Chaos, Solitons and Fractals 98 (2017) 158–172

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Frontier

A prediction method based on wavelet transform and multiple models fusion for chaotic time series

Tian Zhongda[a,*], Li Shujiang[a], Wang Yanhong[a], Sha Yi[b]

[a] College of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China
[b] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Article history:
Received 4 June 2016
Accepted 7 March 2017
Available online 21 March 2017

Keywords:
Chaotic time series
Prediction
Wavelet transform
Multiple models
Fusion

ABSTRACT

In order to improve the prediction accuracy of chaotic time series, a prediction method based on wavelet transform and multiple models fusion is proposed. The chaotic time series is decomposed and reconstructed by wavelet transform, and approximate components and detail components are obtained. According to different characteristics of each component, least squares support vector machine (LSSVM) is used as predictive model for approximation components. At the same time, an improved free search algorithm is utilized for predictive model parameters optimization. Auto regressive integrated moving average model (ARIMA) is used as predictive model for detail components. The multiple prediction model predictive values are fusion by Gauss–Markov algorithm, the error variance of predicted results after fusion is less than the single model, the prediction accuracy is improved. The simulation results are compared through two typical chaotic time series include Lorenz time series and Mackey–Glass time series. The simulation results show that the prediction method in this paper has a better prediction.

1. Introduction

Chaos is a kind of irregular movement, which is widely existed in nature. It is a complex behavior generated by a certain nonlinear dynamical system. The characteristics of chaotic signal make it more and more important in the field of signal processing, communication, control, social economy, and biomedicine, etc. With the research of chaos theory and application technology, the modeling and prediction of chaotic system has become an important research topic in the field of information processing in recent years [1,2]. Chaotic time series prediction is an important research topic in chaotic systems. It is widely used in weather prediction [3], power load forecast [4] financial stocks forecast [5], traffic flow prediction [6], runoff forecast [7] and etc. In nature and human activities, a lot of time series are nonlinear and even chaotic, so how to accurately predict the chaotic time series is very important.

The prediction of chaotic time series can be regarded as an inverse problem in the study of dynamical systems. The positive problem is to study the various properties of the orbits in the phase space of a given nonlinear dynamical system. The inverse problem is how to construct a nonlinear mapping function to express the original system, which is a series of iterative sequences or a set of observation sequences in the given phase space. This mapping can be seen as a predictive model. Therefore, how to construct the prediction model is a key problem in chaotic time series prediction [8]. In recent years, many scholars have used many methods to model and predict the chaotic time series system. Such as linear prediction model (including AR [9], ARMA [10], ARIMA [11], etc.), neural network model (including RBF neural network [12], echo state [13], extreme learning machines [14], Elman neural network [15]), support vector machine [16], least square support vector machine [17], adaptive filter [18] and etc. At the same time, many scholars have used some combination prediction model to predict the chaotic time series [19–21].

But there is a major problem in the relevant research results at present. The research shows that the chaotic time series contains the noise signal, which will destroy the inherent dynamic characteristics of the system and reduce the prediction accuracy. Therefore, it is necessary to reduce the noise of chaotic time series. Wavelet transform is a good tool to solve the problem of noise reduction. Although some scholars have considered this problem, using wavelet transform, empirical mode decomposition and other methods to reduce the noise of chaotic time series. According to the characteristic of each component, the prediction is carried out by independent [22–25]. However, the results of each model prediction are a direct superposition, so that the prediction errors of each prediction model can be accumulated to the final prediction value, and can not further improve the prediction accuracy. Based on the above discussions, this paper proposes a novel chaotic time

* Corresponding author.
E-mail address: tianzhongda@126.com (T. Zhongda).

http://dx.doi.org/10.1016/j.chaos.2017.03.018
0960-0779/

series prediction method based on wavelet transform and multiple models fusion. Firstly, wavelet decomposition and single reconstruction are carried out for non stationary chaotic time series. The single reconstruction ensures that the signal is consistent with the length of the original signal. The original chaotic time series is filtered into a smooth approximate component sequence through wavelet decomposition and reconstruction. The approximate component reflects the trend of the original time series, so the LSSVM with good prediction ability is used to predict the nonlinear time series. At the same time, an improved free search algorithm (IFS) is used to optimize the parameters of the LSSVM prediction model. For the filtered noise sequence, the detail component has a certain random non-stationary characteristic. Therefore, this paper uses the ARIMA model to predict the detailed components. When calculating the final predictive value, this paper does not directly add the predictive value of each component. The multiple model fusion prediction is carried out by Gauss–Markov estimation algorithm. The predictive error variance after fusion is smaller the single model predictive error variance, which makes the prediction accuracy is further improved. Finally, through the simulation of two typical chaotic time series — Lorenz time series and Mackey–Glass time series show that the proposed prediction method has higher accuracy.

2. Wavelet transform

Wavelet transform uses orthogonal basis to decompose the signal [26]. Discrete wavelet transform is composed of a series of parameters.

$$c_j(k) = \langle X, \varphi_{jk}(t)\rangle, d_j(k) = \langle X, \psi_{jk}(t)\rangle, j, k \in Z \tag{1}$$

Where, $\langle *, * \rangle$ in the above equation is the inner product, $c_j(k)$ as the approximation component, $d_j(k)$ as the detail component, the scaling function $\varphi_{jk}(t)$ is obtained from the mother wavelet $\varphi(t)$ after shifting and stretching.

$$\varphi_{jk}(t) = 2^{-j/2}\varphi(2^{-j}t - k) \tag{2}$$

$\varphi_j(t)$ is a low-pass filter, the low frequency component of the input signal can be separated. Wavelet transform can decompose a signal into a set of detail components with large scale and approximation components with different small scale.

In this paper, the fast discrete orthogonal wavelet transform Mallat algorithm [20] is used for decompose and reconstruct for chaotic time series.

a_j will be considered as a chaotic time series for decomposition, according to decomposition algorithm:

$$a_{j+1} = Ha_j, \quad d_{j+1} = Gd_j, \quad j = 0, 1, 2 \cdots N \tag{3}$$

H is a low pass filter and G is a high pass filter. Through the Mallat algorithm, the original chaotic time series sequence is decomposed into approximation components with low-frequency and detail components with high-frequency. Approximate components can reflect the changing trend and characteristics of time series. The detail components reflect the dynamic factors of disturbance. The chaotic time series after decomposed can be reconstructed by Mallat algorithm, the algorithm is as follows:

$$a_{j-1} = a_jH^* + d_jG^*, j = 0, 1, 2, \cdots N \tag{4}$$

Where, H^* and G^* are dual operator of H and G. Mallat reconstruction algorithm uses two interpolations, which is the input chaotic time series between every two adjacent sequence zero. The algorithm can keep the decomposition and reconstruction of sequence length consistent.

Daubechines wavelet has very good characteristics for non-stationary time series, but the dbN wavelet with different N values has different processing effect. The greater N is, the computation time is longer. Db3 wavelet is used according to experiments and references. The decomposition level is mainly related to SNR (signal to noise ratio), when the SNR is low, the input signal is mainly noise, and then decomposition level should choose bigger. It is conducive to the separation of signal and noise. When the SNR is high, the input signal is mainly signal, then decomposition level shouldn't choose bigger. Decomposition level is too large will lead reconstruction distortion is more serious, the error will be large. In this paper, the three decomposition and reconstruction is used to ensure real-time and prediction precision [27–28].

3. The multiple model fusion prediction

3.1. ARIMA prediction model

The basic idea of ARIMA model for non-stationary time series uses several difference operation to make it become a stationary series, the number of differential is d. The ARMA model with p, q as parameters is used for modeling the stationary time series. The original time series is obtained after inverse transform. ARIMA model prediction equation with p, d, q as parameters can be expressed as:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots \theta_q\varepsilon_{t-q} \tag{5}$$

Where, y_t is the sample value of the time series, φ_t and θ_t are the model parameters, ε_t is independent normally distributed white noise.

After the time series is smoothed, the auto correlation function (ACF) and the partial correlation function (PCF) of the original time series is first calculated. For a time series y_t, there is auto covariance:

$$\gamma_k = \frac{1}{N}\sum_{j=1}^{N-k} y_k y_{t+k} \tag{6}$$

Auto correlation function:

$$\rho = \frac{\gamma_k}{\gamma_0} \tag{7}$$

Partial correlation function:

$$\begin{cases} \alpha_{11} = \rho_1 \\ \alpha_{k+1,k+1} = (\rho_{k+1} - \sum \rho_{k+1-j}\alpha_{kj}) \times (1 - \sum_{j=1}^{k} \rho_j\alpha_{kj})^{-1} \\ \alpha_{k+1,j} = \alpha_{kj} - \alpha_{k+1,k+1} \times \alpha_{k,k-j+1} \end{cases} \tag{8}$$

The model order can be determined through cutoff property of ρ_k, α_k. Parameter identification of time series can be obtained by least squares estimation. Through the parameters estimation of φ_1, φ_2, $\cdots$, φ_p, θ_1, θ_2, $\cdots$, θ_q, it makes the following equation minimum.

$$\sum_{t=1}^{N}\alpha_t^2 = \sum_{t=1}^{N}\left(\theta_q^{-1}(Z)\varphi_p(Z)\nabla^d y_t\right)^2 \tag{9}$$

Where, α_t^2 is the sum of squared residuals, Z is backward shift operator, y_t is original time series, N is the length of time series, ∇^d is d order difference operator, φ_1, φ_2, $\cdots$, φ_p,θ_1, θ_2, $\cdots$, θ_q are parameters to be estimated. The (9) shows that these parameters φ_1, φ_2, $\cdots$, φ_p, θ_1, θ_2, $\cdots$, θ_q can be estimated when the minimum sum of squared residuals are 1obtained. Then y_t can be predicted by (5) through historical data $y_{t-1}, \cdots, y_{t-p}$.

The combination of different parameters p, d and q, the parameters of the optimal model is obtained by AIC (Akaike information criterion) [29]. AIC criterion characterized as stingy principle of concrete, defined as follows:

$$\mathrm{AIC} = -2\ln L + 2n \tag{10}$$

Where, L is a maximum likelihood parameter of model, n is an independent parameter format of model.

