Expert Systems with Applications 37 (2010) 479–489

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An artificial neural network (p, d, q) model for timeseries forecasting

Mehdi Khashei *, Mehdi Bijari

Department of Industrial Engineering, Isfahan University of Technology, Isfahan, Iran

ARTICLE INFO

Keywords:
Artificial neural networks (ANNs)
Auto-regressive integrated moving average (ARIMA)
Time series forecasting

ABSTRACT

Artificial neural networks (ANNs) are flexible computing frameworks and universal approximators that can be applied to a wide range of time series forecasting problems with a high degree of accuracy. However, despite all advantages cited for artificial neural networks, their performance for some real time series is not satisfactory. Improving forecasting especially time series forecasting accuracy is an important yet often difficult task facing forecasters. Both theoretical and empirical findings have indicated that integration of different models can be an effective way of improving upon their predictive performance, especially when the models in the ensemble are quite different. In this paper, a novel hybrid model of artificial neural networks is proposed using auto-regressive integrated moving average (ARIMA) models in order to yield a more accurate forecasting model than artificial neural networks. The empirical results with three well-known real data sets indicate that the proposed model can be an effective way to improve forecasting accuracy achieved by artificial neural networks. Therefore, it can be used as an appropriate alternative model for forecasting task, especially when higher forecasting accuracy is needed.

1. Introduction

Artificial neural networks (ANNs) are one of the most accurate and widely used forecasting models that have enjoyed fruitful applications in forecasting social, economic, engineering, foreign exchange, stock problems, etc. Several distinguishing features of artificial neural networks make them valuable and attractive for a forecasting task. First, as opposed to the traditional model-based methods, artificial neural networks are data-driven self-adaptive methods in that there are few a priori assumptions about the models for problems under study. Second, artificial neural networks can generalize. After learning the data presented to them (a sample), ANNs can often correctly infer the unseen part of a population even if the sample data contain noisy information. Third, ANNs are universal functional approximators. It has been shown that a network can approximate any continuous function to any desired accuracy. Finally, artificial neural networks are nonlinear. The traditional approaches to time series prediction, such as the Box–Jenkins or ARIMA, assume that the time series under study are generated from linear processes. However, they may be inappropriate if the underlying mechanism is nonlinear. In fact, real world systems are often nonlinear (Zhang, Patuwo, & Hu, 1998).

Given the advantages of artificial neural networks, it is not surprising that this methodology has attracted overwhelming attention in time series forecasting. Artificial neural networks have been found to be a viable contender to various traditional time series models (Chen, Yang, Dong, & Abraham, 2005; Giordano, La Rocca, & Perna, 2007; Jain & Kumar, 2007). Lapedes and Farber (1987) report the first attempt to model nonlinear time series with artificial neural networks. De Groot and Wurtz (1991) present a detailed analysis of univariate time series forecasting using feedforward neural networks for two benchmark nonlinear time series. Chakraborty, Mehrotra, Mohan, and Ranka (1992) conduct an empirical study on multivariate time series forecasting with artificial neural networks. Atiya and Shaheen (1999) present a case study of multi-step river flow forecasting. Poli and Jones (1994) propose a stochastic neural network model-based on Kalman filter for nonlinear time series prediction. Cottrell, Girard, Girard, Mangeas, and Muller (1995) and Weigend, Huberman, and Rumelhart (1990) address the issue of network structure for forecasting real world time series. Berardi and Zhang (2003) investigate the bias and variance issue in the time series forecasting context. In addition, several large forecasting competitions (Balkin & Ord, 2000) suggest that neural networks can be a very useful addition to the time series forecasting toolbox.

One of the major developments in neural networks over the last decade is the model combining or ensemble modeling. The basic idea of this multi-model approach is the use of each component model's unique capability to better capture different patterns in the data. Both theoretical and empirical findings have suggested that combining different models can be an effective way to improve the predictive performance of each individual model, especially when the models in the ensemble are quite different (Baxt, 1992; Zhang, 2007). In addition, since it is difficult to completely know the characteristics of the data in a real problem, hybrid

* Corresponding author. Tel.: +98 311 39125501; fax: +98 311 3915526.
E-mail address: Khashei@in.iut.ac.ir (M. Khashei).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.05.044

methodology that has both linear and nonlinear modeling capabilities can be a good strategy for practical use. In the literature, different combination techniques have been proposed in order to overcome the deficiencies of single models and yield more accurate results. The difference between these combination techniques can be described using terminology developed for the classification and neural network literature. Hybrid models can be homogeneous, such as using differently configured neural networks (all multi-layer perceptrons), or heterogeneous, such as with both linear and nonlinear models (Taskaya & Casey, 2005).

In a competitive architecture, the aim is to build appropriate modules to represent different parts of the time series, and to be able to switch control to the most appropriate. For example, a time series may exhibit nonlinear behavior generally, but this may change to linearity depending on the input conditions. Early work on threshold auto-regressive models (TAR) used two different linear AR processes, each of which change control among themselves according to the input values (Tong & Lim, 1980). An alternative is a mixture density model, also known as nonlinear gated expert, which comprises neural networks integrated with a feedforward gating network (Taskaya & Casey, 2005). In a cooperative modular combination, the aim is to combine models to build a complete picture from a number of partial solutions. The assumption is that a model may not be sufficient to represent the complete behavior of a time series, for example, if a time series exhibits both linear and nonlinear patterns during the same time interval, neither linear models nor nonlinear models alone are able to model both components simultaneously. A good exemplar is models that fuse auto-regressive integrated moving average with artificial neural networks. An auto-regressive integrated moving average (ARIMA) process combines three different processes comprising an auto-regressive (AR) function regressed on past values of the process, moving average (MA) function regressed on a purely random process, and an integrated (I) part to make the data series stationary by differencing. In such hybrids, whilst the neural network model deals with nonlinearity, the auto-regressive integrated moving average model deals with the non-stationary linear component (Zhang, 2003).

The literature on this topic has expanded dramatically since the early work of Bates and Granger (1969), Clemen (1989) and Reid (1968) provided a comprehensive review and annotated bibliography in this area. Wedding and Cios (1996) described a combining methodology using radial basis function networks (RBF) and the Box–Jenkins ARIMA models. Luxhoj, Riis, and Stensballe (1996) presented a hybrid econometric and ANN approach for sales forecasting. Ginzburg and Horn (1994) and Pelikan et al. (1992) proposed to combine several feedforward neural networks in order to improve time series forecasting accuracy. Tsaih, Hsu, and Lai (1998) presented a hybrid artificial intelligence (AI) approach that integrated the rule-based systems technique and neural networks to S&P 500 stock index prediction. Voort, Dougherty, and Watson (1996) introduced a hybrid method called KARIMA using a Kohonen self-organizing map and auto-regressive integrated moving average method for short-term prediction. Medeiros and Veiga (2000) consider a hybrid time series forecasting system with neural networks used to control the time-varying parameters of a smooth transition auto-regressive model.

In recent years, more hybrid forecasting models have been proposed, using auto-regressive integrated moving average and artificial neural networks and applied to time series forecasting with good prediction performance. Pai and Lin (2005) proposed a hybrid methodology to exploit the unique strength of ARIMA models and support vector machines (SVMs) for stock prices forecasting. Chen and Wang (2007) constructed a combination model incorporating seasonal auto-regressive integrated moving average (SARIMA) model and SVMs for seasonal time series forecasting. Zhou and Hu (2008) proposed a hybrid modeling and forecasting approach based on Grey and Box–Jenkins auto-regressive moving average (ARMA) models. Armano, Marchesi, and Murru (2005) presented a new hybrid approach that integrated artificial neural network with genetic algorithms (GAs) to stock market forecast.

