Vol 16 No 11, November 2007
1009-1963/2007/16(11)/3262-09

Chinese Physics

© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

Prediction of chaotic time series based on modified minimax probability machine regression*

Sun Jian-Cheng(孙建成)†

School of Electronics, Jiangxi University of Finance and Economics, Nanchang 330013*, China*

(Received 27 June 2007; revised manuscript received 31 July 2007)

Long-term prediction of chaotic time series is very difficult, for the chaos restricts predictability. In this paper a new method is studied to model and predict chaotic time series based on minimax probability machine regression (MPMR). Since the positive global Lyapunov exponents lead the errors to increase exponentially in modelling the chaotic time series, a weighted term is introduced to compensate a cost function. Using mean square error (MSE) and absolute error (AE) as a criterion, simulation results show that the proposed method is more effective and accurate for multistep prediction. It can identify the system characteristics quite well and provide a new way to make long-term predictions of the chaotic time series.

Keywords: minimax probability machine regression (MPMR), time series, prediction, chaos
PACC: 0545

1. Introduction

Prediction for chaotic time series is just to approximate the unknown nonlinear functional mapping of a chaotic signal. The potential in short-term prediction in chaos based models have been widely appreciated and the prediction accuracy of such models has been improved by various techniques. For instance, some traditional methods associated with time series prediction have been proposed such as linear regression and auto regressive integrated moving average (ARIMA) models.[1,2] Nevertheless, in recent years various innovative approaches, such as time-delay embedding,[3,4] wavelets,[5] and global dynamical models,[6] have been proposed to solve the prediction problems associated with nonlinear time series. Furthermore, the more recent studies, on the other hand, place much emphasis on machine learning methods such as general neural networks,[7,8] radial basis function network,[9,10] Gaussian process[11] and support vector machines (SVM).[12−14]

In practice, the training set in conventional approaches, behaving in a nonlinear autoregressive model with exogenous input (NARX) or autonomous case, is involved with the theory of reconstructed embedding phase space in a sense. So the embedding parameters must be determined in advance. However, the reasonable explanation to the choice of embedding parameters is not given in some classical methods. In addition, the sensitivity to initial condition of chaotic system introduced by the largest Lyapunov exponent, namely 'butterfly effect', must be taken into account to solve the prediction problem, for this effect can induce some additional prediction errors.

In this paper, we pay attention to a novel approach to solving the prediction problem by using the minimax probability machine regression (MPMR) method, which has been successfully applied to the classification and regression problem.[15,16] Compared with traditional general approaches such as Gaussian process, the MPMR does not need to make any assumptions of the data distribution which are often invalid and lacking in generality. Especially, the most attractive property of MPMR is that it can explicitly provide a low bound to the probability with which the regression model is within an assigned error of the true regression function when the mean and covariance matrix of the data is known.[16] In addition, we expect the errors resulting from the later iteration to have a less influence by introducing a compensatory component since the largest Lyapunov exponent might introduce the exponential distortion into a chaotic time series.

The remainder of this paper is organized as follows. In Section 2 the theory of reconstruction of phase space is described. In Section 3 the problem

*Project supported by the National Natural Science Foundation of China (Grant No 60602034) and the Natural Science Foundation of Jiangxi Province, China (Grant No 0611031).
†E-mail: Sunjc73@gmail.com

of chaotic time series prediction is considered by using the MPMR. In Section 4, the prediction simulated by using MPMR is described. Finally, the summary and conclusion are given in Section 5.

2. Reconstruction of phase space

2.1. Time series and dynamical system

In common cases, a discrete scalar time series $\{x_i, i = 1, 2, \cdots, N\}$ is observed in a chaotic dynamical system. According to the Takens' embedding theorem, the phase space can be reconstructed with the appropriate embedding dimension m and the time delay τ, and the point (reconstructed vector) in the reconstructed phase space can be written as

$$\boldsymbol{x}_i = \left[x_{i-(m-1)\tau}, x_{i-(m-2)\tau}, \cdots, x_i\right]. \tag{1}$$

For simplicity, we assume $\tau = 1$ in the following discussion. According to Takens' embedding theorem, a map $F : \boldsymbol{R}^{m+1} \mapsto \boldsymbol{R}^{m+1}$ exists that transforms the current reconstructed state point $\boldsymbol{x}_i$ to the next state point $\boldsymbol{x}_{i+1}$, i.e.

$$\boldsymbol{x}_{i+1} = F\left(\boldsymbol{x}_i\right). \tag{2}$$

Furthermore, the Takens's embedding theorem ensures that such a smooth map $\boldsymbol{f} : \boldsymbol{R}^m \to \boldsymbol{R}$ exists that

$$x_{i+1} = f\left(\boldsymbol{x}_i\right). \tag{3}$$

In Eqs.(2) and (3), the function $F(\cdot)$ represents the state change of dynamical system; $\boldsymbol{x}_{i+1}$ is the observed value at time $i+1$; $f(\cdot)$ denotes the observation function.

2.2. Choice of embedding parameters for prediction

Reconstruction of a phase space by using the time delay coordinate method needs to estimate the phase space parameters including embedding dimension m and time delay τ. Varieties of heuristic techniques of estimating the embedding parameters have been proposed. In this paper, we pay attention to the most representative ones. For the estimation of the embedding dimension, the most commonly used technique is the false nearest neighbour method.[17] For the estimation of delay time, average mutual information[18] at a varying sample rate is computed and the first minimum is taken as the appropriate sample rate.

