Correlation between detrended fluctuation analysis and the Lempel–Ziv complexity in nonlinear time series analysis*

Tang You-Fu(唐友福)[a)b)], Liu Shu-Lin(刘树林)[a)†], Jiang Rui-Hong(姜锐红)[a)], and Liu Ying-Hui(刘颖慧)[a)]

[a)] *School of Mechatronics Engineering and Automation, Shanghai University, Shanghai* 200072, *China*
[b)] *School of Mechanical Science and Engineering, Northeast Petroleum University, Daqing* 163318, *China*

(Received 18 July 2012; revised manuscript received 28 August 2012)

We study the correlation between detrended fluctuation analysis (DFA) and the Lempel–Ziv complexity (LZC) in nonlinear time series analysis in this paper. Typical dynamic systems including a logistic map and a Duffing model are investigated. Moreover, the influence of Gaussian random noise on both the DFA and LZC are analyzed. The results show a high correlation between the DFA and LZC, which can quantify the non-stationarity and the nonlinearity of the time series, respectively. With the enhancement of the random component, the exponent α and the normalized complexity index C show increasing trends. In addition, C is found to be more sensitive to the fluctuation in the nonlinear time series than α. Finally, the correlation between the DFA and LZC is applied to the extraction of vibration signals for a reciprocating compressor gas valve, and an effective fault diagnosis result is obtained.

Keywords: nonlinear time series, detrended fluctuation analysis, Lempel–Ziv complexity, correlation coefficient

PACS: 05.45.–a **DOI:** 10.1088/1674-1056/22/3/030504

1. Introduction

Time series analysis plays an important role in both theoretical analysis and practical applications in numerous fields.[1] Conventional time series analysis techniques including model estimation and Fourier transformation have been proven to be effective for the analysis of linear and stationary processes.[2–4] However, the time series, either from a physical measurement or from numerical modeling, most likely represents a nonlinear and non-stationary process.[5] Therefore, conventional techniques may bring some misleading results for the analysis of a nonlinear time series. Since Lorenz firstly described chaos and tried to model weather patterns using a system of coupled differential equations known as the Lorenz equations, the chaos theory has greatly promoted the development of nonlinear time series analysis.[6,7] Meanwhile, time-frequency analysis techniques such as short time Fourier transformation (SFTF), Wigner–Ville transformation (WVT), wavelet transformation (WT), and Hilbert–Huang transformation (HHT) have been developed and widely applied to non-stationary time series analysis, which has attracted more and more attention in the past few decades.[8]

Nonlinearity and non-stationarity are two important properties of a time series. For many applications, the analysis of nonlinearity and non-stationarity in an observed time series is a fundamental problem to be solved before the data can be further processed. In the case of nonlinearity detection, some typical characteristics including Lyapunov exponents, correlation dimension, entropy, and complexity have been proposed and well applied.[9] Among them, the Lempel–Ziv complexity (LZC) has been extensively used to calculate the information content for time series analysis involving nonlinear dynamics, such as coding, data compression, and the generation of test signals.[10] The LZC is much better than the traditional spectral analysis and the time-frequency analysis, and can be used to detect the long-range correlations embedded in a non-stationary time series. Moreover, the LZC also avoids the spurious detection of apparent long-range correlations that are artifacts of non-stationarity.[11] In recent years, the detrended fluctuation analysis (DFA) method proposed by Peng *et al.* has become a widely used technique for the detection of long-range correlations in a non-stationary time series[12,13] and has been successfully applied to diverse fields, such as DNA sequences, heart rate dynamics, long-time weather records, economic time series, and neuron spiking.[14]

At present, work on the nonlinearity and the non-stationarity of time series analysis are relatively independent.[15,16] In fact, nonlinearity and non-stationarity vary with the dynamic behavior of a system, so there may be some correlations between them. To verify this correlation, the scaling exponent α of the DFA and the normalized complexity index C of the LZC are investigated in this paper, which respectively quantify the non-stationarity and the nonlinearity of the time series. First, some nonlinear time series generated by typical dynamic systems including a logistic map, Duffing model, Lorenz model, etc. are taken as the simulation data,

*Project supported by the National Natural Science Foundation of China (Grant No. 51175316) and the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20103108110006).

