International Journal of Database Theory and Application
Vol.6, No.6 (2013), pp.75-84
http://dx.doi.org/10.14257/ijdta.2013.6.6.07

Research on Multi-Input Complex System based on Phase Reconstruction

Jianming Sun

Harbin University of Commerce, Harbin
sjm@hrbcu.edu.au

Abstract

This paper uses phase space reconstruction as the basis of the multi – input nonlinear method. It is obvious that for a system with multiple variables, it is necessary to choose the involved variable before reconstruction, and identify the reconstruction parameter, after determining the reconstructed variable, in order to complete the basic reconstruction. Therefore, based on the selection of the nonlinear correlation, this paper introduces the method for choosing the correct input variable at first , and then introduces some commonly used methods to identify the reconstruction parameter, such as mutual information method, auto – correlation method and average displacement method etc.. Furthermore, it specially introduces the C – C method. By carrying out the multivariate combination forecasting simulation for time series of the Lorentz Equation and comparing the reconstruction phase diagram of the multivariate phase space, this paper verifies the accuracy of selecting reconstructed input vector based on nonlinear correlation and the effectiveness of using C- C method to decide the reconstruction parameter.

Keywords: *Phase space reconstruction; multi – input; complex system; reconstruction variable*

1. Introduction

The basic characteristic of complex system is whole emergence, and it is different from the characteristic of linear add up each components. It is generally believed that the complex system is constituted by numerous components that existed complex interaction, simply add the grouped behavior to obtain the system behavior is impractical. In order to study the complex system, it is necessary to grasp its integrity, as well as consider the interaction between each component [1, 2]. In general, by analyzing the evolution rule of the observed time series, we can understand the dynamic features of the complex system. Therefore, in order to well analyze the system, it is essential to synchronously measure the time series for all variables in the system. In fact, it is difficult to realize. Furthermore, too much variable information will bring information disaster for our following analysis. For example, the forecasting model establishment will become too large, and it will cause noise, redundant information, increase the difficulty of modeling, as well as bring large errors while analyzing and predicting the time series at later stage. Obviously, this will require us to make decision for the considerable variable, that is, select the variable with useful information for helping us well understand the system, and at the same time, the process to understand will not be over complicated.

For a real complex system, after identifying the observed variable and measuring by the corresponding sensor, we will obtain the single variable or multi variable time series with equal time interval. Before the advent of the reconstruction theory, people always analyzes

ISSN: 2005-4270 IJDTA

the time evolution of the system from these sequences directly, obviously, it will have significant limitation. Because the time series comprehensively reflect the interaction between various physical factors existed in the system, and it contains all of variable traces involved in the movement. Although, formally, time series is random, it may contain sequence with essential movement information in the system. While for a complex system, in order to show its nature of movement, three dimension system is necessary. Therefore, it is necessary to expand the time series to at least three or higher dimensional vector space in order to fully bring out the nature of the information for the time series, that is, the reconstruction phase space of the time series [3].

Vector space reconstruction is firstly put forward by Chaotic Effect in the study of the Dynamic System. It is an effective way to recover the track of dynamic system, and by utilizing the recovered track, it is able to decide whether there is chaos existed in a nonlinear discrete and continuous dynamical system or not. At the same time, it is obvious that the construction of the phase space will contribute to calculate the characteristic quantity of the dynamic system, such as fractal dimension, Lyapunov index and so on. Obviously, utilizing these characteristic quantities could help us to identify the nature characteristics of things. At present, the application area of the reconstruction phase space becomes more and more widely use, and it is no longer limited to the study of chaotic phenomena, on the other hand, it can be used to study the causal relationship between various things in the nonlinear complex system, such as the law of development, that means the past, present and future development trends of the things. While forecasting is depend on things' regular pattern to reveal their future states. The usual method is to establish the model according to the issue, and then solve the model, finally, make calculation by using the solved model. With the development of the related scientific theory, phase space reconstruction has become more and more useful in the sample selection when establishing the prediction model.

2. Model Approach

Multi input phase space reconstruction is developed based on the univariate input phase space reconstruction. In this section, we make a brief overview of the thoughts of the univariate reconstruction at first, and then provide a detailed introduction for the multivariate phase space reconstruction.

