International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:1, No:2, 2007

Detecting the Nonlinearity in Time Series from Continuous Dynamic Systems based on Delay Vector Variance Method

Shumin Hou, Yourong Li, Sanxing Zhao

***Abstract*—**Much time series data is generally from continuous dynamic system. Firstly, this paper studies the detection of the nonlinearity of time series from continuous dynamics systems by applying the Phase-randomized surrogate algorithm. Then, the Delay Vector Variance (DVV) method is introduced into nonlinearity test. The results show that under the different sampling conditions, the opposite detection of nonlinearity is obtained via using traditional test statistics methods, which include the third-order autocovariance and the asymmetry due to time reversal. Whereas the DVV method can perform well on determining nonlinear of Lorenz signal. It indicates that the proposed method can describe the continuous dynamics signal effectively.

***Keywords*—**Nonlinearity; Time series; continuous dynamics system; DVV method

I. Introduction

DETECTING the nature of time series has received considerable attention in recent years. Much time series are generated by complicated systems for which it is impossible to solve or even set up the equations governing the dynamics, it is quite commonly assumed without further proof that such time series display significant nonlinearity. Consequently the analysis is carried out by advanced numerical algorithms borrowed from nonlinear dynamics, whereas the repertory of more traditional linear tools is largely neglected. While for many systems the assumption of nonlinearity may be correct in principle, it has for specific cases to be shown explicitly that employing nonlinear tools and models is justified and useful. If an experimental time series of limited length and finite precision is given, it may be impossible to distinguish between nonlinear dynamics and linear dynamics involving stochastic components. In addition, these time series data are commonly from continuous dynamic system. Thus, the sampling space can influence to determine the nature of time series. It is for this reason that tests for nonlinearity are important tools in time series analysis. Currently the technique of surrogate data testing [1] is one of the most popular approaches to nonlinearity testing.

Being motivated by statistical hypothesis testing, this technique presents an indirect way of detecting nonlinearity; as a consequence of this a failure to detect nonlinearity does not disprove nonlinearity, but may also result from an inappropriate choice of the test statistic. Recently more weak points of surrogate data testing were observed, first and foremost the effect that a rejection of the null hypothesis does not necessarily prove nonlinearity, but still may be a consequence of other properties of the time series such as non-stationary [2]. There are also considerable problems with artifacts occurring in the process of generating the surrogate data sets [3,4].

This paper studies the nonlinearity in time series, which are under different sampling conditions. Then, a novel test statistic for detecting the linear or nonlinear nature (Delay Vector Variance or 'DVV' method) is introduced. The simulation results demonstrate that DVV is more excellent to analyze the nonlinearity in time series from continuous dynamic system than traditional test statistic.

Authors are with Wuhan University of Science & Technology, 430081 Wuhan, P.R.China. e-mail: zhenduan@hotmail.com

II. Surrogate Data Method and Nonlinear Hypothesis Testing

A. Description of Phase-randomized surrogate data testing

There exist many methods for generating surrogates [1]. Three classes of surrogates are now often in use: random-shuffle surrogates, random-phase surrogates and Gaussian-scaled random-phase surrogates.

A comprehensive account of phase-randomized surrogate data testing for nonlinearity can be found in [5, 6]. The method was originally motivated by the search for deterministic chaos in experimental data. Chaos cannot be proven directly in a time series. However, chaos requires nonlinear dynamics. So, as stated above, a more modest but also more realistic goal is to search for nonlinearity in the data. Without nonlinearity, chaos can be excluded. If there is nonlinearity in the data, chaos has not been proven, but at least one necessary condition has been established. The phase-randomized, surrogate data technique tries to establish the presence of nonlinearity by excluding a reasonable alternative, which is called the "null hypothesis". For this purpose a (usually nonlinear) discriminating statistic Q (for instance, correlation dimension; Lyapunov exponents, entropy, nonlinear predictability) is needed. The null hypothesis is rejected by demonstrating that the observed value of Q for the data is very unlikely when the null hypothesis is true. This requires that the distribution of Q under the null hypothesis (mean value and standard deviation) be known. These values are obtained from the surrogate data. These surrogate data share with the original data only those properties (for instance, linear structure) which are defined by the null hypothesis. If many realizations of the experimental data are available, the distributions of Q for the experimental and surrogate data can be compared directly using conventional statistical tests such as a t-test or Mann-Whitney U-test. Alternatively, it is possible to generate many surrogate data from a single experimental time series and estimates a z-score (which is also referred to as "number of sigmas") [7].