†Corresponding author. E-mail: lsl346@shu.edu.cn

© 2013 Chinese Physical Society and IOP Publishing Ltd

and the correlation between the DFA and LZC of the nonlinear time series is investigated. The simulation results indicate that both the DFA and LZC can clearly reflect the period, the quasi-period, and the chaos states of the dynamic systems. Moreover, the correlation coefficient of the DFA and LZC for each nonlinear time series represents a significant or high correlation. Then, the influence of the Gaussian random noise on the correlation of the DFA and LZC is also analyzed, which shows that the values of scaling exponent α and normalized complexity index C similarly decrease with the increase of the signal to noise ratio (SNR). Moreover, C is more sensitive to the fluctuation of the nonlinear time series than α. Finally, the vibration signals of a reciprocating compressor are studied, the result further identifies the high correlation between the DFA and LZC in detecting the nonlinearity and the non-stationarity of the time series.

2. DFA and LZC

2.1. Algorithm of DFA

The DFA can be used to analyze the long-range correlation of a non-stationary time series. To facilitate our discussion, a brief description of the DFA algorithm is given below. The method consists of four main steps.[17]

Step 1 For a time series $x(i)$ $(i = 1,2,\ldots,N)$, we determine a new series $y(k)$ $(k = 1,2,\ldots,N)$

$$y(k) = \sum_{i=1}^{k} [x(i) - \langle x \rangle], \tag{1}$$

where $\langle x \rangle$ means the average of $x(i)$. It is noted that the subtraction is not compulsory, since it can be eliminated by the detrending procedure in step 3.

Step 2 For each given positive integer s, the series $y(k)$ can be divided into N_s non-overlapping local segments with equal length s which is called the scale index. The points inside a given interval are fitted by a polynomial line. Assuming that the order of the fitting polynomial is p, we can get the p order trend of each data interval, noted as $y_s(k)$. Then, a detrended variation function $\Delta y_s(k)$ $(k = 1,2,\ldots,N)$ is obtained by subtracting the integrated data that is given by the fit. Explicitly, we define

$$\Delta y_s(k) = y(k) - y_s(k). \tag{2}$$

Step 3 Finally, we calculate the root-mean-square fluctuation $F(s)$ inside the interval by

$$F(s) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [\Delta y_s(k)]^2}. \tag{3}$$

Step 4 Now change the size of the scale index s, set a certain step length increment Δs, and then repeat the operations above. The curve of the fluctuation function $F(s)$ varying with the scale index s will be obtained. Usually, the average $F(s)$ behaves as

$$F(s) \propto s^{\alpha}, \tag{4}$$

where the parameter α is called the scaling exponent, and represents the long-range correlation properties of the time series. Generally, if $\alpha = 0.5$, there is no correlation, and the time series is an uncorrelated one (random noise); if $\alpha < 0.5$, the time series is anti-correlated; if $\alpha > 0.5$, there are positive correlations in the time series.[18]

At present, there is no unified approach to determining the size of the non-overlapping segment s when the scaling exponent α is calculated. Usually, it is defined based on experience, so it is easily influenced by randomness and subjectivity. Moreover, for the same time series, the scaling exponent α will vary greatly for different parameters of s, and even leads to the wrong information.

In order to ensure the rationality and the correctness of the scaling exponent α, the Gaussian random time series is taken as an example, and the law of different parameters is calculated. The length of samples N is respectively selected as 1000, 2000, ..., 50000. The scaling parameter s is respectively selected as [100, 200], [100, 400], [100, 600], [100, 800], and [100, 1000]. The curves of scaling exponent α for the Gaussian random time series under different parameters N and s are shown in Fig. 1.

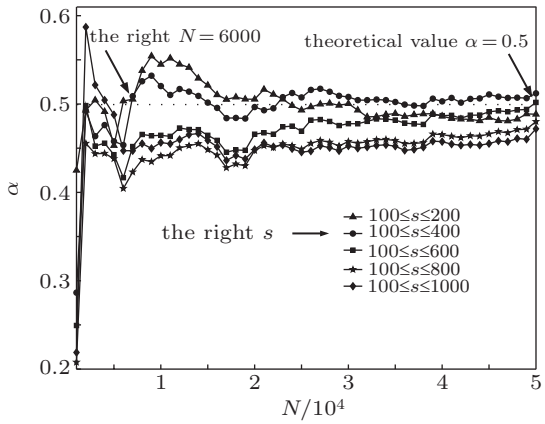


Fig. 1. Scaling exponent α for Gaussian series with different parameters.

