Advances in Computer Science Research, volume 65
3rd International Conference on Communications, Information Management and Network Security (CIMNS 2018)

Prediction of Pumping Unit Well System Efficiency Based on Chaotic Time Series Method

Chaodong Tan[1, 3, *], Zhenhua Peng[2], Gang Feng[1], Ruogu Yang[3] and Lijie Gao[1]
[1.] China University of Petroleum (Beijing), Beijing Changping 102249, China
[2.] Engineering Technology Institute of Sinopec Northwest Oil Field Branch, Urumqi Xinjiang, 830011, China
[3.] Beijing Yadan Petroleum Technology Development Co., Ltd., Beijing Changping 102200, China
[*]Corresponding author

***Abstract*—The system efficiency prediction of pumping unit well always plays an important part in dynamic analysis of oil field production. The system efficiency of pumping unit well is restricted by a variety of factors, which is extremely complex. Ideal effect cannot be realized when applying structural causal models to forecast it. The chaotic time series forecasting model of system efficiency for pumping unit well is established in this paper; C-C method is used to determine the embedding dimension of phase space reconstruction parameters and delay time; the maximum Lyapunov exponent, which is figured out through small-data method, is used to detect the chaos of time series; the system efficiency of pumping unit well is predicted through univariate and multivariate time series forecasting method. The experiment verifies that chaotic time series forecasting method can be used in accurate system efficiency forecasting.**

Keywords*— pumping unit well; system efficiency; chaotic time series; phase space reconstruction; C-C method; forecasting

I. INTRODUCTION

The system efficiency of pumping unit well is a ratio of effective power and motor input power, which can directly reflect the whole operating efficiency of pumping unit well, so it is an important working condition index in pumping unit system. System efficiency prediction of pumping unit well always plays an important part in dynamic analysis of oil field production. Therefore, research on system efficiency prediction of pumping unit well never cease[1]. The so-called forecasting is predicting what might happen in the future by analyzing the historical data. If the system efficiency can be accurately predicted, preventive measures can be taken to achieve the purpose of energy-saving and cost-reducing. However, the production system of pumping unit well is a mechanic-electric-hydraulic coupled system with flexible slim and long rod to transfer energy. The invisible downhole pump movement and complex oil flow are both obstacles for people to know the system condition. Besides, oil well is a highly complex and non-linear open system, which not only follows various linear or nonlinear deterministic equation of seepage under complex media, but also can be artificially intervened, therefore, the changes of system efficiency in pumping unit well is complex with various constraint factors, besides, there are intricate relations between those factors. It is hard to achieve an ideal effect when using structural causal models in forecasting the system.

Chaos is a form of non-periodic and highly random motion which is widely existed. According to the historical data and current data, time series analysis is a way to find the internal characteristics and rules of the system and then build a digital model as accurately as possible. Chaotic time series analysis is useful in analyzing the chaotic and disordered random-noise-like time series. Therefore, a new system efficiency predicting method of pumping unit well is proposed in this paper, that is, the forecasting method based on chaotic time series.

II. THE ESTABLISHMENT OF SYSTEM EFFICIENCY PREDICTION MODEL BASED ON CHAOTIC TIME SERIES

Chaotic time series analysis includes chaotic identification and forecasting of various time series. Because chaotic phenomenon is conditional, any nonlinear system should be discriminated before chaotic time series analysis. The basis of chaotic time series analysis is phase space reconstruction theory because the distinction calculation of the system and the establishment of the predicting model are both carried out in phase space[2]. Firstly, C-C method is used to calculate the delay time and the embedding dimensions; secondly, the delay time and the embedding dimensions are combined to reconstruct the phase space; thirdly; the small-data method is used to calculate the max. Lyapunov exponent[3], if the Lyapunov exponent is greater than 0, it means that the chaos characteristic exists in this system; lastly, adding-weight one-rank local-region method is used to forecast the chaotic time series.

Theoretically, if the embedding dimension is large enough, the original dynamic system can be reconstructed by univariate time series. However, practically, it cannot be reconstructed only by univariate time series. Therefore, the theory of phase space reconstruction with multivariate time series is put forward in the references[3-4], in which, the experiment proved that when using multivariate time series, the calculation effect is better than using univariate time series. In this paper, we use chaos forecasting method of both univariate and multivariate time series to build the forecasting model of pumping unit production system.