In practical application, we have found that the embedding parameters estimated by the above methods are not the optimal values with respect to the prediction performance. Thus, we choose the optimal embedding parameters by utilizing the cross-validation technique in a neighbouring region of the value estimated by the methods in Refs.[8,15].

3. Proposed algorithm for multi-step prediction

3.1. Regression model

Equation (3) is actually the estimated nonlinear autoregressive model of the input time series. The objective of modelling the time series is to find a true functional form f^* based on the training set pairs $[\boldsymbol{x}_i, y_i], i = 1, 2, \cdots, N$, where $y_i = x_{i+1}$ is the output of the model, and $\boldsymbol{x}_i$ is the input of the model. So the relationship between $\boldsymbol{x}_i$ and y_i is shown as follows:

$$y_i = f^*\left(\boldsymbol{x}_i\right) + \rho_i, \tag{4}$$

where ρ_i represents the noise or fitting error with an expected value of $E[\rho] = \bar{\rho} = 0$ and some finite variance $\mathrm{Var}[\rho] = \sigma_\rho^2$.

In this case, given a hypothesis space $\boldsymbol{H}$ of functions from $\boldsymbol{R}^m$ to $\boldsymbol{R}$, the strategy of MPMR is to find a model $\hat{f} \in \boldsymbol{H}$ that maximizes the minimum probability with an accuracy of $\pm\varepsilon$. The fitting error from the point of view of probability is defined as follows:

$$\Omega_f = \inf_{(x,y)\sim(\bar{x},\bar{y},\Sigma)} \Pr\left\{|f\left(\boldsymbol{x}\right) - y| < \varepsilon\right\}, \tag{5}$$

where $\Sigma = \mathrm{cov}\left(\boldsymbol{x}_i, y_i\right) = \mathrm{cov}(x_{i-(m-1)}, x_{i-(m-2)}, \cdots, x_i, x_{i+1})$.

$\Sigma \in \boldsymbol{R}^{(m+1)\times(m+1)}$ is the covariance matrix of the random vector $\boldsymbol{x}_i$ and the random variable $y_i = f^*\left(\boldsymbol{x}_i\right) + \rho_i$. Consequently, the model $\hat{f}$ that we are looking for must satisfy

$$\Omega_{\hat{f}} \geq \Omega_f \ \text{ for all } f \in \boldsymbol{H}.$$

In the linear model case, $\boldsymbol{H}$ contains all the functions that are linear combinations of the input vectors:

$$f\left(x\right) = \sum_{i=1}^{m} \beta_i x_i + \beta_0 = \beta^{\mathrm{T}} x + \beta_0, \tag{6}$$

where $\beta \in \boldsymbol{R}^m$, and $\beta_0 \in \boldsymbol{R}$.

For a chaotic system, the effect of Lyapunov exponent on the fitting error should be taken into account. The Lyapunov exponent measures the rate at which nearby orbits converge or diverge. If it is positive, then the distance between nearby orbits grows exponentially in time, and the system exhibits sensitive dependence on initial conditions. In Fig.1(a), θ_0 is the initial distance between two nearby points; as the dynamical system evolves, it will turn into θ_i. The relationship between θ_i and θ_0 is shown as follows:

$$\theta_i = \theta_0 \mathrm{e}^{\lambda_{\max} \mathrm{i} \Delta t}, \tag{7}$$

where $\lambda_{\max}$ is the largest Lyapunov exponent and the Δt sampling interval. We introduce a compensatory component for weighting the fitting error resulting from the largest Lyapunov exponent. Therefore, Eq.(5) can be changed into the following form:

$$\Omega_f = \inf_{(x,y)\sim(\bar{x},\bar{y},\Sigma)} \Pr\left\{|f(\boldsymbol{x}) - y| < \varepsilon p\right\}, \tag{8}$$

where p is the compensatory variable and $p(i) = \left(\mathrm{e}^{\lambda_{\max}\Delta t}\right)^{-(i-m-1)}$, $i = m+1, m+2, \cdots, N$. With this weighting variable, the errors for later iteration are given to be less credible due to the divergence of trajectories, and a small error is magnified proportionally to the largest Lyapunov exponent. Assuming $f(x) = x$, Fig.1(b) shows the prediction accurate within $\pm\varepsilon * p$.

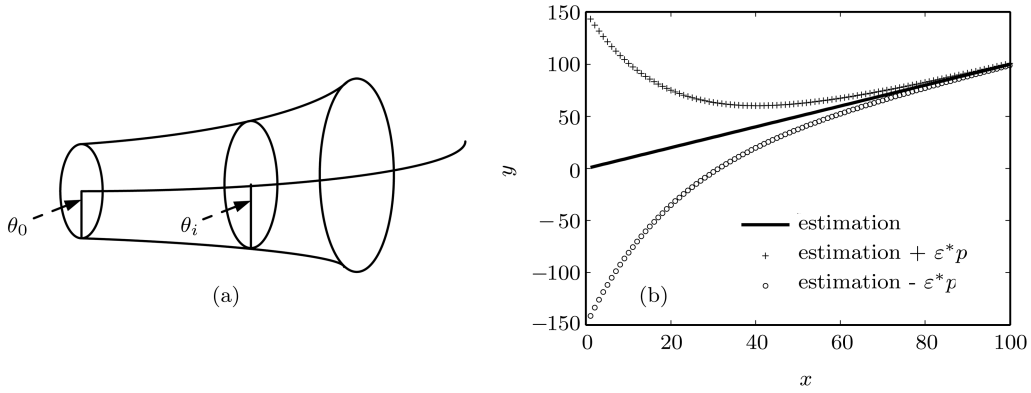


Fig.1. Effect of largest Lyapunov exponent. (a) is for the evolution of dynamical system, and (b) for the prediction accurate within $\pm\varepsilon * p$.