Figure 1 clearly shows that the scaling exponent α of the Gaussian random time series finally tends to the theoretical value 0.5 with the increase of N. The fluctuation performance of the scaling exponent α is basically the same for different scale parameters s, but it is nearest to the theoretical value 0.5 when $100 \leq s \leq 400$. Moreover, in order to reduce the running time and keep the scaling exponent α near to the the theoretical value 0.5, the minimal length of the samples is selected as 6000. In theory, the best scaling exponent α will be obtained if a large enough N is set. However, it will also consume more

running time. So a suitable scaling parameter s should be selected. In this study, we select $N = 6000$ and $s = [100, 400]$ as the suitable combinatorial parameters for calculating the scaling exponent α of the nonlinear time series.

2.2. Algorithm of LZC

The LZC analysis is based on the coarse graining of the measurements. Before the LZC measure is calculated, the time series must be transformed into a finite symbol sequence. Generally, an arbitrary time series $\{x(i)|i = 1, 2, \ldots, n\}$ is converted into a binary series. By comparing with the threshold, $x(i)$ is converted into a 0–1 series $\{S(i)|i = 1, 2, \ldots, n\}$ as follows:[19]

$$p(i) = \begin{cases} 0, & x(i) < x_{\text{ave}}, \\ 1, & x(i) \geq x_{\text{ave}}, \end{cases} \tag{5}$$

where x_{ave} is the mean of the time series of $x(i)$.

Then, the LZC measure can be estimated by using the algorithm shown in Fig. 2.

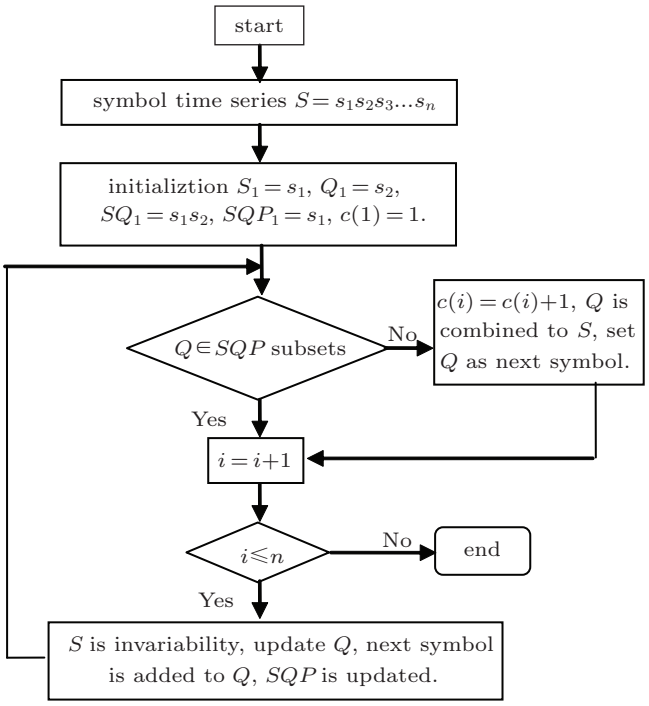


Fig. 2. Principle diagram of LZC.

The above algorithm is repeated until Q is the last character. The measure of complexity is $c(n)$. In order to obtain a complexity measure which is independent of the series length, $c(n)$ must be normalized. If the length of the sequence is n and the number of different symbols in the symbol set is 2, it has been proven that the upper bound of $c(n)$ is given by

$$\lim_{n\to\infty} c(n) = b(n) = \frac{n}{\log_2(n)}, \tag{6}$$

and $c(n)$ can be normalized via $b(n)$

$$0 \leq C = \frac{c(n)}{b(n)} \leq 1, \tag{7}$$

where the normalized complexity index C is called the Lempel–Ziv complexity. Obviously, equation (7) can be established only if sample length n is large enough. Reference [20] has given the experience valve $n \geq 3600$. So the sample length $N = 6000$ is selected in the following study.