Goh, Lim, and Peh (2003) use an ensemble of boosted Elman networks for predicting drug dissolution profiles. Yu, Wang, and Lai (2005) proposed a novel nonlinear ensemble forecasting model integrating generalized linear auto regression (GLAR) with artificial neural networks in order to obtain accurate prediction in foreign exchange market. Kim and Shin (2007) investigated the effectiveness of a hybrid approach based on the artificial neural networks for time series properties, such as the adaptive time delay neural networks (ATNNs) and the time delay neural networks (TDNNs), with the genetic algorithms in detecting temporal patterns for stock market prediction tasks. Tseng, Yu, and Tzeng (2002) proposed using a hybrid model called SARIMABP that combines the seasonal auto-regressive integrated moving average (SARIMA) model and the back-propagation neural network model to predict seasonal time series data. Khashei, Hejazi, and Bijari (2008) based on the basic concepts of artificial neural networks, proposed a new hybrid model in order to overcome the data limitation of neural networks and yield more accurate forecasting model, especially in incomplete data situations.

In this paper, auto-regressive integrated moving average models are applied to construct a new hybrid model in order to yield more accurate model than artificial neural networks. In our proposed model, the future value of a time series is considered as nonlinear function of several past observations and random errors, such ARIMA models. Therefore, in the fist phase, an auto-regressive integrated moving average model is used in order to generate the necessary data from under study time series. Then, in the second phase, a neural network is used to model the generated data by ARIMA model, and to predict the future value of time series. Three well-known data sets – the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data – are used in this paper in order to show the appropriateness and effectiveness of the proposed model to time series forecasting. The rest of the paper is organized as follows. In the next section, the basic concepts and modeling approaches of the auto-regressive integrated moving average (ARIMA) and artificial neural networks (ANNs) are briefly reviewed. In Section 3, the formulation of the proposed model is introduced. In Section 4, the proposed model is applied to time series forecasting and its performance is compared with those of other forecasting models. Section 5 contains the concluding remarks.

2. Artificial neural networks (ANNs) and auto-regressive integrated moving average (ARIMA) models

In this section, the basic concepts and modeling approaches of the artificial neural networks (ANNs) and auto-regressive integrated moving average (ARIMA) models for time series forecasting are briefly reviewed.

2.1. The ANN approach to time series modeling

Recently, computational intelligence systems and among them artificial neural networks (ANNs), which in fact are model free dynamics, has been used widely for approximation functions and forecasting. One of the most significant advantages of the ANN models over other classes of nonlinear models is that ANNs are universal approximators that can approximate a large class of functions with a high degree of accuracy (Chen, Leung, & Hazem, 2003; Zhang & Min Qi, 2005). Their power comes from the parallel processing of the information from the data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting (Zhang et al., 1998). The model is characterized by a network of three layers of simple processing units connected by acyclic links (Fig. 1). The relationship between the output (y_t) and the inputs $(y_{t-1}, \ldots, y_{t-p})$ has the following mathematical representation:

$$y_t = w_0 + \sum_{j=1}^{q} w_j \cdot g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} \cdot y_{t-i}\right) + \varepsilon_t, \tag{1}$$

where, $w_{i,j}(i = 0, 1, 2, \ldots, p, j = 1, 2, \ldots, q)$ and $w_j(j = 0, 1, 2, \ldots, q)$ are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes. Activation functions can take several forms. The type of activation function is indicated by the situation of the neuron within the network. In the majority of cases input layer neurons do not have an activation function, as their role is to transfer the inputs to the hidden layer. The most widely used activation function for the output layer is the linear function as non-linear activation function may introduce distortion to the predicated output. The logistic and hyperbolic functions are often used as the hidden layer transfer function that are shown in Eqs. (2) and (3), respectively. Other activation functions can also be used such as linear and quadratic, each with a variety of modeling applications.

$$Sig(x) = \frac{1}{1 + exp(-x)}, \tag{2}$$

$$Tanh(x) = \frac{1 - exp(-2x)}{1 + exp(-2x)}. \tag{3}$$

Hence, the ANN model of (1), in fact, performs a nonlinear functional mapping from past observations to the future value y_t, i.e.,

$$y_t = f(y_{t-1}, \ldots, y_{t-p}, w) + \varepsilon_t, \tag{4}$$

where, w is a vector of all parameters and $f(\cdot)$ is a function determined by the network structure and connection weights. Thus, the neural network is equivalent to a nonlinear auto-regressive model. The simple network given by (1) is surprisingly powerful in that it is able to approximate the arbitrary function as the number of hidden nodes when q is sufficiently large. In practice, simple network structure that has a small number of hidden nodes often works well in out-of-sample forecasting. This may be due to the overfitting effect typically found in the neural network modeling process. An overfitted model has a good fit to the sample used for model building but has poor generalizability to data out of the sample (Demuth & Beale, 2004).

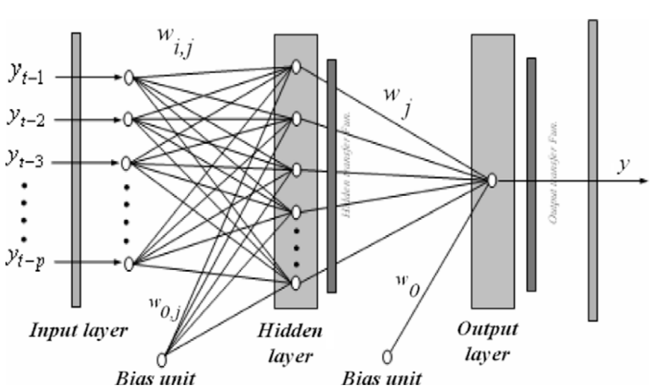

Fig. 1. Neural network structure $(\boldsymbol{N}^{(p-q-1)})$.

The choice of q is data-dependent and there is no systematic rule in deciding this parameter. In addition to choosing an appropriate number of hidden nodes, another important task of ANN modeling of a time series is the selection of the number of lagged observations, p, and the dimension of the input vector. This is perhaps the most important parameter to be estimated in an ANN model because it plays a major role in determining the (nonlinear) autocorrelation structure of the time series.

There exist many different approaches such as the pruning algorithm, the polynomial time algorithm, the canonical decomposition technique, and the network information criterion for finding the optimal architecture of an ANN (Khashei, 2005). These approaches can be generally categorized as follows: (i) Empirical or statistical methods that are used to study the effect of internal parameters and choose appropriate values for them based on the performance of model (Benardos & Vosniakos, 2002; Ma & Khorasani, 2003). The most systematic and general of these methods utilizes the principles from Taguchi's design of experiments (Ross, 1996). (ii) Hybrid methods such as fuzzy inference (Leski & Czogala, 1999) where the ANN can be interpreted as an adaptive fuzzy system or it can operate on fuzzy instead of real numbers. (iii) Constructive and/or pruning algorithms that, respectively, add and/or remove neurons from an initial architecture using a previously specified criterion to indicate how ANN performance is affected by the changes (Balkin & Ord, 2000; Islam & Murase, 2001; Jiang & Wah, 2003). The basic rules are that neurons are added when training is slow or when the mean squared error is larger than a specified value. In opposite, neurons are removed when a change in a neuron's value does not correspond to a change in the network's response or when the weight values that are associated with this neuron remain constant for a large number of training epochs (Marin, Varo, & Guerrero, 2007). (iv). Evolutionary strategies that search over topology space by varying the number of hidden layers and hidden neurons through application of genetic operators (Castillo, Merelo, Prieto, Rivas, & Romero, 2000; Lee & Kang, 2007) and evaluation of the different architectures according to an objective function (Arifovic & Gencay, 2001; Benardos & Vosniakos, 2007).