$$Z = \frac{\left|Q_d - \langle Q_s \rangle\right|}{\sigma_s} \tag{1}$$

In this formula Q_d is the value of Q for the experimental data, $\langle Q_s \rangle$ is the mean of Q for the surrogate data set and σ_s is the standard deviation of Q for the surrogate data. Thus the z-score expresses how many standard deviations ("sigmas") Q of the experimental data deviates from the average Q for the surrogates. Assuming that Qs has a normal distribution, the null hypothesis can be rejected for two sided testing at a significance level of $p < 0.05$ when $z > 1.96$. The proposed method is quite general in that it allows one to test different null hypotheses, and use any discriminating statistic deemed appropriate. In this study we will use three different statistics: third-order autocovariance and the asymmetry due to time reversal and DVV (described in Section III). The null hypothesis tested is that the data can be explained by a stationary linear Gaussian model. This implies that all the relevant information is contained in the power spectrum and in the [circular] autocorrelation function. The end result of this procedure is a surrogate that has exactly the same power spectrum as the original data, but with random phases.

We have opted for random-phase surrogate approach, since it has been observed to yield more superior results compared to other methods. This type of surrogate time series retains the signal distribution and amplitude spectrum[2] of the original time series, and takes into account a possibly nonlinear and static observation function due to the measurement process. The stages of specific procedure are:

1. Input the experimental time series $s(t_i)$, $i = 1,\cdots,N$ into a complex array:

$$z(n) = s(n) + iy(n) \tag{2}$$

where $s(n)$ is the original data $s(t_i)$, $y(n) = 0$, $n = 1,\cdots,N$

2. Construct the discrete Fourier Transform:

$$Z(m) = S(m) + iY(m) = \frac{1}{N}\sum_{n=1}^{N} z_n e^{-2\pi i(m-1)(n-1)/N} \tag{3}$$

3. Construct a set of random phase:

$$\phi_m \in [0,\pi] , \; m = 2,3,\cdots,\frac{N}{2} \tag{4}$$

4. Apply the randomized phases to the Fourier transformed data:

$$Z(m)' = \begin{cases} Z(m) & m = 1, m = \frac{N}{2}+1 \\ |Z(m)|e^{i\phi_m} & m = 2,3,\cdots,\frac{N}{2} \\ |Z(N-m+2)|e^{-i\phi_{N-m+2}} & m = \frac{N}{2}+2,\frac{N}{2}+3,\cdots,N \end{cases} \tag{5}$$

5. Construct the inverse Fourier Transform of $Z(m)'$ and get the surrogates:

$$z(n)' = s(n)' + iy(n)' \quad \frac{1}{N}\sum_{n=1}^{N} Z_m' e^{2\pi i(m-1)(n-1)/N} \tag{6}$$

For the simulations in this article, 99 surrogates have been generated for each of the time series under study.

B. *Traditional nonlinearity metrics*

To undertake the performance comparison between the proposed DVV method and other nonlinearity analysis methods, we have implemented two traditional measures of nonlinearity, which have also been used in [8], namely

the third-order autocovariance (C3):

$$\beta_{c3} = \frac{1}{N}\sum_{n=2\tau+1}^{N}(s_n \cdot s_{n-\tau} \cdot s_{n-2\tau}) \tag{7}$$

and a measure of the deviation due to time reversibility (REV):

$$\beta_{rev} = \frac{1}{N}\sum_{n=\tau+1}^{N}(s_n - s_{n-\tau})^3 \tag{8}$$

Where τ is a time lag for which simplicity and convenient comparison is set to unity in all simulations. In combination with the surrogate data strategy, both measures yield two-tailed tests for nonlinearity.

III. The Delay Vector Variance Method for Detecting Nonlinearity in Time Series

A. *The delay vector variance method*

We introduce a novel analysis of a time series which examines the predictability of a time series by virtue of the observation of the variability of the targets.

