illustrate the error and correlation coefficient for the prediction step from 1 to 20 for both noisy series at 10% NL. Several conclusions can be drawn from these panels. First, MSEs increase and correlation coefficients decrease as expected with the increase of prediction lead time across all three approaches and both data sets. Then, NLP and SVM have similar prediction performance within short prediction lead time. With the increased lead time, SVM gradually outperforms NLP in terms of both performance indices, which is consistent with the previous studies.[13,14] Finally, compared with NLP and SVM approaches, significantly lower MSE and higher correlation coefficient are obtained from the SSA-FuzzyEn approach, particularly for long prediction lead time. The results confirm the negative effect of noise on NLP

as well as the effectiveness of SSA FuzzyEn to improve the prediction quality.

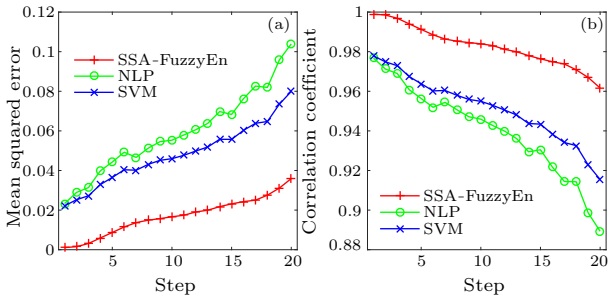

Fig. 6. Multi-step prediction mean squared errors (a) and correlation coefficients (b) for the noisy Lorenz series with 10% NL using SSA FuzzyEn (red) and NLP (green).

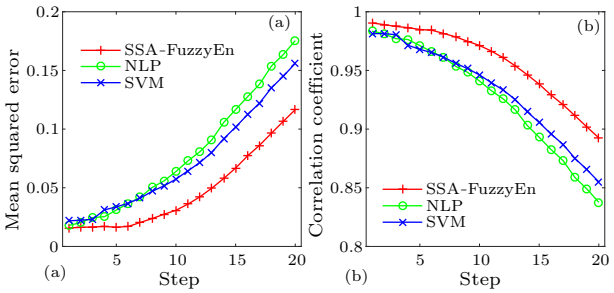

Fig. 7. Multi-step prediction mean squared errors (a) and correlation coefficients (b) for the noisy Mackey–Glass series with 10% NL using SSA FuzzyEn (red), NLP (green) and SVM (blue).

In summary, we have presented a hybrid singular spectrum analysis and fuzzy entropy approach for noisy time series analysis. In our proposed approach, SSA decomposes the noisy time series into its constituent components, while FuzzEn adaptively differentiates the optimal dominant components from the noise. To demonstrate the effectiveness of the method, we applied it to reconstruct the attractors of Lorenz and Mackey–Glass system from contaminated data, and to improve the prediction quality of both noisy series. The results confirm the effectiveness of SSA FuzzyEn in denoising and its superiority to the original local approximation and SVM methods in multi-step prediction. In this study, only Gaussian additive noise is considered. As is well known, both outlier and intrinsic noise can degrade the prediction performance. In addition, the influences of a moving window on prediction error and finite sample size on entropy estimation are the other two fundamental issues in nonlinear time series analyses.[15,16] The effect of these factors on SSA FuzzyEn should be further investigated. The robust singular value decomposition and the ensemble prediction model are potentially useful to tackle these issues in SSA FuzzyEn as well as other nonlinear time series approaches.

References

[1] Vautard R, Yiou P and Ghil M 1992 *Physica* D **58** 95
[2] Xie H B, Guo T, Sivakumar B, Liew A W C and Dokos S 2014 *Proc. R. Soc.* A **470** 210409
[3] Hossein H 2007 *J. Data Sci.* **5** 239
[4] Pilgram B and Schappacher W 1998 *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **8** 571
[5] Alharbi N and Hassani H 2016 *J. Franklin Inst.* **353** 1
[6] Xie H B, Chen W, He W X and Liu H 2011 *Appl. Soft Comput.* **11** 2871
[7] Xie H B, Guo J Y and Zheng Y P 2010 *Ann. Biomed. Eng.* **38** 1483
[8] Kong D R and Xie H B 2011 *Chin. Phys. Lett.* **28** 090502
[9] Xie H B and Dokos S 2013 *Chaos* **23** 023131
[10] Farmer J D and Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
[11] Lim T P and Puthusserypady S 2007 *Phys. Lett.* A **365** 309
[12] Casdagli M 1989 *Physica* D **35** 335
[13] Li H C and Zhang J S 2005 *Chin. Phys. Lett.* **22** 2776
[14] Cui W Z, Zhu C C, Bao W X and Liu J H 2004 *Acta Phys. Sin.* **53** 3303 (in Chinese)
[15] Yang Z L, Tian Y, Zhang G T and Lin K Y 2005 *Power Syst. Technol.* **29** 31 (in Chinese)
[16] Yang Z L, Feng Y, Xiong D F, Chen X and Zhang J 2015 *Trans. Tianjing Univ.* **21** 299