Although many different approaches exist in order to find the optimal architecture of an ANN, these methods are usually quite complex in nature and are difficult to implement (Zhang et al., 1998). Furthermore, none of these methods can guarantee the optimal solution for all real forecasting problems. To date, there is no simple clear-cut method for determination of these parameters and the usual procedure is to test numerous networks with varying numbers of input and hidden units (p, q), estimate generalization error for each and select the network with the lowest generalization error (Hosseini, Luo, & Reynolds, 2006). Once a network structure (p, q) is specified, the network is ready for training a process of parameter estimation. The parameters are estimated such that the cost function of neural network is minimized. Cost function is an overall accuracy criterion such as the following mean squared error:

$$E = \frac{1}{N}\sum_{n=1}^{N}(e_i)^2$$
$$= \frac{1}{N}\sum_{n=1}^{N}\left(y_t - \left(w_0 + \sum_{j=1}^{Q} w_j g\left(w_{0j} + \sum_{i=1}^{P} w_{i,j} y_{t-i}\right)\right)\right)^2, \tag{5}$$

where, N is the number of error terms. This minimization is done with some efficient nonlinear optimization algorithms other than the basic backpropagation training algorithm (Rumelhart & McClelland, 1986), in which the parameters of the neural network, $w_{i,j}$, are changed by an amount $\Delta w_{i,j}$, according to the following formula:

$$\Delta w_{i,j} = -\eta \frac{\partial E}{\partial w_{i,j}}, \tag{6}$$

where, the parameter η is the learning rate and $\frac{\partial E}{\partial w_{i,j}}$ is the partial derivative of the function E with respect to the weight $w_{i,j}$. This derivative is commonly computed in two passes. In the forward pass, an input vector from the training set is applied to the input units of the network and is propagated through the network, layer by layer, producing the final output. During the backward pass, the output of the network is compared with the desired output and the resulting error is then propagated backward through the network, adjusting the weights accordingly. To speed up the learning process, while avoiding the instability of the algorithm (Rumelhart & McClelland, 1986) introduced a momentum term δ in Eq. (6), thus obtaining the following learning rule:

$$\Delta w_{i,j}(t+1) = -\eta \frac{\partial E}{\partial w_{i,j}} + \delta \Delta w_{i,j}(t), \tag{7}$$

The momentum term may also be helpful to prevent the learning process from being trapped into poor local minima, and is usually chosen in the interval [0; 1]. Finally, the estimated model is evaluated using a separate hold-out sample that is not exposed to the training process.

2.2. The auto-regressive integrated moving average models

For more than half a century, auto-regressive integrated moving average (ARIMA) models have dominated many areas of time series forecasting. In an ARIMA (p, d, q) model, the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates the time series with the mean μ has the form:

$$\phi(B)\nabla^d(y_t - \mu) = \theta(B)a_t, \tag{8}$$

where, y_t and a_t are the actual value and random error at time period t, respectively; $\phi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i, \theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in B of degree p and $q, \phi_i (i = 1, 2, \ldots, p)$ and $\theta_j (j = 1, 2, \ldots, q)$ are model parameters, $\nabla = (1 - B), B$ is the backward shift operator, p and q are integers and often referred to as orders of the model, and d is an integer and often referred to as order of differencing. Random errors, a_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

The Box and Jenkins (1976) methodology includes three iterative steps of model identification, parameter estimation, and diagnostic checking. The basic idea of model identification is that if a time series is generated from an ARIMA process, it should have some theoretical autocorrelation properties. By matching the empirical autocorrelation patterns with the theoretical ones, it is often possible to identify one or several potential models for the given time series. Box and Jenkins (1976) proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data as the basic tools to identify the order of the ARIMA model. Some other order selection methods have been proposed based on validity criteria, the information-theoretic approaches such as the Akaike's information criterion (AIC) (Shibata, 1976) and the minimum description length (MDL) (Hurvich & Tsai, 1989; Jones, 1975; Ljung, 1987). In addition, in recent years different approaches based on intelligent paradigms, such as neural networks (Hwang, 2001), genetic algorithms (Minerva & Poli, 2001; Ong, Huang, & Tzeng, 2005) or fuzzy system (Haseyama & Kitajima, 2001) have been proposed to improve the accuracy of order selection of ARIMA models.

In the identification step, data transformation is often required to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model used for forecasting. A stationary time series is characterized by statistical characteristics such as the mean and the autocorrelation structure being constant over time. When the observed time series presents trend and heteroscedasticity, differencing and power transformation are applied to the data to remove the trend and to stabilize the variance before an ARIMA model can be fitted. Once a tentative model is identified, estimation of the model parameters is straightforward. The parameters are estimated such that an overall measure of errors is minimized. This can be accomplished using a nonlinear optimization procedure. The last step in model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors, a_t, are satisfied.

Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentatively entertained model to the historical data. If the model is not adequate, a new tentative model should be identified, which will again be followed by the steps of parameter estimation and model verification. Diagnostic information may help suggest alternative model(s). This three-step model building process is typically repeated several times until a satisfactory model is finally selected. The final selected model can then be used for prediction purposes.

3. Formulation of the proposed model

Despite the numerous time series models available, the accuracy of time series forecasting currently is fundamental to many decision processes, and hence, never research into ways of improving the effectiveness of forecasting models been given up. Many researches in time series forecasting have been argued that predictive performance improves in combined models. In hybrid models, the aim is to reduce the risk of using an inappropriate model by combining several models to reduce the risk of failure and obtain results that are more accurate. Typically, this is done because the underlying process cannot easily be determined. The motivation for combining models comes from the assumption that either one cannot identify the true data generating process or that a single model may not be sufficient to identify all the characteristics of the time series.

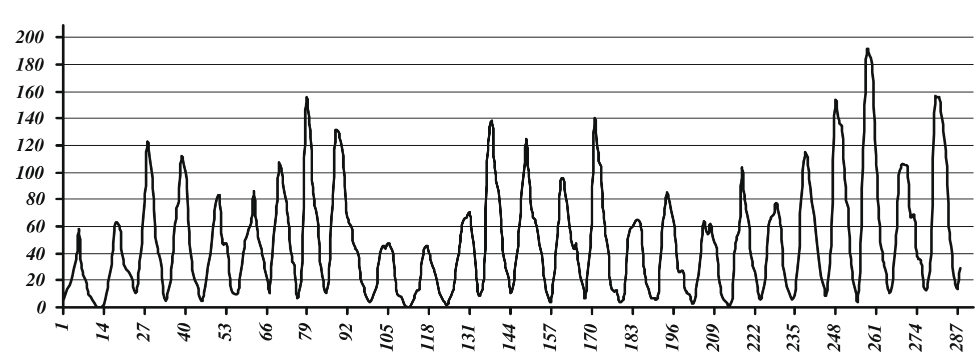

Fig. 2. Sunspot series (1700–1987).

In this paper, a novel hybrid model of artificial neural networks is proposed in order to yield more accurate results using the auto regressive integrated moving average models. In our proposed model, based on Box and Jenkins (1976) methodology in linear modeling, a time series is considered as nonlinear function of several past observations and random errors as follows:

$$y_t = f[(z_{t-1}, z_{t-2}, \ldots, z_{t-m}), (e_{t-1}, e_{t-2}, \ldots, e_{t-n})], \tag{9}$$

where f is a nonlinear function determined by the neural network, $z_t = (1-B)^d(y_t - \mu)$, e_t is the residual at time t and m and n are integers. So, in the first stage, an auto-regressive integrated moving average model is used in order to generate the residuals (e_t).

In second stage, a neural network is used in order to model the nonlinear and linear relationships existing in residuals and original data. Thus,

$$z_t = w_0 + \sum_{j=1}^{Q} w_j \cdot g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} \cdot z_{t-i} + \sum_{i=p+1}^{p+q} w_{i,j} \cdot e_{t+p-i}\right) + \varepsilon_t, \tag{10}$$

where, $w_{i,j}(i = 0, 1, 2, \ldots, p+q, j = 1, 2, \ldots, Q)$ and $w_j(j = 0, 1, 2, \ldots, Q)$ are connection weights; p, q, Q are integers, which are determined in design process of final neural network.