A. Phase Space Reconstruction

Through phase space reconstruction, the multivariate system state can be known and with no need to know the mathematical model of the production dynamic system., which makes it possible for the experimental data to be analyzed and processed by nonlinear dynamic method. Besides, the method of delay-coordinate reconstruction[3-4] is generally used.

In the production dynamic system of pumping unit well, the most important observation is the time series of system efficiency, which is set as $\{x_n\}$, n=1, 2, …, N. N means the overall length of the sequence. Delay time τ and embedding dimension m are used for phase-space reconstruction

$$\begin{bmatrix} x(1) & x(2) & \cdots & x(M) \\ x(1+\tau) & x(2+\tau) & \cdots & x(M+\tau) \\ \vdots & \vdots & & \vdots \\ x(1+(m-1)\tau) & x(2+(m-1)\tau) & \cdots & x(M+(m-1)\tau) \end{bmatrix} \quad (1)$$

In which the phase points can be expressed as

$$X(n)=[x(n),x(n+\tau),\ldots x(n+(m-1)\tau)];\ n=1,2,\ldots M \quad (2)$$

$$M = N-(M-1)\tau \quad (3)$$

X(n) can describe the original system, if suitable delay time and embedding dimension are chosen. The X(n) that structured by x(n) is called phase-space reconstruction. C-C method that put forward in reference[4] is applied in this paper to calculate the delay time and embedding dimension.

B. *The Determination of Delay Time and Embedding Dimension*

Work out the optimal delay time τ and embedding window w by using correlation integral method. Work out embedding dimension by using the formula $\tau_w=(m-1)*\tau$.

The original system efficiency time series $\{x_n\}$, n=1, 2, …, N is divided into t non-intersecting time subsequences. Length l=[N/t] ([·] means rounding)

$$\begin{aligned} &\{ x(1), x(t+1), x(2t+1),\ldots\} \\ &\{ x(2), x(t+2), x(2t+2),\ldots\} \\ &\{ x(t), x(t+t), x(2t+t),\ldots \} \end{aligned} \quad (4)$$

Respectively calculate the statistical magnitude of each subsequence S(m, N, r, t)

$$S(m,N,r,t)=\frac{1}{t}\sum_{l=1}^{t}\{C_l(m,N/t,r,t)-[C_l(1,N/t,r,t)^m]\} \quad (5)$$

In the calculation, C_l means the correlation integral of the first subsequence. Distance between any two points in phase space that smaller than that of neighbor radius can be defined as

$$C(m,N,r,t)=\frac{2}{M(M-1)}\sum_{1\le i<j\le M}\theta(r-\|X_i-X_j\|_\infty) \quad (6)$$

r means neighbor radius, M=N-(m-1)t m ans the number
e
of the points in phase space, N means the whole length of time series, θ(·) means Heaviside unit function (defined as θ(x)=0, if x< 0; θ(x)=, if x≥0)

When

$$N\to\infty$$

$$S(m,r,t)=\frac{1}{t}\sum_{l=1}^{t}\{C_l(m,r,t)-[C_l(1,r,\tau t)^m]\} \quad (7)$$

According to BDS statistic, if time series is independent and identically distributed, when N→∞, S(m, r, t) will constantly be 0. S(m, r, t) ~t can reflect the auto-correlation characteristic of time series. The first time when it is greater than zero or when it has the minimal distance difference to all the radius r, it means that the points in phase space are almost even-distributed and the reconstructed attractor orbits are fully expanded in phase space.

Defined as

$$\Delta S(m,t)=\max\{S(m,\gamma_i,t)\}-\min\{S(m,\gamma_i,t)\} \quad (8)$$

ΔS(m, t) can be used to measure the first zero point of S(m,r,t) ~t or the first minimal point of ΔS(m, t) ~t.

By applying BDS statistics, the reasonable estimation of N, m, r can be gained. Take N=3000, m=2, 3, 4, 5, r=k·σ/2, k=1, 2, 3, 4, σ is the mean square error of the time series. So the equation is as follows

$$\begin{aligned} \bar{S}(t) &= \frac{1}{16}\sum_{m=2}^{5}\sum_{j=1}^{4} S(m,r,t) \\ \Delta\bar{S}(t) &= \frac{1}{4}\sum_{m=2}^{5}\Delta S(m,t) \\ S_{cor}(t) &= \Delta\bar{S}(t)+\left|\bar{S}(t)\right| \end{aligned} \quad (9)$$

According to C-C method, the first zero point of S(t) or the first minimal point of is ΔS(t) the optimal delay time τ. The minimum point of S_{cor} (t) is the width of the embedding window τ_w[5]. The embedding dimension m can be gotten from $\tau_w=(m-1)*\tau$.