2.1. Univariate Phase Space Reconstruction

The phase space of a complex nonlinear system may have a high number of dimensions, and in most cases, it is difficult to know the exact number of dimensions. Usually, the common way is to expand the given time series in the real system to three or higher dimensional space, so as to sufficiently find the information in the time series, this is called the time series phase space reconstruction issue. Obviously, the purpose to reconstruct the phase space is to show us how to reconstruct attractor according the limited time series.

For an n – dimension autonomous dynamic system

$$\begin{cases} \dot{x}_1 = f_1(x_1, x_2, \cdots, x_n) \\ \dot{x}_2 = f_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ \dot{x}_n = f_n(x_1, x_2, \cdots, x_n) \end{cases} \tag{1}$$

It is important to perform differentiation for equation (1) to eliminate $x_2, \cdots, x_n$, and obtain n order differential equation for x_1

$$x_1^n = f\left(x_1, \dot{x}_1, \ddot{x}_1, \cdots, x_1^{(n-1)}\right) \tag{2}$$

Therefore, the coordinates of the state space will be replaced by the all-order derivatives of x_1, that is $\left(x_1, \dot{x}_1, \ddot{x}_1, \cdots, x_1^{(n-1)}\right)$, or $\left(x_1, \dot{x}_1, \ddot{x}_1, \cdots, x_1^{(n-1)}, x_1^{(n)}\right)$. Such replacements will not loss the evolution information of the dynamic system. The derivatives could be written as the difference between the evolution sequences in different time, while τ is called delay parameter.

$$x(t), x(t+\tau), x(t+2\tau), \cdots, x\left(t+(n-1)\tau\right) \tag{3}$$

Because $x(t+\tau) = x(t) + \dot{x}(t) \cdot \tau$, and $x(t+\tau)$ is the derivative result of $x(t)$. Similarly, $x(t+2\tau)$ can be considered as the result of the second order derivative, by analogy, we can consider each value in above equation as the result of all order derivatives of x. Therefore, it is reasonable to reconstruct the system phase space based on the above equation.

2.2. Multivariate Phase Space Reconstruction

By reconstructing m dimensional phase space via some observed time series in the system, $\{x_i | i = 1, 2, \cdots, N\}$, we can obtain a set of phase space vector

$$X_i = \left\{x_i, x_{i+\tau}, \cdots, x_{i+(m-1)\tau}\right\}, i = 1, 2, \cdots, M \qquad X_i \in R^m \tag{4}$$

Where, τ is time delay; $m \geq 2d + 1$, d represents the number of independent variable in the system; $M = N - (m-1)\tau$. Obviously, the key to reconstruct the phase space depends on the embedded dimension m and the selection of time delay τ. It is important to expand the thought of univariate reconstruction to the multivariate space. By the interaction with M – variables, a multivariate complex system, $\{x_i\}_{i+1}^{M}$, at t time, could perform multivariate phase space reconstruction according to the following equation[4-6].

$$V(t) = \begin{bmatrix} x_{1,t}, x_{1,t-\tau_1}, \cdots, x_{1,t-(m_1-1)\tau_1} \\ x_{2,t}, x_{2,t-\tau_2}, \cdots, x_{2,t-(m_2-1)\tau_2} \\ \vdots \\ x_{M,t}, x_{M,t-\tau_M}, \cdots, x_{M,t-(m_M-1)\tau_M} \end{bmatrix} \tag{5}$$

m_i is the embedded number of dimension for the its variable x_i, while τ_i is its delay time.

Appropriate reconstruction parameter is a guarantee to effectively perform the multivariate reconstruction. Since the advent of reconstruction theory, many parameter determination methods have been proposed, however, there is no accepted effective method so far. For the multivariate time series, reconstruction parameter selection continuous to use the reconstruction parameter selection method of univariate time series.

3. Algorithm and Implementation

In order to construct phase space by time series, it is very important to identify parameter m and choose the appropriate sample interval τ. Theoretically, the selection of τ can be almost random. However, in the real system, the selection of τ should be identified by repeated trials. If τ is too small, then the track of the phase space tends to be a straight line; If τ is too large, then the data point will be concentrated in a small region of the phase space, it is difficult to obtain the partial structure of the attractors from the reconstructed phase space. Many people have put forward various schemes to select t τ : One method think, in order to make each component independently of each other in the constructed vector (that means lower correlation), it is necessary to find a τ value that make the related function become zero. However, this method cannot give us an appropriate time interval τ in any cases.

