Seasonal Time Series Data Forecasting by Using Neural Networks Multiscale Autoregressive Model

Suhartono, B.S.S. Ulama and A.J. Endharta
Department of Statistics, Faculty of Mathematics and Natural Sciences,
Institute Technology Sepuluh Nopember, Surabaya 60111, Indonesia

Abstract: Problem statement: The aim of this research was to study further some latest progress of wavelet transform for time series forecasting, particularly about Neural Networks Multiscale Autoregressive (NN-MAR). **Approach:** There were three main issues that be considered further in this research. The first was some properties of scale and wavelet coefficients from Maximal Overlap Discrete Wavelet Transform (MODWT) decomposition, particularly at seasonal time series data. The second focused on the development of model building procedures of NN-MAR based on the properties of scale and wavelet coefficients. Then, the third was empirical study about the implementation of the proposed procedure and comparison study about the forecast accuracy of NN-MAR to other forecasting models. **Results:** The results showed that MODWT at seasonal time series data also has seasonal pattern for scale coefficient, whereas the wavelet coefficients are stationer. The result of model building procedure development yielded a new proposed procedure of NN-MAR model for seasonal time series forecasting. In general, this procedure accommodated input lags of scale and wavelet coefficients and other additional seasonal lags. In addition, the result showed that the proposed procedure works well for determining the best NN-MAR model for seasonal time series forecasting. **Conclusion:** The comparison study of forecast accuracy showed that the NN-MAR model yields better forecast than MAR and ARIMA models.

Key words: Neural networks, multiscale, MODWT, NN-MAR, seasonal, time series

INTRODUCTION

Recently, neural network has been proposed in many researches about different kinds of statistical analysis. There are many types of neural network applied to solve many problems. For examples, Feedforward Neural Network (FFNN) is applied in electricity demand forecasting Taylor *et al*. (2006), General Regression Neural Network (GRNN) is used in exchange rates forecasting and Recurrent Neural Network (RNN) has been applied in detecting changes in autocorellated process for quality monitoring. Different from those previous researches, here, the predictors or the inputs are not the lags of the variables or the data variables, but they are the coefficients from wavelet transformation.

A new development related with wavelet transformation application for time series analysis is proposed. As an overview this can be seen in Nason and von Sachs (1999). At the beginning, most wavelet research for time series analysis is focused on periodogram or scalogram analysis of periodicities and cycles evaluation (Priestley, 1996; Morettin, 1997; Gao, 1997; Percival and Walden, 2000). Bjorn (1995); Soltani *et al*. (2000) and Renaud *et al*. (2003) are some first researcher groups discussing wavelet for time series prediction based on autoregressive model. In this case, wavelet transformation gives good decomposition from a signal or time series, so that the structure can be evaluated by parametric or nonparametric models.

WNN is a neural network with wavelet function used in processing in transfer function. In time series forecasting cases, input used in WNN is wavelet coefficients in certain time and resolution. Recently, there are some articles about WNN for time series forecasting and filtering, such as Bashir and El-Hawary (2000); Renaud *et al*. (2003); Murtagh *et al*. (2004) and Chen *et al*. (2006).

Wavelet transformation that is mostly used for time series forecasting is Maximal Overlap Discrete Wavelet Transform (MODWT). The use of MODWT is to solve the limitation of Discrete Wavelet Transform (DWT), that requires $N = 2^J$ where J is positive integer. In practice, time series data rarely

Corresponding Author: Suhartono, Department of Statistics, Faculty of Mathematics and Natural Sciences,
Institute Technology Sepuluh Nopember, Surabaya 60111, Indonesia

fulfill those numbers, which are two powered with a positive integer.

Some present researches related with WNN for time series forecasting usually focus on how to determine the best WNN model which is appropriate for time series forecasting. The aim of this research is to develop an accurate procedure for WNN modeling of seasonal time series data and to compare the forecast accuracy with Multiscale Autoregressive (MAR) and ARIMA models.