Establish the model according to the theory mentioned above and realize the algorithm through programming, take the actual data of the pumping unit well system efficiency as the object to gain the delay time and the embedding width.

It is shown in Figure 1 that in the first zero point of $\bar{S}(t)$, the delay time τ=9; in the first minimum point of $\bar{S}(t)$, the delay time τ=3; and in the minimum point of S_{cor} (t), the embedding width is τ_w=96. By applying Lorenz chaotic system, it is verified that the delay time determined by the first minimum point of $\bar{S}(t)$ is more stable, so the actual system efficiency of the pumping unit well is τ=3, m=33.

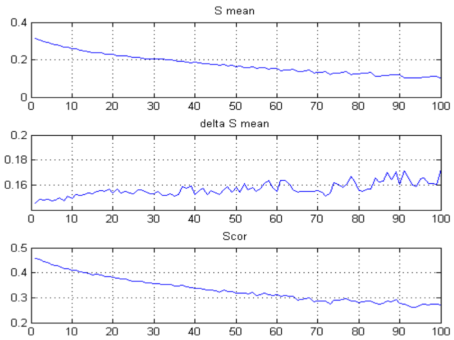

FIGURE I. RECONSTRUCTING THE SYSTEM EFFICIENCY X-COMPONENT THROUGH C-C ALGORITHM

III. IDENTIFICATION OF CHAOTIC CHARACTERISTIC FOR PUMPING UNIT SYSTEM EFFICIENCY

Chaotic motion has sensitivity to the initial conditions. The distance between adjacent orbitals from two adjacent initial points will exponentially separate from each other over time. This geometrical property can be described quantitatively. It is verified that as long as Lyapunov exponent is greater than 0, the system has chaos characteristic[6], which becomes one of the main quantitative methods for chaos identification. In this paper, the small-data method mentioned in reference [5] which is more suitable for small sample set is used to calculate the maximum Lyapunov exponent.

After the phase-space reconstruction, find out the nearest neighbor points for each phase point in system efficiency time series. For example, if the nearest neighbor point of phase point X_j is $X_{(\eta(j))}$, then the minimum distance will be

$$d_j(0) = \min_j \left\| X_j - \eta(j) \right\|, \quad [j - \eta(j)] > P \tag{10}$$

What needs to be pointed out is that average period of the series P $|j-\eta(j)| > P$ is used to separate the neighboring points transiently to ensure that they can run in different orbits[5-7].

Figure out the distance $d_j(i)$ of every neighboring point after i iscrete time steps.

$$d_j(i) = \left| X_{j+i} - X_{\eta(j)+i} \right|, \quad (i = 1,2,\cdots \min(M - j, M - \eta(j))) \tag{11}$$

$$\lambda(i) = \frac{1}{i\Delta t}\frac{1}{(M-i)}\sum_{j=1}^{M-i} \ln \frac{d_j(i)}{d_j(0)} \tag{12}$$

Among which Δt is the period of the time series.

According to the definition of Lyapunov exponent

$$d_j(i) = d_j(0) \cdot e^{\lambda_L (i\Delta t)} \tag{13}$$

After taking logarithm on both sides

$$\ln d_j(i) = \ln d_j(0) + \lambda_L (i\Delta t) S \tag{14}$$

So far a series of straight lines is gained from $\ln d_j(i)$ changing with the discrete time steps. The slope of the straight line is λ_L Δt. The max. Lyapunov exponent λ_L can be gotten by averaging the slope of the straight line.

Therefore, fixing i, computing the average number of $d_j(i)$ that corresponded by all of the j and then divided by Δt, the y(i) can be gained

$$y(i) = \frac{1}{q\Delta t}\sum_{j=1}^{q} \ln d_j(i), \quad d_j(i) \neq 0 \tag{15}$$

q is nonzero number of $d_j(i)$. Select a linear region in the curve of y(i)~i and make a regression line through least square method. The slope of this straight line is the max. Laypunov exponent λ_L.