For a given embedding dimension *m*, the mean target variance σ^{*2} is computed over all sets Ω_k . A set Ω_k is generated by grouping those delay vectors (DVs) that are within a certain distance to *x(k)*, which is varied in a manner standardized with respect to the distribution of pairwise distances between DVs. The threshold scales automatically with the embedding dimension *m*, as well as with the dynamical range of the time series at hand, and thus, the complete range of pairwise distances is examined. The proposed DVV method can be summarized as follows for a given embedding dimension *m* [9]:

1. The 99 surrogates are generated by the use of random-phase surrogate method. Then, the reconstructed vector $X = \{x(k) \mid k = 1,\cdots N\}$ can be obtained.

2. The mean u_d and standard deviation σ_d are computed over all pairwise distances between DVs, $\|x(i)-x(j)\|\ i \neq j$

3. The sets Ω_k ($k=1,\cdots N$) are generated such that $\Omega_k = \{x(i)\|x(k)-x(i)\| \le \tau_d\}$, i.e., sets which consists of all DVs that lie closer to *x(k)* than a certain distance τ_d ,taken from the interval $[\min\{0, u_d - n_d\sigma_d\}; u_d + n_d\sigma_d]$,e.g., uniformly spaced, where n_d is a parameter controlling the span over which to perform the DVV analysis.

4. For every set Ω_k , the variance of the corresponding targets σ_k^2 is computed. The average over all sets Ω_k , normalized by the variance of the time series σ_x^2 yields the measure of unpredictability σ^{*2} :

$$\sigma^{*2} = \frac{(1/N)\sum_{k=1}^{N}\sigma_k^2}{\sigma_x^2} \tag{9}$$

5. The linear or nonlinear nature of the time series is examined by performing DVV analyses on both the original and a number of surrogate time series. Directly, these plots can be conveniently combined in a scatter diagram, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, the 'DVV scatter diagram' coincides with the bisector line, and the original time series is likely to be linear, vice versa.

B. Simulations

To verify the proposed method, we employ surrogate data method to research the nature of Lorenz chaos time series under different sampling conditions. Lorenz signal is produced by eq.(9):

$$\begin{cases} \dfrac{dx}{dt} = \delta(y-x) \\ \dfrac{dy}{dt} = rx - y - xz \\ \dfrac{dz}{dt} = -bz + xy \end{cases} \tag{10}$$

where, δ , b and r are parameters:

$$\delta = 10, b = \frac{8}{3}, r = 28$$

1) Over-sampling conditions

According Shannon sampling theorem, the over-sampling frequency is beyond to double main frequency $f_{s\max}$ of time series signal. When sampling frequency $f_s = 100Hz$, the Lorenz time series time-domain wave and frequency wave is shown in Fig.1(a) and Fig.1(b), alternatively.

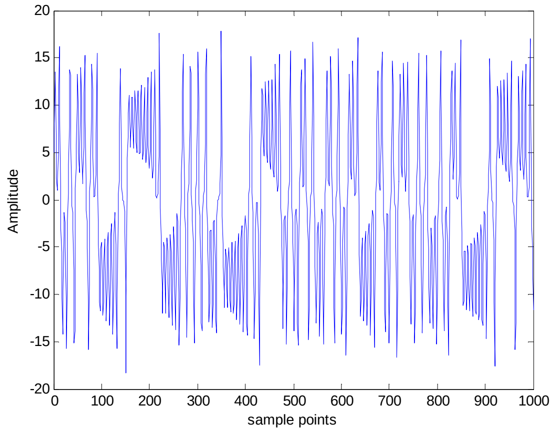

Fig. 1(a) Lorenz time series time-domain wave

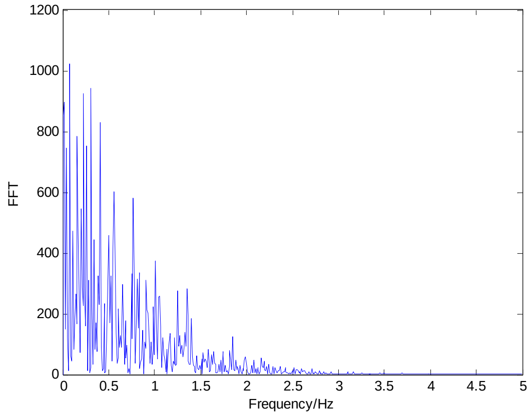

Fig. 1(b) Lorenz time series frequency wave
Fig.1 Under oversampling condition (1000 samples)

2) Under-sampling conditions

According Shannon sampling theorem, the Under-sampling frequency is much less than double main frequency $f_{s\max}$ of time series signal. When sampling frequency $f_s = 1Hz$, the corresponding Lorenz time series time-domain wave and frequency wave is shown in Fig.2 (a) and Fig.2 (b).