In the nonlinear model case, the general basis function is as follows:

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \beta_i \varphi_i(\boldsymbol{x}) + \beta_0. \tag{9}$$

Similar to the SVM, the kernelized version of the MPMR can also be constructed by mapping the input vectors into a high-dimensional feature space $\boldsymbol{R}^{\mathrm{h}}$, where the linear regression in $\boldsymbol{R}^{\mathrm{h}}$ corresponds to a nonlinear regression in the original input space. According to the mapping function $\varphi_i(\boldsymbol{x}) = K(\boldsymbol{x}_i, \boldsymbol{x})$, where $K : \boldsymbol{R}^m \times \boldsymbol{R}^m \mapsto \boldsymbol{R}$ is the kernel representation, Eq.(9) can be rewritten as

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \beta_i K_i(\boldsymbol{x}_i, \boldsymbol{x}) + \beta_0. \tag{10}$$

3.2. Prediction with MPMR

According to the above mentioned regression model, the MPMR can be realized by utilizing the minimax probability machine classification (MPMC). Before the modified MPMR is proposed, we give an overview on the MPMC. Let $\{\boldsymbol{u}_i\}_{i=1}^{N_u}$ and $\{\boldsymbol{v}_i\}_{i=1}^{N_v}$ represent the training samples for the class $\boldsymbol{u} \in \boldsymbol{R}^m$ and class $\boldsymbol{v} \in \boldsymbol{R}^m$ respectively. We assume that the point $\boldsymbol{u}_i$ is generated from a bounded distribution that has a mean $\bar{\boldsymbol{u}}$ and a covariance matrix $\Sigma_{\boldsymbol{u}}$ which is written as $\boldsymbol{u} \sim (\bar{\boldsymbol{u}}, \Sigma_{\boldsymbol{u}})$. Similarly, $\boldsymbol{v} \sim (\bar{\boldsymbol{v}}, \Sigma_{\boldsymbol{v}})$ is defined. For the nonlinear case, input data are mapped into a high-dimensional space by using the mapping function $\varphi : \boldsymbol{R}^m \mapsto \boldsymbol{R}^{\mathrm{h}}$ as follows: $\boldsymbol{u} \mapsto \varphi(\boldsymbol{u}) \sim \left(\bar{\varphi}(\boldsymbol{u}), \Sigma_{\varphi(\boldsymbol{u})}\right)$ and $\boldsymbol{v} \mapsto \varphi(\boldsymbol{v}) \sim \left(\bar{\varphi}(\boldsymbol{v}), \Sigma_{\varphi(\boldsymbol{v})}\right)$. Consequently, the decision hyperplane in $\boldsymbol{R}^{\mathrm{h}}$ is then given by $\boldsymbol{a}^{\mathrm{T}}\varphi(\boldsymbol{z}) = b$ with $\boldsymbol{a}, \varphi(\boldsymbol{z}) \in \boldsymbol{R}^{\mathrm{h}}$ and $b \in \boldsymbol{R}$. The following optimization problem needs solving in $\boldsymbol{R}^{\mathrm{h}}$:

$$\min_{a} \sqrt{\boldsymbol{a}^{\mathrm{T}} \Sigma_{\varphi(\boldsymbol{u})} \boldsymbol{a}} + \sqrt{\boldsymbol{a}^{\mathrm{T}} \Sigma_{\varphi(\boldsymbol{v})} \boldsymbol{a}} \text{ s.t. } \boldsymbol{a}^{\mathrm{T}} \left(\bar{\varphi}(\boldsymbol{u}) - \bar{\varphi}(\boldsymbol{v})\right) = 1. \tag{11}$$

According to the kernel trick method, we can

avoid using φ explicitly. Thus the kernel version of decision hyperplane for binary classifier can be derived as the following form:

$$c = \text{sign}\left[\sum_{i=1}^{N_u+N_v} \gamma_i K(\boldsymbol{z}_i, \boldsymbol{z}) + b\right], \tag{12}$$

where $K(\boldsymbol{z}_i, \boldsymbol{z}) = \varphi(\boldsymbol{z}_i)\varphi(\boldsymbol{z})$ and γ_i and b are obtained by solving the following optimization problem:

$$\min_{\gamma}\left\{\left\|\frac{\boldsymbol{K_u}}{\sqrt{N_u}}\gamma\right\|_2 + \left\|\frac{\tilde{\boldsymbol{K}}_{\boldsymbol{v}}}{\sqrt{N_v}}\gamma\right\|_2\right\} \text{ s.t. } \gamma^{\mathrm{T}}\left(\tilde{\boldsymbol{k}}_{\boldsymbol{u}} - \tilde{\boldsymbol{k}}_{\boldsymbol{v}}\right) \tag{13}$$

with $\tilde{\boldsymbol{K}}$ and $\tilde{\boldsymbol{k}}$ being the parameters for the Gram matrix and kernel function respectively.

