Neurocomputing 73 (2010) 2540–2553

Contents lists available at ScienceDirect

Neurocomputing

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

Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks

Muhammad Ardalani-Farsa, Saeed Zolfaghari *

Department of Mechanical and Industrial Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada M5B 2K3

A R T I C L E I N F O

Article history:
Received 10 August 2009
Received in revised form
3 February 2010
Accepted 10 June 2010
Communicated by G.P. Zhang
Available online 25 June 2010

Keywords:
Chaos theory
Embedding theorem
Elman neural network
NARX neural network
Residual analysis
Nonlinear time series prediction

A B S T R A C T

Residual analysis using hybrid Elman–NARX neural network along with embedding theorem is used to analyze and predict chaotic time series. Using embedding theorem, the embedding parameters are determined and the time series is reconstructed into proper phase space points. The embedded phase space points are fed into an Elman neural network and trained. The residual of predicted time series is analyzed, and it was observed that residuals demonstrate chaotic behaviour. The residuals are considered as a new chaotic time series and reconstructed according to embedding theorem. A new Elman neural network is trained to predict the future value of the residual time series. The residual analysis is repeated several times. Finally, a NARX network is used to capture the relationship among the predicted value of original time series and residuals and original time series. The method is applied to Mackey–Glass and Lorenz equations which produce chaotic time series, and to a real life chaotic time series, Sunspot time series, to evaluate the validity of the proposed technique. Numerical experimental results confirm that the proposed method can predict the chaotic time series more effectively and accurately when compared with the existing prediction methods.

1. Introduction

Over the last several decades, prediction of chaotic time series has been a popular and challenging subject. Chaos theory as a new area of mathematics has been used to analyze chaotic systems and draw the hidden information from random-look data produced by chaotic systems. Chaotic time series are deterministic systems and inherit a high degree of complexity. Although, chaotic time series show the characteristic of dynamical systems as random, in the embedding phase space they present deterministic behaviour [66].

The chaos theory, as an essential part of nonlinear theory, has provided an appropriate tool to illustrate the characteristics of the dynamical systems and predict the trend of complex systems. There are four fundamental characteristics for chaotic systems: *aperiodic* that is the same state will not be repeated, *bounded* meaning that neighbour states remain within a finite range and does not approach infinity, *deterministic* that there is a governing rule with no random term to predict the future state of the system, and *sensitivity to initial conditions* meaning that small difference in initial conditions will cause two points close to each other diverge as the state of system progress [62].

Takens' [53] embedding theorem is an essential element of chaotic time series analysis. A set of single observations from a chaotic system can be reconstructed into a series of D-dimensional vectors with two parameters of time delay and dimension. Based on Takens' theorem, if dimension is large enough, the reconstructed vectors exhibit many of the significant entities of the time series [12].

Prediction of nonlinear time series is a useful method to evaluate characteristics of dynamical systems. Prediction of chaotic time series have been observed in the areas of marketing system [44], foreign exchange rate [5], signal processing [22], supply chain management [61], traffic flow [39], power load [48], weather forecast [31], Sunspot prediction [42] and many others. Due to the importance of these fields, the interests in a robust technique to predict chaotic time series have been increased.

A number of techniques to predict chaotic time series have been introduced in the literature. A method of local modelling was proposed by McNames [36]. Wichard and Ogorzalek [60] described the use of ensemble methods to build proper models for chaotic time series prediction. Zhang et al. [67] proposed a multi-dimension prediction method using Lyapunov exponents.

Artificial neural networks (ANNs) have been also employed independently or as an auxiliary tool to predict chaotic time series. ANNs are nonlinear methods which mimic nerve system [68]. They have functions of self-organizing, data-driven, self-study, self-adaptive and associated memory [64,68]. ANNs can learn from patterns and capture hidden functional relationships in a given data even if the functional relationships are not known or difficult to identify [64]. Using the training methods, an ANN can be trained to identify the underlying correlation between the

* Corresponding author.
E-mail address: zolfaghari@ryerson.ca (S. Zolfaghari).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.004

inputs and outputs. Later, the unseen inputs can be fed to the trained ANN to generate appropriate outputs [11,17,41,43,64].

A number of researchers have utilized ANN to predict chaotic time series. Multi-layer perceptron neural network (MLP) has been used by Liu et al. [28] and Park et al. [42]. Zhang and Man [66], Tenti [55], Ma et al. [32], and Assaad et al. [2] have utilized recurrent neural network (RNN). Nonlinear Autoregressive model with eXogenous input (NARX) has been also applied to chaotic time series perdition by Menezes and Barreto [38] and Diaconescu [8].

Some other artificial intelligence (AI) methods such as radial basis function network (RBF) [54,46], self-organizing map (SOM) [3,26,51], support vector machine (SVM) [47,58], fuzzy and neuro-fuzzy [14,65,24], and wavelet neural networks (WNN) [13,59] among the others are used in the literature to forecast chaotic time series.

In prediction methods, the analysis of residuals is underestimated. In some occasions, residuals are not due to randomness; therefore, residuals show high correlation meaning that the prediction model has not completely captured the characteristics of the system. There are occasions that residuals inherit the characteristic of original system.

This study investigates the contribution of residual analysis to increase the performance of prediction method. The proposed method utilizes the embedding theorem to "unfold" the chaotic time series and reconstruct the phase space points. Two well-known dynamic neural networks, Elman and NARX networks are selected for training purposes. An Elman Neural Network is trained using gradient descent with momentum and adaptive learning rate backpropagation algorithm to predict the future value of the obtained phase space points and accordingly the original time series. Normally, the residuals of predicted time series show high degree correlation and demonstrate chaotic behaviour. Therefore, the residuals are considered as a new chaotic time series and similarly analyzed and predicted. The residual analysis is repeated several times. Finally, a NARX neural network is trained to capture the relationship among the predicted value of original time series, residuals and original time series. The weights and biases of NARX neural network is kept to predict the future value of original time series. The block diagram in Fig. 1 demonstrates the methodology developed in this study to forecast chaotic time series.

The paper is organized as follows. Section 2 describes chaotic time series and the method to reconstruct a time series. Section 3 briefly discusses dynamic neural networks. Section 4 illustrates the proposed prediction technique in detail. In Section 5, the prediction performance of the proposed technique on three well-known chaotic time series, the Mackey–Glass, Lorenz and Sunspot, are studied. Finally, conclusions are given in Section 6.

2. Chaotic time series and embedding theorem

Many natural systems show nonlinear or chaotic behaviour. Using chaos theory, these systems have been described by mathematical equations. For a chaotic system, the phase space is defined as a vector space R^n with each point in the phase space being described by a n-dimensional vector $s(t)$, which is required to obtain the progression of the system [40]. $s(t)$ is defined as

$$s(t) = [s_1(t), s_2(t), s_3(t), \ldots, s_n(t)] \tag{1}$$

where t is an index for the time series and n is the dimension of vector space R^n. With the use of the nonlinear function F: $R^n \rightarrow R^n$, which describes the system, the future value of the system at time $t+\tau$ can be determined by

$$s(t) \rightarrow F(s(t)) = s(t+\tau) \tag{2}$$

A small change in the state of the system, $s(t)$, will substantially influence the trend of the system and after several iterations, the system becomes unforeseeable. This behaviour of dynamical systems is known as sensitivity to initial conditions or butterfly effect [12,31].

The progression of a non-random system creates a trajectory named an attractor. Takens' [53] embedding theorem states that because the value of $s(t)$ and its components, $s_1(t)$, $s_2(t)$, $s_3(t)$,… in a chaotic system are unknown; if one is able to observe a single quantity or variable $x(t)$ from this dynamical system, then the attractor can be unfolded from this set of observed samples [1]. This means that if a single quantity $x(t)$ is observed from a chaotic system, the reconstructed dynamics of a system $Y(t)=[x(t), x(t-T), x(t-2T)\ldots]$, with T defined as time delay, is geometrically similar to the original attractor. Therefore, if a dynamical system $s(t)\rightarrow s(t+1)$ exists, then sequential order of reconstructed phase space points $Y(t)\rightarrow Y(t+1)$ follows the unknown dynamics of $s(t)\rightarrow s(t+1)$. Therefore, the behaviour of the actual system is reflected in the observed time series generated from the system [21].

2.1. Determining chaos in time series

In analyzing time series, an important step is to determine the characteristic of the data. The following methods have been used to differentiate periodic or random data from chaotic data.

2.1.1. Fourier transform

Fourier transform can be used to identify chaos in a given time series. It is common to plot power spectrum instead of frequency spectrum. The power spectrum spikes at frequencies that characterize the system for periodic data, and will be approximately zero for the others. The broadband power spectrum with broad peak proves the existence of chaos in the time series [12].

2.1.2. Lyapunov exponent

An important characteristic of chaotic systems which is defined as the butterfly effect is the high sensitivity of the system to the initial conditions. If the high sensitivity to initial conditions is detected in a system, the system can be considered chaotic [63]. Largest Lyapunov exponent is the most practical method to identify chaotic behaviour in a system. Lyapunov exponent quantify the divergence of neighbouring trajectories. Positive Lyapunov exponent proves the existence of chaos in the system [12].

2.1.3. Hurst exponent

Hurst is known for introducing Hurst exponent as a measurement for the predictability of a time series [16]. Hurst exponent is derived using R/S analysis. Hurst exponents can change between 0 and 1. Hurst exponent of 0.5 shows a random walk. A Hurst exponent between 0.5 and 1 proves the presence of chaos in the system.