If the max. Laypunov exponent of time series $\lambda_L>0$, it means that the evolutionary track of this time series is divergent and it has a chaotic characteristic. Therefore, it's max. forecasting time scale T_f can be predicted, and its relation with the max. Laypunov exponent can be expressed as follows

$$T_f = 1/\lambda_L \tag{16}$$

According to the theory mentioned above, programming to use small-data method for getting the max. Laypunov exponent. Take an actual system efficiency data of pumping unit well as the object, setting the delay time as 3 and the embedding dimension as 33, the result is shown in Figure 2.

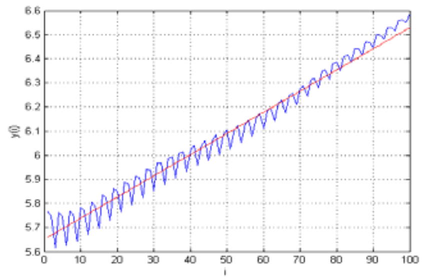

FIGURE II. THE MAX. LAYPUNOV EXPONENT GAINED BY SMALL-DATA METHOD

Fit the section with good linearity through least square method, evaluate the slope and then get the max. Laypunov exponent 0.088, which verifies that the time series of this system efficiency definitely has chaotic characteristics and system efficiency over the next 114 days can be predicted.

IV. CHAOTIC TIME SERIES FORECASTING

The adding-weight one-rank local-region method is used to forecast system efficiency of pumping unit well {xn}. Spatial distance from the central point to other points in the phase is an important parameter. The accuracy of the prediction often depends on several points that close to the central point.

Therefore, The prediction precision and the ability of noise elimination are improved by introducing space distance of the central point. The principle is as follows.

Set the proximal point of the central point X_M, X_{Mi}, i=1, 2, …, q, and the distance between them is d_i, assuming that d_m is the minimum number in d_i, the weight number of definition point is

$$P_i = \frac{\exp[-a(d_i - d_m)]}{\sum_{j=1}^{q} \exp[-a(d_i - d_m)]} \tag{17}$$

Among which, a is a parameter and in general, a=1.

In adding-weight one-rank local-region method, the first order linear fitting method is used to approach to the evolutionary trend of the phase point

$$x_{M_{i+1}} = a + b x_{M_i}(j) \tag{18}$$

When each phase point is decomposed into each dimension component, there will be

$$x_{M_i} = a + b x_{M_i}(j), \quad j = 1, 2, \cdots m \tag{19}$$

Among which, $X_{Mi+1}(j)$ and X_{Mi} mean j-dimension component of phase point $X(M_i+1)$ and phase point $X(M_i)$, m is embedded dimension, that is, each phase point has m dimension component.

Using weighted least square method to get the parameter values of optimal a and b, then the problem is transformed into

$$\sum_{i=1}^{k} P_i \cdot \{\sum_{j=1}^{m} [x_{M_{i+1}}(j) - a - b x_{M_i}(j)]^2\} \to \min \tag{20}$$

Regarding the above equation as a function of a and b, in order to get the minimum value, evaluating the partial derivatives of a and b to get the value of a and b.

$$\left\{\begin{aligned} &\sum_{i=1}^{k} P_i \cdot \left\{\sum_{j=1}^{m} [x_{M_{i+1}}(j) - a - b x_{M_i}(j)]\right\} = 0 \\ &\sum_{i=1}^{k} P_i \cdot \left\{\sum_{j=1}^{m} [x_{M_{i+1}}(j) - a - b x_{M_i}(j)] \cdot x_{M_i}(j)\right\} = 0 \end{aligned}\right. \Rightarrow \tag{21}$$

$$\left\{\begin{aligned} &\sum_{i=1}^{k} P_i \cdot \left\{\sum_{j=1}^{m} [x_{M_{i+1}}(j) - a - b \sum_{i=1}^{k} P_i \sum_{i=1}^{k} x_{M_i}(j)]\right\} = 0 \\ &\sum_{i=1}^{k} P_i \cdot \left\{\sum_{j=1}^{m} x_{M_{i+1}}(j) \cdot x_{M_i}(j) - a \sum_{i=1}^{k} P_i x_{M_i}(j) - b \sum_{i=1}^{k} P_i x_{M_i}(j)^2\right\} = 0 \end{aligned}\right.$$

Get the prediction formula of the phase point evolution and forecast the evolution of the phase point in the next step, at last, the one dimension component is the predicted value of the time series.