For the sake of using the above mentioned MPMC method to formulate our modified MPMR, two classes of points $\boldsymbol{u} \in \boldsymbol{R}^m$ and $\boldsymbol{v} \in \boldsymbol{R}^m$ must be created from the original learning data set. Namely, original scalar chaotic time series data point $[\boldsymbol{x}(i), y(i)]$ for $i = 1, 2, \cdots, N$ in Eq.(3) turns into two $(m+1)$-dimensional vectors by shifting y value by $\pm\varepsilon p$ as expressed in the following form:

$$\begin{aligned}\boldsymbol{u}_i &= \left(y_i + \varepsilon p, x_{i-(m-1)}, x_{i-(m-2)}, \cdots, x_i\right)\\ &\qquad i = 1, \cdots, N,\\ \boldsymbol{v}_i &= \left(y_i - \varepsilon p, x_{i-(m-1)}, x_{i-(m-2)}, \cdots, x_i\right)\\ &\qquad i = 1, \cdots, N,\end{aligned}$$

where p has the same meaning as that defined in Eq.(8).

With these two classes of training points set, we can obtain γ and b by solving Eq.(13). Thus, with reference to Eq.(12), the classification boundary of the binary MPM classifier is given by

$$\sum_{i=1}^{2N} \gamma_i K(\boldsymbol{z}_i, \boldsymbol{z}) + b = 0. \tag{14}$$

Here this classification boundary is interpreted as a regression surface that is accurate within $\pm\varepsilon p$ (see Fig.1(b)).

From Eq.(3), we can see that the resulting dynamic model $f(\cdot)$ is obtained for the single step prediction. For the multistep prediction, the prediction time series is generated by regarding the prediction one, obtained by single step prediction, as a new input to realize the iterative procedure. This procedure can be formulated as follows:

$$\begin{aligned}\tilde{x}_{i+1} &= f(\hat{\boldsymbol{x}}_i)\\ &= f\left(\hat{x}_{i-(m-1)\tau}, \hat{x}_{i-(m-2)\tau}, \cdots, \hat{x}_i\right)\end{aligned} \tag{15}$$

with

$$\hat{x}_i = \begin{cases} x_i, & 0 \le i \le m,\\ f\left(x_{i-(m-1)\tau}, x_{i-(m-2)\tau}, \cdots, x_i\right) & i > m.\end{cases}$$

Equation (15) states that $\tilde{x}_{i+1}$ is the $(i-m)$th iteration of the predictive part of the map (for $i > m$), i.e.

$$\tilde{x}_{i+1} = f(f(\cdots f(\boldsymbol{x}_m))) = (f(\boldsymbol{x}_m))^{i-m}. \tag{16}$$

4. Simulation and discussion

In this section, we present the application of our algorithm to a dynamical system (Section 4.1) and a set of experimental data (Section 4.2). The test systems that we choose are the Rössler systems with and without the noise. We then describe the application of this method to experimental recordings of the chaotic laser data utilized in the 1992 Santa Fe time series competition.[19]

The prediction is different from that from the dynamic modelling of chaotic time series since the former addresses only the instantaneous prediction error, while the latter is interested in the long-term behaviour. In order to quantitatively investigate the performance of the algorithm, absolute error (AE) and mean square error (MSE) are computed from the following formulas.

$$E_{\text{abserr}}(k) = \hat{x}_i - x_i, \tag{17}$$

$$E_{\text{MSE}} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{k=1}^{N}|\hat{x}_{i,k} - x_{i,k},|^2, \tag{18}$$

where $\hat{x}_i$ and x_i are the evaluated point and true point, respectively; M is the number of independent experiments; N is the number of available samples in each experiment.

4.1. Dynamical equations

The Rössler equations are described as follows:

$$\begin{aligned}\dot{x} &= -y - z,\\ \dot{y} &= x + ay,\\ \dot{z} &= b + z(x - c).\end{aligned} \tag{19}$$

A Rössler attractor is generated by taking $a = 0.45$, $b = 2.0$, and $c = 4.0$. A trajectory has been generated for initial condition $[0.1; -0.1; 0.02]$ by using a chaos function of TSTOOL toolbox (this toolbox can be free to download from

www.physik3.gwdg.de/tstool/indexed.html). After the transients have died down to a time series of 2000 points, the time series is divided into two subsets referred to as the training subset and the test subset. We use the points 1–700 for training, while the points 701–2000 for testing. For estimating the performance of modified MPMR, we compare the simulation results from the classical local weighted method (LWP) and least squares support vector machines (LS-SVM).

The choosing of embedding parameters is a key problem for prediction. As mentioned in Section 2, the most commonly used techniques are the false nearest neighbour method and average mutual information method. However, we find that the values estimated by using these methods are not optimal to our prediction problem. In order to acquire the better performance, we utilize the cross-validation technique by changing the combination (m, τ) within a range of $[10, 30] \times [1, 10]$ for x-component data of the system. With respect to our MPMR method, the relationship between MSE and (m, τ) can be obtained, and it is shown in Fig.2. Therefore, the optimal value $(20, 1)$ for (m, τ) is chosen in the sense of minimal MSE. Similarly, $(60, 1)$ is chosen for the y-component and z-component data.