Autocorrelation method [9] is a well-developed method to identify time delay τ , its basis thought is to pick up the linear dependence of the time series. For a time series, it is important to write its related function at first, and then draw the autocorrelation function image about time τ , as the autocorrelation function reduced to $1-\frac{1}{e}$ of the initial value, the obtained time is the time delay τ of the reconstruction phase space. The application of Autocorrelation function method is only because the calculation is simple. It describes the linear correlation for data sequences, and naturally, it is not suitable for nonlinear analysis,

Average displacement method is geometric method of phase space reconstruction. According to practice results, the corresponding time delay τ , which waveform slope of the selected metric value firstly reduce to below 40% of its initial slope, is the value to find. However, this method has a disadvantage of randomness, and an advantage of significant geometric meaning. Mutual information is an effective way to estimate the delay parameter of the phase space reconstruction, from the information theory point of view, this method not only able to analyze linear system, but also analyze the nonlinear system. It is suitable for nonlinear analysis in theory. Show firstly put forwards to use the mutual information's first time achieved minimum time lag as the time delay of the phase space reconstruction. Fraser provides recursive algorithm of mutual information calculation [1, 8]. Compared with autocorrelation method, although the calculation of the mutual information method is slightly more complicated, it is able to efficiently extract the nonlinear characteristic in the sequence. Obviously, its calculation results are significantly superior to autocorrelation method. Next, we will introduce a kind of method to calculate both embedding dimension and delayed time, this improved method called C-C method [11, 12],

For the observed time series, $\{x_i | i = 1, 2, \cdots, N\}$, with a group of length, N, it is necessary to perform delay phase space reconstruction according to the Tokens Theorem. Therefore, we define the correlation integral of the embedding time series as:

$$C(m, N, r, t) = \frac{2}{L(L-1)} \sum_{1 \le i \le j \prec L} \theta(r - d_{ij}) \quad (5)$$

The correlation integral is a function used to characterize the cumulative distribution, and it calculate the probability that the distance between any two points, $d_{ij} = |Y_i - Y_j|$, is less than radium r in the reconstructed phase space . $i = 1, \cdots, L = N - (m-1)$, L is the number of phase points in the m dimension phase space, r is search radius, $r \succ 0$, t is delay time τ , $\theta(x)$ is Heaviside function: if $x \prec 0$, $\theta(x) = 0$; if $x \ge 0, \theta(x) = 1$. $d_{ij} = \|x_i - x_j\|_{(\infty)}$ is generally characterized by the maximum norm.

Define the statistic test quantity as:

$$S_1(m, n, r, t) = C(m, n, r, t) - C^m(1, n, r, t) \quad (6)$$

For Equation (6), it is necessary to split the initial time series into t disjoint sub – time series, and then adopt partition average strategy, separately calculate the test statistics of each time series, $S_2(m,n,r,t)$, when $n \to \infty$,

$$S_2(m,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s(m,r,t) - C_s^m(1,r,t)\right] \tag{7}$$

Specific operation: it is important to divide the time series $\{x_i \mid i = 1,2,\cdots,N\}$ into t disjoint time series, for general natural number t ,

We have

$$\begin{aligned}&\{x_1, x_{t+1}, x_{2t+1}, \cdots\}\\&\{x_2, x_{2+1}, x_{2t+2}, \cdots\}\\&\cdots\cdots\\&\{x_t, x_{2t}, x_{3t}, \cdots\}\end{aligned} \tag{8}$$

The length of sub – sequence for each time series is $l = N/t$, defines $S(m,N,r,t)$ as

$$S(m,N,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s\left(m,\frac{N}{t}r,t\right) - C_s^m\left(1,\frac{N}{t},r,t\right)\right] \tag{9}$$

Let $N \to \infty$, we have

$$S(m,r,t) = \frac{1}{t}\sum_{s=1}^{t}\left[C_s(m,r,t) - C_s^m(1,r,t)\right] \tag{10}$$