MATERIALS AND METHODS

Data: The number of tourist arrivals to Bali through Ngurah Rai airport, from January 1986 until April 2008, is used as a case study. The in-samples are first 216 observations and the last 16 observations are used as the out-sample dataset. The analysis starts by applying MODWT decomposition to the data. Based on the scale and wavelet coefficients pattern, then the proposed of WNN model building procedure for time series data forecasting will be developed. This procedure is the improvement of general FFNN model building procedure for time series data forecasting. In this new procedure, the determination of the inputs in WNN model is done by using wavelet coefficient lags and the boundary effects. Whereas, the selection of the best WNN model is done by employing a combination between the inferential statistics for the addition contribution in forward scheme for selecting the optimum number of neurons in the hidden layer and Wald test in backward scheme for determining the optimum input unit.

Wavelets and prediction: Wavelet means small wave, whereas by contrast, sinus and cosines are big waves (Percival and Walden, 2000). A function $\psi(.)$ is defined as wavelet if it satisfies:

$$\int_{-\infty}^{\infty} \psi(u)du = 0 \tag{1}$$

$$\int_{-\infty}^{\infty} \psi^2(u)du = 1 \tag{2}$$

Commonly, wavelets are functions that have characteristic as in Eq. 1. If it is integrated on $(-\infty, \infty)$ the result is zero and the integration of the quadrate of function $\psi(.)$ equals to 1 as written in Eq. 2.

There are two functions in wavelet transform, i.e., scale function (father wavelet) and mother wavelet. These two functions give a function family that can be used for reconstructing a signal. Some wavelet families are Haar wavelet (the oldest and simplest wavelet), Meyer wavelet, Daubechies wavelet, Mexican hat wavelet, Coiflet wavelet and last assymetric wavelet (Daubechies, 1992).

Scale and wavelet equations: Scale equation or dilate equation shows scale function ϕ experiencing contraction and translation (Debnath, 2001), which is written as:

$$\phi(t) = \sqrt{2}\sum_{l=0}^{L-1} g_l \phi(2t-l) \tag{3}$$

where, $\phi(2t-l)$ is scale function $\phi(t)$ experiencing contraction or translation in time axis with l steps with scale filter coefficient g_l. Wavelet function ψ is defined as:

$$\psi(t) = \sqrt{2}\sum_{l=0}^{L-1} (-1)^l g_l \phi(2t+1-L/2+1) \tag{4}$$

Coefficient g_l must satisfy conditions:

$$\sum_{l=0}^{L-1} g_l = \sqrt{2} \quad \text{dan} \quad \sum_{l=0}^{L-1} (-1)^l l^m g_l = 0 \tag{5}$$
$$\text{for } m = 0,1,\cdots,(L/2)-1$$

and:

$$\sum_{l=0}^{L-1} g_l g_{l+2m} = 0, \quad m \neq 0 \quad \text{for } m = 1,\cdots,(L/2)-1 \tag{6}$$

and:

$$\sum_{l=0}^{L-1} g_l^2 = 1 \tag{7}$$

The relationship between coefficients h_l and g_l is $h_l = (-1)^l g_{l-1}$, or it can be written as $g_l \equiv (-1)^{l+1} h_{l-1}$.

Maximal Overlap Discrete Wavelet Transform (MODWT): One of modifications from Discrete Wavelet Transform (DWT) is Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT has been discussed in wavelet literatures with some names, such as undecimated-Discrete Wavelet Transform (DWT), Shift invariant DWT, wavelet frames, translation DWT, non decimated DWT. Percival and Walden (2000) stated that essentially those names are the same with MODWT which have connotation as 'mod DWT' or modified DWT. This is the reason of this research using Maximal Overlap Discrete Wavelet Transform (MODWT) term.

DWT suppose the data satisfy 2^j. In real world most time series data has the length that is not following this multiplication. MODWT has the advantage, which can eliminate the presence of data reduction to the half (down sampling). So that in MODWT there are N wavelet and scale coefficients in each levels of MODWT (Percival and Walden, 2000).