We first investigate the performance under the noise-free condition. Owing to the sightlessness of high-dimensional space, we can only draw the 3D phase space portrait for the Rössler system to estimate the performance intuitively. In Fig.3, the original attractor (Fig.3(a)) is compared with the reconstructed one by using the MPMR method. The similarity between Fig.3(a) and Fig.3(b) suggests that the two time series represent two trajectories lying on the same attractor. With the development of prediction lead time, there appears the difference between Fig.3(a) and Fig.3(b), and the prediction accuracy is reduced, which implies that the chaotic system is sensitive to the initial condition, namely, butterfly effect.

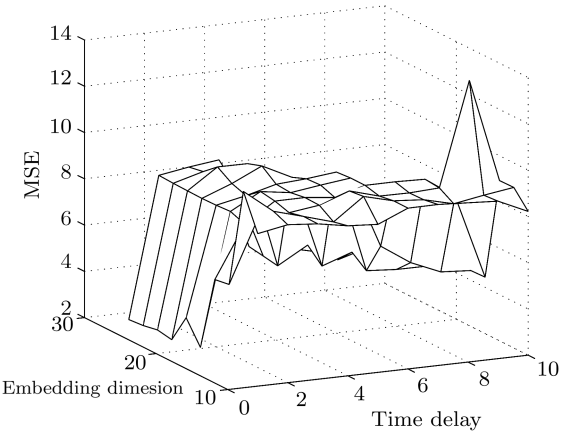


Fig.2. MSE versus embedding dimension and time delay.

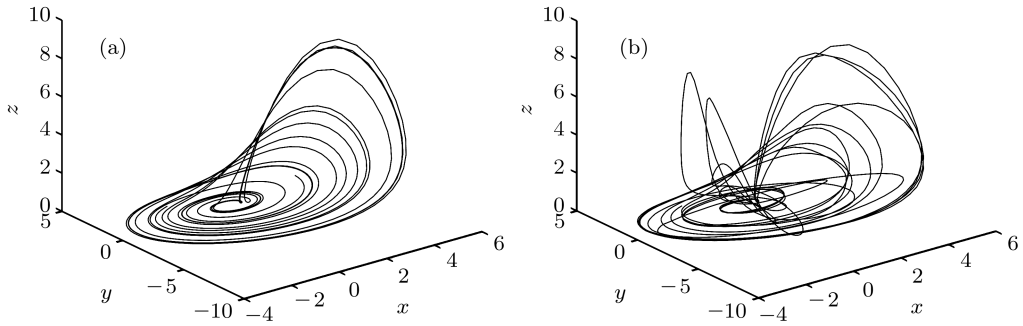


Fig.3. Reconstructed phase space portrait of Rössler time series: (a) the origin time series and (b) prediction time series of MPMR.

For the x-component data of Rössler system, the prediction performances of our MPMR, LS-SVM and LWP methods are shown in Figs.4(a)–4(c) under the noise-free condition, respectively.

We can see that about 400 anterior steps can be predicted accurately for MPMR and LWP methods. Specially, LS-SVM method can predict 400 steps perfectly. With the prediction time growing, the prediction error increases accordingly. When prediction time is greater than 400 steps, the prediction errors from the LS-SVM and LWP methods increase remarkably. At the same time, better and more acceptable results for predicting about 1000 steps can be reached by using the MPMR method as shown in Fig.4(a). The

error mainly comes from two sources. One is the essential character of the chaotic system, that is, the chaotic time series that can be predicted only in a short range due to the 'Butterfly Effect'. The other, obviously, is the algorithm in which the system can be learnt accurately only within a certain principle, and the prediction errors will be cumulated with prediction step increasing.