3) Fit-sampling conditions

When sampling frequency don't include above two conditions, it is the fit-sampling conditions. We get the Lorenz time series by applying $f_s = 10Hz$. Its waves time-domain wave and frequency wave is shown in Fig.3 (a) and Fig.3 (b), respectively.

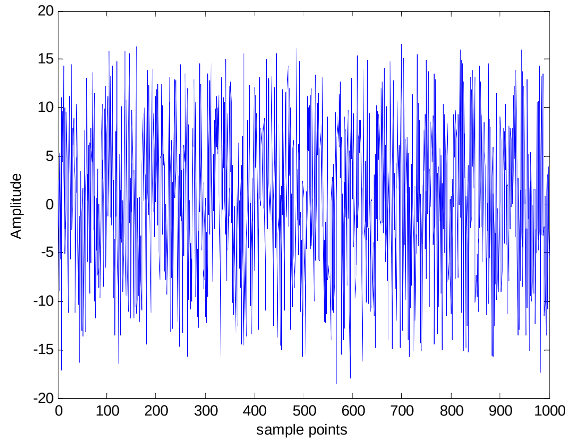

Fig. 2(a) Lorenz time series time-domain wave

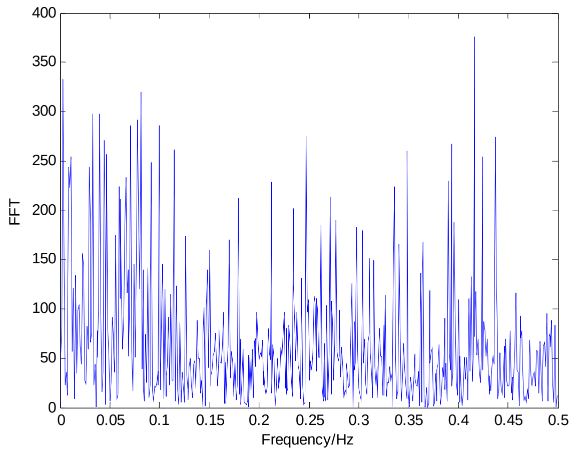

Fig. 2(b) Lorenz time series frequency wave
Fig. 2 Under under-sampling condition (1000 samples)

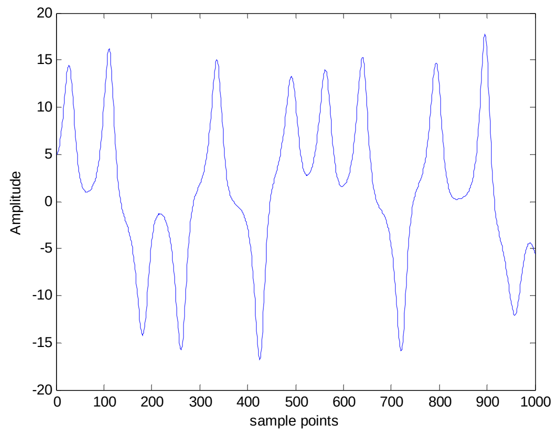

Fig. 3 (a) Lorenz time series time-domain wave

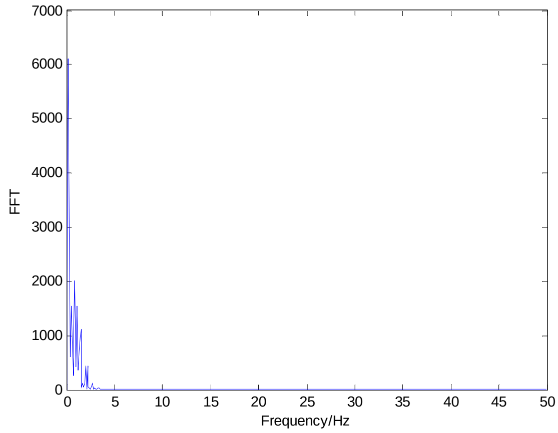

Fig. 3(b) Lorenz time series frequency wave
Fig. 3 Under fit-sampling condition (1000 samples)