3. Correlation between DFA and LZC of nonlinear time series

3.1. Correlation between DFA and LZC of the logistic map

A real time series is usually nonlinear and non-stationary, based on which it is difficult to establish a nonlinear dynamical model. The scaling exponent α and the complexity index C for the real time series can not be effectively obtained either, and consequently the correlation of them can not be accurately described. In order to identify the feasibility and the effectiveness of the proposed approach, the typical nonlinear dynamical systems are taken as examples to investigate the correlation between the DFA and LZC. Consider the following logistic map:[21]

$$x_{n+1} = \lambda x_n (1 - x_n), \tag{8}$$

where $x_n \in [0, 1]$, and λ is the control parameter, $\lambda \in [2.8, 4]$, the step size is 0.01. The sample length of the nonlinear time series is $N = 6000$. Figure 3 shows the logistic map with a different parameter λ.

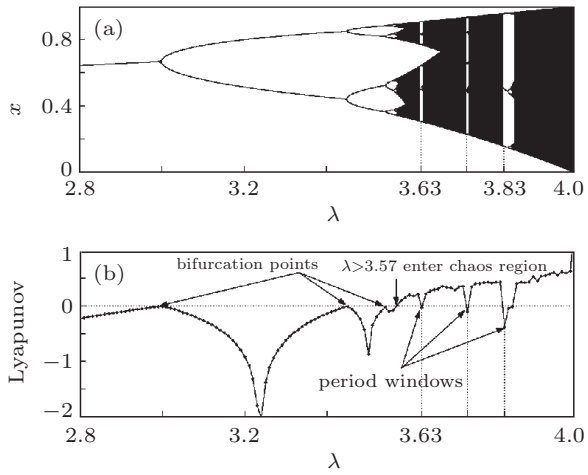


Fig. 3. State of the logistic map versus parameter λ: (a) bifurcation diagram, (b) Lyapunov exponent.

In Fig. 3, the dynamic behavior of a logistic system with a different control parameter λ is shown. When $2.8 \leq \lambda < 3.57$, the Lyapunov exponent is less than zero, the logistic system is periodic or quasi-periodic in this region. When $3.57 < \lambda < 4$, the Lyapunov exponent is positive, which shows that the system enters the chaos region. In the chaos region, the regions with Lyapunov exponents less than zero for $\lambda = 3.63$, $\lambda = 3.74$, and $\lambda = 3.83$ are called period windows.

According to the algorithm mentioned above, the scaling exponent α and the complexity index C of the logistic system with a different parameter λ can be calculated. The result is shown in Fig. 4.

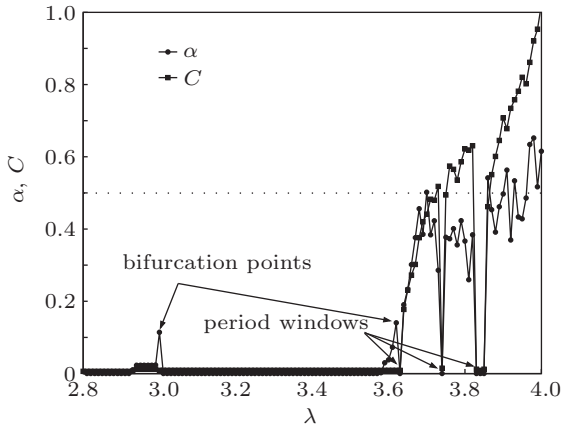


Fig. 4. Scaling exponent α and complexity index C of the logistic system versus parameter λ.

Figure 4 indicates that the trends of α and C are very synchronous. When $2.8 \leq \lambda < 3.57$, the logistic system enters the periodic or quasi-periodic region, and most values are close to zero. However, there are some non-zero α in the bifurcation points. When $3.57 < \lambda < 4$, the logistic system enters the chaos region, and the scaling exponent α tends to 1, the complexity index C tends to be around 0.5. This shows that the time series of the dynamic system is completely random. In addition, when $\lambda = 3.63$, $\lambda = 3.74$, and $\lambda = 3.83$, the logistic system enters the period windows, in the meantime, α and C are also close to zero. By comparing Fig. 4 with Fig. 3, the feature laws of the scaling exponent α and the complexity index C are rather similar to that of the Lyapunov exponent, so they can also be used as identification for the dynamic behavior of the logistic map.

