

A New Hybrid Linear-Nonlinear Model Based on Decomposition of Discrete Wavelet Transform for Time Series Forecasting

Warut Pannakkong^(✉) and Van-Nam Huynh

School of Knowledge Science, Japan Advanced Institute of Science and Technology,
Ishikawa, Japan
{warut,huynh}@jaist.ac.jp

Abstract. Time series forecasting research area generally aims at improving prediction accuracy. Discrete wavelet transform (DWT) has been applied to time series for decomposing it into approximation and detail. Nevertheless, typically, the property of the approximation and the detail are presumed as either linear or nonlinear. Actually, the purpose of the DWT is not decomposing the original time series into linear and nonlinear time series. Hence, this paper develops a new hybrid model of autoregressive integrated moving average (ARIMA), artificial neural network (ANN), and the DWT without prior assumption on linear and nonlinear property of the approximation and the detail. The different Khashei and Bijari's hybrid models involving the ARIMA and the ANN are built for the approximation and the detail in order to extract their both linear and nonlinear components and fit the relationship between the components as the function instead of additive relationship. Finally, the forecasted approximation and detail are combined to obtain final forecasting. The prediction capability of the proposed model is examined with two well-known time series: the sunspot and the Canadian lynx time series. The results show that the proposed model has the best performance in all two data sets and all three measures (i.e. MSE, MAE and MAPE).

Keywords: Hybrid model · Time series forecasting · Autoregressive integrated moving average (ARIMA) · Artificial neural network (ANN) · Discrete wavelet transform (DWT)

1 Introduction

For decades, time series forecasting research area contributes several real-world applications in their prediction and decision making support [1]. This research area attempts to achieve better prediction accuracy by developing effective forecasting models. Traditionally, autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) were developed and widely applied in time series forecasting. The ARIMA has an advantage in dealing with

both stationary and non-stationary time series but it presumes that the relationship between inputs (e.g. historical time series) and outputs (e.g. future time series) is linear. On the other hand, the ANN has no such assumption. However, there is no a universal forecasting model that has the best performance in all situations. Hence, applying sole single forecasting model to time series prediction is not adequate to predict real-world time series [2].

Discrete wavelet transform (DWT), a transformation technique for signals, is adapted to transform time series into approximation (trend) and detail (noise) before further analysis. With the DWT, the prediction accuracy of the ARIMA and the ANN has been improved in many applications such as short term load [3]; electrical price [4, 5]; groundwater level [6]; river discharge [7–9]; hourly flood [10]; rainfall and runoff [11, 12].

Moreover, Khandelwal et al. [13] has developed the hybrid model of the ARIMA and the ANN (Zhang’s model [2]) with the DWT. Nevertheless, this hybrid model considers the approximation as only a nonlinear component; in fact, there is no theoretical prove whether the approximation is linear or nonlinear. In addition, the additive relationship between linear and nonlinear components is assumed in the final forecasting step.

This study proposes a new hybrid model that can capture both linear and nonlinear components of the approximation and the detail, and has no assumption on relationship between linear and nonlinear components. Firstly, the discrete wavelet transform (DWT) is used to decompose the time series. Then, the hybrid model of ARIMA and ANN are constructed for the approximation and the detail to extract their linear and nonlinear components. Eventually, the final prediction is the combination of the predicted approximation and detail.

The rest of this paper is organized as follows. In Sect. 2, the ARIMA, the ANN, and the DWT are briefly explained. In Sect. 3, the proposed model is presented. The experiments and their results are shown and interpreted in Sect. 4. Finally, Sect. 5 provides the conclusions.