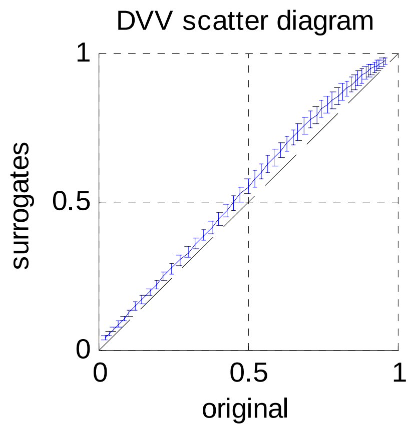

Fig. 4 DVV scatter diagrams of Lorenz time series under oversampling condition

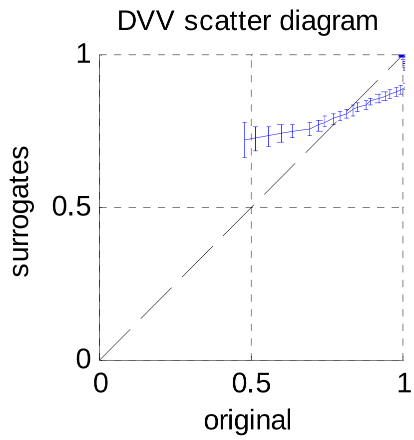

Fig. 5 DVV scatter diagrams of Lorenz time series under under-sampling condition

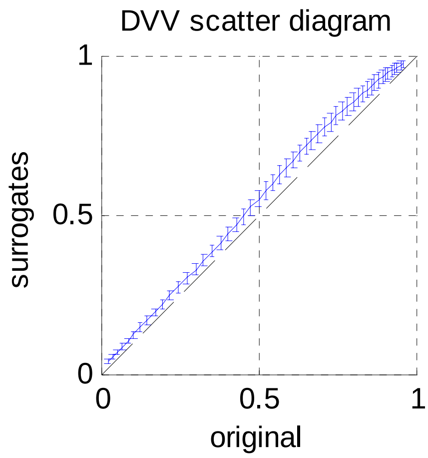

Fig. 6 DVV scatter diagrams of Lorenz time series under fit-sampling condition

TABLE I
RESULTS OF THE RANK TESTS

	C3	REV	DVV
1Hz	25	48	100
	linearity	linearity	nonlinearity
10Hz	1	39	99
	nonlinearity	linearity	nonlinearity
100Hz	3	53	97
	linearity	linearity	nonlinearity

Significant rejections of the null hypothesis at the level of 0.05

The test statistics are computed for the original and 99 surrogate time series, after the non-parametric rank-based (two-tailed) testing is applied. The quantitative results from the nonlinearity tests are summarized in Table I.

According to the simulated test results in tab.1, the false detection results of continuous chaos dynamics systems are obtained via employing the traditional test statistics. Especially under no-fitted sampling condition, linearity correlation in Lorenz signal is beyond larger than nonlinearity component. Consequently, the nature of Lorenz signal is principal linearity component (particularly in limited length chaos time series). Moreover, the traditional test statistics mainly describe the consistency of linearity characteristic in time series. Accordingly, it is difficult to judge the nonlinearity in time series under no fitted sampling conditions. This phenomenon also exists in other continuous chaos dynamics systems (such as Rossler system and Duffer system). On the other hand, the DVV method can successfully describe the nonlinearity of Lorenz signal under the different sampling conditions. The corresponding DVV scatter diagrams are shown in Fig. 4~Fig.6 (the error bars denote the upper and lower quartiles, of which only one in three is shown). Qualitatively, it is clear that the deviation from the bisector line, and thus, from the null hypothesis of linearity, is strongest for the Lorenz signal of under-sampling conditions.

IV. CONCLUSION

In this paper, we employ random-phase surrogate method in generating Lorenz surrogate sets. The proposed DVV method has been shown to perform well on Lorenz time series under different sampling spacing. A comparison has been made between the proposed method and several traditional nonlinearity analysis techniques, namely REV and C3. Consequently, DVV method is readily employed to detect the nonlinearity in continuous chaos dynamics systems. Moreover, this paper suggests that under no-fitted sampling conditions, it is best to apply nonlinear values as test statistics for detecting nonlinearity. It can avoid spurious detection result.

Overall, the proposed DVV method seems to correctly detect the presence of nonlinearity in a wider variety of signals than C3 and REV.