In order to quantify the similarity between the curves of α and C, a correlation coefficient is proposed. The equation is given by

$$\rho = \frac{\sum_{i=1}^{n} \alpha_i C_i}{\sqrt{\sum_{i=1}^{n} \alpha_i^2 \sum_{i=1}^{n} C_i^2}}, \tag{9}$$

where the correlation coefficient $\rho \in [-1, 1]$. When $\rho = 0$, it indicates that there is no correlation between α and C. The greater the absolute value of ρ, the higher the correlation between α and C.

According to Eq. (9), the correlation coefficient between the scaling exponent α and the complexity index C for the logistic map is calculated, and the result is 0.9561. It shows that the nonlinearity and the non-stationarity of the logistic time series are highly correlated.

3.2. Correlation between DFA and LZC of the Duffing model

Consider the following Duffing model:[22]

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + r\frac{\mathrm{d}x}{\mathrm{d}t} - x + x^3 = F\cos(2\pi f_0 t), \tag{10}$$

where $F = 14$, $f_0 = 0.164$, $r \in [1.2, 1.6]$, and r is the control parameter; the step size is 0.01. The sample length of the nonlinear time series is $N = 6000$. Figure 5(a) shows the bifurcation diagram of the Duffing model with a different parameter r. Figure 5(b) shows the Lyapunov exponent of the Duffing model with a different parameter r.

Figure 5 shows the dynamic behavior of the Duffing model with a different control parameter r. When $1.2 \leq r < 1.215$ and $1.225 \leq r < 1.255$, the Lyapunov exponent is positive, and the Duffing model is in the chaos state in this region. Moreover, between the two regions, the Lyapunov exponent is less than zero, which corresponds to the period window. When $1.255 \leq r < 1.4$, the Lyapunov exponent changes from positive to negative, the system leaves the chaos region and enters the quasi-period state. When $1.4 \leq r < 1.6$, the system is in the period state.

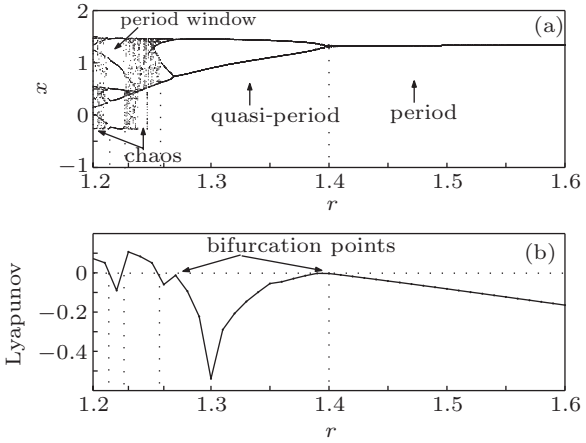


Fig. 5. State of the Duffing system versus parameter r: (a) bifurcation diagram, (b) Lyapunov exponent.

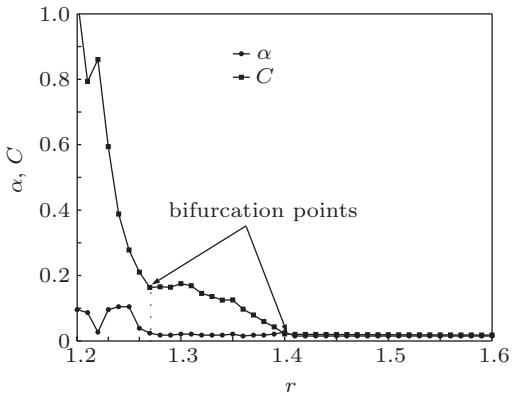


Fig. 6. Scaling exponent α and complexity index C of the Duffing system versus parameter r.

According to the algorithms of DFA and LZC, the scaling exponent α and the complexity index C of the Duffing model with a different parameter r can be calculated. Compared with Fig. 5(b), figure 6 shows that both α and C decrease with the increase of parameter r, and they also match the dynamic behavior of the Duffing model. Moreover, the trends of α and C are very synchronous. The correlation coefficient of them is 0.7371, which represents that the nonlinearity and the non-stationarity of the Duffing time series are significantly correlated.