The relationship between the test statistics and t reflects the autocorrelation of the time series. If the time series is independent and identically distribute, then for fixed m , t , as $N \to \infty$, for all r , the $S(m,r,t)$ is identically equal to zero. However, the real sequences are limited, and the sequence elements may be relevant, it is obvious that the actual $S(m,r,t)$ that we obtained will not equal to zero. Therefore, the null point of $S(m,r,t)$ or the point with minimum difference of all search radius can be selected as the maximum time interval for the partial part. That means these points are almost uniformly distributed. We choose two radius with the maximum and minimum corresponding value, and define its difference value is $\Delta S(m,t) = \max\{S(m,r_i,t)\} - \min\{S(m,r_j,t)\}$, it is obvious that the maximum time of the partial part should be the null point of the $S(m,r,t)$ or the minimum value of its difference. However, the null point of $S(m,r,t)$ should be equal respond to almost all m , t , while the minimum value of $\Delta S(m,t)$ should be equal respond to all m .

According to the statistic results,

$$\bar{S}(t) = \frac{1}{16}\sum_{m=2}^{5}\sum_{j=1}^{4} S(m,r_j,t) \tag{11}$$

$$\Delta\bar{S}(t) = \frac{1}{4}\sum_{m=2}^{5}\Delta S(m,t) \tag{12}$$

$$S_{cor}(t) = \Delta\bar{S}(t) + \left|\bar{S}(t)\right| \tag{13}$$

From the above three equation, we look for the corresponding timing for the first $\bar{S}(t)=0$ or $\Delta\bar{S}(t)=\min\{\Delta\bar{S}(t)\}$ as the delay time τ_i. At the same time, look for corresponding time of the minimum point of $S_{cor}(t)$ as the first overall maximum time delay window, $\tau=T_w$, in the time series.

The major steps show as follow：

1)Read in data；

2)Calculate the standard deviation of the time series；

3)Let r varies from $sigma/2$ to $2sigma$， m varies s from 2 to5， t varies from 1 to 200；

4)partition the time series to t disjoint sub – sequence by calling disjoint function, and call correlation integral function to calculate $C(1,N,r,t)$;

5) Reconstruct phase space for sub – sequence by calling reconstitution function, and call correlation integral function to calculate $C(m,N,r,t)$;

6) Calculate $C(m,N,r,t)-C(1,N,r,t)^m$， and make summation for t to obtain s_t3；

7) According to the calculation of $S(m,r_j,t)=\frac{1}{t}\sum_{s=1}^{t}\left[C_s(m,r_j,t)-C_s^m(1,r_j,t)\right]$, obtained s_t2(j)， and make summation for r_j , and obtained s_t1；

8)Assign s_t1 to s_t0(m)， and make summation for m， obtained s_t；

9)According to equation $\bar{S}(t)=\frac{1}{16}\sum_{m=2}^{5}\sum_{j=1}^{4}S(m,r_j,t)$, obtained s(t)；

10)At the same time, utilizing the obtained s(t) ， according to equation $\Delta S(m,t)=\max\{S(m,r_i,t)\}-\min\{S(m,r_j,t)\}$ and $S_{cor}(t)=\Delta\bar{S}(t)+|\bar{S}(t)|$ ， find corresponding delt_s(t) and s_cor(t) separately;

11) Mapping according to the obtained results.

4. Multivariate Forecasting Simulations

It is important to select the time series of three variable Lorentz equations in order to perform simulation, and its equation expression is as follows:

$$\begin{aligned}\dot{x}&=-\sigma(x-y)\\ \dot{y}&=rx-xz-y\\ \dot{z}&=xy-bz\end{aligned}\qquad(14)$$

In the above equation, parameter $\sigma=10$， $b=8/3$， $r=28$, the integration step $h=0.01$, by applying four step Runge – Kutta Method to generate time series. Figure 1 shows the three dimension space diagram of Lorentz equation. It is necessary to get rid of preceding 20000 points, and take the next 2500 points as the chaotic time series for the experiment.