2.1.4. Fractal dimension

Another method to identify the existence of chaos in a system is Fractal dimension. Non-integer fractal dimension shows that the system is chaotic. Correlation dimension is one of the most common Fractal dimension used in literature [12,52]. If a sphere of radius R is centred on a specific point in D-dimensional space, then the mean of points in the sphere, $C(R)$, excluding the center point can be calculated. A plot of $C(R)$ versus R should give an approximately straight line whose slope is d_c , the correlation dimension. d_c with integer value shows that the attractor is a

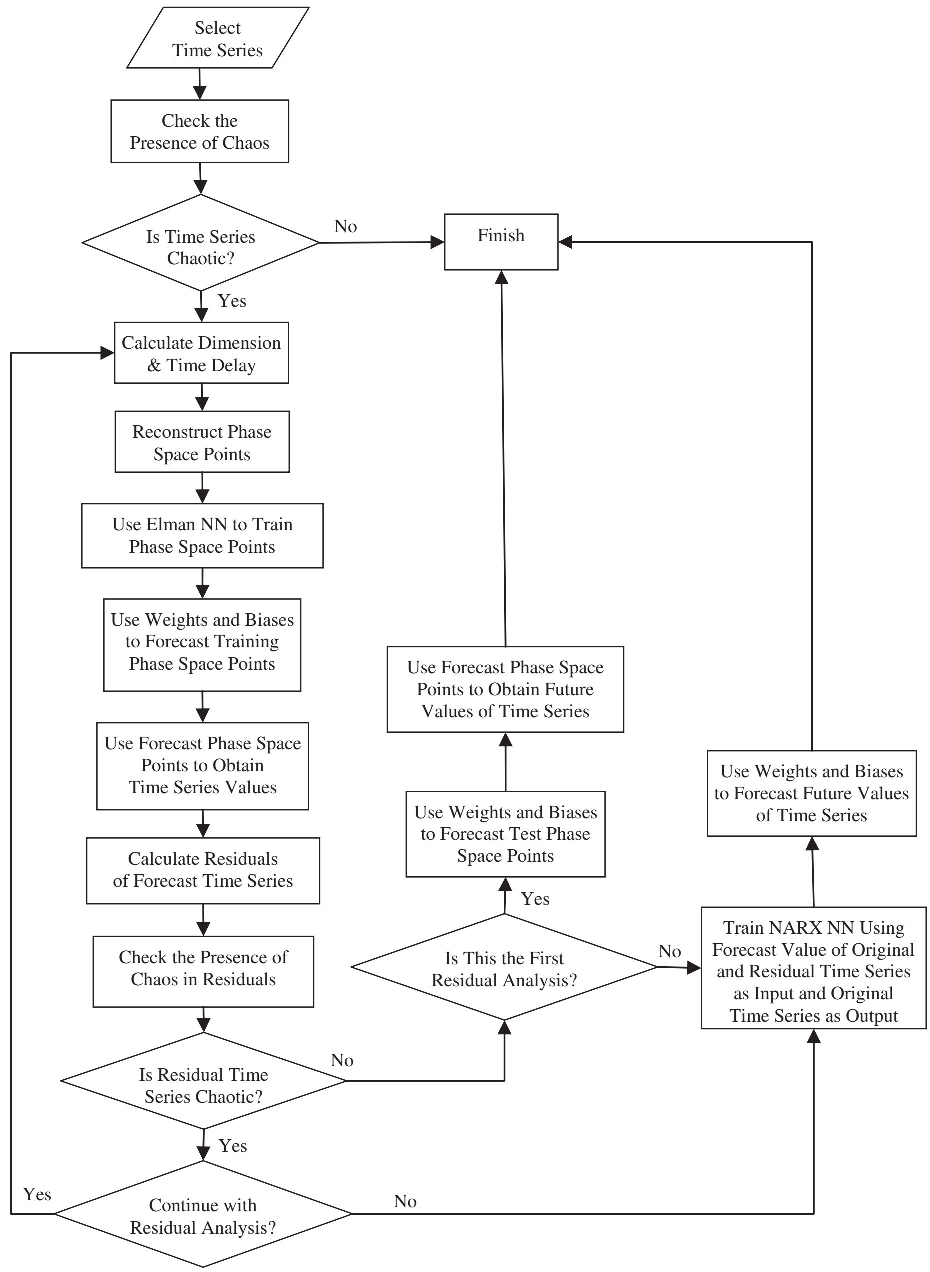

Fig. 1. A block diagram demonstrating the proposed methodology.

simple geometric object. A non-integer value for d_c indicates that the attractor is strange and the system is chaotic [12,18].

2.2. Embedding a time series

In the absence of a governing equation for the system, the phase space points are generated from the original time series using an embedding method [40]. Embedding theorem proposes that if an appropriate dimension is selected for a chaotic time series, the data can be reconstructed and the unfolded data can reveal the hidden information of the system.

A chaotic time series can be embedded in multi-dimensional space by plotting data point $Y(t)$, $t=0,1,2,\ldots,N-DT-1$ where $Y(t)=[x(t), x(t-T), \ldots, x(t-(D-1)T)]$, N is the length of the original time series, D is embedding dimension of the time series, and T is time delay [68]. The method of embedding observed data in a D-dimensional space is known as phase space reconstruction [1].

2.3. Embedding parameters

From an observed time series $x(t)$, phase space points $Y(t)=[x(t), x(t-T), \ldots, x(t-(D-1)T]$ can be generated, where T is the time delay, and D is the embedding dimension [68]. To apply Taken's theorem efficiently, appropriate selection of the embedding dimension, D and time delay, T, are required [12].

To select proper time delay, T, low correlation between nearby elements in the phase space points should be provided. Therefore, the first minimum of the average mutual information function can be selected [12]. The average mutual information will be used to give a precise definition to the notion that measurements $x(t)$ at time t are connected in an information theoretic fashion to measurements $x(t+T)$ at time $t+T$. The joint probability density for measurements at time t and $t+T$ resulting in values $x(t)$ and $x(t+T)$ and the individual probability densities for the measurements at time t and $t+T$ is required to calculate the average mutual information [1]. When T becomes large, the chaotic behaviour of the system makes the measurements $x(t)$ and $x(t+T)$ become independent in a practical sense, and the average mutual information value will tend to zero.

To choose a suitable embedding dimension, D, it is important to understand the impact of embedding dimension on the structure of attractor. As embedding dimension increases, the attractor unfolds. When attractor is unfolded completely, the same point on the attractor will not cross itself [12].

The method of false nearest-neighbours suggests that, when the attractor's trajectories come to cross, two neighbouring phase space points stand extremely apart from each other in the successive order of the embedded time series. Cao [4] has implemented this idea and has developed a method to estimate an effective embedding dimension [12].

3. Dynamic neural networks

Artificial neural networks (ANNs) are nonlinear methods which simulates nerve system [68]. Although a neural network model can predict nonlinear systems, it does not provide satisfactory results in direct prediction of chaotic time series.

Neural networks consist of extensive classes of various architectures. They are composed of a number of interconnecting elements named neurons. Neural networks can be categorized into dynamic and static networks. Static or feedforward networks do not have any delay or feedback elements. On the contrary, the output of dynamic networks depends on the current, previous inputs or outputs of the network [15]. In this study, two dynamic neural networks, Elman neural network and NARX neural network are used to develop the proposed prediction method.

3.1. Elman neural network

Elman networks belong to the class of recurrent neural networks (RNN) architecture. Elman networks are two-layer networks and contain additional feedback connections from the output of the hidden layer to the input layer [9]. Fig. 2 demonstrates that an Elman neural network contains four units of input, hidden, recurrent and output units. The recurrent units which transfer the previous state of the hidden units to the input layer are recognized as an one-step time delay [66].

At a certain time, the previous state of the hidden units and the present inputs are fed to the Elman network as inputs. With this arrangement, the network can be treated as a feedforward neural network and can process the inputs to generate the outputs. When the training starts, the recurrent units are receiving the feedbacks from the current state of the hidden units and the data will be kept for the next training step [66].

The primary input to the Elman network is $U(k)$ and the network output is $Y(k)$. The architecture of Elman network can be written mathematically by

$$y_k = f_o\left[\sum_{o=1}^{No} b_o + \sum_{h=1}^{Nh} w_{ho} f_h\left(b_h + \sum_{i=0}^{Ni} w_{ih} u_i + \sum_{j=0}^{Nh} w_{jh} a_h(k-1)\right)\right] \tag{3}$$

where w_{ih}, w_{jh} and $w_{ho}, i=1,2,\ldots,N_i;\ j,h=1,2,\ldots,N_h;\ o=1,2,\ldots,N_o$ are the weights of the connections between the input and hidden units, between the recurrent and the hidden units, and between the hidden and the output units, respectively. b_h and b_o are biases of hidden units and output units, and $f_h(\cdot)$ and $f_o(\cdot)$ are hidden and output functions, respectively [66].

3.2. NARX neural network

Nonlinear AutoRegressive with eXogenous input (NARX) neural network is another class of ANNs which are suitable to model nonlinear systems and time series [15]. NARX network is a dynamic neural network and contain recurrent feedbacks from several layers of the network to the input layer [27,29,40]. NARX can be mathematically represented by

$$y(n+1) = f[y(n),\ldots,y(n-d_y); u(n),\ldots,u(n-d_u)] \tag{4}$$

where $u(n) \in \Re$ and $y(n) \in \Re$ are the input and output of the model at discrete time step n respectively, and $d_u \geq 1$ is the input and $d_y \geq 1$, $(d_y \geq d_u)$ is the output delay. The function $f(\cdot)$ is generally unknown and can be approximated [38]. For the NARX architecture shown in Fig. 3 with one time series as input and one time series as output, the general NARX network equation can be written as

$$y(n+1) = f_o\left[b_o + \sum_{h=1}^{Nh} w_{ho} f_h\left(b_h + \sum_{i=0}^{d_u} w_{ih} u(n-i) + \sum_{j=0}^{d_y} w_{jh} y(n-j)\right)\right] \tag{5}$$

where w_{ih}, w_{jh} and $w_{ho}, i=1,2,\ldots,d_u, j=1,2,\ldots,d_y, h=1,2,\ldots,Nh$ are the weights, b_h and b_o are biases and $f_h(\cdot)$ and $f_o(\cdot)$ are hidden and output functions [38].