The essence of ARIMA model is combined with the ARMA model and difference operation. It shows that any non-stationary time series by proper order difference can realize the difference stationary. ARIMA model is suitable for predict the detail components of time series after wavelet decomposed.

3.2. IFS optimized LSSVM prediction model

LSSVM through a nonlinear mapping function φ, the sample space is mapped into a high-dimensional or even infinite dimensional feature space. In this feature space, there is $y(x) = \boldsymbol{w}\varphi(x) + b$, $\boldsymbol{w}$ is the weight coefficient vector, b is a constant bias. Optimal $\boldsymbol{w}$ and b can be obtained by minimizing the objective function.

$$\min_{w,b,e} J(\boldsymbol{w}, e) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{1}{2}\gamma\sum_{k=1}^{N} {e_k}^2 \tag{11}$$

Lagrange function is established for solving above constrained optimization problem.

$$L(\boldsymbol{w}, b, e; \boldsymbol{\alpha}) = J(\boldsymbol{w}, e) - \sum_{k=1}^{N}\alpha_k\{\boldsymbol{w}^T\varphi(\boldsymbol{D}_k) + b + e_k - y_k\} \tag{12}$$

Where, a_k is a Lagrange multiplier, the Lagrange function for extreme value, and the above optimization problem can be transformed into solving the linear equations.

According to Mercer conditions, the presence of mapping function $\varphi()$ and kernel function $K()$ satisfy the following equation:

$$K\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) = \varphi(\boldsymbol{X}_i)^T\varphi(\boldsymbol{X}_j) \tag{13}$$

RBF kernel function is a typical local kernel function for time series prediction, that is

$$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \exp\left(\frac{-\left\|\boldsymbol{X}_i - \boldsymbol{X}_j\right\|^2}{2\sigma^2}\right) \tag{14}$$

Where σ^2 is the width of RBF kernel function.

Assuming the input time series is $x(k)$. $x(k)$ is transformed into the matrix form as the following (15).

$$\boldsymbol{X} = \begin{bmatrix} x(1) & x(2) & \cdots & x(m) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-m) & x(N-m+1) & \cdots & x(N-1) \end{bmatrix} \tag{15}$$

Where, N is the total length of the input time series, m is the embedding dimension, p is prediction steps. Therefore, the output matrix $\boldsymbol{Y}$ can be expressed as (16).

$$\boldsymbol{Y} = \begin{bmatrix} x(m+1) & x(m+2) & \cdots & x(m+p) \\ \vdots & \vdots & \ddots & \vdots \\ x(m+p) & x(m+p+1) & \cdots & x(m+p+p-1) \end{bmatrix} \tag{16}$$

The input matrix is generated according to (15). The output matrix is generated according to (16). The Modeling and training of LSSVM prediction model is completed. The key parameters of LSSVM algorithm are regularization parameter γ and radial basis kernel width σ^2. However, there is no standard on how to determine γ and σ^2 in theory. And at present, how to select the appropriate γ and σ^2 does not have a consistent method [30]. At the same time, this paper found time series embedding dimension m also has influence on prediction accuracy. Therefore, how to select these parameters is very important. In order to determine the optimal parameters of the prediction model, an improved free search algorithm is used to solve the parameters optimization problem.

As a new optimization method based on the population optimization, free search (FS) algorithm was proposed in 2005 [31]. The FS algorithm has excellent global search capability, has been widely used in many optimization problems [32,33]. In the literature [34], an improved free search algorithm is proposed, and the performance of the FS algorithm is improved.

Standard free search algorithm is put forward to find the maximum objective function as the optimized object. The objective function in time series prediction is needed to obtain minimum value. The standard free search algorithm pheromone updating algorithm modified as followed equations to achieve the minimum value.

$$f_{tj} = f(x_{tji}),\ f_j = \min(f_{tj}) \tag{17}$$

$$PH_j = \min(f_j)/f_j \tag{18}$$

Search radius R_{ji} is an important parameter in free search algorithm. R_{ji} decides the quality of search process. It is a constant in standard free search algorithm. If the search radius is broad, the individual search range is wide, required a longer time, and has low convergence accuracy. If the search radius is too small, it is easy to fall into local optimum. Therefore, this paper uses the following dynamic search radius method as (19). In the process of optimization search, radius decreased. Its initial value is $R_j(0) = 1$. While the step of search increased, the radius decreased.

$$R_j(t) = \begin{cases} R_{\min} + R_j(t-1) \times \lambda & R_j(t) \geq R_{\min} \\ R_{\min} & R_j(t) < R_{\min} \end{cases} \tag{19}$$

The sensitivity parameter also has a significant influence on the performance of free search algorithm. The sensitivity parameters in standard algorithm have randomness. If the sensitivity will be properly decreased in the early of algorithm, the individual can perform a global search in other neighborhood. The sensitivity will be increased in the later of algorithm, the individual can perform a local search in the optimal value of population. Then, the convergence speed of the algorithm is improved. So the modification of sensitivity parameters in this paper as follows.

$$SE_j = SE_{\min} + \Delta SE_j \tag{20}$$

$$\Delta SE_j = (SE_{\max} - SE_{\min}) \times \frac{t - G}{G} \tag{21}$$

In summary, the implementation steps of improved free search algorithm optimized LSSVM (IFS-LSSVM) prediction model can be described as follows.

Step 1: Initialization

1.1 The parameters initialization of algorithm includes population size M, maximum algebra G, search step length T. The parameters to be optimized as γ, $\sigma^2$and m. Normalize the input time series;

1.2 According to algorithm, the initial population is obtained;

1.3 Initialize searching: according to the initial value, the initial pheromone is obtained. The initial pheromone is also released. Then the initial search results are obtained.

Step 2 Searching process

2.1 According to the (20) and (21), the sensitivity is calculated;

2.2 According to algorithm, the new starting point is determined;

2.3 Recording RMSE value of the error between the predicted value and the actual value as the fitness function. Fitness function is defined as follows.

$$f_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y - \widehat{y})^2} \tag{22}$$

2.4 The pheromone is calculated according to (17). The pheromones are released according to (18). Then the search results can be obtained;
2.5 Select and retain the best individuals;
2.6 Adjust the search radius according to (19).
Step 3: Termination condition judgment. Judge the termination condition, if it satisfied, then output optimal parameters γ, σ^2 and m.

3.3. Multiple models fusion based on Gauss–Markov estimation

The components of chaotic time series after wavelet decomposition and reconstruction are predicted by ARIMA and IFS-LSSVM. In order to improve the prediction accuracy of the final prediction value, this paper has not taken the method of adding the prediction value of each prediction model directly. The final prediction accuracy is improved through the Gauss–Markov estimation for multiple model fusion.

The essence of Gauss–Markov estimation is a special case of weighted least square estimation [35]. Its function is to make the error variance of the fusion result less than the variance of the error of the single model prediction result. Algorithm is as follows.

Assuming there is L prediction models. The estimated value of the ith model is $\hat{z}_i$, the actual value is y, prediction error is v_i, variance is γ. The fusion model is

$$\hat{\boldsymbol{Z}} = \boldsymbol{H}y + \boldsymbol{V}, i = 1, 2, \cdots, L \tag{23}$$

Where $\hat{\boldsymbol{Z}} = \begin{bmatrix} \hat{z}_1 \\ \vdots \\ \hat{z}_L \end{bmatrix}$, $\boldsymbol{H} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, $\boldsymbol{V} = \begin{bmatrix} v_1 \\ \vdots \\ v_L \end{bmatrix}$

Variance matrix of error $\boldsymbol{R}$ is

$$\boldsymbol{R} = diag(\sigma_1^2, \cdots, \sigma_L^2) \tag{24}$$

Minimum performance index of estimated value $\hat{y}$ is

$$J = (\hat{\boldsymbol{Z}} - \boldsymbol{H}\hat{y})^T \boldsymbol{W} (\hat{\boldsymbol{Z}} - \boldsymbol{H}\hat{y}) \tag{25}$$

$\boldsymbol{W}$is a symmetric positive definite weight matrix with $L \times L$ order. Let

$$\frac{\partial J}{\partial \hat{y}} = -2\boldsymbol{H}^T \boldsymbol{W} (\hat{\boldsymbol{Z}} - \boldsymbol{H}\hat{y}) = 0 \tag{26}$$

The estimated value $\hat{y}$ is calculated as

$$\hat{y} = (\boldsymbol{H}^T \boldsymbol{W} H)^{-1} \boldsymbol{H}^T \boldsymbol{W} \hat{Z} \tag{27}$$

The $\hat{y}$ is Gauss–Markov estimated value when $\boldsymbol{W} = \boldsymbol{R}^{-1}$ in (27), that is

$$\hat{y} = (\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{R}^{-1} \hat{\boldsymbol{Z}} = \left[\sum_{i=1}^{L} \frac{1}{\sigma^2} \right]^{-1} \left[\sum_{i=1}^{L} \frac{1}{\sigma_i^2} \hat{z}_i \right] \tag{28}$$

It can prove the error variance of Gauss–Markov estimated value $\hat{y}$ is less than the error variance of any single model

The error variance matrix of Gauss–Markov estimated value $\hat{y}$ is

$$P_{GM} = (\boldsymbol{H}^T \boldsymbol{R}^{-1} \boldsymbol{H})^{-1} = \left(\sum_{i=1}^{L} \frac{1}{\sigma_i^2} \right)^{-1} \tag{29}$$

There is ${\sigma_i}^2 > 0$ in (29),

$$\frac{1}{{\sigma_1}^2} + \cdots + \frac{1}{{\sigma_i}^2} > \frac{1}{{\sigma_i}^2}, i = 1, 2, \cdots, L \tag{30}$$

Then

$$P_{GM} < {\sigma_i}^2, i = 1, 2, \cdots, L \tag{31}$$

The Eq. (31) shows Gauss–Markov estimation algorithm can make the variance of the prediction error after multiple model fusion less than the variance of the single model. That means the prediction accuracy of the single model is improved.

4. The prediction method based on wavelet transform and multiple models fusion

In this paper, the prediction method based on wavelet transform and multiple model fusion includes modelling process and prediction process. Fig. 1 is the modelling process.