2 Preliminaries

2.1 Autoregressive Integrated Moving Average (ARIMA)

The autoregressive integrated moving average (ARIMA) is a popular forecasting model for decades due to its capability in handling both stationary and nonstationary time series [1]. However, The ARIMA assumes the relationship between predicted and historical time series as linear relationship. The ARIMA consists of three parts: autoregressive (AR), integration (I), and moving average (MA). In the situation that the time series is nonstationary, the time series is transformed by differencing in integration (I) step. The mathematical expression of the ARIMA can be written as:

$$\phi_p(B)(1-B)^d y_t = c + \theta_q(B)a_t \quad (1)$$

where y_t and a_t denote the time series and random error in period t respectively, $\phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i$, $\theta_q(B) = 1 - \sum_{j=1}^q \theta_j B^j$, B denotes the backward shift

operator defined as $B^i y_t = y_{t-i}$, ϕ_i and θ_j denote the parameters of AR and MA respectively, p and q denote the orders of AR and MA respectively, and d denotes the degree of differencing, and c denotes the constant.

2.2 Artificial Neural Network (ANN)

The artificial neural network (ANN) is an artificial intelligent imitating biological neurons, and it is good at nonlinear modeling [14]. The ANN is widely used in time series forecasting because it is more flexible than the ARIMA in capturing relationship between predicted and historical time series without assumption. Typically, the ANN consists of three types of layer: input, hidden, and output layers. There are nodes in each layer. Normally, the architects choose the number of the layers and the nodes by their intuition in the problem and trial and error. Nevertheless, a feed-forward neural network that has only one hidden layer has been tested that it can be considered as a universal approximator [15]. The mathematical expression of the feed-forward neural network [16] can be written as:

$$y_t = f \left(b_h + \sum_{h=1}^R w_h g \left(b_{i,h} + \sum_{i=1}^Q w_{i,h} p_i \right) \right) \quad (2)$$

where y_t denotes the time series at period t , $b_{i,h}$ and b_h denote the biases of hidden and output layers, f and g denote the transfer functions which are typically linear and nonlinear functions respectively, $w_{i,h}$ and w_h denote the connection weights between the layers, Q and R denote the numbers of the input nodes and the hidden nodes respectively.

In this paper, the feed-forward neural network that has only one hidden layer and Levenberg-Marquardt algorithm with Bayesian regularization training algorithm [17] is applied in the experiments.

2.3 Discrete Wavelet Transform (DWT)

The wavelet transform is a tool for simultaneously analysis of both time and frequency of signals [18]. After the analysis, the original signal is decomposed into low frequency (approximation) and high frequency (detail) by applying low and high frequency pass filters. In case of multiple decomposition level, the approximation and the detail in the next level are the decomposition of the approximation in the previous level. In fact, there are two main categories of the wavelet transform such as continuous and discrete wavelet transforms. Nevertheless, in real-word applications, the time series are discrete and appropriate to be decomposed by the discrete wavelet transform (DWT) as:

$$\begin{aligned} y_t &= A_J(t) + \sum_{j=1}^J D_j(t) \\ &= \sum_{k=1}^K c_{J,k} \phi_{J,k}(t) + \sum_{j=1}^J \sum_{k=1}^K d_{j,k} \psi_{j,k}(t) \end{aligned} \quad (3)$$

where y_t denotes the time series in period t ; $A_J(t)$ denotes the approximation in the highest decomposition level (J); $D_j(t)$ denotes the detail in decomposition level j ; $c_{j,k}$ and $d_{j,k}$ denote the coefficients of the approximation and detail respectively, in decomposition level j and in period k ; $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ denote low (approximation) and high (detail) pass filters respectively, in decomposition level j and at period k ; K denotes the length of the time series; J denotes the highest level of decomposition.

3 Proposed Forecasting Model

The main objective of developing the proposed model is to obtain the advantage of both the ARIMA and the ANN in fitting linear and nonlinear components from the time series without presuming the characteristic of the approximation and the detail as either linear or nonlinear. The proposed model can be divided into three steps: decomposition of the time series, capturing linear and nonlinear components, and final forecasting (Fig. 1).

In the first step, the actual time series (y_t) is decomposed by the DWT with Daubechies wavelet basis function in order to obtain the approximation (y_t^{app}), which presents the trend, and the detail (y_t^{det}), which is the difference between actual value and the trend. The pattern of the detail reveals the seasonality, the white noise, etc.