V. INSTANCES ANALYSES

By applying the prediction model of chaotic time series that mentioned above, the data of system efficiency of a well in an oil field are analyzed. Except the abnormal data, there are 1000 data.The changing features of system efficiency time sequences are shown in Figure 3.

Programming to realize the prediction model of the chaotic time series. Firstly, based on C-C method, figure out the delay time τ=3 and embedding dimension m=33. Adding-weight one-rank local-region method is used to forecast the chaotic time series. The last 60 samples are used for verifying the prediction model. Comparison between the actual data and the predicted data of system efficiency time series is shown in Figure 4.

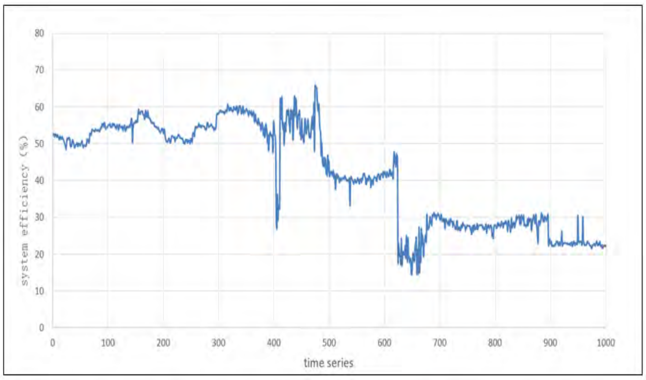

FIGURE III. CHANGING CURVE OF ORIGINAL SYSTEM EFFICIENCY DATA

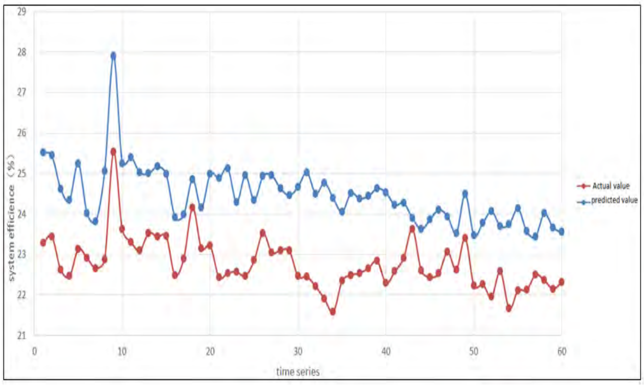

FIGURE IV. COMPARISON DIAGRAM OF THE ACTUAL DATA AND PREDICTED DATA

It can be known from the diagram that the prediction curve coincides well with the actual measured curve. However, there are individual predicted points that greatly differ from the actual points. Average relative error cannot truly reflect the predicting accuracy of the model. In order to objectively evaluate the advantages and disadvantages of the prediction model, the root-mean-square error (RMSE) is applied to evaluate the model in this paper. RMSE can sensitively reflect the large or small error in the predicted values. Thus it can reflect the accuracy of the prediction result. After computing, the square error of the time series analysis model is 0.0438.

According to the calculation above, the accuracy of the prediction of univariate chaotic time series still can be improved. Therefore, multivariate chaotic time series prediction is applied to improve accuracy. Choosing the balance rate of the pumping unit, yield of the pumping unit well and electricity consumption of 100-meter-ton liquid as influence factors in reconstruction of the phase space and figure out the embedding time and embedding dimension. Since the balance rate τ1=3, =32, yield τ2=15, =7 and electricity consumption of 100-meter-ton liquid τ3=4, =13, when forecasting the system efficiency, the input sample is 33+32+7+13=85 dimensions. Taking τ=3, 85 to reconstruct the phase space again and forecasting the last 60 data to contrast with actual data. The prediction result of multivariate chaotic time series is shown in Fig.5. It can be known from the diagram that the multivariate prediction curve coincides well with the actual data curve, which is more accurate than using univariate chaotic time series prediction method.

In order to evaluate the advantages and disadvantages of the prediction model more objectively, the RMSE is applied to evaluate the model in this paper.