Firstly, the training samples of chaotic time series are decomposed and reconstructed by three layers of db3 wavelet. Then, one low frequency approximate component Ca and three high frequency detail components Cd1, Cd2 and Cd3 are obtained. ARIMA is chosen as prediction model for detail components. IFS-LSSVM is chosen as prediction model for approximate component. Through the error of each single model, the output error variance matrix $\boldsymbol{R}$ is calculated.

Fig. 2 is the prediction process. Similar to the modelling process, the testing samples of chaotic time series are decomposed and reconstructed by db3 wavelet with three layers. One low frequency approximate component Ca and three high frequency detail components Cd1, Cd2 and Cd3 are obtained. The approximate component is predicted through IFS-LSSVM. The detail components are predicted by ARIMA. Assuming prediction value of each model is $\hat{y}_1$, $\hat{y}_2$, $\hat{y}_3$ and $\hat{y}_4$. According to Eq. (28), the final fusion prediction value $\hat{y}$ can be obtained by Gauss–Markov estimation algorithm.

5. Simulation

In order to verify the effectiveness of the proposed prediction method, it is applied to Lorenz time series and Mackey–Glass time series prediction. In order to compare the prediction effect, the prediction method based on wavelet decomposition and reconstruction without Gauss–Markov fusion, prediction method in [17] is chosen for comparison. The prediction effect in [17] and direct superposition without Gauss–Markov fusion in [22–25] is better than LSSVM, ARIMA, ESN, Elman neural network and other prediction methods. So this paper directly chooses these two methods to compare prediction performance.

Root mean square error (RMSE), mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) are used as the prediction evaluation indicators. The actual value of the time series at moment i is y_i, prediction value is $\widehat{y}_i$, the length of time series is N, the 3 indicators are defined as follows,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2} \tag{32}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y}_i| \tag{33}$$

$$\text{SMAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \widehat{y}_i}{y_i + \widehat{y}_i} \right| \tag{34}$$

5.1. Lorenz time series prediction

As one of the most classical autonomous chaotic models with three dimensional and two degree polynomial, the research on the Lorenz system runs through the development of the chaotic time series prediction. So it is very essential and important to study the Lorenz system. This paper takes Lorenz time series as an example to verify the accuracy of the proposed prediction model. The expression of Lorenz time series is

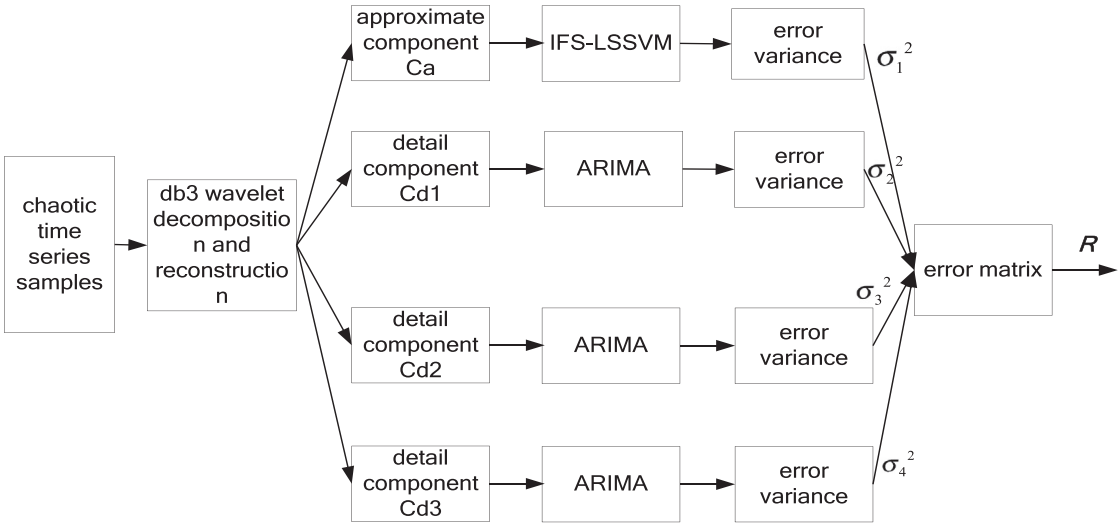

Fig. 1. The modelling process of wavelet transform and multiple models fusion.

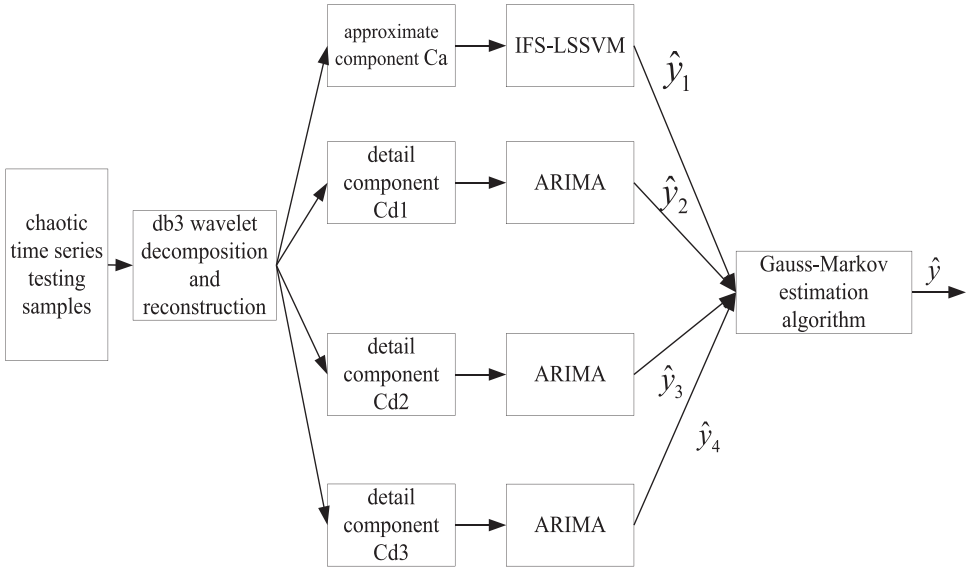

Fig. 2. The prediction process of wavelet transform and multiple models fusion.

$$\begin{aligned}\frac{dx}{dt} &= -\alpha(x-y),\\ \frac{dy}{dt} &= -xz+cx-y,\\ \frac{dz}{dt} &= xy-bz.\end{aligned} \tag{35}$$

The Eq. (35) will produce chaotic characteristics when $\alpha = 10$, $b = 8/3$, $c = 28$, $x(0) = -1$, $y(0) = 0$, $z(0) = 1$. The four order Runge–Kutta algorithm is used to generate the chaotic time series with 1000 groups data. Sampling interval ΔT is 0.05. Fig. 3 is a three-dimensional image of Lorenz chaotic time series. Fig. 4 is 1000 groups Lorenz time series on x, y and z axis.

Three layer decomposition and reconstruction of Lorenz time series by using db3 wavelet. Fig. 5 is the approximate component and the detail components on the x axis of the Lorenz series. Fig. 6 is the approximate component and the detail components on the y axis of the Lorenz series. Fig. 7 is the approximate component and the detail components on the z axis of the Lorenz series.

In order to eliminate the influence of the initial value on the prediction effect, the first 100 groups data of the approximate component and the detail components on x, y and z axis of Lorenz time series are excluded. The total 500 groups data from 101 to 600 as a training sample. A total of 200 groups of data from 601 to 800 as testing set, which used for validation for prediction accuracy. Firstly, the ARIMA prediction model is built according to the AIC criteria. Modeling results are shown in Table 1 (x_d1 indicates the detail component cd1 on the x axis, other variables are similar meanings).

Table 1
Modeling results of detail components.

Components	Model	AIC
x_d1	ARIMA(5,1,5)	−3.131
x_d2	ARIMA(10,1,8)	−3.759
x_d3	ARIMA(10,1,10)	−2.610
y_d1	ARIMA(7,1,7)	−3.812
y_d2	ARIMA(10,1,8)	−2.645
y_d3	ARIMA(10,1,10)	−1.789
z_d1	ARIMA(7,1,7)	−3.703
z_d2	ARIMA(8,1,9)	−2.246
z_d3	ARIMA(9,1,10)	−1.558

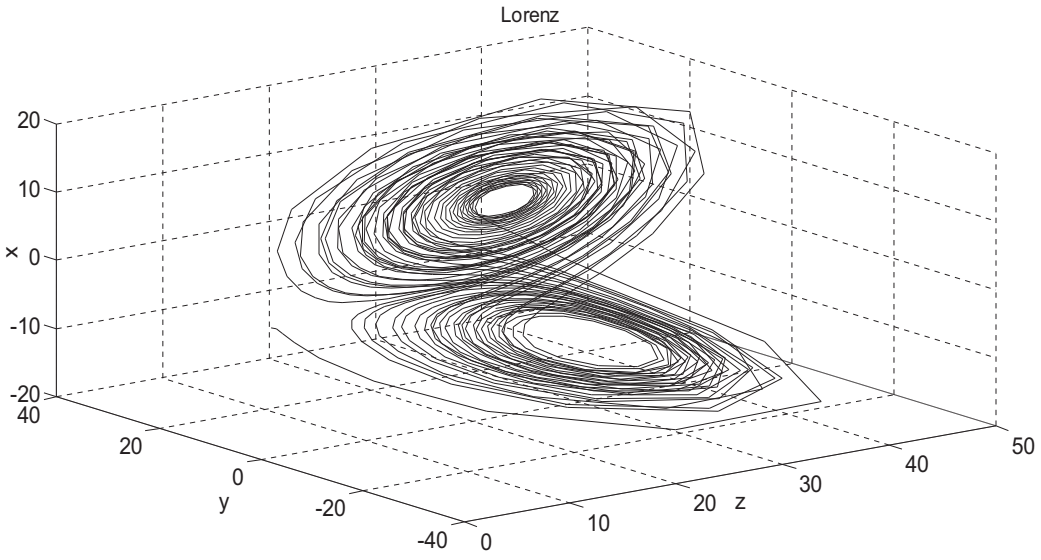

Fig. 3. Three-dimensional image of Lorenz chaotic time series.