Instead of applying a forecasting model to time series consisting of both trend and noise, using different forecasting models to separately predict the trend from the approximation and predict the noise (e.g. seasonality and white noise) from

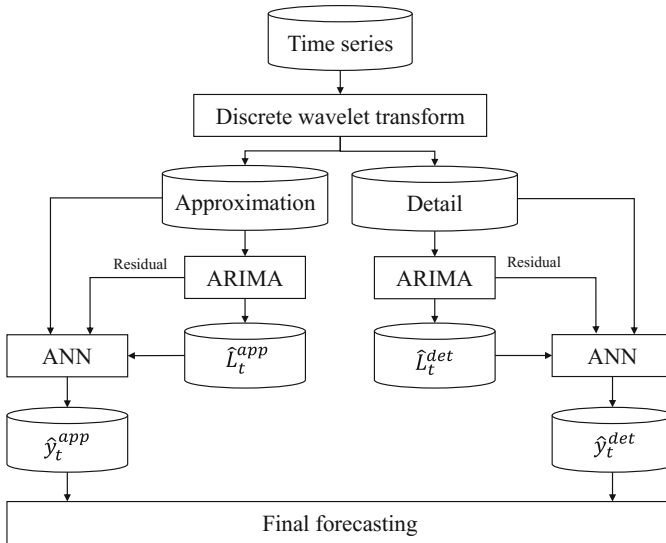


Fig. 1. The proposed forecasting model

the detail can provide better prediction results because each forecasting model deal with only either trend or noise, not both of them simultaneously.

In the second step, the Khashei and Bijari's hybrid model of the ARIMA and the ANN [19] is applied to both the approximation and the detail. This step contributes the new approach that does not make either linear or nonlinear assumption to the property of the approximation and the detail, and does not assume additive relationship between their linear and nonlinear components as well.

Generally, the Khashei and Bijari's model forecasts the future time series at period t (\hat{y}_t) by using linear (\hat{L}_t) and nonlinear (\hat{N}_t) components as:

$$\hat{y}_t = f(\hat{L}_t, \hat{N}_t) \tag{4}$$

The linear component (\hat{L}_t) is the result of adopting the ARIMA to the actual time series (y_t). After that, the residual of the ARIMA (e_t) can be computed as:

$$e_t = y_t - \hat{L}_t \tag{5}$$

In case of the nonlinear component (\hat{N}_t), it can be obtained from the ANN that has the lagged values of both the time series (y_t) and the ARIMA residual (e_t) as its inputs:

$$\hat{N}_t^1 = f^1(e_{t-1}, e_{t-2}, \dots, e_{t-n}) \tag{6}$$

$$\hat{N}_t^2 = f^2(y_{t-1}, y_{t-2}, \dots, y_{t-m}) \tag{7}$$

where f^1 and f^2 denote the function fitted by the ANN, n and m is total included lagged periods.

In the proposed model, the Khashei and Bijari's model is separately built for the approximation and the detail as:

$$\hat{y}_t^{app} = f^{app}(\hat{L}_t^{app}, \hat{N}_t^{app}) \tag{8}$$

$$\hat{y}_t^{det} = f^{det}(\hat{L}_t^{det}, \hat{N}_t^{det}) \tag{9}$$

where \hat{y}_t^{app} and \hat{y}_t^{det} denote the forecasted approximation and detail respectively, in period t ; f^{app} and f^{det} denote the function fitted by the ANN; \hat{L}_t^{app} and \hat{L}_t^{det} denote the linear components of the approximation and the detail respectively, in period t ; \hat{N}_t^{app} and \hat{N}_t^{det} denote the nonlinear components of the approximation and the detail respectively, in period t .