3.3. Correlation between DFA and LZC of other typical nonlinear systems

Similarly, the approach proposed above can be extended to other typical nonlinear dynamical systems, such as the Lorenz, Rossler, Henon, and Chen systems. The correlation coefficients between the DFA and LZC of the typical systems are shown in Table 1.

Table 1. Correlation coefficients between the DFA and LZC of a typical nonlinear time series.

System	Lorenz	Rossler	Henon	Chen	Logistic	Duffing
Parameter	$b\in[23,28]$	$c\in[2.5,6.5]$	$a\in[0,1.4]$	$b\in[2,5.5]$	$\lambda\in[2.8,4]$	$r\in[1.2,1.6]$
ρ	0.5751	0.6119	0.8924	0.5856	0.9561	0.7371
Correlation	significant	significant	high	significant	high	significant

4. Influence of random noise on the correlation between DFA and LZC

Random noise is a special nonlinear time series. The non-stationary and the nonlinearity arising from a dynamic system may be correlated with the noisy fluctuation. So the effect of random noise on the scaling exponent α and the complexity index C should be taken into consideration. A time series $x(t)$ which contains the Gaussian noise is taken as an example, and it is given by

$$s(t) = 0.1\sin(20\pi t) + 0.3\sin(80\pi t), \tag{11}$$

$$x(t) = s(t) + e(t), \tag{12}$$

where $s(t)$ is the original signal without noise, and $e(t)$ is the Gaussian random noise.

As is well known, the signal to noise ratio (SNR) measures the energy ratio between the original signal and the noise signal. The smaller the value, the greater the noise. The SNR is defined as

$$\mathrm{SNR} = 10\log_{10}\frac{\sum s(n)^2}{\sum e(n)^2}, \tag{13}$$

where the SNR is in units of dB.

The SNR of time series $x(t)$ is set to be 10 dB, 11 dB, ..., 70 dB, respectively. Figure 7 shows the time waveform of $x(t)$ with a different SNR. According to the approach proposed above, the scaling exponent α and the complexity index C of $x(t)$ are calculated. The result is shown in Fig. 8.

Figure 8 illustrates that both the scaling exponent α and the complexity index C of $x(t)$ gradually decrease with the increase of the SNR. However, C decays faster than α. This shows that the complexity index C is more sensitive to the noise. When the SNR is close to 10 dB, α and C tend to 0.5 and 1, respectively. This represents that when the ratio of the random components is increasing, the time series $x(t)$ becomes more nonlinear and non-stationary. In addition, according to Eq. (9), the correlation coefficient between α and C of $x(t)$ with a different SNR can be calculated. The result is 0.8096, which represents a high correlation.

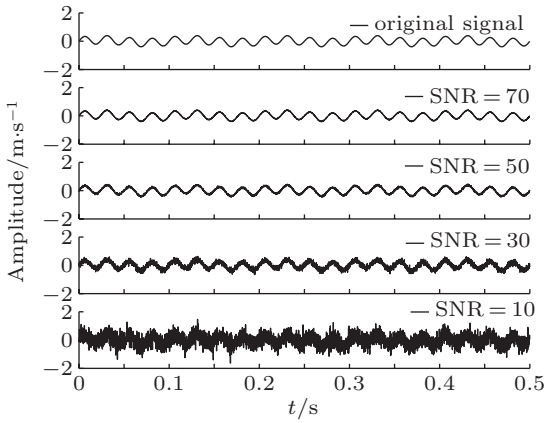


Fig. 7. Time waveform of $x(t)$ with a different SNR.

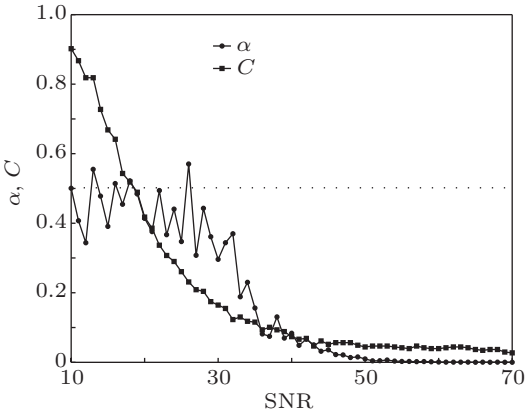


Fig. 8. Scaling exponent α and complexity index C of $x(t)$ versus the SNR.