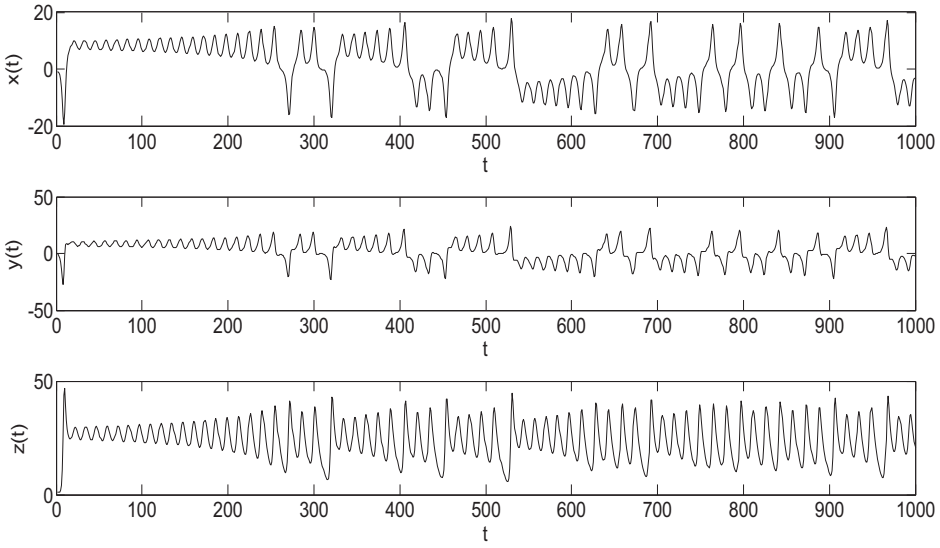

Fig. 4. 1000 groups Lorenz time series on *x*, *y* and *z* axis.

When ARIMA prediction model of detail components is built, 200 groups test data are simulated and verified. Fig. 8 is the predicted contrast curve of 3 detail components on the *x* axis. Fig. 9 is the predicted contrast curve of 3 detail components on the *y* axis. Fig. 10 is the predicted contrast curve of 3 detail components on the *z* axis. From the figures, it can be found that the predicted value of the prediction model is in good accordance with the actual value ("o represents the actual value," + line represents the predicted value).

The approximate components of Lorenz time series on *x*, *y* and *z* axis are trained and modeled based on IFS-LSSVM algorithm. The parameters of IFS algorithm: the population number is 50, maximum number of iterations is 100, λ is 0.9, δ is 0.95, $R_{\min}$ is 0.01. The variation range of the variables to be optimized as $\gamma \in [0.01, 1000)$, $\sigma^2 \in [0.01, 1000]$, $m \in [5, 50)$. Fig. 11 is fitness curve of approximate components of Lorenz time series on *x*, *y* and *z* axis. The optimization results are shown in Table 2 (*x*_a in the table represents the approximate component on the *x* axis and the other variables are similar).

Table 2
The optimization results.

Components	M	γ	σ^2
*x*_a	38	61.48	10.36
*y*_a	27	12.78	161.39
*z*_a	31	14.10	0.09

IFS optimized LSSVM is used as prediction model, 200 groups of test data are simulated and verified. Fig. 12 is the predicted contrast curve of the approximate components of Lorenz time series on the *x* axis. Fig. 13 is the predicted contrast curve of the approximate components of Lorenz time series on the *y* axis. Fig. 14 is the predicted contrast curve of the approximate components of Lorenz time series on the *z* axis ('o' line represents the actual value, '+' line represents the predicted value).

After the approximate component and the detail components prediction model of Lorenz time series is built, the variance matrix of Gauss-Markov estimation error of Lorenz time series on *x*, *y* and

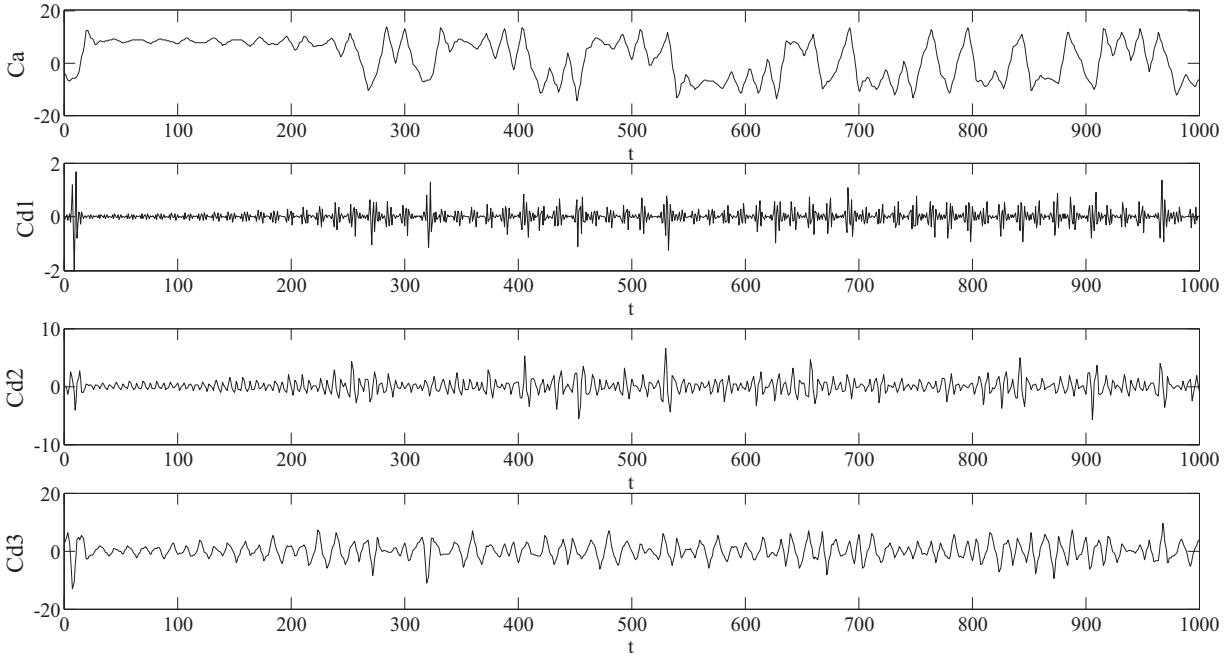

Fig. 5. The approximate component and the detail components on the *x* axis of Lorenz series.

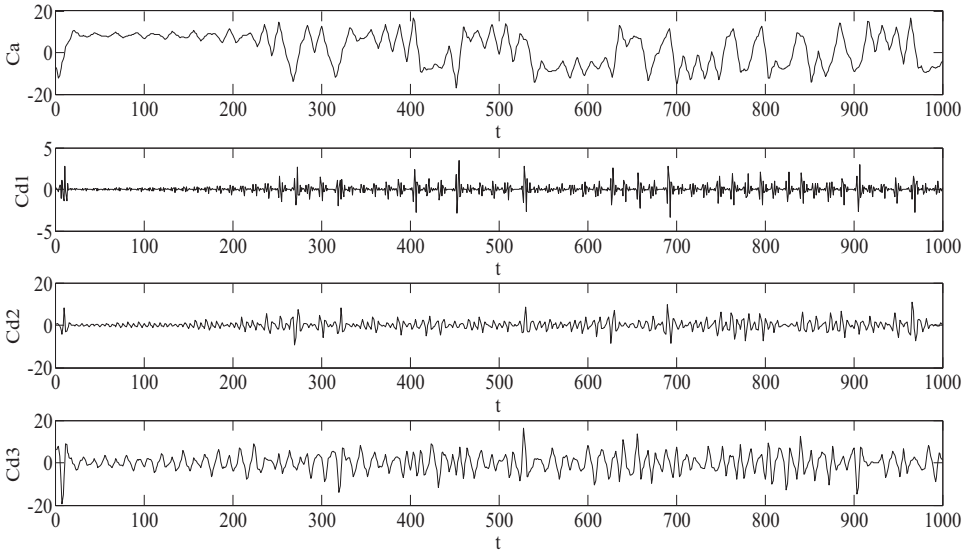

Fig. 6. The approximate component and the detail components on the *y* axis of Lorenz series.

z axis can be calculated as followed.

$$\boldsymbol{R}_x = \begin{bmatrix} 0.0022 & & & \\ & 0.0058 & & \\ & & 0.0231 & \\ & & & 0.0868 \end{bmatrix} \tag{36}$$

$$\boldsymbol{R}_y = \begin{bmatrix} 0.0138 & & & \\ & 0.0230 & & \\ & & 0.0728 & \\ & & & 0.1763 \end{bmatrix} \tag{37}$$

$$\boldsymbol{R}_z = \begin{bmatrix} 0.0274 & & & \\ & 0.0233 & & \\ & & 0.0929 & \\ & & & 0.1696 \end{bmatrix} \tag{38}$$

After the predictive value of the approximate component, the detail components and the Gauss-Markov estimation error variance matrix is obtained, multiple model fusion is carried out according to (28). The predicted value of Lorenz time series on *x*, *y* and *z* axis after multiple model fusion and the actual value contrast curve is shown in Fig. 15 ('o' line represents the actual value, '+' line represents the predicted value). Fig. 16 is the error distribution.

In order to compare the prediction effect, the prediction method in this paper is compared with combination kernel function LSSVM in [17] and direct superposition after wavelet decomposition and reconstruction without Gauss–Markov fusion in [23]. Predictive performance indexes are shown in Tables 3, 4 and 5. The algorithm runs 10 times, the average value is taken as results.

From Table 3 to Table 5, it can know that the chaotic time series prediction method in this paper is better than direct superposition without Gauss–Markov fusion and the method of literature [17].