The linear components (\hat{L}_t^{app} and \hat{L}_t^{det}) denote the result of applying the ARIMA to y_t^{app} and y_t^{det} respectively. Then, the ARIMA residual of the approximation (e_t^{app}) and the detail (e_t^{det}) can be mathematically expressed as:

$$e_t^{app} = y_t^{app} - \hat{L}_t^{app} \tag{10}$$

$$e_t^{det} = y_t^{det} - \hat{L}_t^{det} \tag{11}$$

For the nonlinear components (\hat{N}_t^{app} and \hat{N}_t^{det}), they can be produced from the ANN as:

$$\hat{N}_t^{app1} = f^{app1}(e_{t-1}^{app}, e_{t-2}^{app}, \dots, e_{t-n_1}^{app}) \tag{12}$$

$$\hat{N}_t^{app2} = f^{app2}(y_{t-1}^{app}, y_{t-2}^{app}, \dots, y_{t-m_1}^{app}) \quad (13)$$

$$\hat{N}_t^{det1} = f^{det1}(e_{t-1}^{det}, e_{t-2}^{det}, \dots, e_{t-n_2}^{det}) \quad (14)$$

$$\hat{N}_t^{det2} = f^{det2}(y_{t-1}^{det}, y_{t-2}^{det}, \dots, y_{t-m_2}^{det}) \quad (15)$$

where f^{app1} , f^{app2} , f^{det1} , and f^{det2} denote functions fitted by the ANN; n_1 , n_2 , m_1 , and m_2 denote total lagged periods that are identified by trial and error in the experiments.

After that, the forecasted approximation (\hat{y}_t^{app}) and the forecasted detail (\hat{y}_t^{det}) can be obtained as:

$$\hat{y}_t^{app} = f^{app}(\hat{L}_t^{app}, e_{t-1}^{app}, e_{t-2}^{app}, \dots, e_{t-n_1}^{app}, y_{t-1}^{app}, y_{t-2}^{app}, \dots, y_{t-m_1}^{app}) \quad (16)$$

$$\hat{y}_t^{det} = f^{det}(\hat{L}_t^{det}, e_{t-1}^{det}, e_{t-2}^{det}, \dots, e_{t-n_2}^{det}, y_{t-1}^{det}, y_{t-2}^{det}, \dots, y_{t-m_2}^{det}) \quad (17)$$

Finally, the final forecasting step is performed by combining of the forecasted approximation (\hat{y}_t^{app}) and the forecasted detail (\hat{y}_t^{det}) as:

$$\hat{y}_t = \hat{y}_t^{app} + \hat{y}_t^{det} \quad (18)$$

In sum, rather than applying the Khashei and Bijari's model direct to the time series, the DWT is used at first to transform the time series into the approximation (trend) and the detail (noise). Then, without assuming linear or nonlinear properties of the approximation and the detail, the Khashei and Bijari's model is adopted to both of them. After the specific Khashei and Bijari's models have been separately built to capture the trend and the noise, they would give the better forecasting result because the different Khashei and Bijari's models concentrate on only either trend or noise (not both of them simultaneously). In addition, the relationship of the linear and nonlinear components is defined as the function instead of additive relationship. Finally, the final forecasting is performed by additive combination between forecasted approximation the detail because the relationship between them is additive as well.

4 Experiments and Results

To assess forecasting capability of the proposed model, two well-known time series (Table 1) are used as case studies such as Wolf's sunspot (Fig. 2) and Canadian lynx (Fig. 3). The measures of forecasting performance used in this paper are mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The performance of the proposed model is

Table 1. Detail of time series and experiment

Time series	Size (total, training, test)
Sunspot (1700–1987)	(288, 221, 67)
Canadian lynx (1821–1934)	(114, 100, 14)