It must be noted that any set of above–mentioned variables $\{e_i(i = t-1, \ldots, t-n)\}$ or $\{z_i(i = t-1, ,t-m)\}$ may be deleted in design process of final neural network. This maybe related to the underlying data generating process and the existing linear and nonlinear structures in data. For example, if data only consist of pure nonlinear structure, then the residuals will only contain the nonlinear relationship. Because the ARIMA is a linear model and does not able to model nonlinear relationship. Therefore, the set of residuals $\{e_i(i = t-1, \ldots, t-n)\}$ variables maybe deleted against other of those variables.

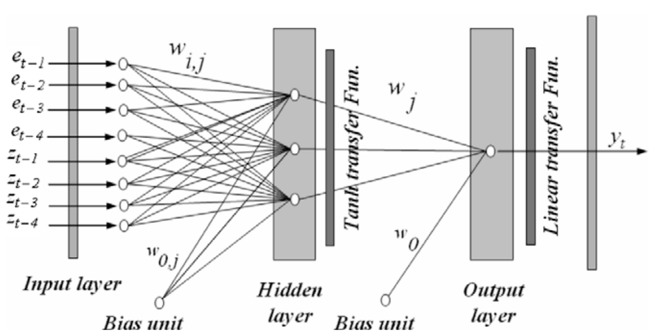

Fig. 3. Structure of the best-fitted network (sunspot data case), $\boldsymbol{N}^{(8-3-1)}$.

As previously mentioned, in building auto-regressive integrated moving average as well as artificial neural networks models, subjective judgment of the model order as well as the model adequacy is often needed. It is possible that suboptimal models will be used in the hybrid model. For example, the current practice of Box–Jenkins methodology focuses on the low order autocorrelation. A model is considered adequate if low order autocorrelations are not significant even though significant autocorrelations of higher order still exist. This suboptimality may not affect the usefulness of the hybrid model. Granger (1989) has pointed out that for a hybrid model to produce superior forecasts, the component model should be suboptimal. In general, it has been observed that it is more effective to combine individual forecasts that are based on different information sets (Granger, 1989).

4. Application of the hybrid model to exchange rate forecasting

In this section, three well-known data sets – the Wolf's sunspot data, the Canadian lynx data, and the British pound/United States dollar exchange rate data – are used in order to demonstrate the appropriateness and effectiveness of the proposed model. These time series come from different areas and have different statistical characteristics. They have been widely studied in the statistical as well as the neural network literature (Zhang, 2003). Both linear and nonlinear models have been applied to these data sets, although more or less nonlinearities have been found in these series. Only the one-step-ahead forecasting is considered.

4.1. The Wolf's sunspot data forecasts

The sunspot series is record of the annual activity of spots visible on the face of the sun and the number of groups into which

Table 1
Comparison of the performance of the proposed model with those of other forecasting models (Sunspot data set).

Model	35 Points ahead		67 Points ahead	
	MAE	MSE	MAE	MSE
Auto-regressive integrated moving average (ARIMA)	11.319	216.965	13.033739	306.08217
Artificial neural networks (ANNs)	10.243	205.302	13.544365	351.19366
Zhang's hybrid model	10.831	186.827	12.780186	280.15956
Our proposed model	**8.944**	**125.812**	**12.117994**	**234.206103**

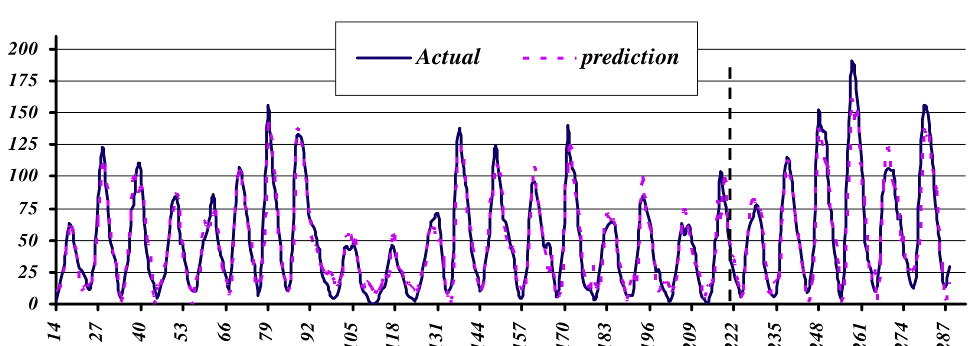

Fig. 4. Results obtained from the proposed model for sunspot data set.

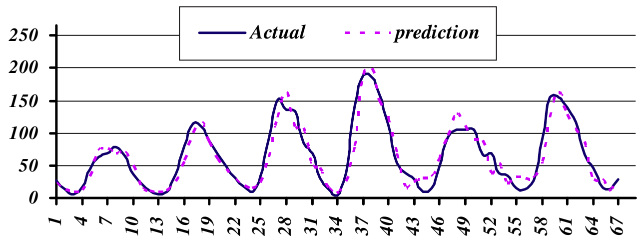

Fig. 5. ARIMA model prediction of sunspot data (test sample).

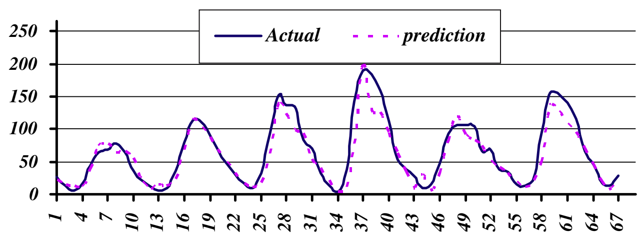

Fig. 6. ANN model prediction of sunspot data (test sample).

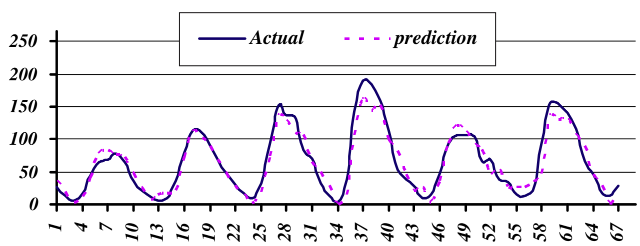

Fig. 7. Proposed model prediction of sunspot data (test sample).

they cluster. The sunspot data, which is considered in this investigation, contains the annual number of sunspots from 1700 to 1987, giving a total of 288 observations. The study of sunspot activity has practical importance to geophysicists, environment scientists, and climatologists (Hipel & McLeod, 1994). The data series is regarded as nonlinear and non-Gaussian and is often used to evaluate the effectiveness of nonlinear models (Ghiassi & Saidane, 2005). The plot of this time series (Fig. 2) also suggests that there is a cyclical pattern with a mean cycle of about 11 years (Zhang, 2003). The sunspot data has been extensively studied with a vast variety of linear and nonlinear time series models including ARIMA and ANNs. To assess the forecasting performance of proposed model, the sunspot data set is divided into two samples of training and testing. The training data set, 221 observations (1700–1920), is exclusively used in order to formulate the model and then the test

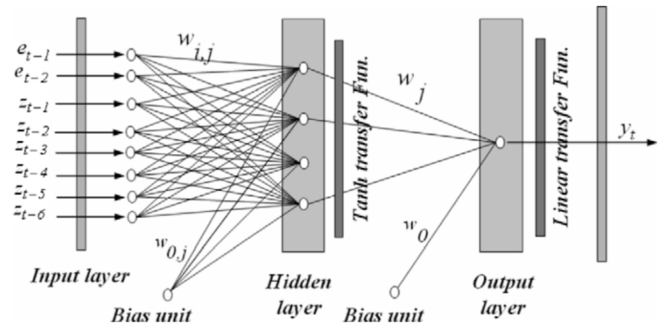

Fig. 9. Structure of the best-fitted network (lynx data case), $\boldsymbol{N}^{(8-4-1)}$.

sample, the last 67 observations (1921–1987), is used in order to evaluate the performance of the established model.

