System for Prediction of Non Stationary Time Series based on the Wavelet Radial Bases Function Neural Network Model

Heni Kusdarwati, Samingun Handoyo

Departement of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Indonesia

Article Info	ABSTRACT
	This paper proposes and examines the performance of a hybrid model called the wavelet radial bases function neural networks (WRBFNN). The model will be compared its performance with the wavelet feed forward neural networks (WFFN model by developing a prediction or forecasting system that considers two types of input formats: input9 and input17, and also considers 4 types of non-stationary time series data. The MODWT transform is used to generate wavelet and smooth coefficients, in which several elements of both coefficients are chosen in a particular way to serve as inputs to the NN model in both RBFNN and FFNN models. The performance of both WRBFNN and WFFNN models is evaluated by using MAPE and MSE value indicators, while the computation process of the two models is compared using two indicators, many epoch, and length of training. In stationary benchmark data, all models have a performance with very high accuracy. The WRBFNN9 model is the most superior model in nonstationary data containing linear trend elements, while the WFFNN17 model performs best on non-stationary data with the non-linear trend and seasonal elements. In terms of speed in computing, the WRBFNN model is superior with a much smaller number of epochs and much shorter training time.

Corresponding Author:

Heni Kusdarwati,
Departement of Statistics,
Faculty of mathematics and Natural Sciences,
Universitas Brawijaya,
Gdg Graha Sainta lt.2, FMIPA UB, Jln. Veteran Malang 65145, Indonesia.
Email: kusdarwati_h l@yahoo.com

1. INTRODUCTION

In the real world, there are many observations collected at certain time intervals such as year, month, week, day, hour, even up to the smallest interval unit. The set of observations is referred to as time series data. The most popular method of time series modeling is the ARMA model. In the ARMA model identification process, time series data must be in a stationary condition. The stationary data is an assumption that must be satisfied in classical time series modeling [1]. Prior to model identification, if the time series data modeled is non-stationary, the data must be Box-Cox transformed so that the data has a constant variance [2]. The selection of the suitable transformation is a complex problem and is usually done by trial and error [3].

One of the important steps in ARMA modeling is parameter estimation to get the best model. The parameter estimation method of the ARMA model typically uses the maximum likelihood (MLE) method [1], but some researchers today propose the estimation of ARMA model parameters using semiparametric and nonparametric [4], [5], or using a combined method of MLE and artificial intelligence [6]. When the best model has been obtained and then the model is used for prediction or forecasting purposes, sometimes the model must be transformed back to produce a prediction value [7], [8]. Thus, the forecasting with the classical time series model for non-stationary data is not a simple task.

Wavelet theory is a very potential theory to be used in solving various problems such as signal processing, medicine, data compression, geophysics, astronomy and nonparametric statistics [9], [10]. For example, the application of wavelet transforms to tomato-fruit recognition by Sabrol and Kumar [11], while Kumar, *et al*. [12] applies hybrid method between wavelet and LSB to the digital watermarking approach. Applied wavelet transformation methods in the field of Statistics are the most commonly used for prediction or forecasting time series data as performed by Soltani [13] and Renaud [14].

Neural networks (NN) model is another example of a nonparametric model that has a flexible functional form, yet contains several parameters that can not be interpreted as in the parametric model [15]. The application of the NN model for time series predictions containing seasonal elements and trending elements is done by Zhang and Qi [16]. Multi-layer perceptron (MLP) architecture is widely used for non-linear and non-stationary time series data prediction, while the commonly used learning method is feed-forward NN (FFNN) as did by Kajitani *et al*. [17]. The radial bases function NN (RBFNN) architecture resembles MLP but it applies the clustering method on the hidden layer unit. The RBFNN can also be used to forecast non-stationary time series with shorter training processes [18].

Several studies with wavelet and NN combinations were initiated by the research community of wavelet and NN. One of the major problems in NN modeling in time series data is the need for selecting a proper initial data processing. The combination of wavelets, as an initial processing method and NN as a method that processes inputs into an output, produces a hybrid model known as Wavelet Neural networks (WNN) [19]-[25]. The application of the WNN model for time series forecasting is one of the most interesting research topics in the fields of mathematics, statistics, and computer science. In general, WNN is neural networks with wavelet functions used in processing in transfer functions. In the case of time series forecasting, the inputs used in WNN are wavelet coefficients at a given resolution. To date, some articles have been discussed in detail with regard to WNN modeling for non-stationary time series forecasting, some of which are Chen *et al*. [19], Subanar and Suhartono [20], and El-Sousy [21]. The articles use the FFNN training algorithm so that the resulting model is specifically called WFFNN.