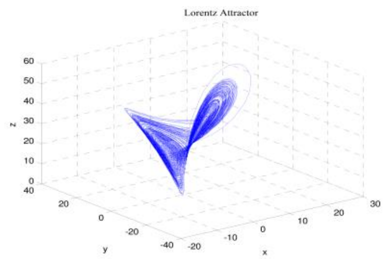
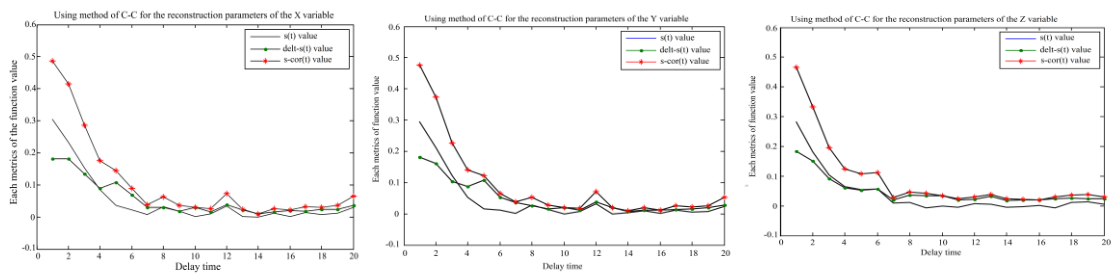


Figure 1. Lorentz Attractor x, y, z Three Variable

As reconstructing the phase space, according to C – C method, we obtained the value of each metric function in the equation (11)、(12)、(13), their value are shown in Figure 2. From Figure 2, the delay of three variables are $\tau_x = 4$, $\tau_y = 4$, $\tau_z = 5$, time delay window is $t_x = 14$, $t_y = 14$, $t_z = 16$, thus, we can obtain their embedding dimension, which is $d_x = 5$, $d_y = 5$, $d_z = 5$ separately.



(a) (b) (c)

(a) Delay time and time window of x based on C-C method (b) Delay time and time window of y based on C-C method (c) Delay time and time window of z based on C-C method

Figure 2. Using C-C method for the Reconstruction Parameter of XYZ variable in the Lorentz Equation

Figure 3 shows three dimension phase diagram of each variable after reconstruction, Figure 4 is three-dimensional display of three variable values after reconstruction. By comparing Figures 1 and 4, the reconstructed parameter selected by C – C method is able to well restore the trace of motive power system, and provide a better platform to analyze the characteristic of system dynamic in the future.

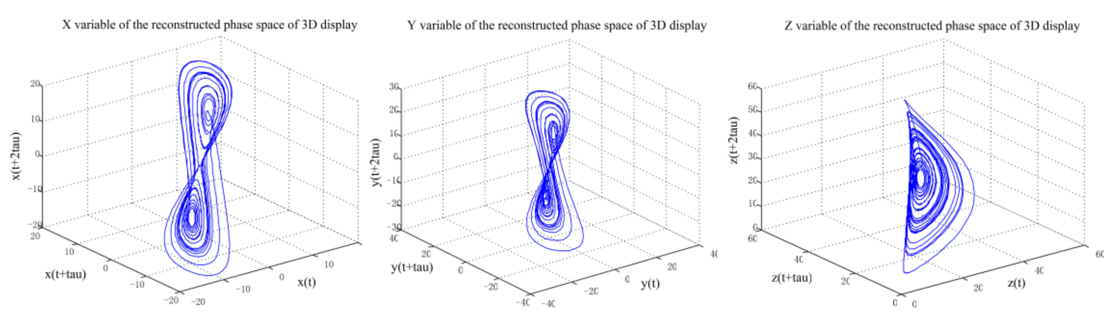
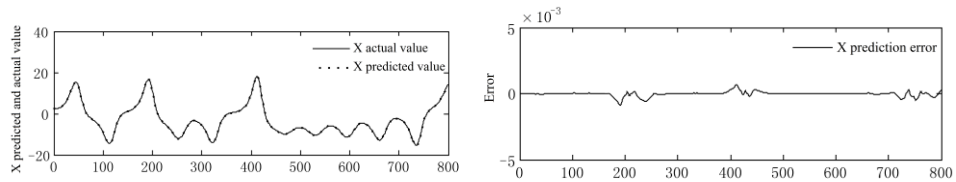
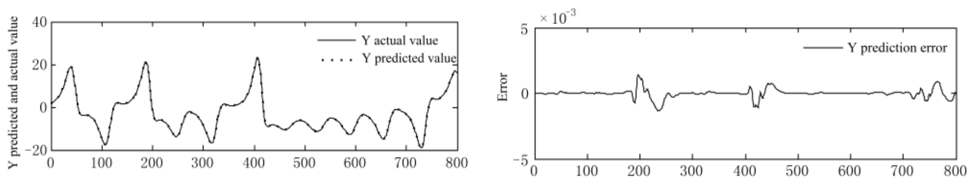