If there is time series data x, with N-length, the MODWT transformation will give column vectors $w_1, w_2, ..., w_{J_0}$ and v_{J_0}, each with N-length. Vector w_J contains scale coefficients. As in DWT, in MODWT the efficient calculation is done by pyramid algorithm. The smoothing coefficient of signal X is obtained iteratively by multiplying X with scale filter or low pass (g) and wavelet filter or high pass (h). In order to abridge the relationship of DWT and MODWT, wavelet filter and scale filter definitions given by:

Definition 1: (Percival and Walden, 2000): MODWT wavelet filter $\{\tilde{h}_l\}$ through $\tilde{h}_l \equiv h_l / \sqrt{2}$ and MODWT scale filter $\{\tilde{g}_l\}$ through $\{\tilde{g}_l\}\tilde{g}_l \equiv g_l / \sqrt{2}$. So that MODWT wavelet filter must satisfy this equation:

$$\sum_{l=0}^{L-1} \tilde{h}_l = 0, \quad \sum_{l=0}^{L-1} \tilde{h}_l^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} \tilde{h}_l \tilde{h}_{l+2m} = 0 \qquad (8)$$

and the scale filter must accomplish the following equation:

$$\sum_{l=0}^{L-1} \tilde{g}_l = 1, \quad \sum_{l=0}^{L-1} \tilde{g}_l^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} \tilde{g}_l \tilde{g}_{l+2m} = 0 \qquad (9)$$

Time series prediction by using wavelet: Generally, time series forecasting given by using wavelet is a forecasting method that use data preprocessing through wavelet transform, especially through MODWT. By the presence of multiscale decomposition like wavelet, the advantage is automatically separating the data components, such as trend component and irregular component in the data. Thereby, this method could be used for forecasting of stationary data (contain only irregular components) or non-stationary data (contain trend and irregular components).

For example, suppose that stationary signal X = $(X_1, X_2, ..., X_t)$ and assume that value X_{t+1} will be forecasted. The basic idea is to use coefficients that are constructed from the decomposition, i.e., (Renaud *et al.*, 2003):

$$w_{j, t-2^j(k-1)} \quad \text{for } k = 1, 2, ..., A_j, \; j = 1, 2, ..., J$$

and:

$$v_{J, t-2^j(k-1)} \quad \text{for } k = 1, 2, ..., A_j$$

The first step that should be known is how many and which wavelet coefficients that should be used in each scale.

Renaud *et al.* (2003) introduced a process to calculate the forecast at time $(t+1)^{th}$ by using wavelet model as illustrated in Fig. 1. Figure 1 represents the common form of wavelet modeling with level J = 4, order $A_j = 2$ and N = 16. Fig. 1 illustrates that if the 18th data will be forecasted, the input variables are wavelet coefficients in level 1 at t = 17 and t = 15, level 2 at t = 17 and t = 13, level 3 at t = 17 and t = 9, level 4 at t = 17 and t = 1 and smooth coefficient in level 4 at t = 17 and t = 1. Hence, we can conclude that the second input at each level is $t-2^j$.

The basic idea of multiscale decomposition is trend pattern influences Low frequency (L) components that tend to be deterministic; whereas High frequency (H) component is still stochastic. The second point in wavelet modeling for forecasting is about the function used to process the inputs, i.e., wavelet coefficients to forecast at $(t+1)^{th}$ period. Generally, there are two kinds of function that can be used in this input-output processing, such as linear and nonlinear functions.

Renaud *et al.* (2003) developed a linear wavelet model known as Multiscale Autoregressive (MAR) model. Moreover, Renaud *et al.* (2003) also introduced the possibility of the nonlinear model use in input-output processing of wavelet model, especially Feed-Forward Neural Network (FFNN). Furthermore the second model is known as Wavelet Neural Network (WNN) model. These two approaches use the lags of wavelet coefficients as the inputs, i.e. scale and smooth coefficients as in Fig. 1.
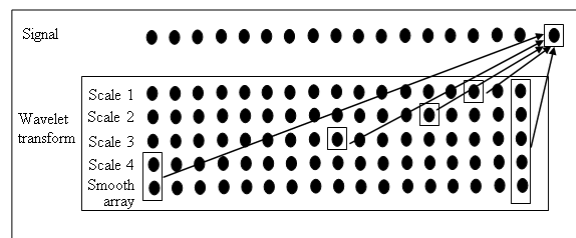

Fig. 1: Wavelet modeling illustration for J = 4 and $A_j = 2$

Multiscale Autoregressive (MAR): An autoregressive process with order p which is known as AR(p) can be written as:

$$\hat{X}_{t+1} = \sum_{k=1}^{p} \hat{\phi}_k X_{t-(k-1)}$$

By using decomposition of wavelet coefficients, Renaud *et al.* (2003) explained that AR prediction in this way could be expanded become Multiscale Autoregressive (MAR) model, i.e.:

$$\hat{X}_{t+1} = \sum_{j=1}^{J}\sum_{k=1}^{A_j} \hat{a}_{j,k} w_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} \hat{a}_{J+1,k} v_{J,t-2^j(k-1)} \qquad (10)$$