In another hand, some researchers who have implemented the hybrid method between wavelet and NN, or hybrid among machine learning methods for time series forecasting ie Bunnoon [22] has forecasted the electricity peak load demand, Poorani and Murugan [23] have forecasted the rising demand for electric vehicles applicable to Indian road conditions, Kamley, *et al*. [24] have measured the performance forecasting of the share market, and the enabling external factors for inflation rate forecasting were conducted by Sari, *et al*. [25]. In the previous hybrid methods that were not a hybrid between wavelet and RBFNN. Both in Burnoon [22], and in Poorani&Murugan [23] combined between wavelet and FFNN, meanwhile both in Kamley, *et al*. [24] and in Sari, *et al*. [25] combined between NN, and fuzzy inferences system. Furthermore, modeling the hybrid between wavelet and RBFNN is focus on this research.

Based on the above description that time series data in the real world is generally non-linear and nonstationary, currently, there is not the hybrid model combined between wavelet and RBFNN for nonstationary time series forecasting, so this study proposes and investigates the performance of a hybrid model called wavelet radial bases function NN (WRBFNN). The model will be compared its performance with the WFFNN model by developed a forecasting system that considers two types of input formats: input9 and input17 in order to investigate the effect of the number of inputs on the model performance, and also 4 types of non-stationary datasets with difference pattern and characteristic that popularly discussed in the nonlinear time series literature as case studies.

2. MAXIMAL OVERLAP DISCRET WAVELET TRANSFROM (MODWT)

Suppose there is a time series data x, size N, then the MODWT transform will produce a column vector w_1, w_2, ..., w_{Jo} and v_{Jo} each of them is N. The vector contains the MODWT wavelet coefficient, while w_{Jo} contains the scale coefficient. The MODWT wavelet filter $\{\tilde{h}_l\}$ is obtained through $\tilde{h}_l = h_l/\sqrt{2}$ and the MODWT scale $\{\tilde{g}_l\}$ obtained through $\tilde{g}_l = g_l/\sqrt{2}$. Thus the condition of a MODWT wavelet filter must satisfy the following equation [9]:

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

Similarly, the scale filter must satisfy the following equation:

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

Where $m = 1,2,\ldots,(L/2) - 1$

The main objective in the MODWT formulation is to define DWT-like transformations, but do not experience difficulties from DWT sensitivity in terms of selecting starting points for a time series. This sensitivity is about the downsampling of the wavelet filter output and the scale filters at each stage of the pyramid algorithm. By defining A which is the matrix containing the filter $\tilde{g}$ and B is the matrix containing the filters [20]. Pyramid algorithm is an efficient calculation algorithm to calculate the scale coefficient and MODWT wavelet coefficients at j-level. Consecutive smoothing coefficients and detailed coefficients in different levels were obtained using pyramid algorithms [10]. Figure 1 illustrates if a data x is decomposed with a wavelet filter and a scale filter will produce wavelet coefficients and scale coefficients. On the first level, second and so on.
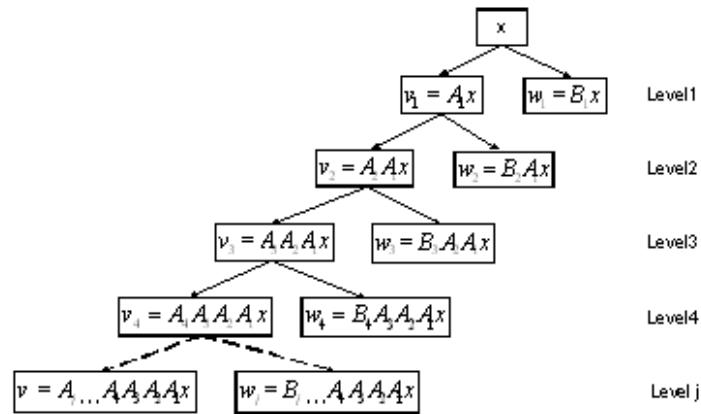

Figure 1. Pyramid algorithm for MODWT

The transformation of v_{j-1} form w_j and v_j use of matrices A_j and B_j that are size NxN is $w_j = B_j v_{j-1}$ and $v_j = A_j v_{j-1}$. Thus, the reconstruction of x at each level are as follows:

Level 1: $x = B_1^T w_1$ and $x = A_1^T v_1$

Level 2: $x = B_2^T A_1^T w_2$ and $x = A_2^T A_1^T v_2$

Level 3: $x = B_3^T A_2^T A_1^T w_3$ and $x = A_3^T A_2^T A_1^T v_3$

Level j: : $x = B_j^T A_{j-1}^T .... A_2^T A_1^T w_j$ and $x = A_j^T ... A_3^T A_2^T A_1^T v_j$. using the information of the reconstruction x on each level above and given $v_o \cong x$ then, it be obtained:

$$x = B_1^T w_1 + A_1^T B_2^T w_2 + \cdots + A_1^T \cdots A_{J_0-1}^T B_{J_0}^T w_{J_0} + A_1^T \cdots A_{J_0-1}^T A_{J_0}^T v_{J_0}$$