Figure 3. XYZ Variables in Lorentz Equation of the Reconstructed Phase Space of 3D Display

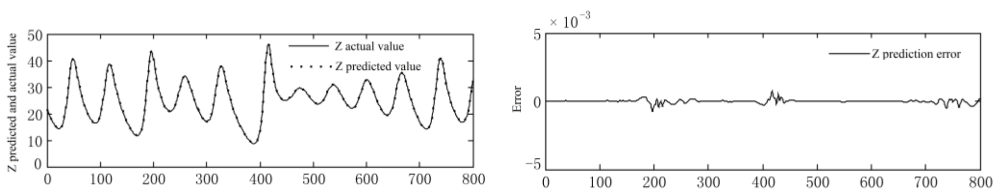
It is useful to establish forecasting model by using newb function in the RBF neural network, we take the first 1000 points reconstructed data for sample study, and make prediction based on the following 800 points reconstructed data, by comparing with the actual value, Figure 4 shows the one step prediction effect diagram based on three variable phase space reconstruction and neural network model.



(a)one step prediction and error curve of x sequence



(b) one step prediction and error curve of y sequence



(c) x one step prediction and error curve of z sequence

Figure 4. One Step Prediction Diagrams of XYZ Variable in Lorentz Equation

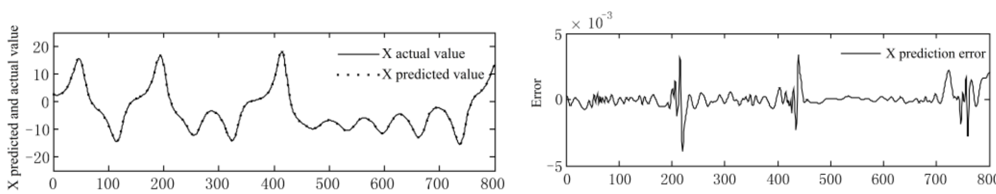
Since the error of one step prediction is 10^{-3} order of magnitude, it is difficult to distinguish the error curve represented by the actual value and the error curve represented by the prediction value. Suppose we have a request for the number of the variable, or there are too many variables, so as to go against our analysis, we can utilize the correlation criterion, introduced by this section, to select the variable. Here take the typical complex system – three variable time series as an example, assume the variable to be studied is X, in reality, and we can only choose one variable with it to form multi input to participate reconstruction. By calculating its linear correlation, its value is $\eta(x,y)=0.3675$, $\eta(x,z)=0.6834$, $\eta(y,z)=0.0129$. Since the nonlinear effect in each component is the characteristic of the complex system, it is significant that only use linear correlation to calculate the result will make a wrong conclusion. Weak linear correlation does not mean the un-correlation exists between two variables, it can only indicate there is no linear relationship exists between them. Therefore, we need an index of nonlinear correlation degree to measure the nonlinear correlation between different variables. Table 1 shows the results of nonlinear correlation between X, Y, Z variable.

Table 1. Correlation of Three Variable in Equation Lorentz

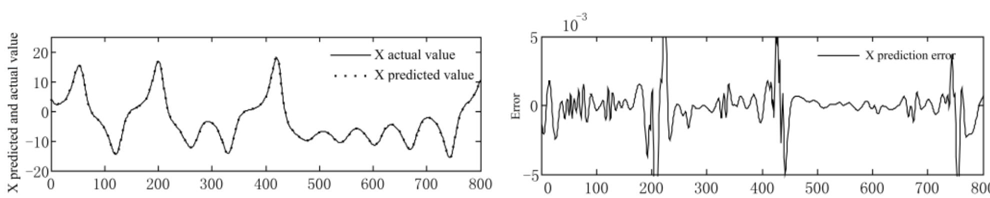
m	5		8	
r_1	0.12	0.16	0.12	0.16
r_2	0.06	0.08	0.06	0.08
$R(x,y,r_1,r_2)$	0.63702	0.65532	0.64568	0.66706
$R(x,z,r_1,r_2)$	0.49261	0.48621	0.50325	0.48751
$R(y,z,r_1,r_2)$	0.60719	0.62822	0.62402	0.63935