Where:
j   = The numbers of level (j = 1,2,…,J)
A_j = Order of MAR model (k = 1,2,…,A_j)
$w_{j,i}$ = Wavelet coefficient value
$v_{j,t}$ = Scale coefficient value
$a_{j,k}$ = MAR coefficient value

Wavelet neural network: Suppose that a stationary signal $X = (X_1, X_2, …, X_t)$ and assume that X_{t+1} will be predicted. The basic idea of wavelet neural network model is the coefficients that are calculated by the decomposition as in Fig. 1 are used as inputs at certain neural network architecture for obtaining the prediction of X_{t+1}. Renaud *et al.* (2003) introduced Multilayer Perceptron (MLP) neural network architecture or known as Feed-Forward Neural Network (FFNN) to process the wavelet coefficients. The architecture of this FFNN consists of one hidden layer with P neurons that is written as:

$$\hat{X}_{N+1} = \sum_{p=1}^{P} \hat{b}_p g \left(\sum_{j=1}^{J}\sum_{k=1}^{A_j} \hat{a}_{j,k,p} w_{j,N-2^j(k-1)} + \sum_{k=1}^{A_{J+1}} \hat{a}_{J+1,k,p} v_{j,N-2^j(k-1)} \right) \qquad (11)$$

where, g is an activation function in hidden layer, which is usually sigmoid logistic. In this FFNN, the activation function in output layer is linear. Furthermore, model in Eq. 11 is known as Wavelet Neural Network (WNN) or Multiresolution Neural Network (MNN).

Procedures: There are four proposed procedures for building WNN model for forecasting non-stationary (in mean) time series, i.e.:

- The inputs are the lags of scale and wavelet coefficients similar to Renaud *et al.* (2003)
- The inputs are the combination between the lags of scale and wavelet coefficients proposed by Renaud *et al.* (2003) and some additional lags that are identified by using stepwise
- The inputs are the lags of scale and wavelet coefficients proposed by Renaud *et al.* (2003) from differencing data
- The inputs are the combination between the lags of scale and wavelet coefficients proposed by Renaud *et al.* (2003) and some additional lags identified by using stepwise from differencing data

In this research, the additional lags are the seasonal lags because of the data pattern. The first and second procedures are used for the stationary data, whereas the third and fourth procedures are used for data that contain a trend. This study only illustrates the fourth procedure. Stepwise method is used to simplify the process in finding the significant inputs. After building WNN model, the results at out-sample dataset are compared to MAR and ARIMA models to find the best model for forecasting the number of tourist arrivals to Bali.