NARX network can be used in multi-time series input and multi-time series output applications. The equation for a general NARX network with two inputs and one output can be written as

$$y(n+1) = f_o\left[b_o + \sum_{h=1}^{N} w_{ho} f_h\left(b_h + \sum_{i1=0}^{d_{u1}} w_{i1h} u_1(n-i1) + \sum_{i2=0}^{d_{u2}} w_{i2h} u_2(n-i2) + \sum_{j=0}^{d_y} w_{jh} y(n-j)\right)\right] \tag{6}$$

where w_{i1h} and $w_{i2h}, i_1=1,2,\ldots,d_{u1}, i_2=1,2,\ldots,d_{u2}$ are the weights of the connections between the first input units and the hidden units and between the second input units and the hidden units, respectively [38].

In the current study, Elman network is trained by gradient descent with momentum and adaptive learning rate backpropagation algorithm [15] and NARX network is trained by Levenberg–Marquardt with Bayesian Regularization method [10,33].

4. Residual analysis and prediction method

Heuristic models which are developed to predict time series mainly overlook the characteristics of residuals. In many occasions heuristic models cannot completely capture the characteristics of the original time series, and the residuals show high correlation. The high correlation of the residuals means that prediction errors are not due to randomness. This study considers this crucial finding and in this section introduces a new technique which employs a combination of residual analysis using hybrid Elman–NARX neural networks and embedding theorem to develop an efficient method to forecast one-step ahead of the chaotic time series.

Based on the observed time series data set, $x(1)$, $x(2)$, $x(3)$, …, $x(N)$, where N is the length of the data set, the presence of chaos is examined by Fourier transforms, Lyapunov exponent, Hurst exponent and Fractal dimensions as mentioned in Section 2.1. When the presence of chaos is detected in the observed time series, then the phase space parameters, embedding dimension D and time delay T, are calculated as described in Section 2.3. Based on the discussion in Section 2.2, $[N-(D-1)T]$ phase space points

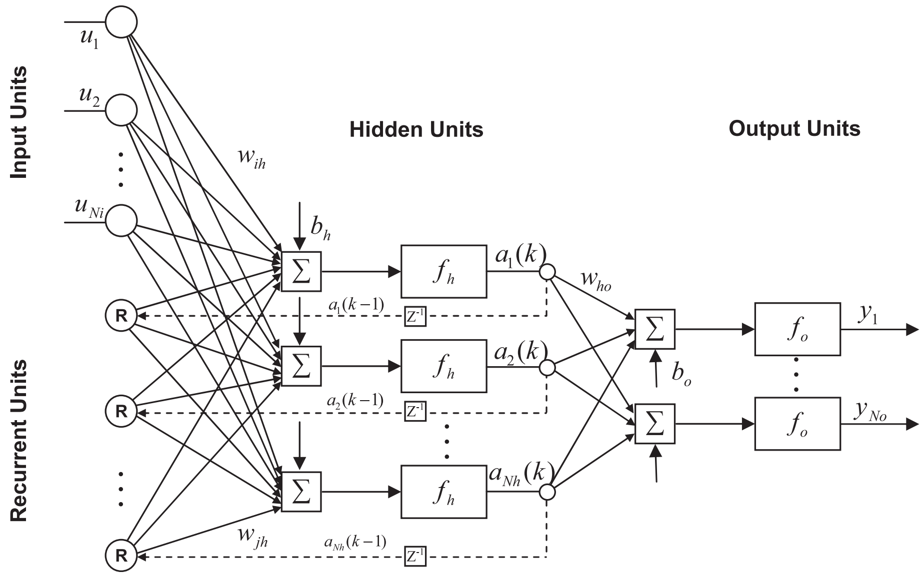

Fig. 2. The architecture of Elman recurrent neural network.

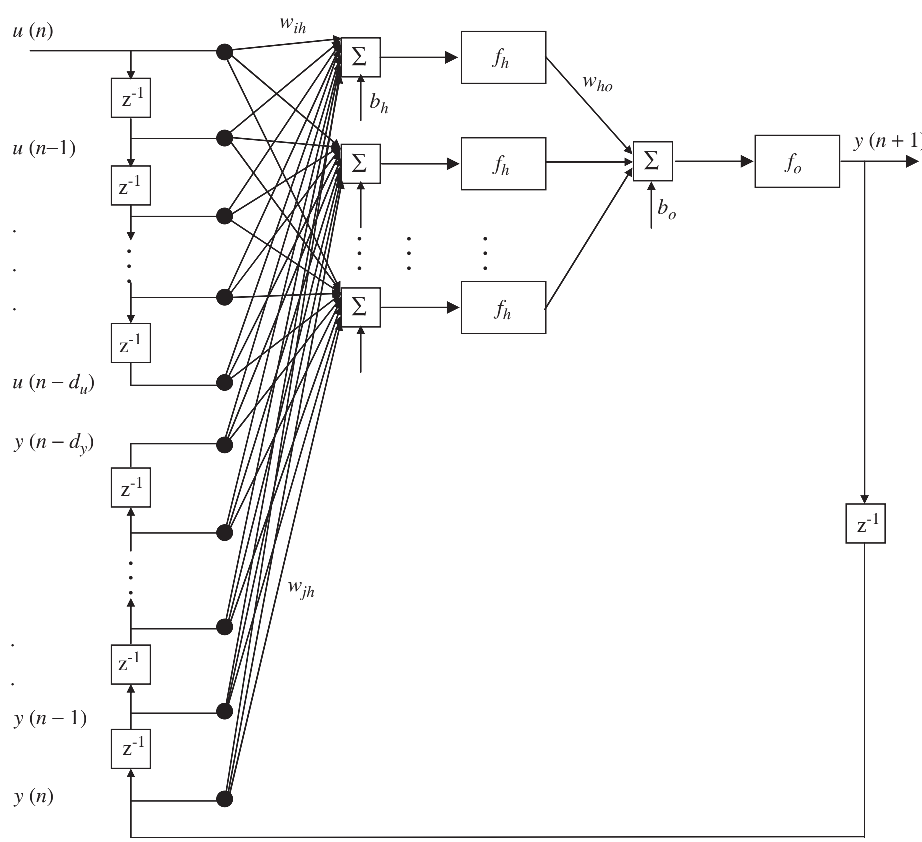

Fig. 3. The architecture of NARX neural network.

$Y(j)$, $j=1\sim[N-(D-1)T]$, are generated as

$$\begin{aligned}
&Y(1)=[x(1),x(1+T),\ldots,x(1+(D-1)T)]\\
&Y(2)=[x(2),x(2+T),\ldots,x(2+(D-1)T)]\\
&\vdots\\
&Y(j)=[x(N-(D-1)T),x(N-(D-1)T+T),\ldots,x(N)]
\end{aligned}\tag{7}$$

An Elman recurrent neural network with D input and D output units is trained using the generated phase space points as inputs, $Y(i)$, $i=1\sim[N-(D-1)T-1]$ and output $Y(k)$, $k=2\sim[N-(D-1)T]$. The input and output of the network are denoted by Eqs. (8) and (9) respectively:

$$\begin{bmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(N-(D-1)T-1) \end{bmatrix}\tag{8}$$

$$\begin{bmatrix} Y(2) \\ Y(3) \\ \vdots \\ Y(N-(D-1)T) \end{bmatrix} \tag{9}$$

After training the Elman network, the weights and biases of the network are kept to forecast the unknown phase space points. When the neural network training is completed, based on the known vector, $Y(N-(D-1)T)$ as the inputs to the trained neural network, the unknown vector, $Y(N-(D-1)T+1)$ is determined. $Y(N-(D-1)T+1)$ is composed of known values $x(N-(D-1)T+1)$, $x(N-(D-1)T+1+T),\ldots,$ $x(N-(D-1)T+1+(D-2)T)$ and unknown value $x(N+1)$. When $Y(N-(D-1)T+1)$ is determined, the only unknown value, $x(N+1)$, can be obtained.

So far, we have found one-step ahead prediction for original time series. It has been indicated earlier that in many occasions heuristic models cannot completely capture the characteristics of the original time series and therefore, the residuals are highly correlated. The residuals resulted from the developed neural network prediction method are analyzed, and it was observed that in some occasions they demonstrate chaotic time series.