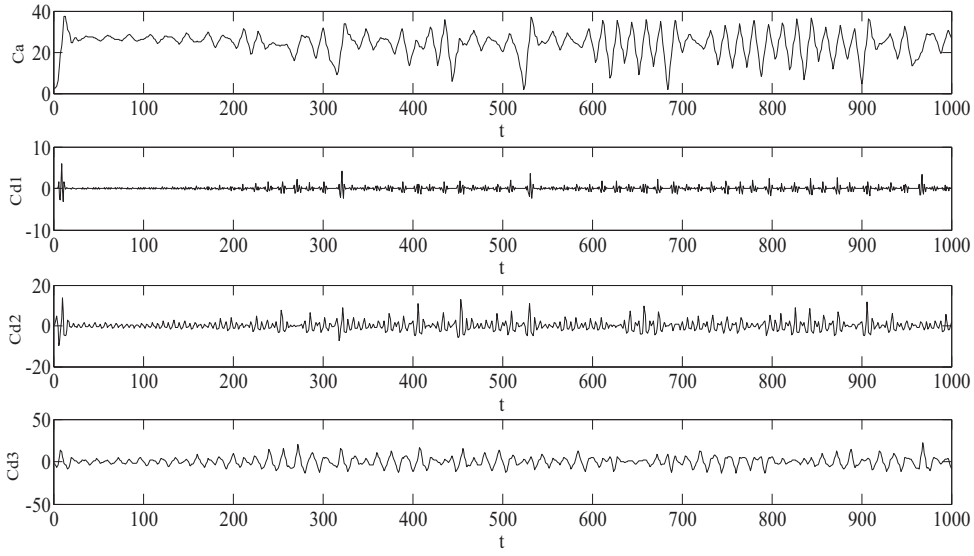

Fig. 7. The approximate component and the detail components on the *z* axis of Lorenz series.

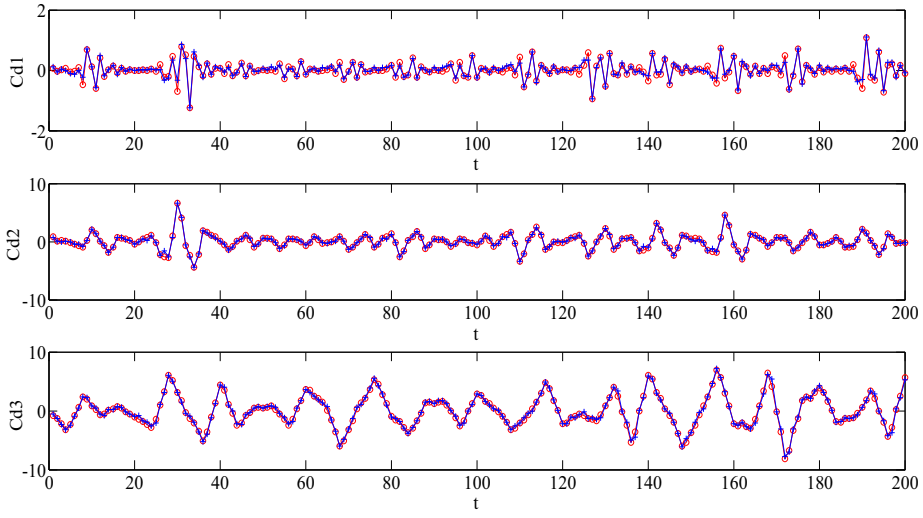

Fig. 8. The predicted contrast curve of 3 detail components on the *x* axis of Lorenz series.

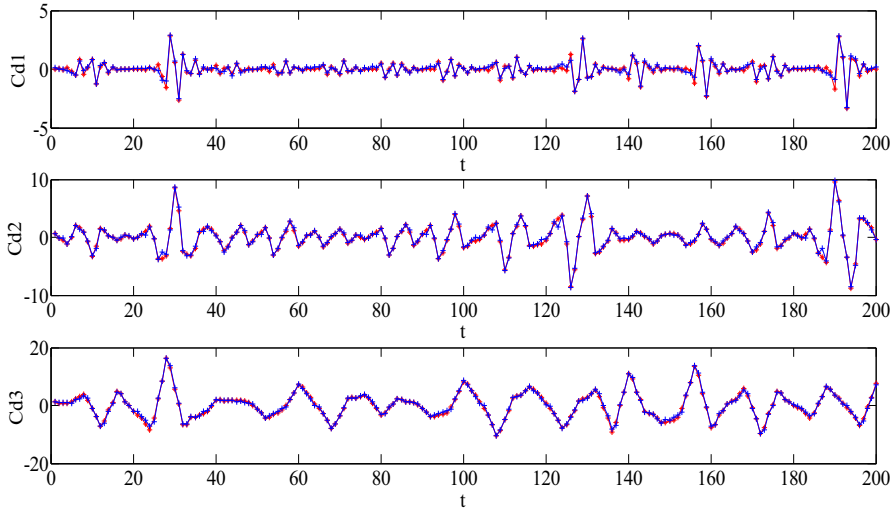

Fig. 9. The predicted contrast curve of 3 detail components on the *y* axis of Lorenz series.

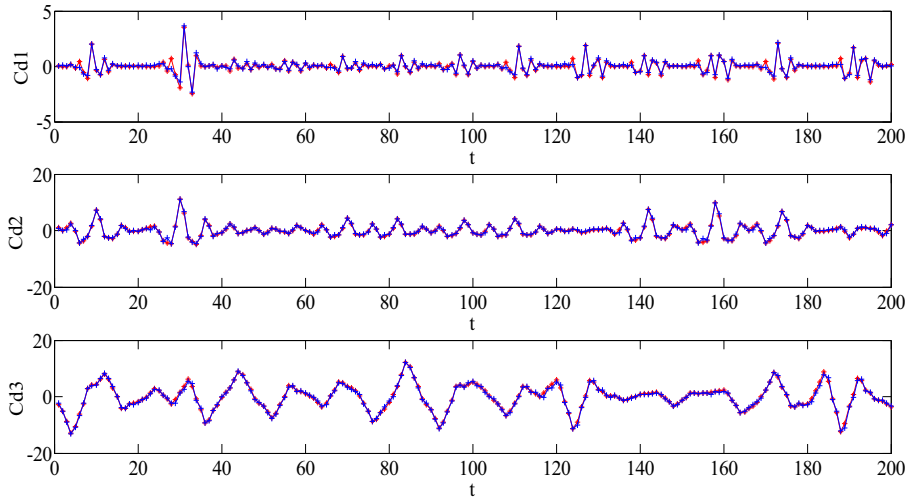

Fig. 10. The predicted contrast curve of 3 detail components on the z axis of Lorenz series.

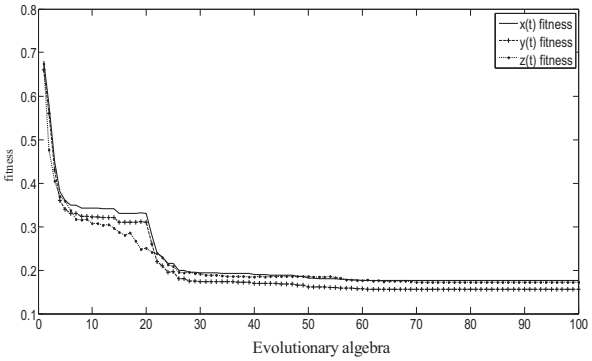

Fig. 11. The fitness curve of approximate components of Lorenz time series on x, y and z axis.

Table 3
Predictive performance indexes of Lorenz time series on x axis.

Prediction methods	RMSE	MAE	SMAPE
Prediction method in this paper	0.0203	0.1240	0.0394
Literature [17]	0.0431	0.1652	0.0485
Direct superposition without Gauss-Markov fusion in [23]	0.0931	0.2304	0.0615

Table 4
Predictive performance indexes of Lorenz time series on y axis.

Prediction methods	RMSE	MAE	SMAPE
Prediction method in this paper	0.0651	0.2181	0.0875
Literature [17]	0.0951	0.3125	0.1745
Direct superposition without Gauss-Markov fusion in [23]	0.2952	0.4206	0.2078

Therefore, the prediction method in this paper has good predictive effect on the Lorenz time series.

5.2. Mackey-Glass time series prediction

In 1977, Mackey and Glass found the chaotic phenomena in time delay systems. Time delayed chaotic systems have attracted wide attention, and are often used as the standard to test the prediction performance of the nonlinear system model. Mackey–Glass chaotic time series is produced by the delay differential equation

Table 5
Predictive performance indexes of Lorenz time series on z axis.

Prediction methods	RMSE	MAE	SMAPE
Prediction method in this paper	0.0579	0.2078	0.0051
Literature [17]	0.1722	0.3015	0.0152
Direct superposition without Gauss-Markov fusion in [23]	0.2981	0.3954	0.0212

Table 6
The modeling results of ARIMA of Mackey–Glass.

Components	Model	AIC
Detail component 1	ARIMA(10,1,9)	−1.720
Detail component 2	ARIMA(5,1,10)	−1.416
Detail component 3	ARIMA(9,1,7)	−1.189

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\Delta)}{1+x^{10}(t-\Delta)} - 0.1x(t) \tag{39}$$

Where, Δ is delay parameter. The (39) presents chaotic characteristics when $\Delta > 17$, its fractal dimension is similar to 2.1. The greater Δ, the more obvious the chaotic characteristics of the system. Fig. 17 shows the 1000 groups of Mackey–Glass time series when Δ is 30. Fig. 18 is the phase diagram when Δ is 30. From the graphs, it can be seen that Mackey–Glass time series have very complex nonlinear chaotic characteristics.

The db3 wavelet is used to reconstruct and decomposition of Mackey–Glass time series, the approximate component and the detail components are obtained and shown in Fig. 19.

The total of 500 groups data from 101 to 600 of approximate component and the detail components of Mackey-Glass time series are used to train prediction model. The total of 200 groups data from 601 to 800 of approximate component and the detail components of Mackey-Glass time series are used as testing data. ARIMA prediction model of detail components are built based on AIC criterion. Through the ARIMA model, the 200 groups of test data are predicted and simulated. The modeling results of ARIMA model are shown in Table 6. Fig. 20 is the comparison curve between the actual value and the predicted value of the detail components ('*' line represents the actual value, '+' line represents the predicted value).