5. Application

In order to verify the effectiveness of the above proposed approach, the reciprocating compressor gas valve is studied,

and the vibration data in four states including normal, gap, fractured, and bad spring valve states are tested respectively. The sampling frequency is 50000 Hz, and the sample length is 20000 points. The time waveforms of the reciprocating compressor gas valve in the four states are shown in Fig. 9. The time waveforms can not effectively show the feature of the reciprocating compressor gas valve in the time domain. The signal contains multiple impact sources, which is typically nonlinear and non-stationary. To apply the DFA and the LZC to the vibration signals of the gas valve in the four states, a sliding window method is proposed as follows.

For a time series, assume that the sample length is N. From the first point, take L points as the sub-series which is defined as a window, and calculate the scaling exponent α and the complexity index C of the sub-series. Then the window is slid backwards, the sliding step is τ points, and the next sub-series is obtained. Repeat the above steps until the last point is reached. In this study, N is set as 20000, L is set as 5000, and τ is 200, so the total number of the sub-series is 75. The scaling exponent α and the complexity index C of the gas valve in the four states are shown in Fig. 10.

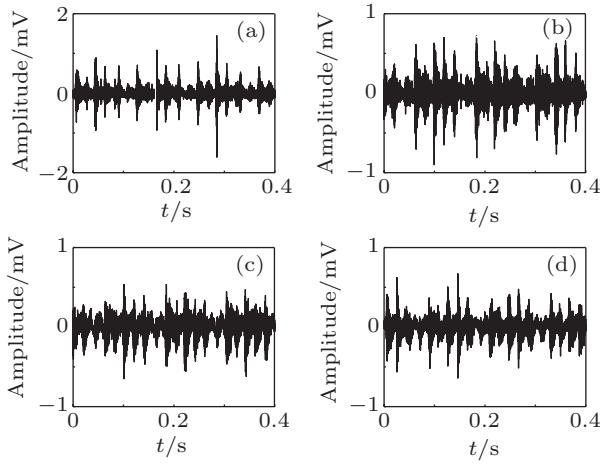


Fig. 9. Time waveforms of the gas valve in the four states: (a) normal valve state, (b) gap valve state, (c) fractured valve state, (d) bad spring valve state.

In Fig. 10, it clearly shows that the scaling exponents α and the complexity indexes C of the gas valve in the four states fluctuate around different mean values. According to Eq. (9), the correlation coefficients between α and C of the gas valve in the four states can be estimated. The detailed results of this analysis are shown in Table 2.

In Table 2, the mean of the scaling exponents α of the four gas valve states increases from 0.7789 to 1.5337, which is much larger than 0.5. It indicates that the non-stationarity of the signals decreases. The mean of the complexity index C decreases from 0.7089 to 0.5816, which represents that the randomness or nonlinearity of the signals also decreases. In addition, there is a significant correlation between α and C of the gas valve in the four states. The correlation coefficient of the normal state is the lowest. This reflects the fact that the nonlinearity and the non-stationarity of the vibration signal in the fault valve states are more significant than those in the normal state. Both the DFA and the LZC can effectively measure the non-stationary and the nonlinear characteristics of the reciprocating compressor gas valve.

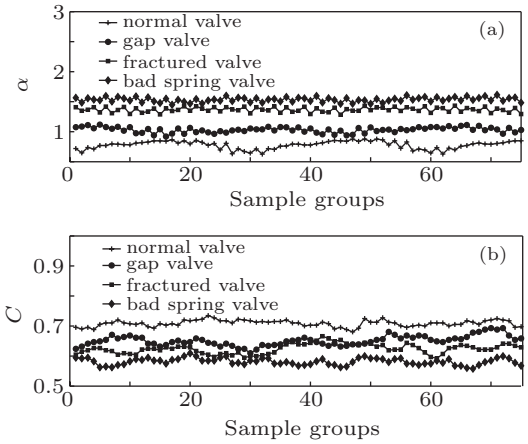


Fig. 10. (a) Scaling exponents α and (b) complexity index C of the gas valve in the four states.