Amount of spot

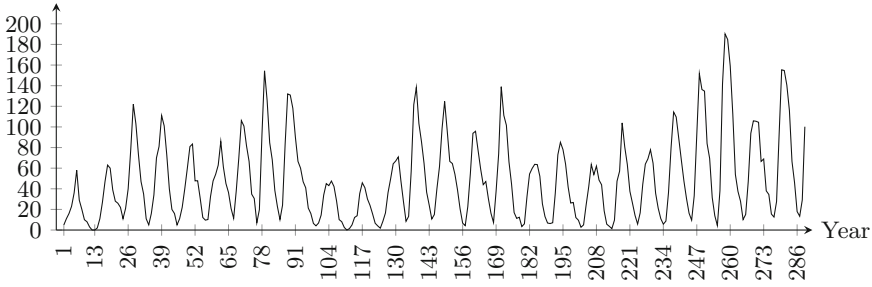


Fig. 2. Sunspot time series (1700–1987)

Amount of lynx

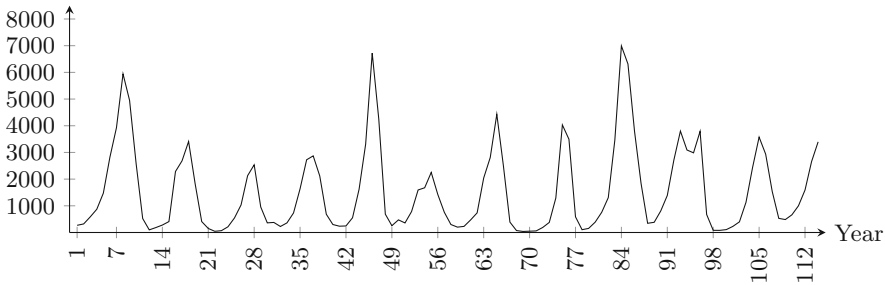


Fig. 3. Canadian lynx time series (1821–1934)

Table 2. Sunspot forecasting result

Model	35 year ahead			67 year ahead		
	MSE	MAE	MAPE	MSE	MAE	MAPE
ARIMA	197.87	10.52	29.17%	323.48	13.25	32.86%
ANN	164.08	9.51	31.76%	413.90	14.19	33.34%
Zhang	156.76	9.63	30.22%	300.88	12.74	32.08%
Khashei and Bijari	127.67	8.60	22.85%	273.15	12.14	25.31%
Khandelwal et al.	144.09	8.21	20.22%	378.00	12.21	21.95%
Proposed model	121.52	5.51	16.21%	206.32	8.07	19.19%

compared with the ARIMA, the ANN, the Zhang’s model, Khashei and Bijari’s model, and Khandelwal et al.’s model.

For the sunspot time series, it contains 288 annual records (1700–1987). The training and test sets are 221 records (1700–1920) and 67 records (1921–1987) respectively. Firstly, the sunspot time series is decomposed by the DWT into the approximation and the detail. Secondly, the ARIMA is applied to both the approximation and the detail. The most fitted ARIMA for the approximation