Stage I: Using the *Eviews* package software, the best-fitted model is a auto-regressive model of order nine, *AR (9)*, which has also been used by many researchers (Hipel & McLeod, 1994; Subba Rao & Sabr, 1984; Zhang, 2003).

Stage II: In order to obtain the optimum network architecture, based on the concepts of artificial neural networks design and using pruning algorithms in *MATLAB 7* package software, different network architectures are evaluated to compare the ANNs performance. The best-fitted network which is selected, and therefore, the architecture which presented the best forecasting accuracy with the test data, is composed of eight inputs, three hidden and one output neurons (in abbreviated form, $\boldsymbol{N}^{(8-3-1)}$). The structure of the best-fitted network is shown in Fig. 3. The performance measures of the proposed model for sunspot data are given in Table 1. The estimated values of proposed model sunspot data sets are plotted in Fig. 4. In addition, the estimated value of ARIMA, ANN, and our proposed models for test data are plotted in Figs. 5–7, respectively.

4.2. The Canadian lynx series forecasts

The lynx series, which is considered in this investigation, contains the number of lynx trapped per year in the Mackenzie River district of Northern Canada. The data set are plotted in Fig. 8, which shows a periodicity of approximately 10 years (Stone & He, 2007). The data set has 114 observations, corresponding to the period of 1821–1934. It has also been extensively analyzed in the time series literature with a focus on the nonlinear modeling (Campbell & Walker, 1977; Cornillon, Imam, & Matzner, 2008; Lin & Pourahmadi, 1998; Tang & Ghosal, 2007) see Wong and Li (2000) for a survey. Following other studies (Subba Rao & Sabr, 1984; Stone & He, 2007; Zhang, 2003), the logarithms (to the base 10) of the data are used in the analysis.

Stage I: As in the previous section, using the *Eviews* package software, the established model is a auto-regressive model of order

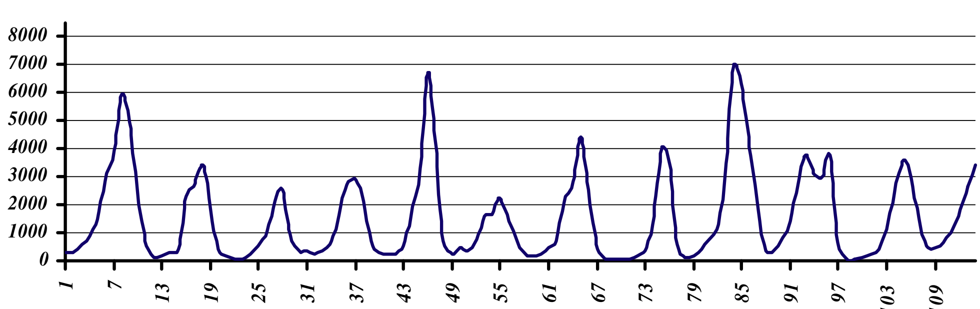

Fig. 8. Canadian lynx data series (1821–1934).

Table 2
Percentage improvement of the proposed model in comparison with those of other forecasting models (Sunspot data set).

Model	35 Points ahead (%)		67 Points ahead (%)	
	MAE	MSE	MAE	MSE
Auto-regressive integrated moving average (ARIMA)	20.98	42.01	7.03	23.48
Artificial neural networks (ANNs)	12.68	38.72	10.53	33.31
Zhang's hybrid model	17.42	32.66	5.18	16.40

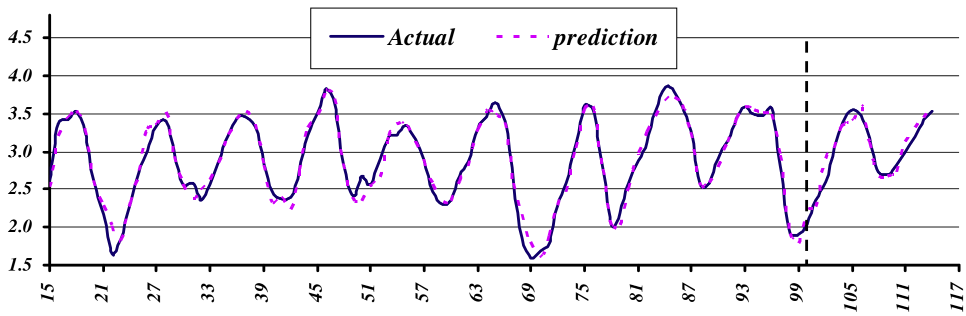

Fig. 10. Results obtained from the proposed model for Canadian lynx data set.

twelve, *AR (12)*, which has also been used by many researchers (Subba Rao & Sabr, 1984; Zhang, 2003).

Stage II: Similar to the previous section, by using pruning algorithms in *MATLAB7* package software, the best-fitted network which is selected, is composed of eight inputs, four hidden and one output neurons ($\boldsymbol{N}^{(8-4-1)}$). The structure of the best-fitted network is shown in Fig. 9. The performance measures of the proposed model for Canadian lynx data are given in Table 2. The estimated values of proposed model for Canadian lynx data set are plotted in Fig. 10. In addition, the estimated value of ARIMA, ANN, and proposed models for test data are plotted in Figs. 11–13, respectively.

4.3. The exchange rate (British pound /US dollar) forecasts

The last data set that is considered in this investigation is the exchange rate between British pound and United States dollar. Predicting exchange rate is an important yet difficult task in international finance. Various linear and nonlinear theoretical models have been developed but few are more successful in out-of-sample forecasting than a simple random walk model. Recent applications

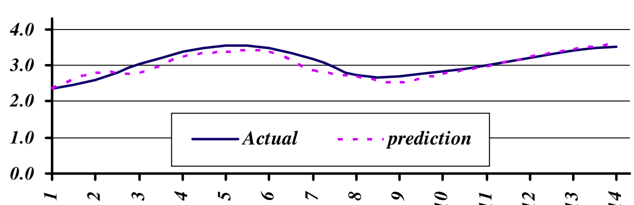

Fig. 11. ARIMA model prediction of lynx data (test sample).

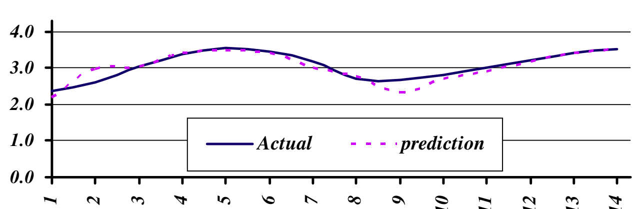

Fig. 12. ANN model prediction of lynx data (test sample).

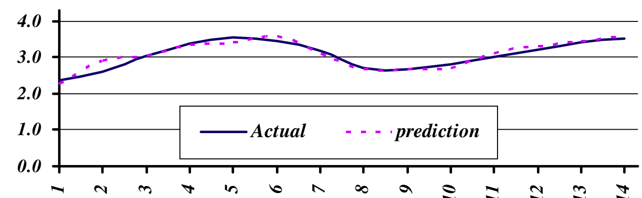

Fig. 13. Proposed model prediction of lynx data (test sample).

of neural networks in this area have yielded mixed results. The data used in this paper contain the weekly observations from 1980 to 1993, giving 731 data points in the time series. The time series plot is given in Fig. 14, which shows numerous changing turning points in the series. In this paper following Meese and Rogoff (1983) and Zhang (2003), the natural logarithmic transformed data is used in the modeling and forecasting analysis.