To investigate the contribution of residual analysis in improving the performance and accuracy of the proposed prediction method, the obtained residuals are considered as a new chaotic time series and analyzed separately. The residuals obtained from original prediction form a time series,$e_1(1)$, $e_1(2),\ldots,e_1(N_1)$, where $N_1=N-(D-1)T-1$ is the length of residual time series. The chaotic behaviour of the residual time series is investigated using the methods mentioned in Section 2. When the presence of chaos is confirmed in the residual time series, the embedding dimension,D_1 and the time delay,T_1 will be determined and the phase space points, $Y_1\ (j_1)$, $j_1=1\sim[N_1-(D_1-1)T_1]$, will be generated as

$$\begin{aligned} Y_1(1) &= [e_1(1),e_1(1+T_1),\ldots,e_1(1+(D_1-1)T_1)] \\ Y_1(2) &= [e_1(2),e_1(2+T_1),\ldots,e_1(2+(D_1-1)T_1)] \\ &\vdots \\ Y_1(j_1) &= [e_1(N_1-(D_1-1)T_1),e_1(N_1-(D_1-1)T_1+T_1),\ldots,e_1(N_1)] \end{aligned} \tag{10}$$

A new Elman network with D_1 input units and D_1 output units is trained using the phase space points as inputs, $Y_1\ (i_1)$, $i_1=1\sim[N_1-(D_1-1)T_1-1]$ and output $Y_1(k_1)$, $k_1=2\sim[N_1-(D_1-1)T_1]$. The input and output of the Elman network are denoted by Eqs. (11) and (12) respectively:

$$\begin{bmatrix} Y_1(1) \\ Y_1(2) \\ \vdots \\ Y_1(N_1-(D_1-1)T_1-1) \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} Y_1(2) \\ Y_1(3) \\ \vdots \\ Y_1(N_1-(D_1-1)T_1) \end{bmatrix} \tag{12}$$

When the network training is completed, based on the known vector, $Y_1(N_1-(D_1-1)T_1)$ as the input to the trained neural network, the unknown vector, $Y_1(N_1-(D_1-1)T_1+1)$ is predicted. $Y_1(N_1-(D_1-1)T_1+1)$ is composed of known values $e_1(N_1-(D_1-1)T_1+1)$, $e_1(N_1-(D_1-1)T_1+1+T_1),\ldots,e_1(N_1-(D_1-1)T_1+1+(D_1-2)T_1)$ and unknown value $e_1(N_1+1)$. When $Y_1(N_1-(D_1-1)T_1+1)$ is predicted, the only unknown value, $e_1(N_1+1)$, can be determined.

To correlate the original and residual time series prediction to original time series, a NARX network is trained. The length of residual time series prediction is $N_1-(D_1-1)T_1-1$; therefore, the last $N_1-(D_1-1)T_1-1$ data from original time series prediction and original time series are selected. The NARX network with two input units and one output unit is trained using the original and residual time series prediction values as input and original time series as output. We denote the prediction values of original time series with $x'(i)$ and residual time series with $e'_1(i)$. Fig. 4 demonstrates the NARX network structure.

To increase the accuracy of the prediction, the residual analysis can be repeated several times. To correlate the original time series prediction and the prediction of residuals in various levels to original time series, a NARX network with $M+1$ input units, where M is the level of residual analysis, and one output unit can be trained.

5. Numerical Experiments

In order to evaluate the proposed methods, the developed technique is applied to two simulated chaotic systems, Mackey–Glass and Lorenz and one real life time series, Sunspot time series. Later, the performance of the prediction method is compared with the results reported in the literature for both Mackey–Glass and Lorenz equations and Sunspot time series.

5.1. Lorenz equations

An evident example of chaotic systems is the weather. Lorenz's works have extensively contributed to the establishment of chaos theory [30]. Lorenz equations are a simplified version of Navier–Stokes equations which are used in fluid mechanics. Lorenz equations are written as

$$\begin{aligned} \frac{dx(t)}{dt} &= \sigma[y(t)-x(t)] \\ \frac{dy(t)}{dt} &= x(t)[r-z(t)]-y(t) \\ \frac{dz(t)}{dt} &= x(t)y(t)-bz(t) \end{aligned} \tag{13}$$

where σ, r, and b are dimensionless parameters and the typical values for these parameters are $\sigma=10$, $r=28$ and $b=8/3$ [14,32]. The x-coordinate of the Lorenz time series is considered for prediction and a time series with a length of 2500 is generated as described in the research work by Ma et al. [32]. The first 1500 samples were used as training data, while the remaining 1000 were used to test the proposed model. The data in time domain is shown in Fig. 5(a). The embedding dimensions $D=3$ and $T=3$ are estimated as explained in Section 4 using TSTOOL [57], add-on toolbox in MATLAB and accordingly phase space points are reconstructed. Fig. 5(b) presents the embedded time series in phase space domain.

Our Elman neural network consists of 3 neurons in the input layer, 6 neurons in the hidden layer, 6 recurrent neurons and 3 neurons in the output layer. The network is trained using gradient descent with momentum and adaptive learning rate backpropagation algorithm. The weights and biases of the network are kept for prediction purposes. Fig. 6 compares the real data and the predicated values for the remaining 1000 test samples using Elman neural network.

The prediction errors for the data selected for the training phase are calculated and analyzed. The methods introduced in Section 2.1 are used to investigate the characteristic of the residuals. The power spectrum for the residuals is broadband and has a broad peak. The largest Lyapunov exponent (LLE) and Fractal dimension (FD) for the time series are calculated using TSTOOL [57] and their values are 0.47 (positive) and 1.513 (non-integer) respectively. Therefore, it is concluded that the residuals show chaotic behaviour.

Based on this finding, the residuals are treated as a new chaotic time series and the prediction method proposed in Section 4is applied to predict the future value of the residuals. The first

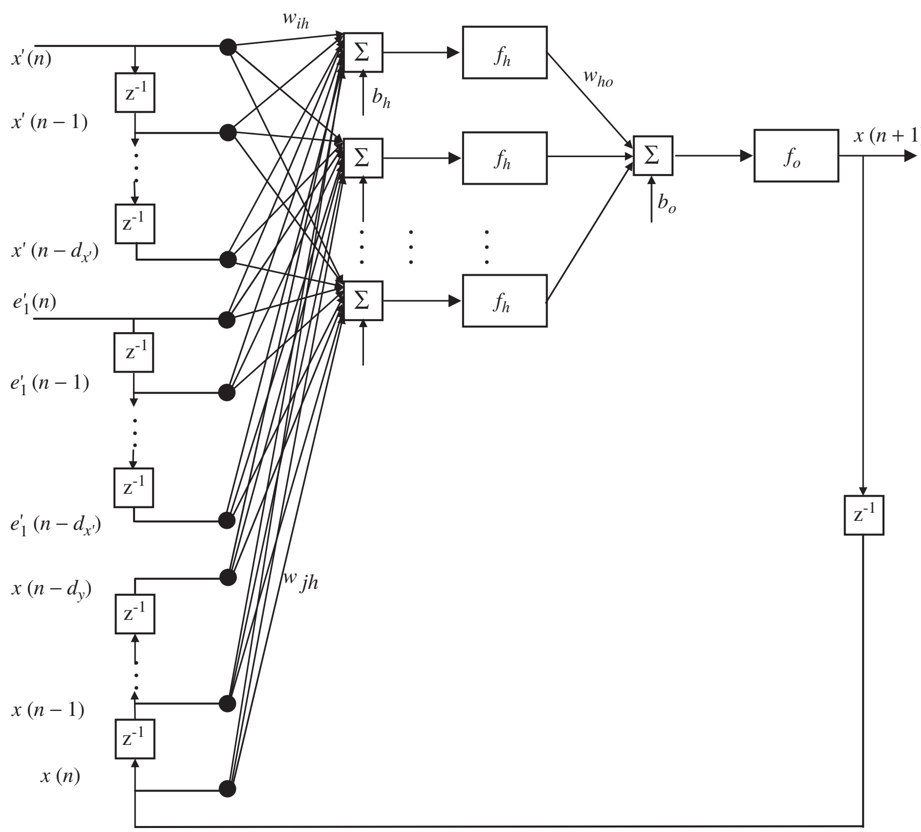

Fig. 4. A NARX network correlating the original and residuals forecasts to original time series.

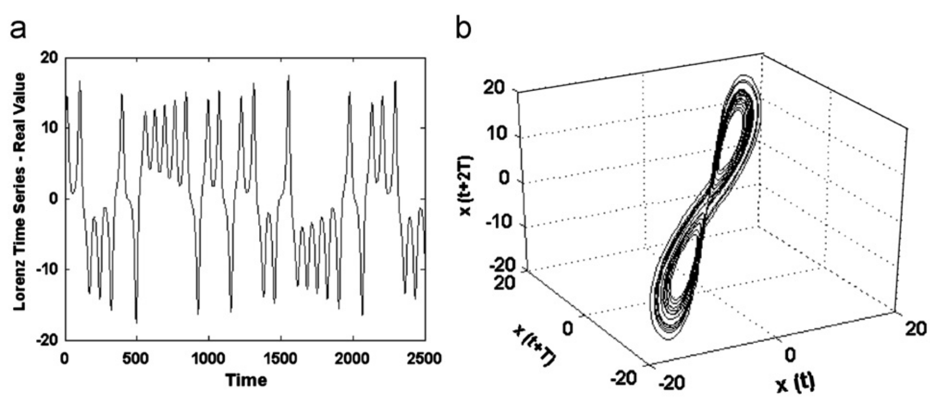

Fig. 5. Lorenz time series: (a) in time domain and (b) in phase space domain with $D=3$ and $T=3$.

$(D-1)T+1$ observations will help to predict the first prediction; therefore, the length of residual time series is $(D-1)T+1$ shorter than original time series. It is evident that residuals do not inherent the same characteristics as the original time series do and they need to be analyzed independently. Therefore, 1493 samples are reconstructed with embedding dimension $D_1=3$ and time delay $T_1=1$.

Fig. 7(a) and (b) show the first residual time series in time domain and embedded phase space respectively. Fig. 8 presents the results of the first order residuals prediction.

Finally, a NARX network is trained using predicted value of original time series and the residuals as input and the original time series as output to correlate the predicted value of original time series and residuals to original time series. Fig. 9 compares the original values of Lorenz time series with their predicted values.

In order to evaluate the prediction performance and compare it with the results reported in the literature, mean squared error (MSE), root mean squared error (RMSE) and normalized mean squared error (NMSE) are calculated according to Eqs. (14),(15) and (16) respectively.