(3)

3.   TIME SERIES PREDICTION USING WAVELET NEURAL NETWORKS

Suppose a stationary signal $X = (x_1, x_2, ..., x_t)$ and it is assumed to be forecast the value x_{t+1}. The basic idea of the wavelet neural network model is to use the coefficients obtained from decomposition such as MODWT to obtain a forecast value with a particular neural network architecture. Kajitani, *et al.* [17] introduced the Multi-Layer Perceptron (MLP) neural network or known as feed-forward neural network (FFNN) to process the wavelet coefficients. The FFNN architecture used that it consists of a hidden layer with P neuron, which is mathematically written as follow:

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

(4)

where *g* is an activation function on the hidden layer, which is usually sigmoid logistics, while the activation function at the output layer is linear function.

Renaud, *et al.* [14] introduce an input processing of a wavelet transform model such as MODWT. The time series forecasting procedure in the t + 1 period with wavelet transform at level J=4, the order Aj=2 and N=17 are illustrated in Figure 3. Based on Figure 3, it is obtained that the value in the 18th period is

predicted using the input processing result MODWT by selecting some scale and smooth coefficients. In the wavelet coefficient of level 1 chosen as input at t=17 and t=15, wavelet coefficient level 2 at t=17 and t=13, wavelet coefficient level 3 at t=17 and t=9, wavelet coefficient level 4 on t=17 and t=1, and smooth level 4 coefficients at t=17 and t=1. So it can be formulated that the second input at each level is in the period of $t - 2^j$.
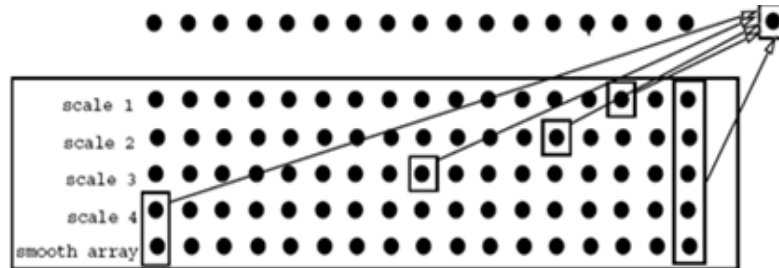

Figure 2. Selection of neural network inputs from wavelet transforms to J=4 and Aj=2 [14]

Renaud *et al*. [14] developed a linear wavelet model known as the MultiscaleAutoregression Model (MAR). In addition, there is also the possibility of using non-linear models in the input-output process of the wavelet model, particularly the Feed-forward Neural Network (FFNN) approach. The second model is then known as Wavelet Neural Network or WNN model. Both approaches above are models that use input lag-lag of wavelet coefficient, that is scale and smooth coefficient as in Figure 2.

The basic idea of multiscale decomposition is the trend of affecting low-frequency components (L), which tend to be deterministic. While the high-frequency component (H) remains stochastic. The second point that must also be understood in wavelet modeling for forecasting is to know the function used to process the input, ie the wavelet coefficients which become output in the form of the forecast value in the period t + 1. In general, there are two kinds of functions that can be used in this input-output process, namely linear functions and non-linear functions [20].

To facilitate an understanding of the WNN model in Equation (4), consider the general architecture of the MLP that has a hidden layer with four neurons, three inputs, and a linear activation function on the output layer, as shown in Figure 2. The network output or $Y(x)$ in this figure is analogous to the predicted value for the period to N + 1, or $\hat{X}_{N+1}$ in equation (1) above. The inputs X_1, X_2, and X_3 correspond to the wavelet coefficients and the smooth coefficients are $w_{J,t-2^J(k-1)}$ and $v_{J,t-2^J(k-1)}$. The weights between input nodes and hidden nodes are $a_{j,k,p}$ whereas the weights between hidden nodes and output nodes are b_p. To obtain optimal weights then the network must be trained by using a particular learning algorithm.
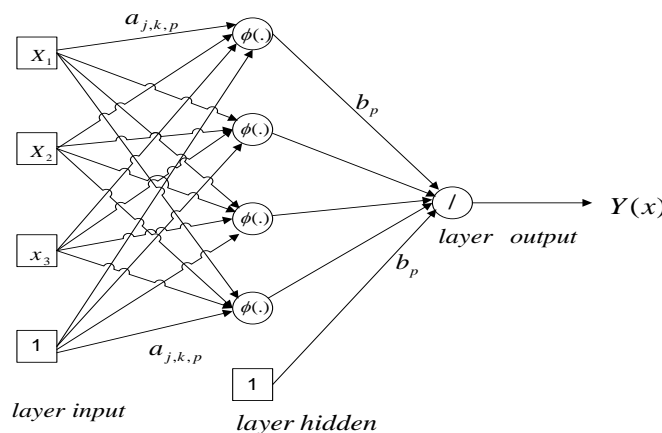