IFS algorithm is used to optimize the parameters of LSSVM prediction model of approximate component. The parameters of the IFS algorithm and the range of parameters to be optimized are the

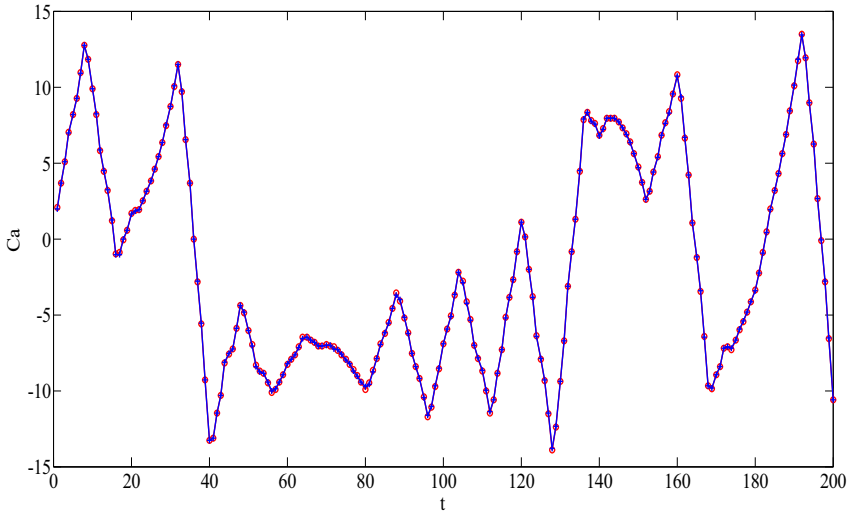

Fig. 12. The predicted contrast curve of the approximate components of Lorenz time series on the *x* axis.

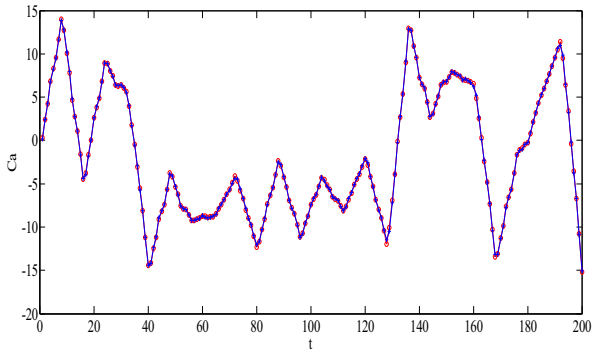

Fig. 13. The predicted contrast curve of the approximate components of Lorenz time series on the *y* axis.

same as Section 5.1. The fitness curve of IFS algorithm is shown in Fig. 21.

Parameter optimization results are: γ is 85.16, σ^2 is 34.06, m is 27. The approximate component of the 200 groups of Mackey–Glass time series testing data are predicted by the IFS-LSSVM prediction model. The prediction comparison results are as shown in Fig. 22.

After the approximate component and the detail components prediction model of Mackey–Glass time series is built, the variance matrix of Gauss–Markov estimation error of Mackey–Glass time series can be calculated.

$$\boldsymbol{R} = \begin{bmatrix} 3.22e^{-7} & & & \\ & 2.51e^{-8} & & \\ & & 6.82e^{-7} & \\ & & & 4.65e^{-6} \end{bmatrix} \tag{40}$$

After the predictive value of the approximate component, the detail components and the Gauss–Markov estimation error variance matrix is obtained, multiple model fusion is carried out according to (28). The predicted value of Mackey–Glass time series after multiple model fusion and the actual value contrast curve is shown in Fig. 23 ('o' line represents the actual value, '+' line represents the predicted value). Fig. 24 is the error distribution.

The prediction method in this paper is compared with combination kernel function LSSVM in [17] and direct superposition after wavelet decomposition and reconstruction without Gauss–Markov fusionin [23]. The algorithm runs 10 times, the average

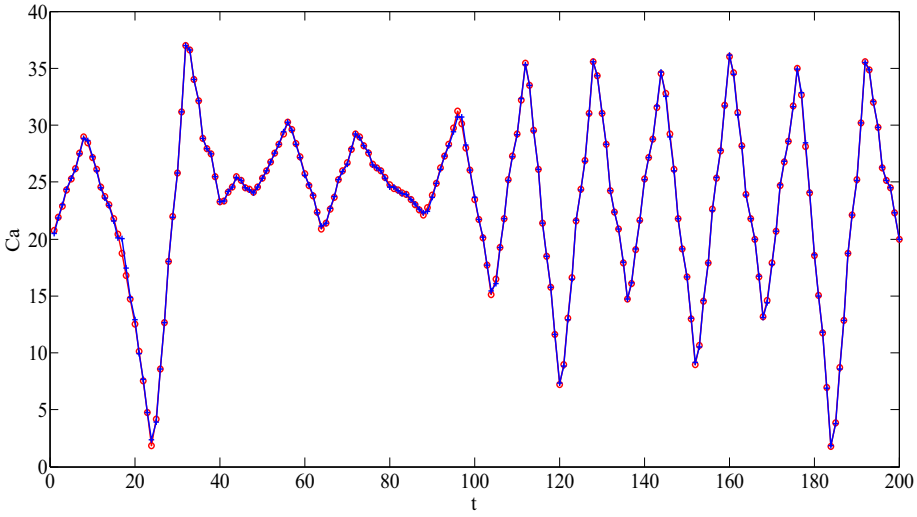

Fig. 14. The predicted contrast curve of the approximate components of Lorenz time series on the *z* axis.

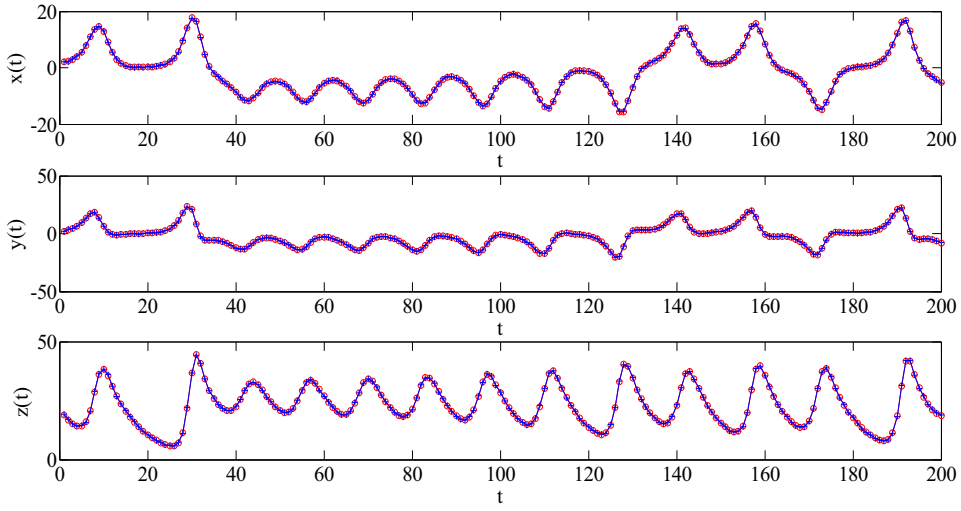

Fig. 15. The predicted and actual value of Lorenz time series on *x*, *y* and *z* axis after multiple model fusion.

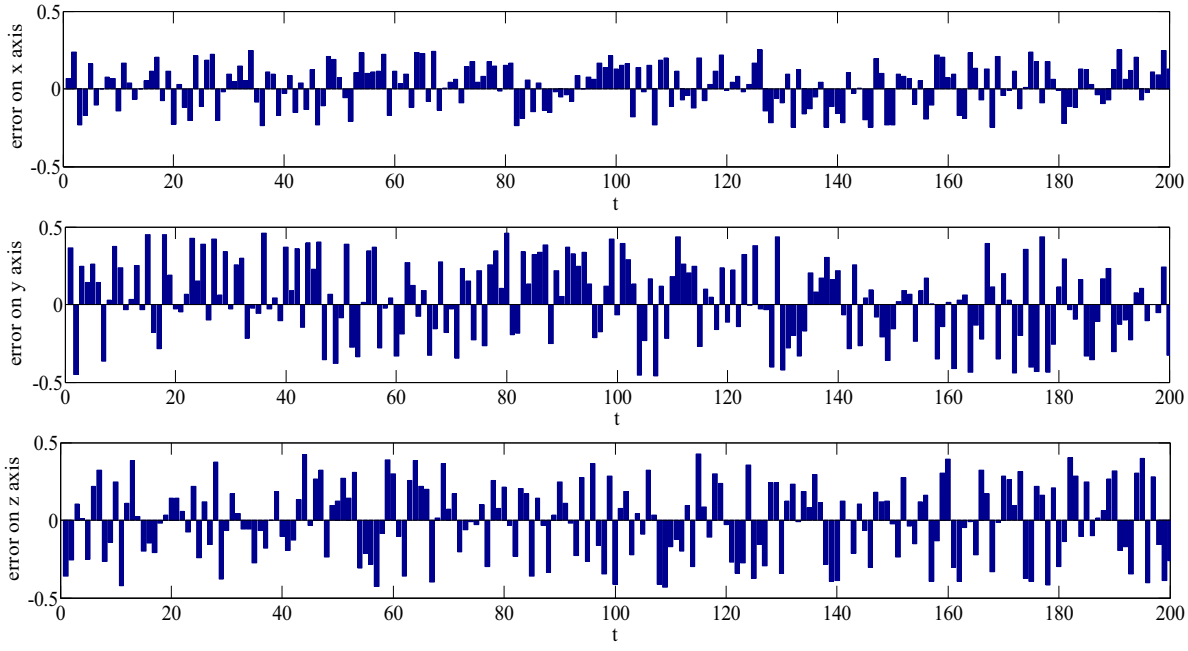

Fig. 16. The predictive error distribution of Lorenz time series.