ACKNOWLEDGEMENTS

Our special thanks are due to Temujin Gautama for his assistance in supplying the Matlab program of DVV. Additional thanks go to the anonymous reviewer for their critical comments. Moreover, this work was supported by grant NO.(2005A04) from Hubei province key lab of machine transmission and manufacturing engineering foundation, and also supported by the research foundation from Wuhan university of Science & Technology.

REFERENCES

[1] T. Schreiber and A. Schmitz, Surrogate time series, Phys. D 142(2000) 346-382.

[2] J. Timmer, The power of surrogate data testing with respect to non-stationarity, Phys. Rev. E 58 (1998) 5153–5156.

[3] A.M. Albano, P.E. Rapp, Phase-randomised surrogates can produce spurious identification of non-random structure, Phys. Lett. A 192(1994) 27-33.

[4] D. Kugiumtzis, Test your surrogate data before you test for nonlinearity, Phys. Rev. E 60 (1999) 2808–2816.

[5] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, Physica D 58 (1992) 77-94.

[6] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank and J.D. Farmer, Using surrogate data to detect nonlinearity in time series, in: Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. Vol. XII, eds. Casdagli and S. Eubank (Addision-Wesley, Reading, MA, 1992) pp. 163-188.

[7] C.J. Stam, J.RM. Pijn , W.S. Pritchard ,Reliable detection of nonlinearity in experimental time series with strong periodic components ,Physica D 112 (1998) 361-380

[8] J. M. Le Caillec, R. Garello, Comparison of statistical indices using third order statistics for nonlinearity detection, Signal Processing 84 (2004) 499 –525

[9] Temujin Gautama,Danilo P. Mandic ,Marc M. Van Hulle, The delay vector variance method for detecting determinism and nonlinearity in time series, Physica D 190 (2004) 167–176

[10] Kanty H, Schreiber T. Nonlinear Time Series Analysis. Cambridge: Cambridge University Press, 1997

[11] T. Schreiber and A. Schmitz, Improved surrogate data for nonlinearity tests, Phys.Rev. Lett. 77(1996) 635-638.

[12] Wu shanyuan, Wang zhaojun Nonparametric statistic method, Publishing House of Higher education, Beijing (1996) 144 150.

[13] D.J. Christini, F.M. Bennett, K.R. Lutchen, H.M. Ahmed, J.M. Hausdorff, and N. Oriol, “Applications of linear and nonlinear time series modeling to heart rate dynamics analysis,” IEEE Trans. Biomed. Eng., vol. 42, pp. 411–415, 1995.

[14] S. Guzzetti, M.G. Signorini, C. Cogliati, S. Mezzetti, A. Porta, S. Cerutti, and A. Malliani, “Non-linear dynamics and chaotic indices in heart rate

variability of normal subjects and heart-transplanted patients," Cardiovasc. Res., vol. 31, pp. 441–446, 1996.
[15] K.K. Ho, G.B. Moody, C.K. Peng, J.E. Mietus, M.G. Larson, D. Levy, and A.L. Goldberger, "Predicting survival in heart
[16] failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indicaes of heart rate dynamics," Circulation, vol. 96, pp. 842–848, 1997.
[17] C.-S. Poon and C.K. Merrill, "Decrease of cardiac choas in congestive heart failure," Nature, vol. 389, pp. 492–495, 1997.
[18] D.P. Mandic and J.A. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms Architecture and Stability, Wiley: Chichester, 2001.
[19] M.C. Casdagli and A.S. Weigend, "Exploring the continuum between deterministic and stochastic modeling," in Time Series Prediction: Forecasting the Future and Understanding the Past, A.S. Weigend and N.A. Gershenfield, Eds. Addison-Wesley, 1994.
[20] H.O.A. Wold, A Study in the Analysis of Stationary Time Series, Almqvist and Wiksell: Uppsala, 1938.
[21] A.S. Weigend and N.A. Gershenfeld, Eds., Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994.

Hou Shumin received a Master's degree in Mechanical Design from the W.U.S.T, Wuhan, P.R.China, and a Ph.D. in Mechanical Design from the W.U.S.T. He is currently a teacher in Wuhan Universtiy of Sci.&Tech.. His research interests include signal processing, fault diagnosis, time series analysis and their application to data mining.