Figure 3. MLP architecture with 3 input nodes, 1 hidden layer with 4 neurons [20]

On the RBFNN the activation function in the hidden layer is Gaussian function, the activation function at the output layer is linear function, and the weight between the input node and the hidden node is 1 or $a_{j,k,p} = 1$[18]. Thus the weight adjustment only occurs on the weights between the hidden node and the output node ie b_p. Based on these properties finally obtained the equation:

$$\widehat{X}_{N+1} = \sum_{p=1}^{P} \widehat{b}_p \, g\left(\sum_{j=1}^{J} \sum_{k=1}^{A_j} w_{j,N-2^j(k-1)} + \sum_{k=1}^{A_J} v_{J,N-2^J(k-1)} \right) \tag{5}$$

which g is a Gaussian function with parameter center (μ) and variance (σ^2). Furthermore, the Model in Equation (5) is called the Wavelet Radial Basis Function Neural Network (WRBFNN). In the WRBFNN model, we need a method to estimate the parameters of Gaussian function distribution. Usually the both parameters of the Gaussian distribution in a given set of data are estimated by the least squares method.

The performance of the system should be evaluated using a measure of accuracy referring to the goodness of a prediction or forecasting system. The accuracy of a model indicates the merit or suitability of the model to predict the value in future periods. There are various measures of accuracy in forecasting, among which are Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE) expressed by the following formula [26].

$$MAPE = {}^{1}\!/_{2} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| x100\% \tag{6}$$

$$MSE = {}^{1}\!/_{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

Both measures of this accuracy, if they have the value near zero then it indicates better prediction model. To select the best prediction model, these both indicators are calculated on the data set testing (out sample).

4. RESEARCH METHODS

In this study built a forecasting system with input processing using MODWT by considering the number of lag as input is N=9 and N=17. MODWT processing results are selected as neural network input using Renauld method [14]. To process the input into the output of the system is done processing with FFNN and RBFNN. There are four types of time series data that have patterns that can be seen in Figure 4 and have characteristics that can be seen in Table 1.
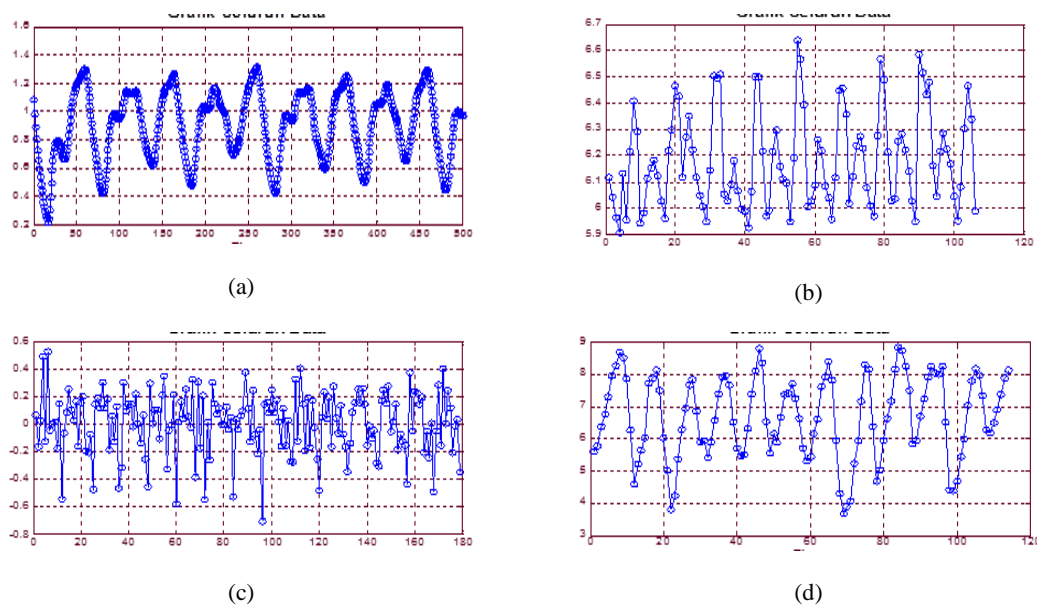

Figure 4. Four pattren of time series data: (a) Chaotic mcglass, (b) Monthly electricity usage, (c) Traffic fatalities, and (d) Canadian lynx

Table 1. Attribute of Dataset used in Thestudy

Name Dataset	Total record	Total train	Total.test	Characteristics
ChoaticMcGlass	500	350	150	Stationary as the benchmark
Electricity Usage	106	80	26	Non stationary on variance
Traffic Fatalities	180	140	40	Non stationary on both mean and variance with linier trend
Canadian Lynx	114	90	24	Non stationary on both mean and variance with non linier trend

The four data are taken from Tong [1] which is the most popular non-linear time series literature to date. The four data that have characteristics as in Table 1 by the researchers are considered capable of representing non-linear and non-stationary time series data patterns that often appear in the real world. Each data is divided into training dataset (70%) and testing dataset (30%). Training dataset is used to build models, while testing dataset is used to select the best model or model validation.