Stage I: In a similar fashion, using the *Eviews* package software, the best-fitted ARIMA model is a random walk model, which has been used by Zhang (2003). It has also been suggested by many studies in the exchange rate literature that a simple random walk is the dominant linear model (Meese & Rogoff, 1983).

Stage II: Similar to the previous sections, using pruning algorithms in *MATLAB 7*package software, the best-fitted network which is selected, is composed of twelve inputs, four hidden and one output neurons ($\boldsymbol{N}^{(12-4-1)}$). The structure of the best-fitted network is shown in Fig. 15. The performance measures of the proposed model for exchange rate data are given in Table 3. The estimated value of proposed model for both test and training data are plotted in Fig. 16. In addition, the estimated value of ARIMA, ANN, and proposed models for test data are plotted in Figs. 17–19, respectively.

4.4. Comparison with other models

In this section, the predictive capabilities of the proposed model are compared with artificial neural networks (ANNs), auto-regressive integrated moving average (ARIMA), and Zhang's hybrid ANNs/ARIMA model (Zhang, 2003) using three well-known real data sets: (1) the Wolf's sunspot data, (2) the Canadian lynx data, and (3) the British pound/US dollar exchange rate data. The MAE

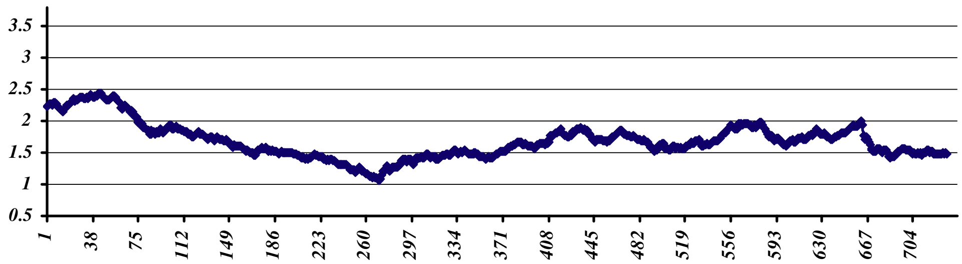

Fig. 14. Weekly British pound against the United States dollar exchange rate series (1980–1993).

(Mean Absolute Error) and MSE (Mean Squared Error), which are computed from the following equations, are employed as performance indicators in order to measure forecasting performance of proposed model in comparison whit those other forecasting models.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|e_i|, \tag{11}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2. \tag{12}$$

In the Wolf's sunspot data forecast case, a subset auto-regressive model of order nine has been found to be the most parsimonious among all ARIMA models that are also found adequate judged by the residual analysis. Many researchers such as Hipel and McLeod (1994), Subba Rao and Sabr (1984) and Zhang (2003) have also used this model. The neural network model used is composed of four inputs, four hidden and one output neurons ($\boldsymbol{N}^{(4-4-1)}$), as also employed by Cottrell et al. (1995), De Groot and Wurtz (1991) and Zhang (2003). Two forecast horizons of 35 and 67 periods are used in order to assess the forecasting performance of models. The forecasting results of above-mentioned models and improvement percentage of the proposed model in comparison with those models for the sunspot data are summarized in Tables 1 and 2, respectively.

Results show that while applying neural networks alone can improve the forecasting accuracy over the ARIMA model in the 35-period horizon, the performance of ANNs is getting worse as time horizon extends to 67 periods. This may suggest that neither the neural network nor the ARIMA model captures all of the patterns in the data and combining two models together can be an effective way in order to overcome this limitation. However, the

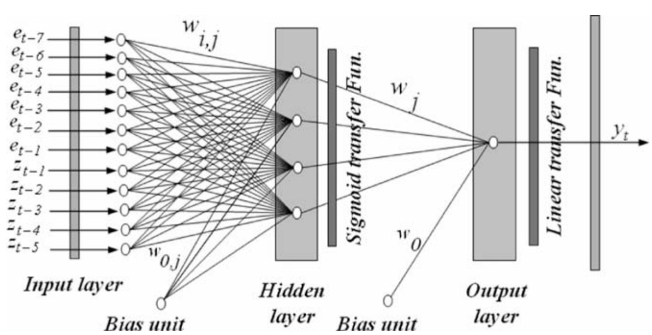

Fig. 15. Structure of the best-fitted network (exchange rate case), $\boldsymbol{N}^{(12-4-1)}$.

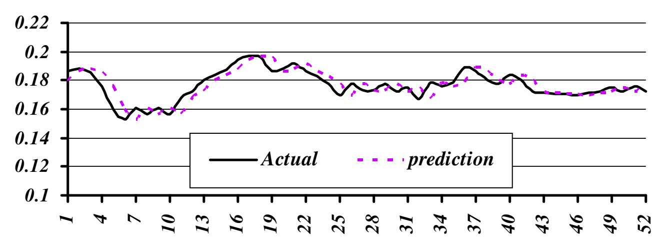

Fig. 17. ARIMA model prediction of exchange rate data set (test sample).

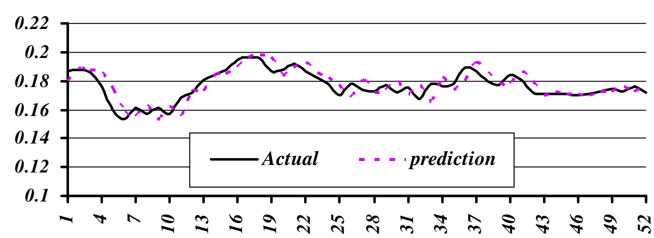

Fig. 18. ANN model prediction of exchange rate data set (test sample).

Table 3
Comparison of the performance of the proposed model with those of other forecasting models (Canadian lynx data).

Model	MAE	MSE
Auto-regressive integrated moving average (ARIMA)	0.112255	0.020486
Artificial neural networks (ANNs)	0.112109	0.020466
Zhang's hybrid model	0.103972	0.017233
Our proposed model	**0.089625**	**0.013609**

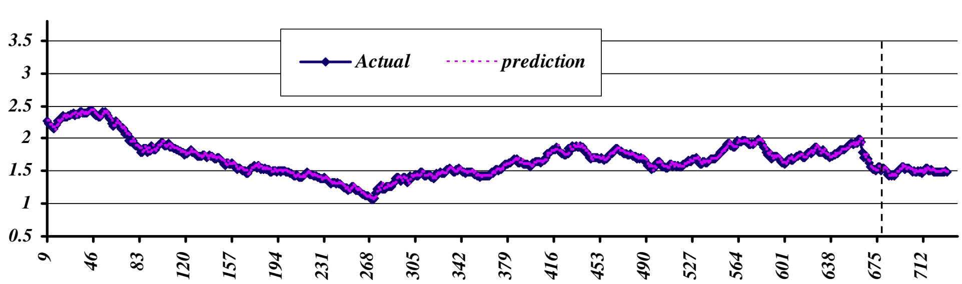

Fig. 16. Results obtained from the proposed model for exchange rate data set.

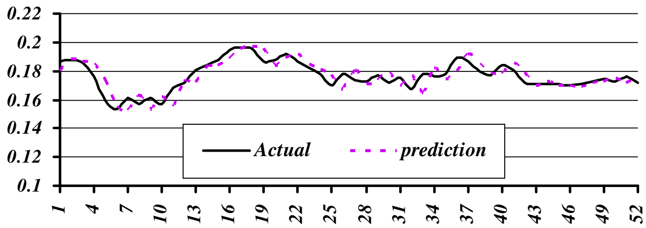

Fig. 19. Proposed model prediction of exchange rate data set (test sample).

Table 4
Percentage improvement of the proposed model in comparison with those of other forecasting models (Canadian lynx data).