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\widehat{y}_i)^2 \tag{14}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\widehat{y}_i)^2} \tag{15}$$

$$NMSE=\left(\frac{\sum_{i=1}^{N}(y_i-\widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\overline{y})^2}\right) \tag{16}$$

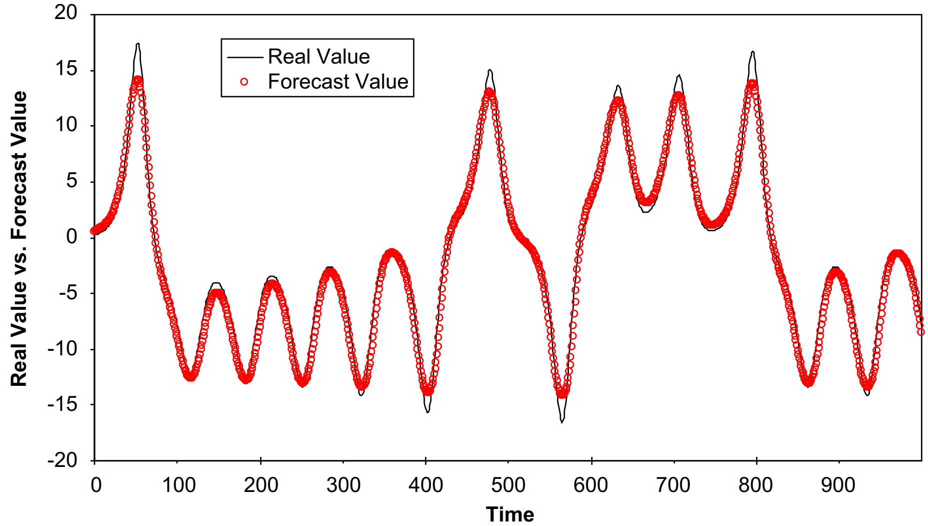

Fig. 6. Lorenz time series – the comparison of the original time series samples and the predicted values based on Elman network.

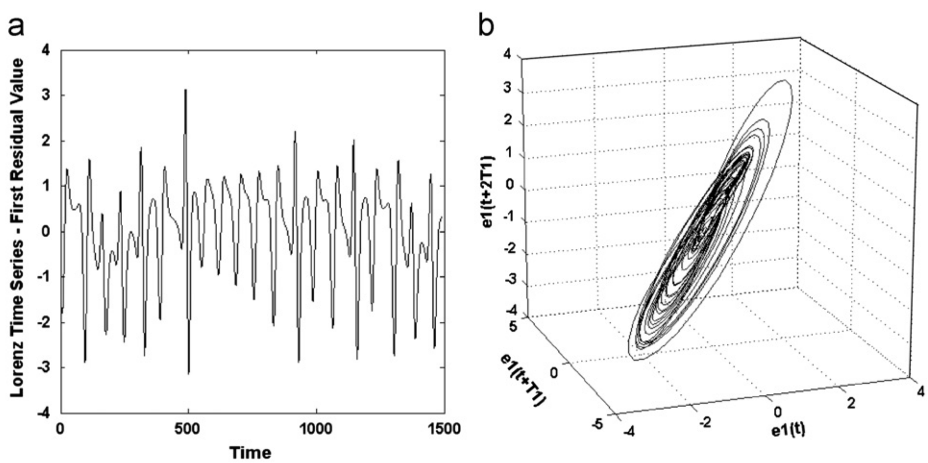

Fig. 7. Lorenz time series L: (a) the first residual value and (b) reconstructed with $D_1=3$ and $T_1=1$.

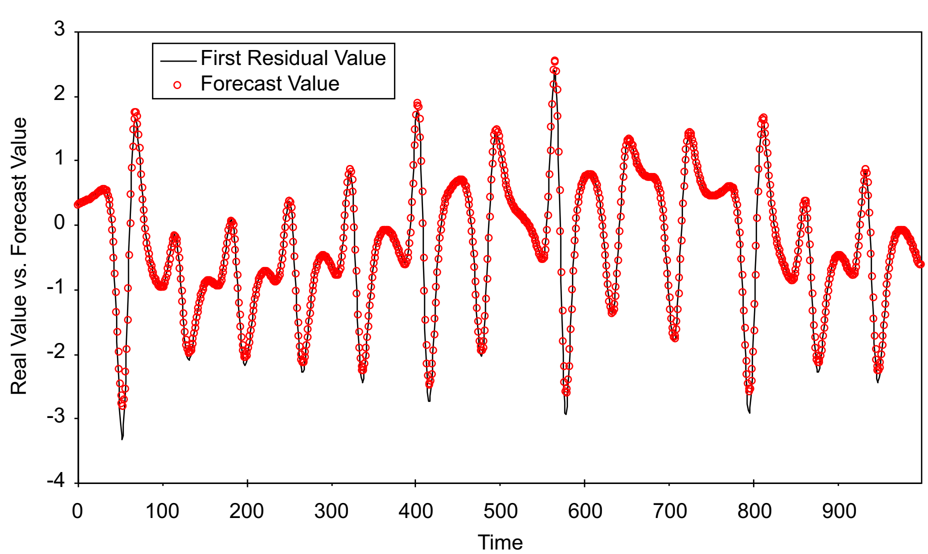

Fig. 8. Lorenz time series – the comparison of the first residual and predicted values.

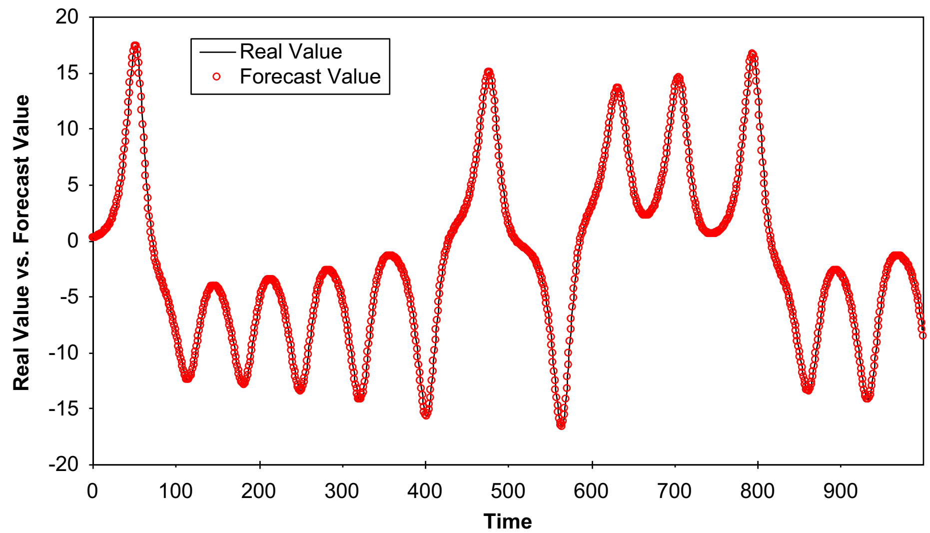

Fig. 9. The comparison between original Lorenz time series and the final predicted values.

Table 1
Comparison between the prediction errors reported in the literature and the proposed method (1000 Lorenz time series test samples).

Prediction method	*Reference*	*Prediction error*		
		MSE	*RMSE*	*NMSE*
Adaptive neuro-fuzzy inference systems (ANFIS)	Jang et al. [20]		0.143	
Pseudo Gaussian – radial basis function	Rojas et al. [45]		0.094	
ARMA – neural network	Rojas et al. [46]		0.0876	
Support vector regression	Martinez-Rego et al. [35]			1.46E-02
TDL-MLP	Martinez-Rego et al. [35]			1.56E-04
DLE-VQIT	Martinez-Rego et al. [35]			2.58E-04
RBF with orthogonal LS	Gholipour et al. [14]			1.41E-09
LLNF-LoLiMot	Gholipour et al. [14]			9.80E-10
CIFCA-ROLSA	Tao and Xiao [54]	1.30E-03		
ROLSA	Tao and Xiao [54]	4.63E-02		
Evolving recurrent neural network (ERNN)	Ma et al. [32]		8.79E-06	9.90E-10
Boosted recurrent neural networks (BRNN)	Assaad et al. [2]			3.77E-03
MDE-RBF	Dhahri and Alimi [7]		1.70E-01	
Wavelet-networks B	Garcia and Alarcon [13]	1.64E-02		
The proposed method	–	1.1731E-8	1.0831E-4	1.983E-10

where y_i, $\hat{y}_i$ and $\bar{y}$ are observed data, predicted data and the average of observed data respectively, and $N$is the length of observed data.

The error indices, NMSE for the 1000 Lorenz time series test samples, for the demonstrated simulation are 0.3219 for the model using only the combination of embedding theorem and artificial intelligence and 0.0836 for the proposed algorithm including residual analysis. The prediction accuracy of the method when applied to Lorenz time series is compared to the prediction performance of the methods presented in the literature. Table 1 presents the comparison between the prediction errors, MSE, RMSE and NMSE reported in the literature and the proposed method for 1000 Lorenz test samples.

This simulation confirms that the proposed method, when it is used to forecast the future values of Lorenz time series, generates better results compared to other prediction methods reported in the literature.

5.2. Mackey–Glass equation

The Mackey–Glass equation originally has been proposed as a model of blood cell regulation [34]. Mackey–Glass has been used in literature as a benchmark model due to its chaotic characteristics. The differential equation leading to the time series is demonstrated in

$$\frac{dx}{dt} = \frac{ax(t-\tau)}{[1+x^c(t-\tau)]} - bx(t) \tag{17}$$

In the Mackey–Glass equation, the delay parameter, τ, determines the characteristic of Eq. (17); i.e. $\tau < 4.43$ produces a fixed point attractor, $4.43 < \tau < 13.3$ stable limit cycle attractor, $13.3 < \tau < 16.8$ double limit cycle attractor and $\tau > 16.8$ chaos.

The parameters are selected according to the previous report by Zhang and Man [66], where the constants are taken to be $a=0.2$, $b=0.1$ and $c=10$ and chaotic time series is generated by time delay $\tau=17$ and initial value $x(0)=1.2$.