The prediction system built has two main processing menus: MODWT wavelet transform and neural network computation. The MODWT menu changes the input time series with many lags 9 and many lags 17 are transformed into scale coefficients and smooth coefficients at a number of levels=4 and autoregressive order (AR)=2. The neural network menu has two models: RBFNN and FFNN which both this neural network model will process the input of the result of the selected MODWT transformation as performed by Renaud (2003) to produce the network output. Furthermore, this network output measured its performance with MAPE and MSE. Figure 5 and Figure 6 illustrate the processes performed on predicted systems that have been built.
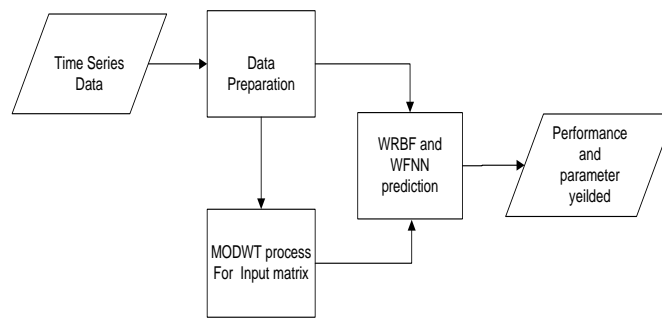

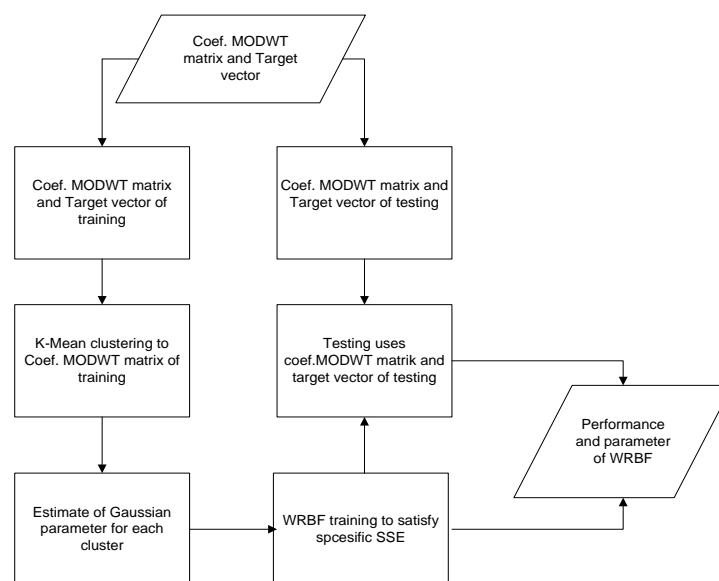Figure 5. The steps of system development

Figure 6. The steps of computation process of WRBFNN model

5. RESULTS AND ANALYSIS

5.1. Pre processing data input and settings neural network parameters

The time series data has a data structure in the form of a row vector in which the sequence number of the observed value t shows the position of the value of the record in the period of time t. In this study consider inputs with many lags 9 (input9) and many lags 17 (input17). In input9 implies that the first 9 observations are used to predict the 10th observational value, the second observation to the 10th observation is used to predict the 11th, and so on. In input17 also implies that the first 17 observations are used to predict the 18th observed value. Therefore, the first step in preparing the data is to transform the vector data structure into a matrices data structure called the pairs of input-output matrix. The matrix at input9 has dimensions (n-9)x10 and the input matrix17 has dimensions (n-17)x18 where n is the number of periods of the time series. In the input-output matrix, the last column is the target vector whereas the previous columns are the input of the system.

The MODWT processing is performed on the system input matrix (all columns other than the last column of the input-output matrix). Each row of the input matrix is transformed MODWT to produce a scale coefficient and a smooth coefficient. In input9, each row with 9 observed values are transformed into 3 rows of scale coefficients (w1, w2, w3) and one row of the smooth coefficient (s). From this transformation, we selected the 9th and 7th values of w1, the 9th and 5th values of w2, the 9th and 1st values of w3, and the 9th and 1st values of s. These values are used as inputs from neural networks. Finally, at input9 after the MODWT transform has an input number of 8 values, whereas at input17 after MODWT transform has 10 input values.