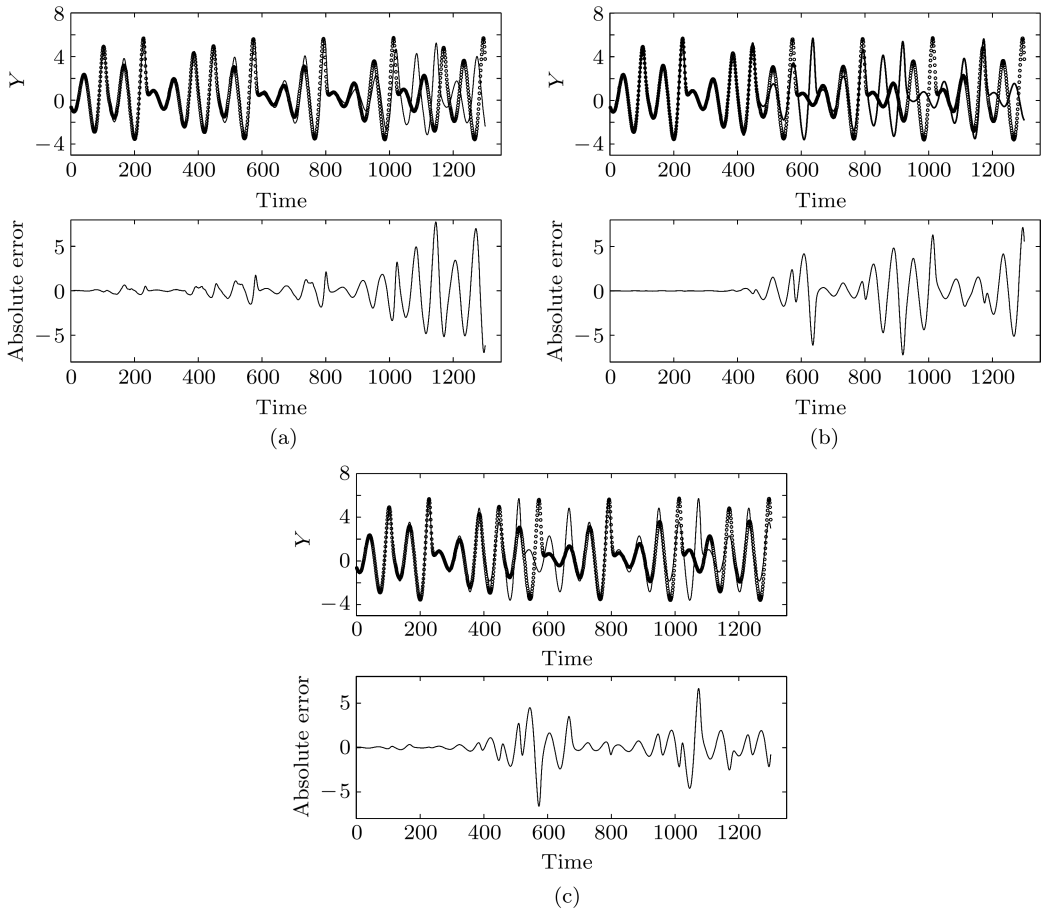


Fig.4. Comparison among the predicted errors of x-component data of Rössler system in noise-free condition, from the MPMR (a), LS-SVM (b), and LWP methods (c), where the circle and continuous line denote the true and training points respectively.

The E_{MSE} can be used to describe how well the model accounts for the variation in the observed data. Under the different initial conditions for a Rössler system, we have carried out eight independent experiments. Figure 5 shows the E_{MSE} versus the prediction time. For 600 steps ahead prediction, the MSE from MPMR method is larger than those from the LS-SVM and LWP methods. Therefore, for a short time prediction (600 steps), the performances of LS-SVM and LWP methods are better. Their improvement is not remarkable, which can be seen in Fig.4(a). In addition, the performance of MPMR method can still be accepted. For the larger prediction lead time case (larger than 600 steps), obviously, the prediction accuracy of the MPMR method is better than those of LS-SVM and LWP methods.

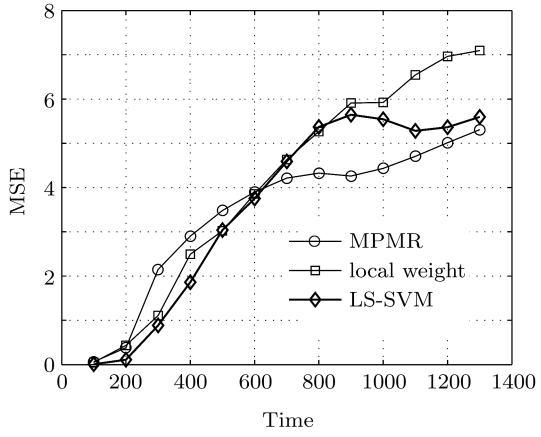


Fig.5. MSE versus lead time from the MPMR, LS-SVM, and LWP methods.

In order to evaluate the prediction performance of our method in the noise case, simulation is carried out by adding a white noise with a Gaussian distribution to a clean time series (SNR = 20 dB). Figures 6(a)–6(c) show the simulation results from the MPMR, LS-SVM and LWP methods, respectively. From these results, we can find that the predicted performances of these methods degrade due to the presence of the noise. Only about 100 steps can be predicted for the MPMR and LS-SVM methods. Compared with the MPMR and LS-SVM methods, the LWP method can still predict about 400 steps despite the higher error existing in the early 100 steps (though the instantaneous prediction error increases, the shape and trend of time series can be accepted). This phenomenon is due to the filtering process present in the LWP method, namely, a weighted average operation. In the MPMR method, it is assumed that the mean and covariance matrix of the distribution that generates the regression data is estimated accurately in advance. However, the estimated error of the mean and covariance matrix increases due to the introducing of the noise that leads the prediction performance to degrade. So we believe that the performance of the MPMR method can be improved by utilizing the filter technique and effective method of estimating the mean and covariance matrix. We must point out that we do not develop the MPMR method in the context of noise in this paper. In fact, the similar research has been done by Strohmann and Grudic[20] who proposed a robust MPMR method to estimate the mean and covariance matrix effectively.