A chaotic time series samples set with length of 1000 is generated by the Eq. (17). The first 500 samples were used as training data, while the remaining 500 were used to test the proposed model. The data in time domain is shown in Fig. 10(a). The original time series is reconstructed with embedding dimension $D=3$ and time delay $T=7$ as shown in Fig. 10(b).

The embedded phase space points are fed into an Elman neural network and trained using gradient descent with momentum and adaptive learning rate backpropagation algorithm. Fig. 11

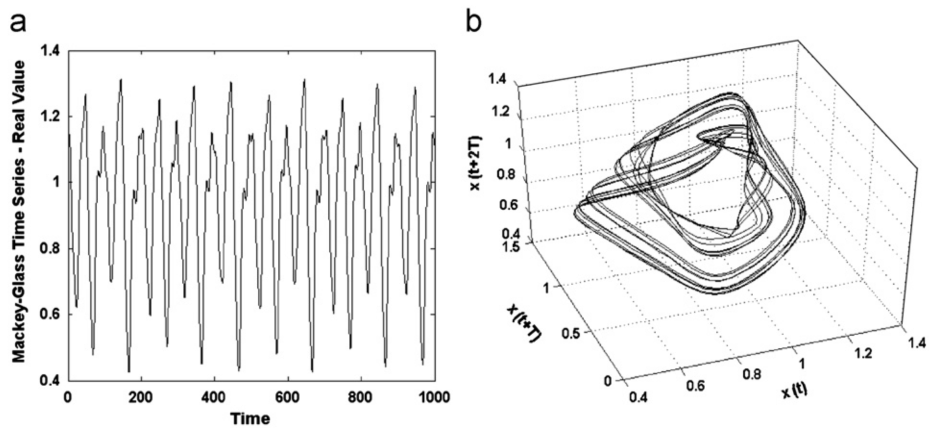

Fig. 10. Mackey–Glass time series: (a) in time domain and (b) in phase space domain reconstructed with $D=3$ and $T=7$.

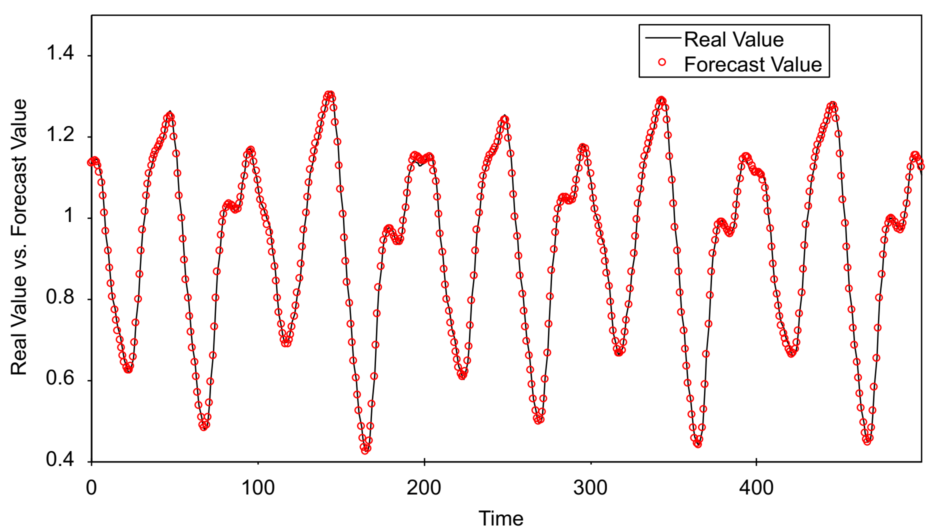

Fig. 11. Mackey–Glass time series – comparing the original time series samples and the predicted values.

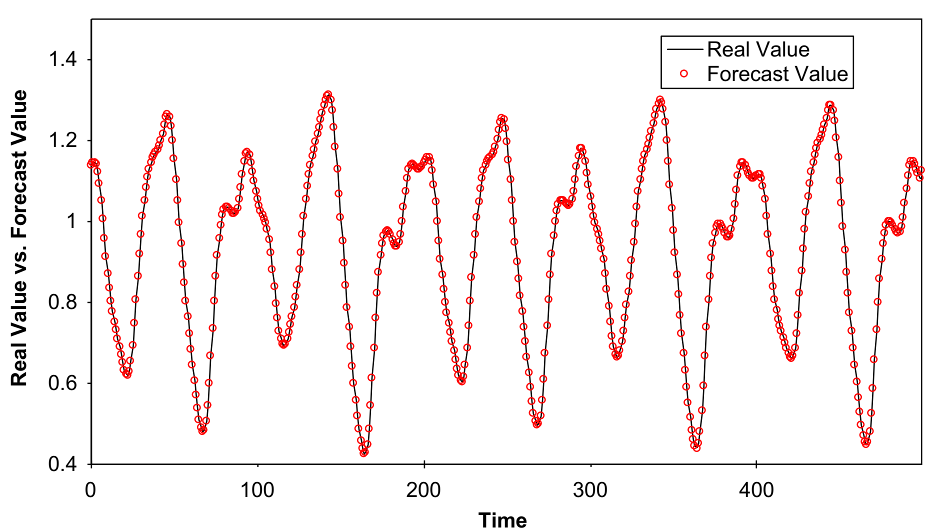

Fig. 12. The comparison between original Mackey–Glass time series and the forecast values.

compares the real data and predicated value for the remaining 500 samples kept for testing.

Two levels of residual analysis are performed to increase the accuracy of the prediction. The first order residuals, 485 samples, are reconstructed with embedding dimension $D_1=3$ and time delay $T_1=3$. Based on broadband power spectrum, positive LLE and non-integer FD, it is concluded that the residuals are chaotic. The reconstructed phase space points are fed into a

Table 2
Comparison between the prediction errors reported in the literature and the proposed method (500 Mackey–Glass time series test samples).

Prediction method	*Reference*	*Prediction error*		
		MSE	*RMSE*	*NMSE*
AutoRegressive model	Rojas et al. [45]		0.19	
Genetic algorithm and fuzzy system – 9 MFs	Kim and Kim [23]		3.8E-02	
ANFIS and fuzzy system	Jang et al. [20]		7.0E-03	
Pseudo Gaussian – radial basis function	Rojas et al. [45]		2.8E-03	
ARMA – neural network	Rojas et al. [46]		2.5E-03	
WP-MLP (A)	Teo et al. [56]	1.25E-06		
RBF-OLS	Gholipour et al. [14]		1.02E-03	
LLNF-LoLiMot	Gholipour et al. [14]		9.61E-04	
ANFIS	Gholipour et al. [14]		1.50E-03	
Neuro-fuzzy system with delay	Zhang et al. [65]			1.26E-03
ERNN	Ma et al. [32]		4.20E-05	3.15E-08
BRNN	Assaad et al. [2]			1.60E-04
Hybrid neural network (HNN)	Inoue et al. [19]			5.30E-02
The proposed method	–	1.3855E-9	3.7222E-5	2.7040E-8

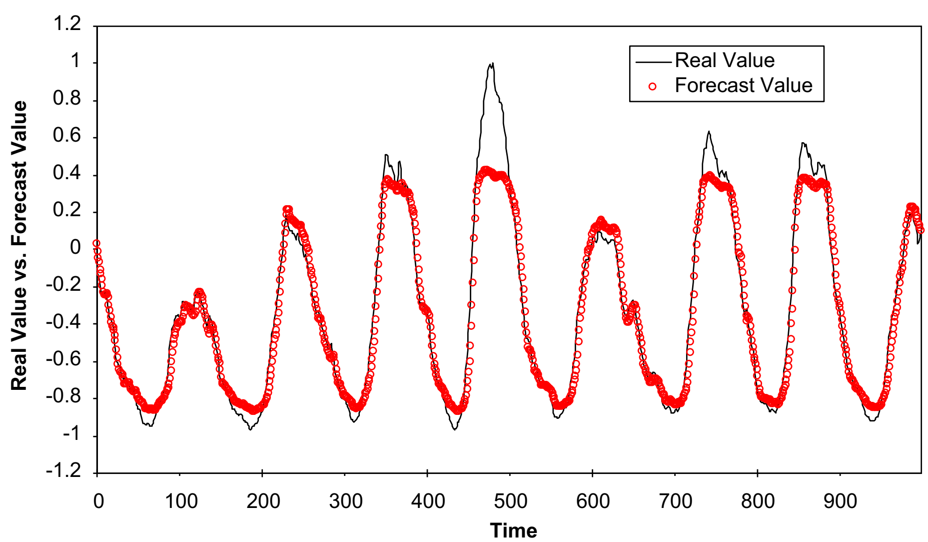

Fig. 13. Sunspot time series – the comparison of the original time series samples and the predicted values based on Elman network.

new Elman neural network. The neural network is set up with 3 neurons in input and output layers and 10 neurons in the hidden layer. The residual analysis is repeated one more time with $D_2=3$ and time delay $T_2=3$ and after confirmation that the residuals are chaotic, the future values of the second order residuals are predicted based on the proposed prediction method. And finally, a NARX network is trained using forecast value of original time series and the residuals as input and the original time series as output. The last neural network is used to capture the relationship among the forecast value of original time series and residuals and original time series. Fig. 12 compares the original values of Mackey–Glass time series with the forecast values.

To evaluate the proposed method, the prediction accuracy of the method when applied to Mackey–Glass time series as benchmarking is compared to the prediction performance of the methods presented in the literature. Table 2 presents the comparison between the prediction errors, MSE, RMSE and NMSE, reported in the literature and the proposed method for 500 Mackey–Glass test samples. The results presented in Table 2 show that the proposed method performs better in prediction of Mackey–Glass chaotic time series when compared to other prediction methods reported in the literature.