In the radial base network, the spread and SSE parameters have a vital role to gain optimal network weight. Initially running the system is done by trial and error against a certain spread value on various SSE values. It aims to get the optimum spread and SSE pair that has the smallest SSE testing value. From the various possible spreads, try to get the best performing spread that is = 0.8.

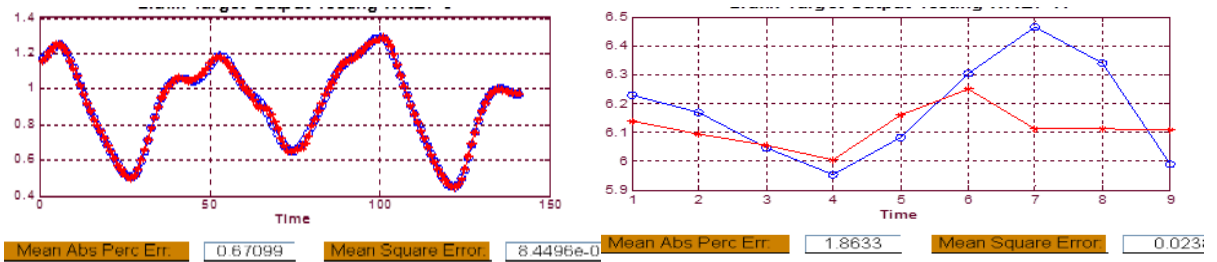Table 2. Pairs of SSE and MSE Intraining and Testingdatafor Mcglassdata with Spread = 0.8

Experiment	SSE Training	MSE Training	SSE Testing	MSE Testing
1	0,89600	0,002628	0,35200	0,002496
2	0,46000	0,001349	0,15900	0,001128
3	0,09800	0,000287	0,06900	0,000489
4	0,04300	0,000126	0,05200	0,000369
5	0,01000	2,93E-05	0,02700	0,000191
6	0,00500	1,47E-05	0,02700	0,000191
7	0,00100	2,93E-06	0,01300	9,22E-05
8	0,00050	1,47E-06	0,01200	8,51E-05
9	0,00010	2,93E-07	0,01400	9,93E-05
10	0,00005	1,47E-07	0,01500	0,000106

Table 2 expresses the pairs of SSE training and SSE testing at spread = 0.8. The value of MSE training or MSE testing is derived from dividing SSE values against the number of training inputs or testing. In this case, MSE training = SSE training divided by 341, while MSE testing = SSE testing divided by 141. Based on Table 2, the lowest MSE testing occurred in the 8th experiment having SSE training = 0.0005. Next SSE value = 0.0005 and spread = 0.8 is used as input parameter on WRBF9 and WRBF17 systems.

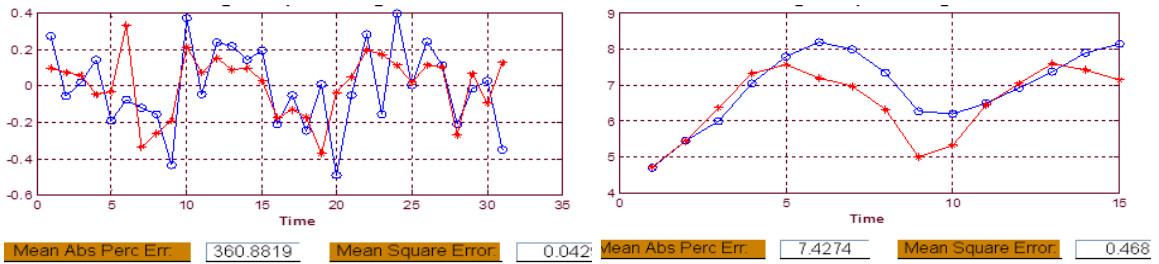### 5.2. Output of WRBFNN and WFFNN models on all four types of datasets

Once network parameters, input training, and input testing are available, then the learning process on neural networks can be run. Suppose the training process on the model WRBF9, the training process on this network is on each epoch formed a neuron. Neurons that have the smallest total errors will be accepted as new neurons, then network errors are re-checked. The iteration will be stopped when the error has reached the specified threshold value, but if the error is still far from the provisions, then the next neuron will be added until the number of neurons is equal to the amount of training input data.