Model	MAE (%)	MSE (%)
Auto-regressive integrated moving average (ARIMA)	20.16	33.57
Artificial neural networks (ANNs)	20.06	33.50
Zhang's hybrid model	13.80	21.03

results of the Zhang's hybrid model (Zhang, 2003) show that; although, the overall forecasting errors of Zhang's hybrid model have been reduced in comparison with ARIMA and ANN, this model may also give worse predictions than either of those, in some specific situations. These results may be occurred due to the assumptions Taskaya and Casey (2005), which have been considered in constructing process of the hybrid model by Zhang (2003).

Our proposed model have yielded more accurate results than Zhang's hybrid model and also both ARIMA and ANN models used separately across two different time horizons and with both error measures. For example in terms of MAE, the percentage improvements of the proposed model over the Zhang's hybrid model, ANN, and ARIMA for 35-period forecasts are 17.42%, 12.68%, and 20.98%, respectively.

In a similar fashion, a subset auto-regressive model of order twelve has been fitted to Canadian lynx data. This is a parsimonious model also used by Subba Rao and Sabr (1984) and Zhang (2003). In addition, a neural network, which is composed of seven inputs, five hidden and one output neurons ($\boldsymbol{N}^{(7-5-1)}$), has been designed to Canadian lynx data set forecast, as also employed by Zhang (2003). The overall forecasting results of above-mentioned models and improvement percentage of the proposed model in comparison with those models for the last 14 years are summarized in Tables 3 and 4, respectively.

Numerical results show that the used neural network gives slightly better forecasts than the ARIMA model and the Zhang's hybrid model, significantly outperform the both of them. However, by applying our proposed model to be obtained more accurate results than Zhang's hybrid model. Our proposed model indicates an 21.03% and 13.80% decrease over the Zhang's hybrid model in MSE and MAE, respectively.

With the exchange rate data set, the best linear ARIMA model is found to be the simple random walk model: $y_t = y_{t-1} + \varepsilon_t$. This is the same finding suggested by many studies in the exchange rate literature (Zhang, 2003) that a simple random walk is the dominant linear model. They claim that the evolution of any exchange rate follows the theory of efficient market hypothesis (EMH) (Timmermann & Granger, 2004). According to this hypothesis, the best prediction value for tomorrow's exchange rate is the current value of the exchange rate and the actual exchange rate follows a random walk (Meese & Rogoff, 1983). A neural network, which is composed of seven inputs, six hidden and one output neurons ($\boldsymbol{N}^{(7-6-1)}$) is designed in order to model the nonlinear patterns, as also employed by Zhang (2003). Three time horizons of 1, 6 and 12 months are used in order to assess the forecasting performance of models. The forecasting results of above-mentioned models and improvement percentage of the proposed model in comparison with those models for the exchange rate data are summarized in Tables 5 and 6, respectively.

Results of the exchange rate data set forecasting indicate that for short-term forecasting (1 month), both neural network and hybrid models are much better in accuracy than the simple random walk model. The ANN model gives a comparable performance to the ARIMA model and Zhang's hybrid model slightly outperforms both ARIMA and ANN models for longer time horizons (6 and 12 month). However, our proposed model significantly outperforms ARIMA, ANN, and Zhang's hybrid models across three different time horizons and with both error measures.

5. Conclusions

Applying quantitative methods for forecasting and assisting investment decision making has become more indispensable in business practices than ever before. Time series forecasting is one of the most important quantitative models that has received considerable amount of attention in the literature. Artificial neural networks (ANNs) have shown to be an effective, general-purpose approach for pattern recognition, classification, clustering, and especially time series prediction with a high degree of accuracy.

Table 5
Comparison of the performance of the proposed model with those of other forecasting models (exchange rate data)*.

Model	1 Month		6 Month		12 Month	
	MAE	MSE	MAE	MSE	MAE	MSE
Auto-regressive integrated moving average	0.005016	3.68493	0.0060447	5.65747	0.0053579	4.52977
Artificial neural networks (ANNs)	0.004218	2.76375	0.0059458	5.71096	0.0052513	4.52657
Zhang's hybrid model	0.004146	2.67259	0.0058823	5.65507	0.0051212	4.35907
Our proposed model	**0.004001**	**2.60937**	**0.0054440**	**4.31643**	**0.0051069**	**3.76399**

* *Note*: All MSE values should be multiplied by 10^{-5}.

Table 6
Percentage improvement of the proposed model in comparison with those of other forecasting models (exchange rate data).

Model	1 Month		6 Month		12 Month	
	MAE	MSE	MAE	MSE	MAE	MSE
Auto-regressive integrated moving average	20.24	29.19	9.94	23.70	4.68	16.91
Artificial neural networks (ANNs)	5.14	5.59	8.44	24.42	2.75	16.85
Zhang's hybrid model	3.50	2.37	7.45	23.67	0.28	13.65

Nevertheless, their performance is not always satisfactory. Theoretical as well empirical evidences in the literature suggest that by using dissimilar models or models that disagree each other strongly, the hybrid model will have lower generalization variance or error. Additionally, because of the possible unstable or changing patterns in the data, using the hybrid method can reduce the model uncertainty, which typically occurred in statistical inference and time series forecasting.

In this paper, the auto-regressive integrated moving average models are applied to propose a new hybrid method for improving the performance of the artificial neural networks to time series forecasting. In our proposed model, based on the Box–Jenkins methodology in linear modeling, a time series is considered as nonlinear function of several past observations and random errors. Therefore, in the first stage, an auto-regressive integrated moving average model is used in order to generate the necessary data, and then a neural network is used to determine a model in order to capture the underlying data generating process and predict the future, using preprocessed data. Empirical results with three well-known real data sets indicate that the proposed model can be an effective way in order to yield more accurate model than traditional artificial neural networks. Thus, it can be used as an appropriate alternative for artificial neural networks, especially when higher forecasting accuracy is needed.

Acknowledgement

The authors wish to express their gratitude to the A. Tavakoli, associated professor of industrial engineering, Isfahan University of Technology.

References

Arifovic, J., & Gencay, R. (2001). Using genetic algorithms to select architecture of a feed-forward artificial neural network. *Physica A, 289*, 574–594.

Armano, G., Marchesi, M., & Murru, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences, 170*, 3–33.

Atiya, F. A., & Shaheen, I. S. (1999). A comparison between neural-network forecasting techniques-case study: River flow forecasting. *IEEE Transactions on Neural Networks, 10*(2).

Balkin, S. D., & Ord, J. K. (2000). Automatic neural network modeling for univariate time series. *International Journal of Forecasting, 16*, 509–515.

Bates, J. M., & Granger, W. J. (1969). The combination of forecasts. *Operation Research, 20*, 451–468.

Baxt, W. G. (1992). Improving the accuracy of an artificial neural network using multiple differently trained networks. *Neural Computation, 4*, 772–780.

Benardos, P. G., & Vosniakos, G. C. (2002). Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. *Robotics and Computer Integrated Manufacturing, 18*, 43–354.

Benardos, P. G., & Vosniakos, G. C. (2007). Optimizing feed-forward artificial neural network architecture. *Engineering Applications of Artificial Intelligence, 20*, 365–382.

Berardi, V. L., & Zhang, G. P. (2003). An empirical investigation of bias and variance in time series forecasting: Modeling considerations and error evaluation. *IEEE Transactions on Neural Networks, 14*(3), 668–679.

Box, P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-day Inc.

Campbell, M. J., & Walker, A. M. (1977). A survey of statistical work on the MacKenzie River series of annual Canadian lynx trappings for the years 1821–1934 and a new analysis. *Journal of Royal Statistical Society Series A, 140*, 411–431.

Castillo, P. A., Merelo, J. J., Prieto, A., Rivas, V., & Romero, G. (2000). GProp: Global optimization of multilayer perceptrons using GA. *Neurocomputing, 35*, 149–163.

Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural Networks, 5*, 961–970.

Chen, A., Leung, M. T., & Hazem, D. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index. *Computers and Operations Research, 30*, 901–923.