At the proposed first new procedure, the selection of the best WNN model is done firstly by determining an appropriate number of neurons in the hidden layer. The starting step before applying the proposed procedure is the determination of the levels or J in MODWT. In this case, all scale and wavelet coefficient lags from MAR(1) and additional seasonal lags which are significant based on stepwise method are used as inputs. Different from linear wavelet model (MAR) that the modeling process was divided into two additive parts, namely modeling the trend by using wavelet coefficients and MAR modeling for the residual by using the wavelet and scale coefficient lags. In this proposed procedure, the modeling of WNN is done simultaneously by using scale and wavelet coefficient lags. This is based on the fact that WNN is nonlinear model expected to be able to catch data characteristics simultaneously by using scale and wavelet coefficients from MODWT. The first proposed procedure for WNN model building for forecasting seasonal time series data can be seen at Fig. 2.
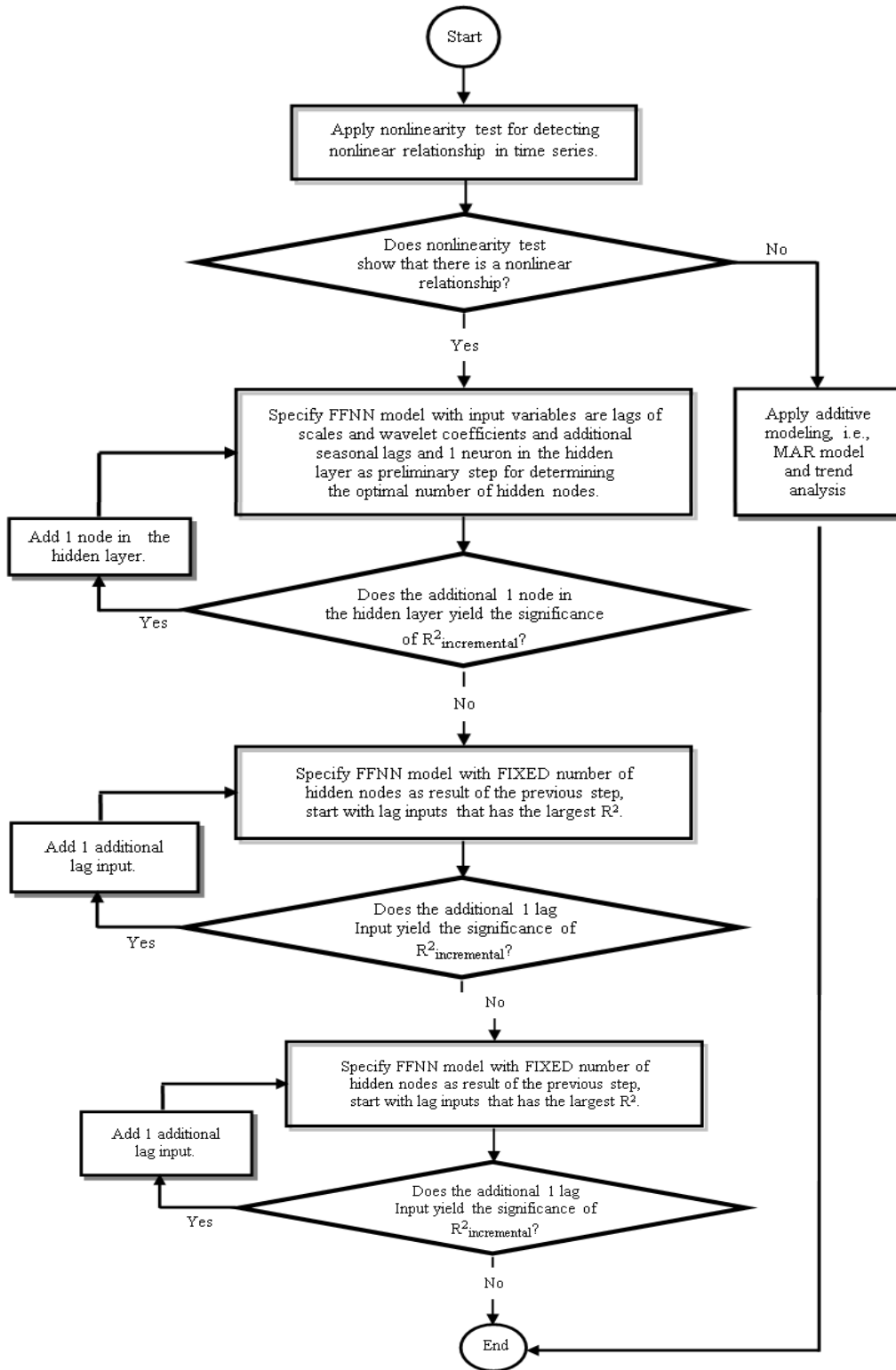
Fig. 2: The procedure for WNN model building for forecasting seasonal time series data using inference combination of $R^2_{incremental}$ and Wald test

RESULTS AND DISCUSSION

The time series plot of the number of tourist arrivals to Bali through Ngurah Rai airport is shown in Fig. 3. The plot shows that the data has seasonal and trend patterns. These data have been analyzed by using MAR and ARIMA models and the results showed that MAR(J = 4;[12,36],[12,36],[36],[0],[0])-Haar yielded better forecast than ARIMA model.

As the starting step, the modeling focuses to determine an appropriate number of neurons in the hidden layer. In this study, scale and wavelet coefficient lag inputs are assumed as lag inputs in nonlinearity test in the first step.