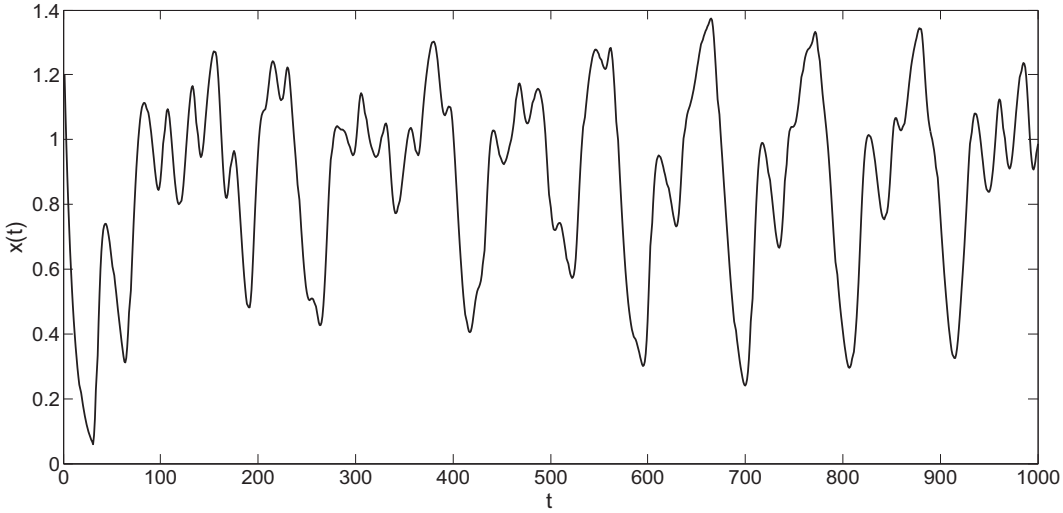

Fig. 17. Mackey–Glass time series.

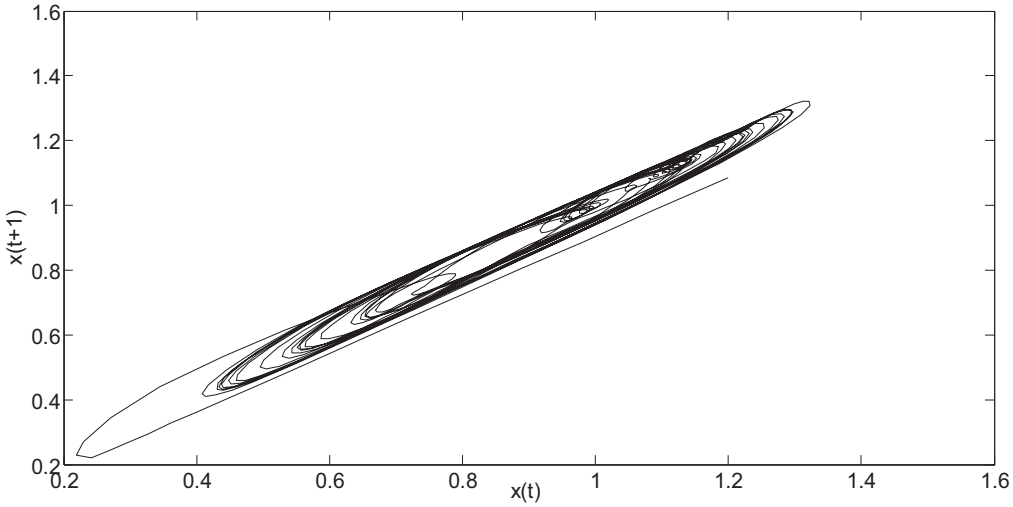

Fig. 18. The phase diagram of Mackey–Glass time series.

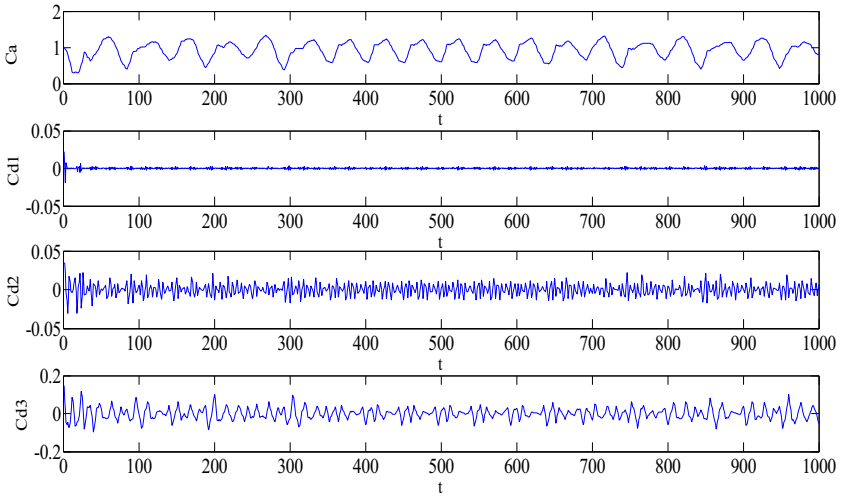

Fig. 19. The approximate component and the detail components of Mackey–Glass time series.

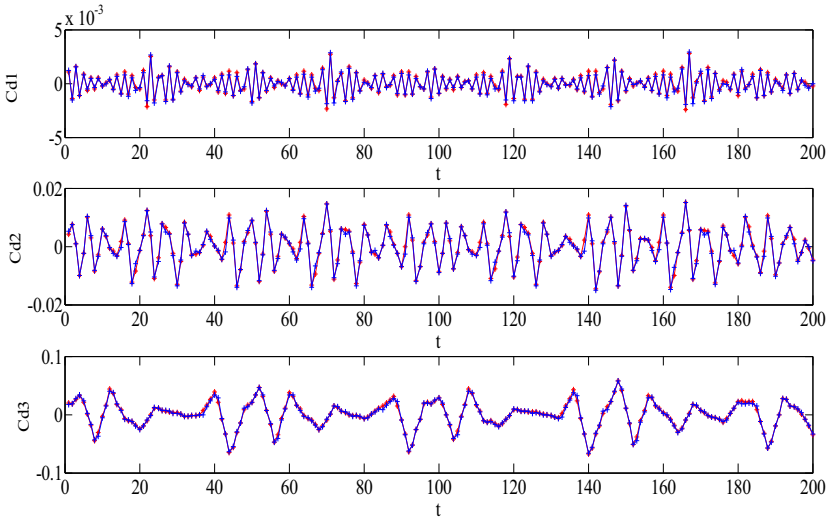

Fig. 20. The comparison curve between the actual value and the predicted value of the detail components of Mackey–Glass.

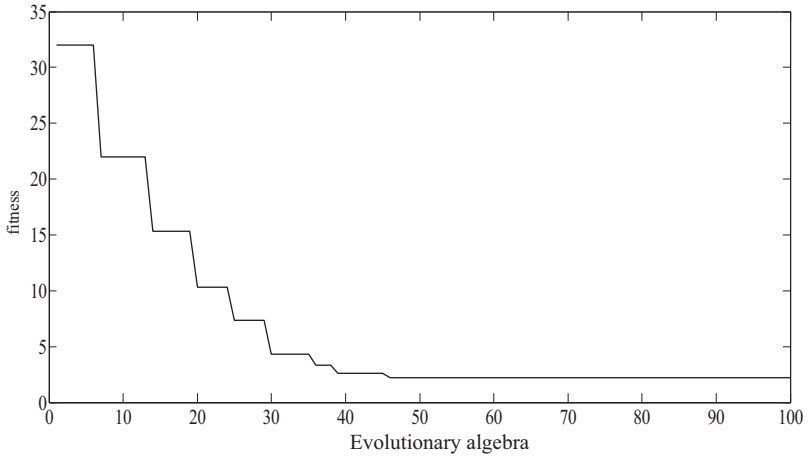

Fig. 21. The fitness curve of IFS algorithm of Mackey–Glass.

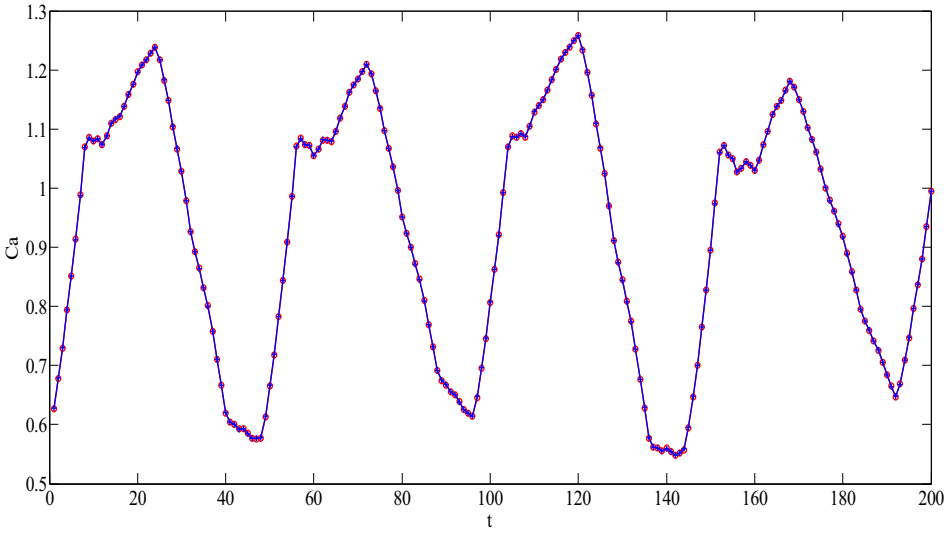

Fig. 22. The predictive value and the actual value of the approximate component of Mackey–Glass time series.

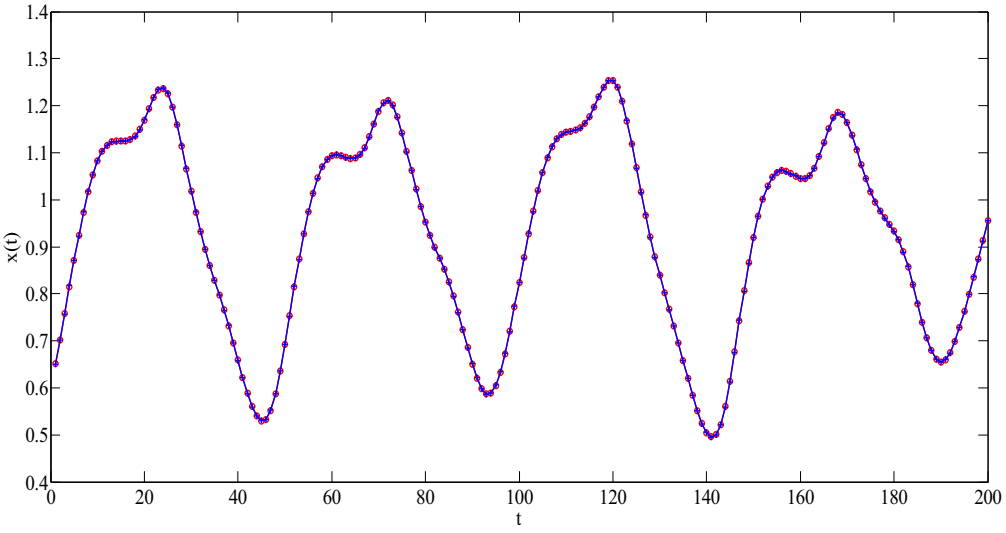

Fig. 23. The predicted value of Mackey–Glass time series after multiple model fusion and the actual value contrast curve.