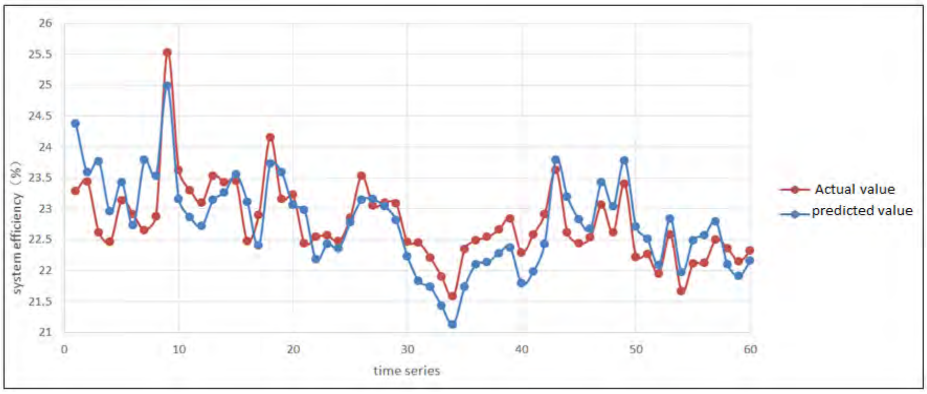

FIGURE V. UMPING EFFICIENCY'S RESULTS OF CHAOTIC TIME SERIES PREDICTION BY MULTIPLE VARIABLES

It can be known from Table 1 that the MSE of univariate is 0.0438, while the MSE of multivariate is 0.0095, which means that the multivariate prediction result is superior to univariate prediction result. Therefore the prediction method of multivariate chaotic time series proposed in this paper is correct and practicable.

TABLEI ERROR OF MEAN SQUARE IN UNIVARIATE AND MULTIVARIATE PREDICTION

——	Univariate	multivariate
MSE	0.0438	0.0095

According to actual calculated result and MSE analysis, based on system efficiency prediction of the actual well, whose accuracy reaches 98%, it means this method can be applied to actual production.

VI. CONCLUSIONS

(1) Prediction model of multivariate chaotic time series is established by analyzing the dynamic features of the system efficiency of complex pumping unit well.

(2) By using the chaotic time series prediction model of system efficiency mentioned in this paper, programming and analyzing the actual efficiency of a well, the max. Laypunov exponent is 0.0088>0, which means that chaotic characteristic exists in the system.

(3) What needs to be pointed out is that specific problems must be analyzed specifically. Some oil fields may not have chaos characteristics because of strong development regularity. Therefore, before applying the chaotic time series forecasting method, the original sequence should be observed and identified to judge whether it is chaotic or no.

(4) The prediction model is applied to an actual well. The prediction result shows that the chaotic time series analysis method can accurately predict the system efficiency of pumping unit well. Besides, the multivariate prediction result is superior to univariate prediction result. Therefore the prediction method of multivariate chaotic time series proposed in this paper is correct and practicable.

REFERENCES

[1] Li Xinlun, Tan Chaodong, Tan Fanfei, etc. Application of ARIMA Model in Tine Series Analysis to Predict the Efficiency of Pumping Well Production System [J], Peak Data Science, 2017, 6(01):41-44.

[2] H.S. Kim, R. Eykholt, J.D. Salas. Nonlinear Dynamics, Delay Times, and Embedding Windows [J]. Physica D, Nonlinear Phenomena, 1999, 127(1-2): 48-60.

[3] Chen Yifeng, Lv Jinhu, Zhou Chuangbing. Slope Displacement Prediction Based on Lyapunov Index Improved Algorithm [J]. Chinese Journal of Rock Mechanics and Engineering, 2016, 05:671-675.

[4] Hu Yu, Chen Tao. Phase Space Reconstruction Technology of Chaotic Attractor Based on C-C Method [J], Journal of Electronic Measurement and Instrumentation, 2012, 05:425-430.

[5] Yang Yongfeng, Wu Minjuan, Gao Zhe, Wu Yafeng, Ren Xingmin. The Parameter Selection of the Largest Lyapunov Index Calculated by the Small Data Method [J] Journal of Vibration, Measurement and Diagnosis, 2012, 03:371-374.

[6] Zhang Hailong, Min Fuhong, Wang Enrong. The Comparison for Lyapunov Exponent Calculation Methods[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2012, 01:5-9.

[7] Li Guohui, Xu Deming, Zhou Shiping. Computing the Largest Lyapunov Exponent from Time Series [J]. Journal of Applied Sciences, 2003, 02:127-131.