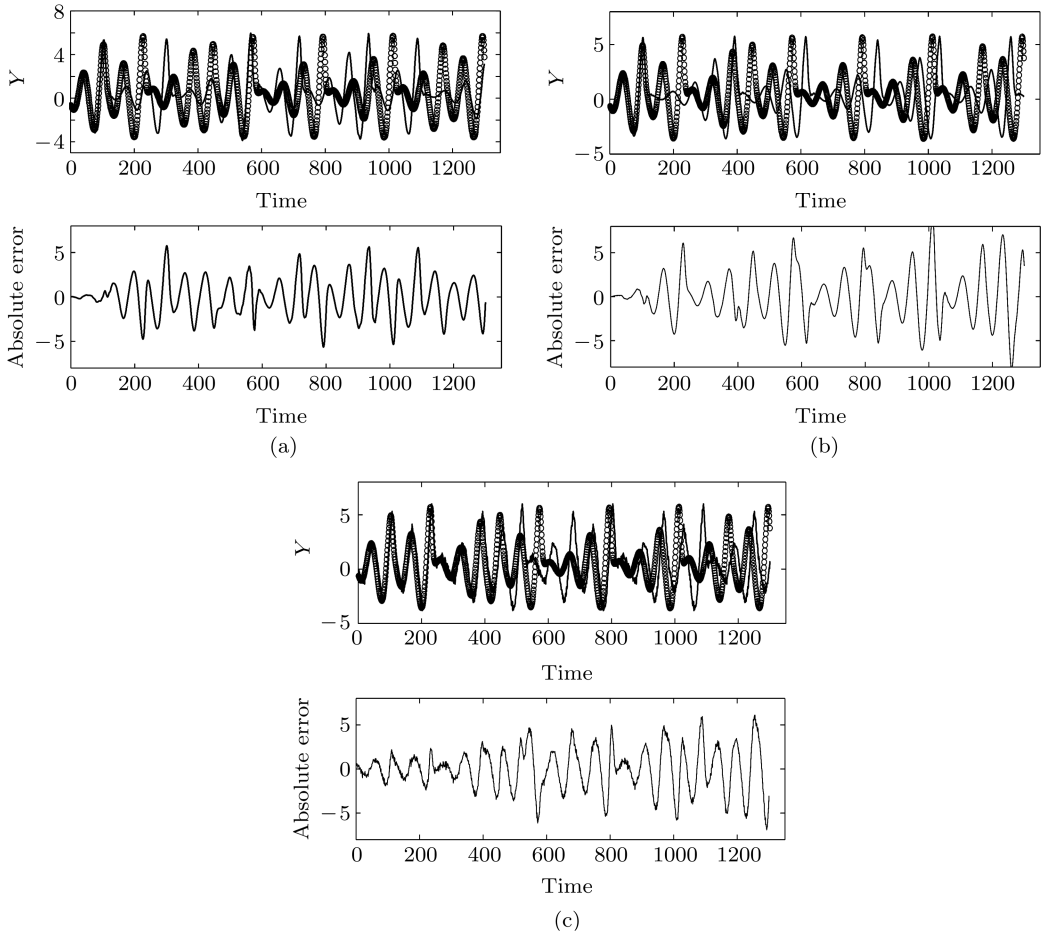


Fig.6. Comparison among the predicted errors of x-component data of Rössler system in the context of noise (SNR = 20 dB) from MPMR (a), LS-SVM (b), and LWP method (c), where the circle and continuous line denote the true and training points respectively.

4.2. Experimental Data

Experimental data, such as of the chaotic laser data and electrocardiographic (ECG) data that are natural phenomena are more complicated than those of the above dynamical system. Here experimental recordings of the chaotic laser data are utilized in the 1992 Santa Fe time series competition. We have chosen the 1100 data points, of which 750 points are used as a training set and the rest as the testing data. The optimal value $(18, 1)$ for (m, τ) is chosen by using a similar method to that described in Section 4.1.

Figures 7(a)–7(c) show the results from the MPMR, LS-SVM and LWP methods, respectively. Compared with LS-SVM and LWP methods, the MPMR method manifests an outstanding prediction ability to predict about 240 steps accurately.