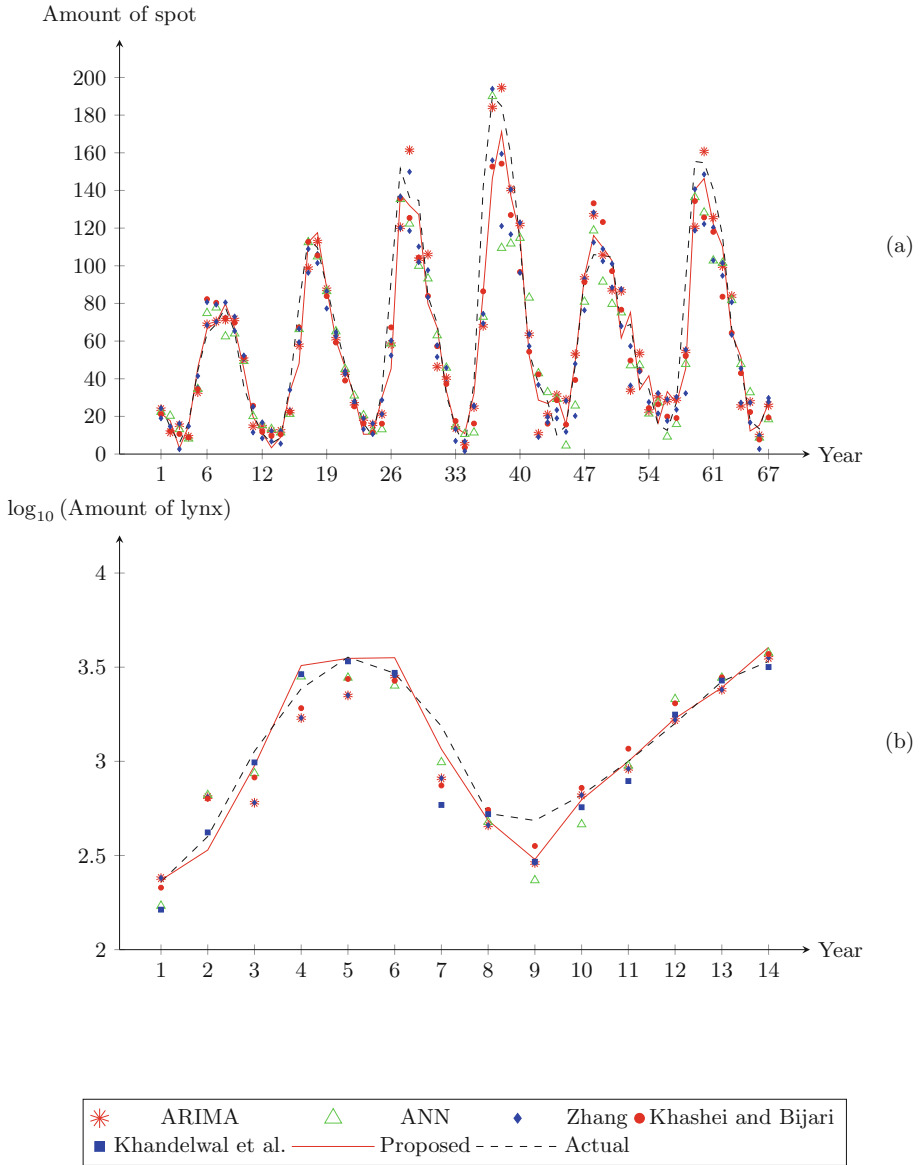


Fig. 4. Forecasted values: (a) Sunspot, (b) Canadian lynx

and the detail are ARIMA(0, 0, 6) and ARIMA(0, 0, 3) respectively. Thirdly, the forecasted approximation and detail are generated by the best fitted ANNs that are ANN(16-1-1) and ANN(9-10-1) respectively. Then, the final forecasting is computed from the combination of the predicted approximation and detail. After obtaining the final forecasting, the performance measures are computed for short term (35 years) and long term (67 years) horizontal predictions (Table 2).

Table 3. Lynx forecasting result

Model	MSE	MAE	MAPE
ARIMA	0.0229	0.1120	3.7062%
ANN	0.0201	0.1165	4.0156%
Zhang	0.0247	0.1083	3.5504%
Khashei and Bijari	0.0160	0.0980	3.2381%
Khandelwal et al.	0.0195	0.0873	2.9737%
Proposed model	0.0071	0.0639	2.1114%

According to the performance comparison, the proposed model has the lowest error in all three measures. The MSE, MAE and MAPE in short term prediction are 121.52, 5.51 and 16.21% respectively. On the other hand, the MSE, MAE and MAPE in long term prediction are 206.32, 8.07 and 19.19% which are higher than the short term prediction because the long term prediction has the highest value at period 37 (see Fig. 4a) that increases the variance causing more prediction error. Nevertheless, the proposed model can still have the best performance because the other models also perform worse. Hence, the proposed model is the best forecasting model for the sunspot time series in both short and long term forecasting.