5.3. *Sunspot time series*

Forecasting solar activity is a challenging area and important topic for various researchers and industries [49]. The Sunspot time series is a good indication of solar activity for solar cycles. The impact of solar activity has been observed on earth, climate, weather, satellites and space missions; therefore, it is critical to forecast Sunspot time series. However, because of the complexity of the system and the lack of a mathematical model, forecasting solar cycle is extremely challenging [14].

The monthly smoothed Sunspot time series has been obtained from the SIDC (World Data Center for the Sunspot Index) [50]. To compare the results with some of the research works published in THE literature, data are selected in the same conditions reported by Ma et al. [32] and Gholipour et al. [14]. Sunspot series from November 1834 to June 2001 (2000 points) are selected and scaled between [0,1]. The first 1000 samples of time series are selected to train and the remainder 1000 samples are kept to test the prediction method. The parameters $D=5$ and $T=1$ are estimated and accordingly phase space points are reconstructed.

The Elman neural network consists of 5 neurons in the input layer, 7 neurons in the hidden layer, 7 recurrent neurons and 5 neurons in the output layer. The network is trained using gradient

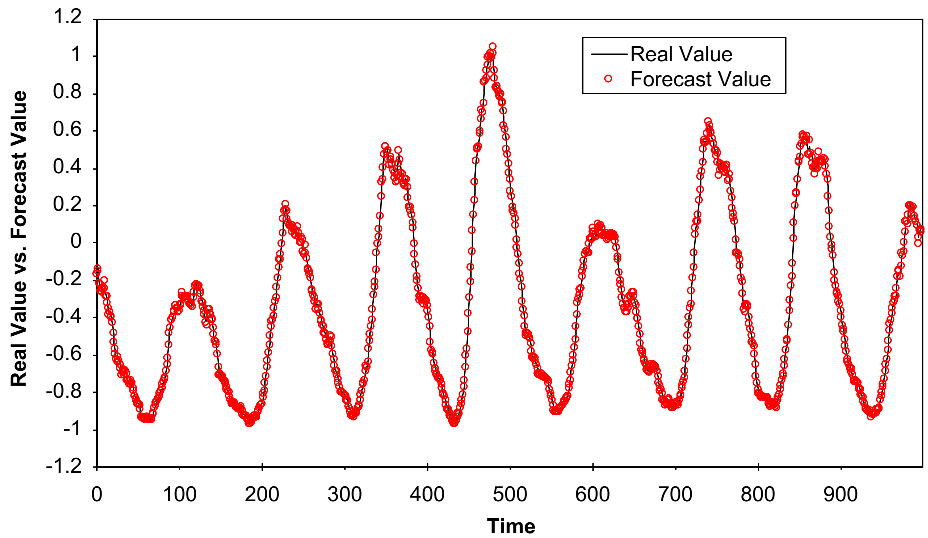

Fig. 14. The comparison between original sunspot time series and the final predicted values.

Table 3
Comparison between the prediction errors reported in the literature and the proposed method (1000 Sunspot time series test samples).

Prediction method	*Reference*	*Prediction error*		
		MSE	*RMSE*	*NMSE*
WP-MLP (A)	Teo et al. [56]			1.25E-01
McNish–Lincoln	McNish and Lincoln [37]			8.00E-02
Sello – nonlinear method	Sello [49]			3.40E-01
Waldmeier	Sello [49]			5.60E-01
Denkmayr	Denkmayr and Cugnon [6]			1.85
RBF-OLS	Gholipour et al. [14]			4.60E-02
LLNF-LoLiMot	Gholipour et al. [14]			3.20E-02
ERNN	Ma et al. [32]		1.29E-02	2.80E-03
Multi-layer perceptron (MLP)	Koskela et al. [25]			9.79E-02
The proposed method	–	1.4078E-4	0.0119	5.9041E-4

descent with momentum and adaptive learning rate backpropagation algorithm. Fig. 13 compares the real data and predicated value for the remaining 1000 test samples using Elman neural network.

The prediction errors for the samples selected for the training are calculated and analyzed. The power spectrum is broadband with a broad peak and the calculated LLE and FD are positive and non-integer respectively. Therefore, the residuals show chaotic behaviour. The 995 residual points, are reconstructed with embedding dimension $D_1=5$ and time delay $T_1=1$. The second Elman neural network is trained using gradient descent with momentum and adaptive learning rate backpropagation algorithm. Another level of residual analysis is performed to increase the accuracy of the prediction. The residual analysis shows that the second order residuals are chaotic too; therefore, 990 residual points, are reconstructed with embedding dimension $D_2=3$ and time delay $T_2=1$. Finally, a NARX network is trained using predicted value of original time series and the residuals as input and the original time series as output to correlate the predicted value of original time series and residuals to original time series. Fig. 14 compares the original values of Sunspot time series with their predicted values.

The error indices, MSE, RMSE and NMSE for the 1000 Sunspot time series test samples are compared with the results reported in the literature and presented in Table 3. The experiment performed for Sunspot time series confirms that the performance of the proposed method is better in prediction of Sunspot chaotic time series when compared to other prediction methods reported in the literature.

6. Conclusion

In this study, a new prediction method was developed using a combination of embedding theorem, residual analysis and hybrid Elman–NARX neural networks to predict chaotic time series. Based on the embedding theorem, the original time series can be unfolded with embedding dimension and time delay and reconstructed into the phase space. An Elman neural network with gradient descent with momentum and adaptive learning rate backpropagation training algorithms was used to predict future values of the embedded phase space points. In some occasions, the residuals of the forecasting method are not random and follow the characteristics of the original time series. The accuracy of the prediction method was further enhanced by analyzing and incorporating the prediction residuals. The residual analysis was repeated several times. Eventually to capture the relationship among the predicted value of original time series, residuals and original time series a NARX neural network is trained using Levenberg–Marquardt with Bayesian Regularization method.

The developed method has been validated for one-step-ahead prediction and the performance of the proposed forecasting method in terms of error index has been compared with the performance of reported forecasting methods in the literature.

The issue of long term forecasting and the evaluation of the performance of the proposed forecasting method in terms of computational complexity of the algorithm and the number of parameters required to be optimized can be considered as an extension to the present study.

In this study, the neural networks' parameters are determined by trial and error. In future work, the application of Design of Experiments for the parameter setting and optimization of a neural network can be tested, and some factors such as the minimum data required to present a reasonably accurate forecasting, training methods, number of epochs, number of hidden neurons, etc. will be set and optimized.

Despite its simplicity in implementation, the proposed model exhibited a superior performance in prediction of chaotic time series. The developed algorithm was applied to two chaotic systems, Mackey–Glass and Lorenz equations and the real life Sunspot time series. It has been concluded that the proposed model can perform more effectively and accurately as compared to the other prediction methods for chaotic time series.

This technique can be further applied to other real world chaotic time series which have been observed in the various fields namely marketing systems, foreign exchange rate, signal processing, supply chain management, traffic flow, power load, and weather forecast. The effectiveness of the proposed model can reduce the risk and increase the productivity of the mentioned areas.

Acknowledgements

The authors wish to thank anonymous reviewers for their valuable suggestions. The financial support of the Natural Sciences and Engineering Research Council Canada is also greatly appreciated.