Chen, K. Y., & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications, 32*, 54–264.

Chen, Y., Yang, B., Dong, J., & Abraham, A. (2005). Time-series forecasting using flexible neural tree model. *Information Sciences, 174*(3–4), 219–235.

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography with discussion. *International Journal of Forecasting, 5*, 559–608.

Cornillon, P., Imam, W., & Matzner, E. (2008). Forecasting time series using principal component analysis with respect to instrumental variables. *Computational Statistics and Data Analysis, 52*, 1269–1280.

Cottrell, M., Girard, B., Girard, Y., Mangeas, M., & Muller, C. (1995). Neural modeling for time series: A statistical stepwise method for weight elimination. *IEEE Transactions on Neural Networks, 6*(6), 355–1364.

De Groot, C., & Wurtz, D. (1991). Analysis of univariate time series with connectionist nets: A case study of two classical examples. *Neurocomputing, 3*, 177–192.

Demuth, H., & Beale, B. (2004). *Neural network toolbox user guide*. Natick: The Math Works Inc.

Ghiassi, M., & Saidane, H. (2005). A dynamic architecture for artificial neural networks. *Neurocomputing, 63*, 97–413.

Ginzburg, I., & Horn, D. (1994). Combined neural networks for time series analysis. *Advance Neural Information Processing Systems, 6*, 224–231.

Giordano, F., La Rocca, M., & Perna, C. (2007). Forecasting nonlinear time series with neural network sieve bootstrap. *Computational Statistics and Data Analysis, 51*, 3871–3884.

Goh, W. Y., Lim, C. P., & Peh, K. K. (2003). Predicting drug dissolution profiles with an ensemble of boosted neural networks: A time series approach. *IEEE Transactions on Neural Networks, 14*(2), 459–463.

Granger, C. W. J. (1989). Combining forecasts – Twenty years later. *Journal of Forecasting, 8*, 167–173.

Haseyama, M., & Kitajima, H. (2001). An ARMA order selection method with fuzzy reasoning. *Signal Process, 81*, 1331–1335.

Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Amsterdam: Elsevier.

Hosseini, H., Luo, D., & Reynolds, K. J. (2006). The comparison of different feed forward neural network architectures for ECG signal diagnosis. *Medical Engineering and Physics, 28*, 372–378.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrica, 76*(2), 297–307.

Hwang, H. B. (2001). Insights into neural-network forecasting time series corresponding to ARMA (p; q) structures. *Omega, 29*, 273–289.

Islam, M. M., & Murase, K. (2001). A new algorithm to design compact two hidden-layer artificial neural networks. *Neural Networks, 14*, 1265–1278.

Jain, A., & Kumar, A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing, 7*, 585–592.

Jiang, X., & Wah, A. H. K. S. (2003). Constructing and training feed-forward neural networks for pattern classification. *Pattern Recognition, 36*, 853–867.

Jones, R. H. (1975). Fitting autoregressions. *Journal of American Statistical Association, 70*(351), 590–592.

Khashei, M. (2005). *Forecasting the Esfahan steel company production price in Tehran metals exchange using artificial neural networks (ANNs)*. Master of Science Thesis, Isfahan University of Technology.

Khashei, M., Hejazi, S. R., & Bijari, M. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems, 159*, 769–786.

Kim, H., & Shin, K. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing, 7*, 569–576.

Lapedes, A., & Farber, R. (1987). *Nonlinear signal processing using neural networks: Prediction and system modeling*. Technical Report LAUR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.

Lee, J., & Kang, S. (2007). GA based meta-modeling of BPN architecture for constrained approximate optimization. *International Journal of Solids and Structures, 44*, 5980–5993.

Leski, J., & Czogala, E. (1999). A new artificial network based fuzzy interference system with moving consequents in if–then rules and selected applications. *Fuzzy Sets and Systems, 108*, 289–297.

Lin, T., & Pourahmadi, M. (1998). Nonparametric and non-linear models and data mining in time series: A case study in the Canadian lynx data. *Applied Statistics, 47*, 87–201.

Ljung, L. (1987). *System identification theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.

Luxhoj, J. T., Riis, J. O., & Stensballe, B. (1996). A hybrid econometric-neural network modeling approach for sales forecasting. *International Journal of Production Economics, 43*, 175–192.

Ma, L., & Khorasani, K. (2003). A new strategy for adaptively constructing multilayer feed-forward neural networks. *Neurocomputing, 51*, 361–385.

Marin, D., Varo, A., & Guerrero, J. E. (2007). Non-linear regression methods in NIRS quantitative analysis. *Talanta, 72*, 28–42.

Medeiros, M. C., & Veiga, A. (2000). A hybrid linear-neural model for time series forecasting. *IEEE Transaction on Neural Networks, 11*(6), 1402–1412.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of samples. *Journal of International Economics, 14*, 3–24.

Minerva, T., & Poli, I. (2001). Building ARMA models with genetic algorithms. *Lecture notes in computer science* (Vol. 2037, pp. 335–342). Springer.

Ong, C. S., Huang, J. J., & Tzeng, G. H. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematical and Computation, 164*(3), 885–912.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega, 33*, 505–597.

Pelikan, E., de Groot, C., & Wurtz, D. (1992). Power consumption in West-Bohemia: Improved forecasts with decorrelating connectionist networks. *Neural Network World, 2*, 701–712.

Poli, I., & Jones, R. D. (1994). A neural net model for prediction. *Journal of American Statistical Association, 89*, 17–121.
Reid, M. J. (1968). Combining three estimates of gross domestic product. *Economica, 35*, 31–444.
Ross, J. P. (1996). *Taguchi techniques for quality engineering*. New York: McGraw-Hill.
Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.
Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika AC-63, 1*, 17–126.
Stone, L., & He, D. (2007). Chaotic oscillations and cycles in multi-trophic ecological systems. *Journal of Theoretical Biology, 248*, 382–390.
Subba Rao, T., & Sabr, M. M. (1984). *An introduction to bispectral analysis and bilinear time series models lecture notes in statistics* (Vol. 24). New York: Springer-Verlag.
Tang, Y., & Ghosal, S. (2007). A consistent nonparametric Bayesian procedure for estimating autoregressive conditional densities. *Computational Statistics and Data Analysis, 51*, 4424–4437.
Taskaya, T., & Casey, M. C. (2005). A comparative study of autoregressive neural network hybrids. *Neural Networks, 18*, 781–789.
Timmermann, A., & Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting, 20*, 15–27.
Tong, H., & Lim, K. S. (1980). Threshold autoregressive, limit cycles and cyclical data. *Journal of the Royal Statistical Society Series B, 42*(3), 245–292.
Tsaih, R., Hsu, Y., & Lai, C. C. (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision Support Systems, 23*, 161–174.
Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change, 69*, 71–87.
Voort, M. V. D., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies, 4*, 307–318.
Wedding, D. K., & Cios, K. J. (1996). Time series forecasting by combining networks, certainty factors, RBFand the Box–Jenkins model. *Neurocomputing, 10*, 149–168.
Weigend, S., Huberman, B. A., & Rumelhart, D. E. (1990). Predicting the future: A connectionist approach. *International Journal of Neural Systems, 1*, 193–209.
Wong, C. S., & Li, W. K. (2000). On a mixture autoregressive model. *Journal of Royal Statistical Society Series B, 62*(1), 91–115.
Yu, L., Wang, S., & Lai, K. K. (2005). A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers and Operations Research, 32*, 2523–2541.
Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175.
Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences, 177*, 5329–5346.
Zhang, G. P., & Qi, G. M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research, 160*, 501–514.
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*, 35–62.
Zhou, Z. J., & Hu, C. H. (2008). An effective hybrid approach based on grey and ARMA for forecasting gyro drift. *Chaos, Solitons and Fractals, 35*, 525–529.