Based on the optimized WRBF9 model, there are 166 hidden nodes. This means that in the hidden layer there are 166 input data into the center of the cluster of Gaussian distribution and each cluster has the same range of spreads = 0.8. The training to obtain the optimal weight is done on the model WRBF9, WFFNN9, and WFFNN17. Output testing is obtained by simulating the input testing data that has been selected from the transformation of MODWT to the optimal network formed by the training process that occurs on each dataset. The system automatically calculates the MAPE and SSE values used to assess model performance.
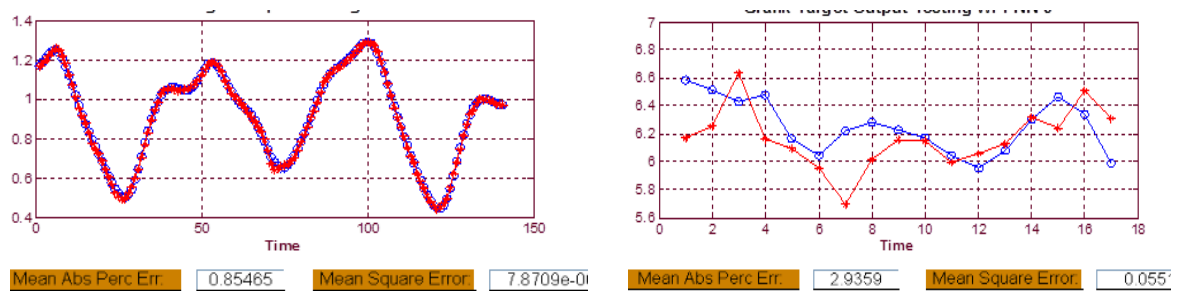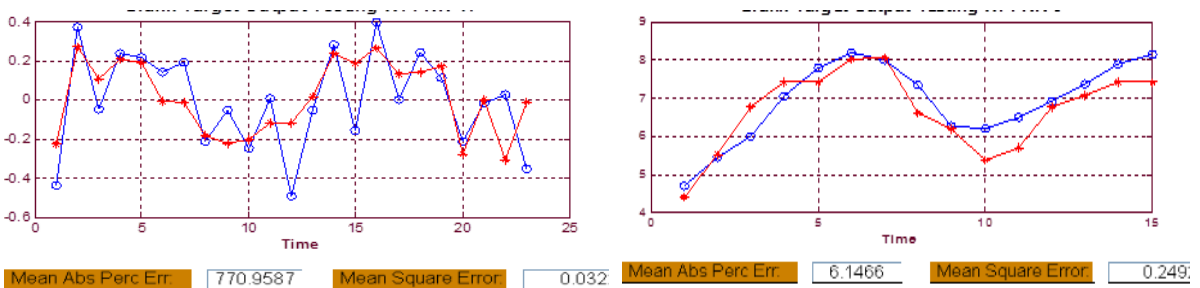
| Mean Abs Perc Err: | 0.67099 | Mean Square Error: | 8.4496e-0 |

(a) WRBF9

| Mean Abs Perc Err: | 1.8633 | Mean Square Error: | 0.023 |

(b) WRBF17

| Mean Abs Perc Err: | 360.8819 | Mean Square Error: | 0.042 |

(c) WRBF9

| Mean Abs Perc Err: | 7.4274 | Mean Square Error: | 0.468 |

(d) WRBF9

Figure 7. Plot target versus output of the system bases on WRBF model (proposed method)

| Mean Abs Perc Err: | 0.85465 | Mean Square Error: | 7.8709e-0 |

(a) WFFNN9

| Mean Abs Perc Err: | 2.9359 | Mean Square Error: | 0.055 |

(b) WFFNN9

| Mean Abs Perc Err: | 770.9587 | Mean Square Error: | 0.032 |

(c) WFFNN17

| Mean Abs Perc Err: | 6.1466 | Mean Square Error: | 0.249 |

(d) WFFNN17

Figure 8. Plot target versus output of system bases on WFNN model

After all optimal models have been obtained both for both input types (input9 and input17) and on all four datasets. To know the goodness of each model in predicting the data testing made a graph between the actual value versus predicted results. Better models between the two input types are exposed in Figure 7 for the WRBF model and Figure 8 for the WFFNN model. In both Figures, it can be seen that both WRBF

and WFFNN models can predict almost perfect testing data ie Figure 7(a) and Figure 8(a). Characteristics of the data in Figure 7(a) (McGlass data) are stationary data in mean and variance.

Based on Figure 7 it can be seen that the input9 data type is better than the input17 data type which in the input17 data produces a better model only on the monthly average electricity usage data. As in Figure 8 exposed that the more complicated the data pattern that is not stationary data on the mean and variance (traffic fatalities and Canadian Lynx data), the WFFNN17 model is a better than WFFNN9 model. But for stationary data in mean and variance or data that is just nonstationary variance, the WFFNN9 model is a better model.

5.3. Performance comparison of WRBF and WFFNN methods on four types of datasets

In this section will be discussed the performance of all systems built namely WRBF9, WRBF17, WFFNN9 and WFFNN17 against all data sets used. Some important indicators used as a basis for comparing it are MSE testing, MAPE testing, Epoch count, and length of the training process. MSE is a standard measure of the accuracy of a forecasting method, while the number of the epoch is proportional to the time required during the learning process. Thus the number of an epoch can be expressed as the effectiveness of a forecasting method.