References

[1] H. Abarbanel, in: Analysis of Observed Chaotic Data, Spinger-Verlag, New York, 1996.
[2] M. Assaad, R. Boné, H. Cardot, Predicting chaotic time series by boosted recurrent neural networks, in: Proceedings of the International Conference on Neural Information Processing, vol. 4233, 2006, pp. 831–840.
[3] G.A. Barreto, Time series prediction with the self-organizing map: a review, in: Pascal Springer-Verlag, Hitzler, Barbara Hammer (Eds.), Perspectives on Neural-Symbolic Integration, 2006.
[4] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D 110 (1997) 43–50.
[5] A. Das, P. Das, Chaotic analysis of the foreign exchange rates, Applied Mathematics and Computation 185 (2007) 388–396.
[6] K. Denkmayr, P. Cugnon, About sunspot number medium-term predictions, in: G. Heckman et al.(Ed.), Solar-Terrestrial Prediction Workshop V, Hiraiso Solar Terrestrial Research Center, Japan1997, p. 103.
[7] H. Dhahri, A.M. Alimi, The modified differential evolution and the RBF (MDE-RBF) neural network for time series prediction, in: Proceedings of the International Joint Conference on Neural Networks, Vancouver, Canada, 2006, pp. 2938–2943.
[8] E. Diaconescu, The use of NARX neural networks to predict chaotic time series, WSEAS Transactions on Computer Research 3 (3) (2008) 182–191.
[9] J.L. Elman, Finding structure time, Cognitive Science 14 (1990) 179–211.
[10] D. Foresee, M. Hagan, Gauss–Newton approximation to Bayesian learning, in: Proceedings of the International Joint Conference on Neural Networks, 1997, pp. 1930–1935.
[11] R.J. Frank, N. Davey, S.P. Hunt, Time series prediction and neural networks, Journal of Intelligent and Robotic Systems 31 (1–3) (2001) 91–103.
[12] C. Frazier, K.M. Kockelman, Chaos Theory & Transportation Systems: an Instructive, Transportation Research Record 1897 (2004) 9–17.
[13] E.S. Garcia-Trevino, V. Alarcon-Aquino, Single-step prediction of chaotic time series using wavelet-networks, in: Proceedings of the Electronics, Robotics and Automotive Mechanics Conference, vol. 1, 2006, pp. 243–248.
[14] A. Gholipour, B.N. Araabi, C. Lucas, Predicting chaotic time series using neural and neurofuzzy models: a comparative study, Neural Processing Letters 24 (2006) 217–239.
[15] M.T. Hagan, H.B. Demuth, M.H. Beale, in: Neural Network Design, PWS Publishing, Boston, 1996.
[16] C.M. Harris, R.W. Todd, S.J. Bungard, R.W. Lovitt, J.G. Morris, D.B. Kell, The dielectric permittivity of microbial suspensions at radio-frequencies: a novel method for the estimation of microbial biomass, Enzyme Microbial Technology 9 (1987) 181–186.
[17] S. Haykin, in: Neural Networks: a Comprehensive Foundation, Prentice Hall PTR, 1998.
[18] R.C. Hilborn, in: Chaos and Nonlinear Dynamics: an Introduction for Scientists and Engineers, 2nd edition, Oxford University Press, Oxford, 2001.
[19] H. Inoue, Y. Fukunaga, H. Narihisa, Efficient hybrid neural network for chaotic time series prediction, in: Proceedings of the International Conference on Artificial Neural Networks, 2001, pp. 712–718.
[20] J.S.R. Jang, C.T. Sun, E. Mizutani, in: Neuro-Fuzzy and Soft Computing, Prentice-Hall, Englewood cliffs, New Jersey, 1997.
[21] D. Kaplan, L. Glass, in: Understanding Nonlinear Dynamics, Springer-Verlag, New York, 1995.
[22] M.B. Kennel, Method to distinguish possible chaos from colored noise and to determine embedding parameters, Physical Review A 46 (1992) 3111–3118.
[23] D. Kim, C. Kim, Forecasting time series with genetic-fuzzy predictor ensemble, IEEE Transactions on Fuzzy Systems 5 (1997) 523–535.
[24] V. Kodogiannis, A. Lolis, Forecasting financial time series using neural network and fuzzy system-based techniques, Neural Computing and Applications 11 (2002) 90–102.
[25] T. Koskela, M. Lehtokangas, J. Saarinen, K. Kaski, Time series prediction with multilayer perceptron, FIR and Elman neural networks, in: Proceedings of the World Congress on Neural Networks, 1996, pp. 491–496.
[26] T. Koskela, M. Varsta, J. Heikkonen, S. Kaski, Time series prediction using recurrent SOM with local linear models, International Journal of Knowledge-based Intelligent Engineering Systems 2 (1) (1998) 60–68.
[27] I.J. Leontaritis, S.A. Billings, Input–output parametric models for nonlinear systems—part I: deterministic nonlinear systems, International Journal of Control 41 (2) (1985) 303–328.
[28] F. Liu, C. Quek, G.S. Ng, Neural network model for time series prediction by reinforcement learning, in: Proceedings of the International Joint Conference on Neural Networks, Montreal, Canada, 2005.
[29] L. Ljung, System Identification: Theory for the User, 2nd Edition, Prentice-Hall, Englewood Cliffs, New Jersey, 1999.
[30] E.N. Lorenz, Deterministic non-periodic flows, Journal of Atmospheric Science 20 (1963) 130–141.
[31] E.N. Lorenz, in: The Essence of Chaos, University of Washington Press, Seattle, 1993.
[32] Q. Ma, Q. Zheng, H. Peng, T. Zhong, L. Xu, Chaotic time series prediction based on evolving recurrent neural networks, in: Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 2007.
[33] D.J.C. MacKay, Bayesian interpolation, Neural Computation 4 (3) (1992) 415–447.
[34] M.C. MacKey, L. Glass, Oscillation and chaos in physiological control systems, Science 197 (4300) (1977) 287–289.
[35] D. Martinez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, A method for time series prediction using a combination of linear models, in: Proceedings of the European Symposium on Artificial Neural Network – Advances in Computational Intelligence and Learning, Bruges, Belgium, 2008.
[36] J. McNames, A nearest trajectory strategy for time series prediction, in: Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, Katholieke Universiteit Leuven, Belgium, 1998, pp. 112–128.
[37] A.G. McNish, J.V. Lincoln, Prediction of sunspot numbers, Transactions American Geophysical Union 30 (1949) 673.
[38] J.M. Menezes, G.A. Barreto, Long-term time series prediction with the NARX network: an empirical evaluation, Neurocomputing 71 (2008) 3335–3343.
[39] A.S. Nair, J.C. Liu, L. Rilett, S. Gupta, Non-linear analysis of traffic flow, in: Proceedings of the IEEE Intelligent Transportation Systems Conference, Oakland, 2001, pp. 25–29.
[40] M. Norgaard, O. Ravn, N.K. Poulsen, L.K. Hansen, in: Neural Networks for Modelling and Control of Dynamic Systems, Springer, Berlin, 2000.
[41] A.K. Palit, D. Popovic, in: Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications, 1st edition, Springer, 2005.
[42] Y.R. Park, T.J. Murray, C. Chen, Predicting sun spots using a layered perceptron neural network, IEEE Transactions on Neural Networks 1 (2) (1996) 501–505.
[43] J.C. Principe, N.R. Euliano, and W.C. Lefebvre, Neural and Adaptive Systems: Fundamentals through Simulations, (John Wiley & Sons, Inc. 1999).
[44] B. Qian, K. Rasheed, Stock market prediction with multiple classifiers, Applied Intelligence 26 (2007) 25–33.
[45] I. Rojas, H. Pomares, J.L. Bernier, J. Ortega, B. Pino, F.J. Pelayo, A. Prieto, Time series analysis using normalized PG-RBF network with regression weights, Neurocomputing 42 (2002) 267–285.
[46] I. Rojas, O. Valenzuelab, F. Rojasa, A. Guillena, L.J. Herreraa, H. Pomaresa, L. Marquezb, M. Pasadasb, Soft-computing techniques and ARMA model for time series prediction, Neurocomputing 71 (2008) 519–537.
[47] N. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey, IEEE Computational Intelligence Magazine 4 (2) (2009) 24–38.
[48] K. Seiji, S. Hiroaki, S. Hiroshi, Development of very-short-term load forecasting based on chaos theory, Electrical Engineering in Japan 148 (2004) 55–63.
[49] S. Sello, Solar cycle forecasting: a nonlinear dynamics approach, Astronomy & Astrophysics 377 (2001) 312–320.

[50] SIDC (World Data Center for the Sunspot Index), 〈http://sidc.oma.be/index.php3〉.
[51] G. Simon, A. Lendasse, M. Cottrell, J.C. Fort, M. Verleysen, Time series forecasting: obtaining long term trends with self-organizing maps, Pattern Recognition Letters 26 (12) (2005) 1795–1808.
[52] J.C. Sprott, in: Chaos and Time Series Analysis, Oxford University Press, Oxford, 2003.
[53] F. Takens, in: Detecting strange attractors in turbulence, in dynamic systems and turbulence, Lecture Notes in Mathematics, Springer Verlag, Berlin, 1981 366–381.
[54] D. Tao, H. Xiao, Chaotic time series prediction based on radial basis function network, in: Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, vol. 1, 2007, pp. 595–599.
[55] P. Tenti, Forecasting foreign exchange rates using recurrent neural networks, Applied Artificial Intelligence 10 (1996) 567–581.
[56] K. Teo, L. Wang, Z. Lin, Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization, Computational Science 2074 (2001) 310–317.
[57] TSTOOL v. 1.11, 〈http://www.physik3.gwdg.de/tstool〉.
[58] X. Wang, H. Zhang, C. Zhang, X. Cai, J. Wang, Prediction of chaotic time series using LS-SVM with automatic parameter selection, in: Proceedings of the Sixth International Conference on Parallel and Distributed Computing – Applications and Technologies, 2005, pp. 962–965.
[59] Y. Wang, W. Jiang, S. Yuan, J. Wang, Forecasting chaotic time series based on improved genetic Wnn, in: Proceedings of the Fourth International Conference on Natural Computation, Jinan, China, 2008.
[60] J.D. Wichard, M. Ogorzalek, Time series prediction with ensemble models, in: Proceedings of the IJCNN, Budapest, Hungary, 2004, pp. 1625–1629.
[61] R. Wilding, Chaos, complexity and supply-chains, Logistics Focus 6 (1998) 8–10.
[62] R. Wilding, Chaos theory: implications for supply chain management, International Journal of Logistics Management 9 (1998) 43–56.
[63] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Physica D 16 (1985) 285–317.
[64] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, International Journal of Forecasting 14 (1) (1998) 35–62.
[65] J. Zhang, H.S. Chung, W. Lo, Chaotic Time series prediction using a neuro-fuzzy system with time-delay coordinates, IEEE Transactions on Knowledge and Data Engineering 20 (7) (2008) 956–964.
[66] J. Zhang, K.F. Man, Time series prediction using recurrent neural network in multi-dimension embedding phase space, IEEE International Conference on Systems, Man, and Cybernetics 2 (1998) 11–14.
[67] J. Zhang, K.F. Man, J.Y. Ke, Time series prediction using Lyapunov exponents in embedding phase space, in: Proceedings of the Fourth International Conference on Signal Processing, vol. 1, 1998, pp. 221–224.
[68] L. Zhang, in: Artificial Neural Network Model and Its Application, Fudan University Press, Shanghai, 1993.

Muhammad Ardalani-Farsa is a PhD candidate in the Department of Mechanical and Industrial Engineering at Ryerson University in Toronto, Canada. He has a BSc in Mechanical Engineering from University of Tehran, Iran (2001) and MASc in Mechanical Engineering from Ryerson University, Toronto, Canada (2006). His research interests are chaos theory, artificial neural networks and other Computational intelligence methods to forecast nonlinear time series.

Saeed Zolfaghari is an Associate Professor at Ryerson University in Toronto, Canada. He received his PhD (1997) from the University of Ottawa, Canada. His research interests include metaheuristics, computational intelligence, productivity improvement, design and planning problems in manufacturing systems, simulation, logistics and transportation planning. He is a senior member of IIE and IEEE, and a member of INFORMS, CORS and PEO.