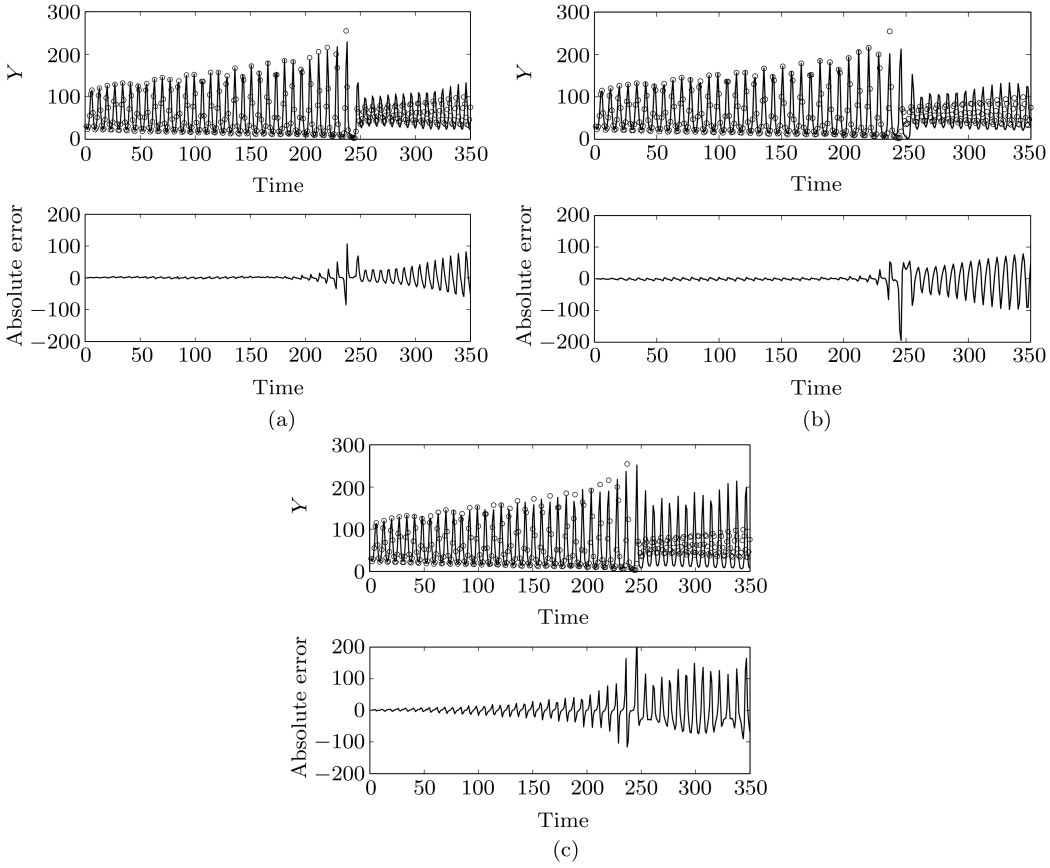


Fig.7. Comparison among the predicted errors of Santa Fe time series in noise free condition, from MPMR (a), LS-SVM (b), and LWP methods (c), where the circle and continuous line denote the true and training points respectively.

5. Conclusion

In this paper we have described a new method to model and predict chaotic time series, based on the MPMR method. And we have investigated the performance of the MPMR method as a global model in chaotic time series predictions compared with the widely used local weighted prediction (LWP) and least squares support vector machines (LS-SVM). To gain a more general and robust conclusions, the analysis is conducted on a chaotic Rössler series and Santa Fe time series competition. The prediction is made by examining trajectories in the reconstructed phase space. The simulation result shows that the MPMR method can learn the chaotic system effectively. Using the absolute error and mean square error as a criterion, the MPMR method exhibits a better prediction performance than LWP and LS-SVM prediction methods under the noise-free condition. For the short-range prediction case (less than 400 steps), the MPMR, LWP and LS-SVM methods can give a remarkably high prediction accuracy. However, the improving of the performance of MPMR method is significant in a long prediction range. Although the prediction accuracies of all the methods, reduce with prediction time developing, the prediction performance of

MPMR method is still better than those of the LWP and LS-SVM methods. For the Santa Fe time series prediction, a similar conclusion can be obtained. In the context of noise, the performances of MPMR and LS-SVM methods degrade dramatically for lack of filter operation. It should be noted that this study is primarily conducted under the noise-free condition. However, these problems could be solved if we use the filter technique and robust method to estimate the mean and covariance matrix. We can conclude that the MPMR method can yield the prediction as good as, if no better than, the widely used LWP and LS-SVM methods in the phase space prediction of chaotic time series.

Acknowledgment

The author wishes to thank Dr Zhou Ya-Tong of Hebei Technical University for introducing the theory of MPMR to me. The TSTOOL toolbox is downloadable from www.physik3.gwdg.de/tstool/ indexe.html.

References

[1] Brillinger D R 1981 *Time Series, Data Analysis and Theory* (New York: McGraw-Hill)
[2] Pandit S M and Yu S M 1983 *Time Series and System Analysis with Applications* (New York: Wiley)
[3] Packard N H, Crutchfield J P, Farmer J D and Shaw R S 1980 *Phys. Rev. Lett.* **45** 712
[4] Takens F In: 1980 *Dynamical Systems and Turbulence* (Berlin: Springer)
[5] Meyer Y and Ryan R D 1991 *Wavelets. Algorithms and Applications* (Philadephia: SIAM)
[6] Kravtsov Y A and Kadtke J B 1997 *Predictability of Complex Dynarnical Systems* (Berlin: Springer)
[7] Zhang J S and Xiao X C 2000 *Chin. Phys. Lett.* **17** 88
[8] Yang G and Meng J E 2005 *Fuzzy Sets and Systems* **150** 331
[9] Omidvar A E 2004 *Chaos, Sol. Fract.* **22** 757
[10] Cowper M R, Mulgrew B and Unsworth C P 2002 *Signal Processing* **82** 775
[11] Brahim-Belhouari S and Bermak A 2004 *Computational Statistics & Data Analysi.* **47** 705
[12] Zhang J S, Dang J L and Li H C 2007 *Acta Phys. Sin.* **56** 67 (in Chinese)
[13] Cui W Z, Zhu C C, BAO W X and Liu J H 2005 *Chin. Phys.* **14** 922
[14] Sun J C, Zhou Y T and Luo J G 2006 *Chin. Phys.* **15** 1208
[15] Lanckriet G R G, Ghaoui L E, Bhattacharyya C and Jordan M I 2002 *Journal of Machine Learning Research* **3** 555
[16] Strohmann T R and Grudic G Z In: 2003 *Advances in Neural Information Processing Systems* **15**
[17] Kennel M B 1992 *Phys. Rev.* A **45** 3403
[18] Fraser A M and Swinney H L *Phys. Rev.* 1986 **33** 1134
[19] Weigend A S and Gershenfeld N A 1994 *Time Series Prediction: Forecasting the Future and Understanding the Past* (Reading, MA: Addison-Wesley)
[20] Strohmann T R and Grudic G 2003 *Technical Report* CU-CS-952-03 (http://csel.cs.colorado.edu/ ∼strohman)