Every proposed procedure is begun by using nonlinearity test, i.e., White test and Terasvirta test. By using scale and wavelet coefficient lags as the inputs as proposed by Renaud *et al*. (2003), the results show that there is a nonlinear relationship between inputs and the output. Hence, it is correct to use a nonlinear model as WNN for forecasting the data. The next step of the fourth procedure is to determine an appropriate number of neurons in the hidden layer. This step is started from one neuron until the additional neuron show does not have significantly contribution.

The results of the selection process of the number of neurons which is appropriate with WNN model using lag inputs proposed by Renaud *et al*. (2003) can be seen in Table 1 for the Daubechies(4) wavelet family or D(4)

and Table 2 for Haar wavelet family. Moreover, the results of forecast accuracy comparison between WNN and MAR could be seen in Table 3.

Based on the results in Table 1 and 2, the first proposed procedure shows that the best WNN model for forecasting the number of tourist arrivals to Bali consists one neuron in the hidden layer for both D(4) and Haar wavelet. In this architecture, the inputs are the lags of scale and wavelet coefficients of MAR(1) and multiplicative seasonal lags which are statistically significant from stepwise methods.
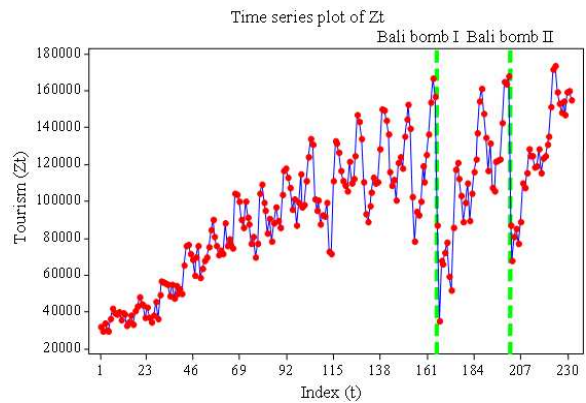

Fig. 3: Plot of the number of tourist arrivals to Bali

Table 1: The result of the first proposed procedure for determining an appropriate number of neurons, using D(4) wavelet

No. of neurons	RMSE of in-sample	RMSE of out-sample	R^2	$R^2_{increment}$	F	p-value
1	0.143446158	0.097281382	0.1500885	-	-	-
2	0.141257894	0.097518635	0.1721394	0.02205	0.805739	0.678619
3	0.141230051	0.097657623	0.1723696	0.00023	0.008134	1
4	0.141225858	0.097705661	0.1724948	0.00013	0.004276	1
5	0.141233214	0.097634864	0.1723066	-0.00019	-0.0062	1
6	0.141250911	0.097472891	0.1721053	-0.00020	-0.00638	1
7	0.141254999	0.097469538	0.1719797	-0.00013	-0.00383	1
8	0.141250312	0.097464803	0.1721176	0.00014	0.004039	1
9	0.141213524	0.097854646	0.1725838	0.00047	0.0131	1
10	0.141263965	0.097378698	0.1719706	-0.00061	-0.01648	1

Table 2: The result of first proposed procedure for determining an appropriate number of neurons, using Haar wavelet

No. of neurons	RMSE of in-sample	RMSE of out-sample	R^2	$R^2_{increment}$	F	p-value
1	0.137605940	0.097762617	0.1649002	-	-	-
2	0.134166526	0.099448277	0.1953890	0.030489	1.146254	0.313262
3	0.132504715	0.094633641	0.2096628	0.014274	0.528266	0.967427
4	0.132422280	0.094433753	0.2106260	0.000963	0.03447	1
5	0.132384239	0.094300145	0.2109797	0.000354	0.012214	1
6	0.132364476	0.094209925	0.2111847	0.000205	0.006823	1
7	0.132308052	0.094155599	0.2116665	0.000482	0.015431	1
8	0.132333238	0.094142760	0.2114571	-0.00021	-0.00644	1
9	0.132367079	0.094338621	0.2111465	-0.00031	-0.00915	1
10	0.132296902	0.094043982	0.2117720	0.000625	0.017656	1

Table 3: The result of forecast accuracy comparison for testing data

Method	Procedure	RMSE of in-sample	RMSE of out-sample	Explanation about the best model
WNN	4 - Haar wavelet	0.1376	0.0978	MAR(1)-Haar, 1 neuron
	4 - Daubechies wavelet	0.1434	0.0973	MAR(1)-D(4), 1 neuron
MAR	MAR	0.1185	0,1141	MAR(J = 4;[12,36],[12,36],[36],0,0)-Haar

If the selection of WNN model is done based on cross-validation principle, then the best model is the model that yields the minimum value of RMSE at testing dataset, i.e., the WNN model that consists of one neuron in the hidden layer both for D(4) and Haar wavelets with RMSE 0.0973 and 0.0978 respectively. Hence, WNN model with one neuron in the hidden layer that uses D(4) wavelet is the best model.