Table 2. Correlation coefficients between the DFA and LZC of the gas valve signal.

State	Normal valve	Gap valve	Fractured valve	Bad spring
$\langle\alpha\rangle$	0.7789	1.0311	1.3671	1.5337
$\langle C\rangle$	0.7089	0.6511	0.6243	0.5816
ρ	0.5399	0.6158	0.6410	0.6705
Correlation	significant	significant	significant	significant

6. Conclusions

In this paper, the correlation between the scaling exponent α of the DFA and the normalized complexity index C of the LZC for a nonlinear time series is investigated. According to the analyses of a typical nonlinear time series, random noise signal, and gas valve vibration signal, the main conclusions are summarized below.

1) The scaling exponents α of the DFA and the normalized complexity indexes C of the LZC for typical dynamic systems including logistic, Duffing, Lorenz, Rossler, Henon, and Chen systems have high correlations. Both of these can be used to identify the dynamic behavior of nonlinear systems.

2) A high correlation also exists between the DFA and the LZC of time series with a different SNR. The presence of random noise in the time series can increase the values of the scaling exponent α and the normalized complexity index C. Compared with α, C is more sensitive to noise.

3) The scaling exponent α and the normalized complexity C can effectively extract the features of the reciprocating compressor gas valve in different states. Both of them can measure the non-stationary and nonlinear characteristics. The correlation coefficient ρ between α and C of the gas valve in the four states is significant.

References

[1] Huang N E, Shen Z and Long S R 1998 *Proc. R. Soc. London, Ser. A* **454** 903

[2] Srinivasan K, Panickar P, Raman G, Kim B H and Williams D R 2009 *J. Sound Vib.* **323** 910

[3] Dwivedi Y and Rao S S 2011 *J. Time Ser. Anal.* **32** 68

[4] Andel J, Netuka I and Ranocha P 2007 *Stat.* **41** 279

[5] Kalvoda T and Hwang Y R 2010 *Sens. Actuators A* **161** 39

[6] Yuan Y, Li Y, Mandicb D P and Yang B J 2009 *Chin. Phys. B* **18** 958

[7] Cong R, Liu S L and Ma R 2008 *Acta Phys. Sin.* **57** 7487 (in Chinese)

[8] Wu T Y, Chen J C and Wang C 2012 *Mech. Syst. Sig. Process* **30** 103

[9] Tang Y F, Liu S L, Lei N, Jiang R H and Liu Y H 2012 *Acta Phys. Sin.* **61** 170504 (in Chinese)

[10] Xie F, Yang R and Zhang B 2012 *Acta Phys. Sin.* **61** 110504 (in Chinese)

[11] Luo X L, Wang J and Han C X 2012 *Chin. Phys. B* **21** 028701

[12] Peng C K, Buldyrev S V and Havlin S 1994 *Phys. Rev. E* **49** 1685

[13] Zhu S S, Xu Z X, Yin K X and Xu Y L 2011 *Chin. Phys. B* **20** 050503

[14] Zhou Y, Yu Z G and Leung Y 2011 *Chin. Phys. B* **20** 090507

[15] Li S S, Lan J and Han B 2012 *Chin. Phys. B* **21** 064601

[16] Burt T and Worrall F 2007 *Hydrocarb Process* **21** 3529

[17] Peng C K, Havlin S, Stanley H E and Goldberger A L 1995 *Chaos* **5** 82

[18] deMoura E P, Vieira A P, Irma M A S and Silva A A 2009 *Mech. Syst. Sig. Process* **23** 682

[19] Daw C S, Finney C E A and Tracy E R 2003 *Rev. Sci. Instrum.* **74** 915

[20] Hong H B and Liang M 2009 *J. Sound Vib.* **320** 452

[21] Hou W, Feng G L and Dong W J 2005 *Acta Phys. Sin.* **54** 3940 (in Chinese)

[22] Wang K, Guan X P, Ding X F and Qiao J M 2010 *Acta Phys. Sin.* **59** 6859 (in Chinese)