The Canadian lynx time series consists of 114 annual records (1821–1934). The training and test sets are 100 records (1821–1920) and 14 records (1921–1934) respectively. The best fitted ARIMAs of the approximation and the detail are ARIMA(0, 0, 5) and ARIMA(2, 0, 0) respectively. The most appropriate ANNs in forecasting the approximation and the detail are ANN(7-9-1) and ANN(6-3-1) respectively. From the performance comparison shown in Table 3, the proposed model gives the best performance in MSE, MAE and MAPE that are 0.0071, 0.0639 and 2.1114% respectively. The most improved measure is MSE. The lower MSE gives more chance to promise lower maximum of error because the MSE is sensitive to a huge error. Therefore, the proposed model has the lowest maximum error which is in period 9 (Fig. 4b).

5 Conclusions

In order to enhance forecasting accuracy in time series prediction, the new hybrid model of the ARIMA, the ANN, and the DWT has been proposed. The proposed model analyses the time series without assuming linear and nonlinear properties on the approximation and the detail, and defines the relationship of the linear and nonlinear components of both the approximation and the detail as the function. The prediction capability of the proposed model is examined with two well-known time series: the sunspot and the Canadian lynx time series. The results show that the proposed model has the best performance in all two data

sets and all three measures (i.e. MSE, MAE and MAPE). The improved performance implies benefit of hybridization of the ARIMA, the ANN, and the DWT in capturing the linear and nonlinear components of the approximation and the detail without prior assumption on their properties.

The limitation in the experiment is that the level of decomposition is one. For future works, the impact of different decomposition levels will be considered. Moreover, the statistical test will be performed to measure the significant level of performance improvement.

References

1. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. *Int. J. Forecast.* **22**(3), 443–473 (2006)
2. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**, 159–175 (2003)
3. Fard, A.K., Akbari-Zadeh, M.R.: A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting. *J. Exp. Theor. Artif. Intell.* **26**(2), 167–182 (2014)
4. Conejo, A.J., Plazas, M.A., Espinola, R., Molina, A.B.: Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Trans. Power Syst.* **20**(2), 1035–1042 (2005)
5. Tan, Z., Zhang, J., Wang, J., Xu, J.: Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Appl. Energ.* **87**(11), 3606–3610 (2010)
6. Adamowski, J., Chan, H.F.: A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **407**(1), 28–40 (2011)
7. Zhou, H.C., Peng, Y., Liang, G.H.: The research of monthly discharge predictor-corrector model based on wavelet decomposition. *Water Resour. Manage.* **22**(2), 217–227 (2008)
8. Wei, S., Zuo, D., Song, J.: Improving prediction accuracy of river discharge time series using a wavelet-NAR artificial neural network. *J. Hydroinformatics* **14**(4), 974–991 (2012)
9. Adamowski, J., Sun, K.: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* **390**(1), 85–91 (2010)
10. Tiwari, M.K., Chatterjee, C.: Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J. Hydrol.* **394**(3), 458–470 (2010)
11. Nourani, V., Komasi, M., Mano, A.: A multivariate ANN-wavelet approach for rainfallruno modeling. *Water Resour. Manage.* **23**(14), 2877–2894 (2009)
12. Partal, T., Kisi, Ö.: Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J. Hydrol.* **342**(1), 199–212 (2007)
13. Khandelwal, I., Adhikari, R., Verma, G.: Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Comput. Sci.* **48**, 173–179 (2015)
14. Zhang, G., Patuwo, B.E., Hu, M.Y.: Forecasting with artificial neural networks: the state of the art. *Int. J. Forecast.* **14**(1), 35–62 (1998)
15. Hecht-Nielsen, R.: Theory of the backpropagation neural network. In: *International Joint Conference on Neural Networks, IJCNN 1989*, pp. 593–605 (1989)

16. Dayhoff, J.A.: *Neural Network Architectures: An Introduction*. MIT press, Cambridge (1995)
17. MacKay, D.J.: A practical bayesian framework for backpropagation networks. *Neural comput.* **4**(3), 448–472 (1992)
18. Fliege, N.J.: *Multirate Digital Signal Processing: Multirate Systems, Filter Banks, Wavelets*. John Willy & Sons, Chichester (1994)
19. Khashei, M., Bijari, M.: A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* **11**(2), 2664–2675 (2011)