From Figure 1, the nonlinear correlation between XY variables is greater than the nonlinear correlation between XZ variables, with limited number of input variable, it is better to select XY variable as the multivariate input to participate phase space reconstruction. Then, using C – C method to calculate its reconstructed parameters and bring them into reconstruction equation to complete the multivariate phase space reconstruction. The reconstructed phased space data of combination XY and XZ is used for single – step prediction of X variable, and its result is shown in Figure 5.

The simulation in both Figure 4 and Figure 5 chooses x, y, z coordinates of the Lorentz equation as the original time series, and perform multivariate phase space reconstruction by using C-C method to obtain embedding dimension and delay time.



(a)Prediction result of combination XY for X variable



(b)Prediction result of combination XZ for X variable

Figure 5. XY and XZ Combination Prediction for X Variable

5. Conclusions

When there are limit number of variables involved in the reconstruction of complex system, or there are too many variables that need us to give up some variables, by calculating the nonlinear correlation between the variables to be studied and the existing variables, we can choose the variable with strong non-correlation as the input vector in the reconstruction. Obviously, by using C – C method to perform multivariate combination prediction simulation for the time series in Lorentz equation and compare the phase diagram of multivariate phase space reconstruction, it is able to confirm the correctness of selecting the reconstructed input vector based on nonlinear correlation as well as the effectiveness of using C-C method to identify reconstruction parameter.

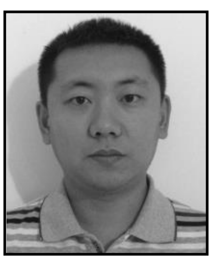
Acknowledgements

This work was supported by The Education Department of Heilongjiang province science and technology research projects (Grant Nos. 12531160).

References

[1] S. Li and X. Tian, "Nonlinear science and complex science", Harbin: Harbin Institute of Technology Press, **(2006)**.
[2] W. Hongli, "Selection of Optimum Successive State Based on Cloud Model in Simulation Reasoning of Complex System Using QSIM", International Journal of Hybrid Information Technology, vol. 6, no. 177, **(2013)**.
[3] Q. Y. Dai, J. D. Wang, R. T. Liu, J. H. Cheng, Y. li and W. Y. Wu, "Inverter Power Monitoring System Based on VRML Virtual Reality Technology", International Journal of Smart Home, vol. 7, no. 187, **(2013)**.
[4] S. Singh and I. Chana, "Introducing Agility in Cloud Based Software Development through ASD", International Journal of u- and e- Service, Science and Technology, vol. 6, no. 191, **(2013)**.
[5] H. Y. Wang, Z. H. Sheng and J. Zhang, "Phased Space Reconstruction of Complex System based on Multivariate Time Series", Journal of Southeast Unviersity, vol. 33, no. 1, **(2003)**.
[6] Y. H. Yue, W. X. Han and G. P. Cheng, "Determination of parameters in the phase space reconstruction of multivariate time series", Control and Decision, vol. 20, **(2005)**, pp. 290.
[7] T. Chen, "Reconstruction of multivariate time series and prediction method", Statistics and Decision, vol. 14, no. 31, **(2009)**.
[8] X. D. Zhang. X. L. Zhu and Z. Bao, "The blind signal separation based on phased study", Science in China (Series E), vol. 32, no. 5, **(2002)**.
[9] Z. B. Lu, Z. M. Cai and K. Y. Jiang, "Determination of Embedding Parameters for Phase Space Reconstruction Based on Improved C-C Method", Journal of System Simulation, vol. 19, **(2007)**, pp. 2527.

Author



Jianming Sun, obtained his Ph.D. at Harbin Engineering University, and the author is mainly engaged in the research of machine vision and automatic control.