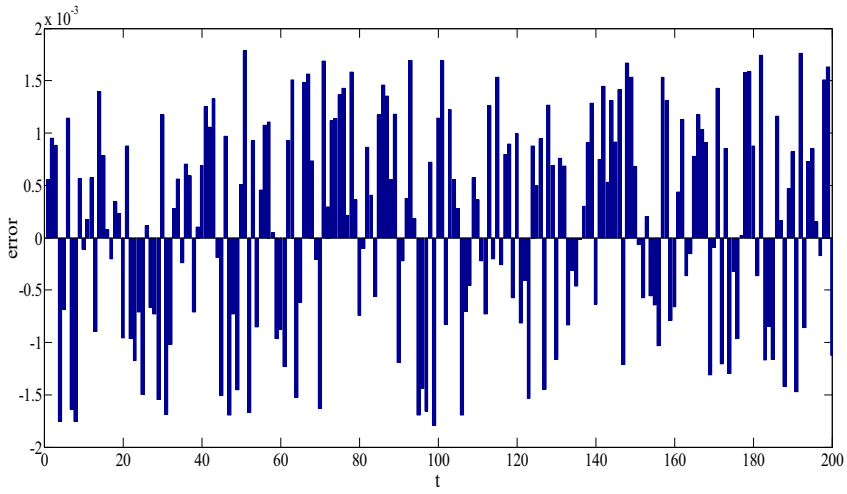

Fig. 24. The predictive error distribution of Mackey–Glass time series.

Table 7
Predictive performance indexes of Mackey–Glass time series.

Prediction methods	RMSE	MAE	SMAPE
Prediction method in this paper	1.059 e−6	9.036 e−6	5.239 e−4
Literature [17]	3.145 e−6	7.452 e−5	10.561 e−4
Direct superposition without Gauss–Markov fusion in [23]	6.431 e−6	1.8 e−4	14.674 e−4

value is taken as results. Predictive performance indexes are shown in Table 7.

From the prediction performance of Figs. 23, 24 and Table 7, it can know that the prediction method of this paper also has a good predictive effect for Mackey–Glass time series.

For two typical chaotic time series of Lorenz and Mackey–Glass, the reasons for the good prediction results are as follows. Firstly, wavelet transform is used to decompose and reconstruct the original time series to approximate component and detail components, and the suitable prediction model is used to predict each component. Secondly, Gauss–Markov estimation algorithm is used to make the fusion of multiple prediction model, and the prediction accuracy is further improved.

6. Conclusion

According to the prediction of chaotic time series, this paper proposed a new prediction method based on wavelet transform and multiple model fusion. The original time series is decomposed and reconstructed by db3 wavelet. LSSVM is chosen as prediction model for the approximate component with noise smoothed. At the same time, the optimization of LSSVM parameters is carried out by the IFS algorithm. ARIMA is chosen as prediction model for the detail components with random and non-stationary characteristics. The prediction results of several prediction models are fusion by the Gauss–Markov estimation algorithm. The error variance of the predicted value is smaller than any sub model, thus improving the accuracy of the prediction model. The simulation results of 2 typical chaotic time series Lorenz and Mackey-Glass time series are studied. The simulation results show that proposed prediction method is effective.

The first future work of this paper is to introduce the kernel principal component analysis into the prediction of chaotic time series. Based on feature extraction of chaotic time series through kernel principal component analysis, correlation and noise of chaotic time series are eliminated. The dimension of the sample space and the computational complexity of the prediction algorithm are reduced. Secondly, the paper will further validate the effectiveness of the proposed prediction method on actual time series with chaotic characteristics, such as short-term wind speed series, network traffic series, sunspot series and etc.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant 11273001, Liaoning Province Doctor Startup Fund under Grant 20141070 and Liaoning Education Department under Grant LGD2016009.

References

[1] Aghababa MP. A Lyapunov-based control scheme for robust stabilization of fractional chaotic systems[J]. Nonlinear Dyn. 2014;78(3):2129–40.
[2] Naze P, Venegeroles R. Number of first-passage times as a measurement of information for weakly chaotic systems[J]. Phys Rev E 2014;90(4):042917.
[3] McGovern A, Rosendahl DH, Brown RA. Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction[J]. Data Min Knowl Discovery 2011;22(1-2):232–58.
[4] Bozic M, Stojanovic M, Stajic Z. Mutual information-based inputs selection for electric load time series forecasting[J]. Entropy 2013;15(3):926–42.
[5] Zhao XJ, Shang PJ, Wang J. Measuring information interactions on the ordinal pattern of stock time series[J]. Phys Rev E 2013;87(2):805–13.
[6] Yang ZS, Zhao S, Bing QH, et al. Research on short-term traffic flow prediction method based on similarity search of time series[J]. Math Prob Eng 2014:184632.
[7] Dutt D, Welsh WD, Vaze J. A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in water source[J]. Water Resour Manage 2012;26(15):4397–415.
[8] Fan GF, Qing S, Wang H, et al. Study on apparent kinetic prediction model of the smelting reduction based on the time-series[J]. Math Prob Eng 2012:720849.
[9] Ding J, Han LL, Chen XM. Time series AR modeling with missing observations based on the polynomial transformation[J]. Math Comput Modell 2010;51(5):527–36.
[10] Toque C, Terraza V. Time series factorial models with uncertainty measures: Applications to ARMA processes and financial data[J]. Commun Statistics Theory Methods 2011;40(9):1533–44.
[11] Khashei M, Bijari M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting[J]. Appl Soft Comput J 2011;11(2):2664–75.
[12] Niu HL, Wang J. Financial time series prediction by a random data-time effective RBF neural network[J]. Soft Comput 2014;18(3):497–508.
[13] Stefan Babinec, Pospichal J. Modular echo state neural networks in time series prediction[J]. Computing Inf 2011;30(2):321–34.
[14] Wang XY, Han M. Online sequential extreme learning machine with kernels for nonstationary time series prediction[J]. Neurocomputing 2014;145:90–7.
[15] Chandra R, Zhang MJ. Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction[J]. Neurocomputing 2012;86:116–123.
[16] Li PX, Tan ZX, Yan LL, et al. Time series prediction of mining subsidence based on a SVM [J]. Mining Sci Technol 2011;21(4):557–62.

[17] Tian ZD, Gao XW, Shi T. Combination kernel function least squares support vector machine for chaotic time series prediction[J]. Acta Phys. Sin 2014:160508.
[18] Zhao HQ, Zeng XP, He ZY. Low-complexity nonlinear adaptive filter based on a pipelined bilinear recurrent neural network[J]. IEEE Trans Neural Netw 2011;22(9):1494–507.
[19] Wang Q, Yang Y. Combined prediction of wind power with chaotic time series analysis[J]. Open Autom Control Syst J 2013;6(1):117–23.
[20] Babu CN, Reddy BE. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data[J]. Appl Soft Comput 2014;23:27–38.
[21] Di CL, Yang XH, Wang XC. A four-stage hybrid model for hydrological time series forecasting[J]. Plos One 2014;25(2):269–81.
[22] Karthikeyan L, Kumar DN. Predictability of nonstationary time series using wavelet and EMD based ARMA models[J]. J Hydrol 2013;502:103–19.
[23] An XL, Jiang DX, Liu C, et al. Wind farm power prediction based on wavelet decomposition and chaotic time series[J]. Expert Syst Appl 2011;38(9):11280–5.
[24] Wang L, Zou F, Hei XH, et al. A hybridization of teachinglearning-based optimization and differential evolution for chaotic time series prediction[J]. Neural Comput Appl 2014;25(6):1407–22.
[25] Guo W, Xu T, Lu ZL. An integrated chaotic time series prediction model based on efficient extreme learning machine and differential evolution[J]. Neural Comput Appl 2016;27(4):883–98.
[26] Pan J, Chen JL, Zi YY, et al. Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive Fourier spectrum segment[J]. Mech Syst Signal Process 2016(72-73):160–83.
[27] Zhang XY, Chen Q, Chen SY, et al. Image compressed sensing based on the double single-layer wavelet transform[J]. J Appl Sci 2013;13(13):2479–82.
[28] Yao X, Herrera L, Ji SC. Characteristic study and time-domain discrete-wavelet–transform based hybrid detection of series DC arc faults[J]. IEEE Trans Power Electron 2014;39(6):3103–15.
[29] Dirick L, Claeskens G, Baesens B. An Akaike information criterion for multiple event mixture cure models[J]. Eur J Oper Res 2015;241(2):449–57.
[30] Tian ZD, Li SJ, Wang YH, et al. A multi-model fusion soft sensor modelling method and its application in rotary kiln calcination zone temperature prediction [J]. Trans Inst Meas Control 2016;38(1):110–24.
[31] Kalin P, Guy L. Free search-a comparative analysis[J]. Inf Science 2005;172:173–93.
[32] Yin JH, Wang YW, Hu JK. Free search with adaptive differential evolution exploitation and quantum-inspired exploration[J]. J Network Comput Appl 2012;35(3):1035–51.
[33] Ayala HVH, Keller P, Morais MDF, et al. Design of heat exchangers using a novel multiobjective free search differential evolution paradigm[J]. Appl Therm Eng 2016;94:170–7.
[34] Gao XW, Tian ZD, Li K. Time-delay prediction method based on maximum Lyapunov exponent and Elman neural network[J]. Sci China (Inf Sci) 2013;43:1042–57.
[35] Buonocore A, Caputo L, Nobile A G Gauss-Markov processes in the presence of a reflecting boundary and applications in neuronal models[J]. Appl Math Comput, 2014, (232): 799–809.