In addition, the result of forecast accuracy comparison between WNN and MAR models at Table 3 shows that WNN model with one hidden neuron that uses D(4) wavelet family yields the most accurate forecast than other models.

CONCLUSION

Based on the results at the previous sections, it can be concluded that there is a difference pattern between scale and wavelet coefficients of MODWT circular decomposition. For non-stationary seasonal time series data, the scale coefficients have non-stationary and seasonal pattern, whereas the wavelet coefficients in each decomposition level tend to have a stationary pattern and the values are around zero. Then, new procedures for building NN-MAR based on these properties of scale and wavelet coefficients are proposed. The empirical results by using data of the number of tourist arrivals to Bali show that the proposed procedure for building a WNN model works well for determining appropriate model architecture. Moreover, the forecast accuracy comparison shows that the proposed procedure using stepwise in the beginning step for determining the lag inputs yields more parsimony model and more accurate forecast than other procedures.

REFERENCES

Bashir, Z. and M.E. El-Hawary, 2000. Short term load forecasting by using wavelet neural networks. Proceeding of the Canadian Conference on Electrical and Computer Engineering, Mar. 7-10, IEEE Xplore Press, Halifax, NS., Canada, pp: 163-166. DOI: 10.1109/CCECE.2000.849691

Bjorn, V., 1995. Multiresolution methods for financial time series prediction. Proceeding of the IEEE/IAFE 1995 Conference Computational Intelligence for Financial Engineering, Apr. 9-11, IEEE Xplore Press, New York, USA., pp: 97-97. DOI: 10.1109/CIFER.1995.495258

Chen, Y., B. Yang and J. Dong, 2006. Time-series prediction using a local wavelet neural network. Neurocomputing, 69: 449-465. DOI: 10.1016/j.neucom.2005.02.006

Daubechies, I., 1992. Ten Lectures on Wavelets. 1st Edn., SIAM: Society for Industrial and Applied Mathematics, USA., ISBN: 0898712742, pp: 377.

Debnath, L., 2001. Wavelet Transform and their Application. 1st Edn., Birkhhauser Boston, Boston, ISBN: 0817642048, pp: 565.

Gao, H.Y., 1997. Choice of thresholds for wavelet shrinkage estimate of the spectrum. J. Time Ser. Anal., 18: 231-251. DOI: 10.1111/1467-9892.00048

Morettin, P.A., 1997. Wavelets in statistics. Resenhas, 3: 211-272.

Murtagh, F., J.L. Starckand and O. Renaud, 2004. On neuro-wavelet modeling. Dec. Support Syst., 37: 475-484. DOI: 10.1016/S0167-9236(03)00092-7

Nason, G.P. and R. von Sachs, 1999. Wavelets in time series analysis. Phil. Trans. R. Soc. Lond. A., 357: 2511-2526. DOI: 10.1098/rsta.1999.0445

Percival, D.B. and A.T. Walden, 2000. Wavelets Methods for Time Series Analysis. 1st Edn., Cambridge University Press, Cambridge, ISBN: 0521640687, pp: 620.

Priestley, M.B., 1996. Wavelets and time-dependent spectral analysis. J. Time Ser. Anal., 17: 85-104. DOI: 10.1111/j.1467-9892.1996.tb00266.x

Renaud, O., J.L. Stark and F. Murtagh, 2003. Prediction based on a multiscale decomposition. Int. J. Wavelets Multiresolut. Inform. Process., 1: 217-232.

Soltani, S., D. Boichu, P. Simard and S. Canu, 2000. The long-term memory prediction by multiscale decomposition. Sign. Process., 80: 2195-2205. DOI: 10.1016/S0165-1684(00)00077-3

Taylor, J.W., L.M. Menezes and P.E. McSharry, 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. Int. J. Forecast., 22: 1-16. DOI: 10.1016/j.ijforecast.2005.06.006