Table 3 states the MAPE testing and the MSE testing value of each model in all four types of data sets.

Table 3. MAPE Testing and MSE Testing Values of Four Models

MODEL	MAPE				MSE			
	McGlass	Electricity	Traffic	Lynx	McGlass	Electricity	Traffic	Lynx
WRBF9	0.671	2.366	360.882	7.427	0.0000084	0.032503	0.042911	0.468970
WRBF17	0.826	1.863	208.034	15.424	0.0001068	0.023873	0.046859	1.263800
WFFNN9	0.855	2.936	869.875	6.147	0.0000787	0.055146	0.068727	0.249220
WFFNN17	1.275	3.596	770.958	12.512	0.0002993	0.066937	0.032220	0.956370

Based on Table 3, it is exposed that for McGlass data the WRBF9 method performs best, and the WRBF method is generally superior to WFFNN. The selection of input numbers also greatly influences the performance of a method, the input 9, in this case, performs better. Thus for stationer data, the WRBF method performs better, but basically, both WRBF and WFFNN methods can be used to predict stationer data with high accuracy.

In Electricity data, the data have not constant variance. The WRBF17 method performs best and in general WRBF method is superior to WFFNN method. In the case of this type of data, input format17 has better performance, although exposed to differences that are not too large. This condition due to the lack of a lot of observation that is 17 observations on input9, and only 9 observations on input17, which quantitatively input17 has the amount of testing data about 50% of input9. Researchers believe the difference in MSE testing will be more evident if the proportion of data testing for both input formats is almost balanced.

In Traffic fatalities data that is not stationary which variance not constant and contains trend elements, WFFNN17 method has the best performance where there is a big difference of MSE testing value between WFFNN17 and WFFNN9 but MSE testing on WRBF9 and WRBF17 is the relatively small difference. These results indicate that in this type of data large numbers of input will contribute significantly to the improved performance of the WFFNN method, but it is not for the WRBF method.

In Canadian Lynx data that is non-stationary and contains a non-linear trend, WFFNN9 method performs best. However, this condition is not necessarily applicable when the proportion of data testing between input 9 and input 17 is almost balanced. The obvious thing is that the WFFNN method is superior to WRBF.

Table 4 shows the number of epochs and the length of time in the training process of each method on four types of data sets.

Table 4. Number of Epoch and Length of Time (Seconds) in the Training Process

MODEL	Numbers of Epoch				Length of Training Time (seconds)			
	McGlass	Electricity	Traffic	Lynx	McGlass	Electricity	Traffic	Lynx
WRBF9	166	41	25	33	7.31	1.11	1.03	0.95
WRBF17	155	31	27	44	7.06	0.64	0.74	0.83
WFFNN9	931	420	251	804	108.34	5.51	3.85	10.52
WFFNN17	1238	100	69	152	101.60	1.71	1.62	1.85

Based on Table 4 overall WRBF method has much smaller epoch number than WFFNN method. This means that the WRBF method has a much faster computation process than the WFFNN method. In the WFFNN method the selection of input numbers also greatly influences the length of the training process in which input17 tends to have many smaller epochs or shorter training time durations

6. CONCLUSION

Based on the results and discussion that have been done in the previous section, it can be concluded that WRBF and WFFNN Method can be used for prediction of McGlass chaotic time series which is non-linear but has mean and variance constant with high accuracy ie MSE value less than 0.0005. However, WRBF method is superior to WFFNN method. The WRBF9 method has the best performance to predict this data with MSE testing = 0.000084. WRBF method will be superior to WFFNN method when applied to stationary data type or non-stationary data type with a simple pattern. WFFNN method will be superior to WRBF method when applied to non-stationary data with a complex pattern, eg stationary data in mean and non-constant variance, or non-stationary data and nonlinear in trend element. Selection of the number of input elements is very influential on the performance of the model, especially in the little data testing will lead to the value of sensitive MSE testing. In the WFFNN method, the selection of input numbers should receive more careful attention. In general, the WRBF method has a much smaller epoch number than the WFFNN method, so the time required for the computation process is much shorter. In future research, it is necessary to experiment on the data set with a large number of observations. In addition, it is also necessary to try various transformations to stationary data on various characteristics of time series data nonstationary.

REFERENCES

[1] H. Tong, "Non-linear Time Series: A Dynamical Systems Approach", Oxford University Press. Oxford. 1990.
[2] P.J. Brockwell, and R.A. Davis, "Nonstationary and seasonal time series models", *Introduction to Time Series and Forecasting*, Springer International Publishing, 2016, pp. 157-193.
[3] N.E. Huang, *et al*., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, The Royal Society, vol. 454, no. 1971, 1998.
[4] C. Velasco, "Gaussian Semiparametric Estimation of Non-stationary Time Series", *Journal of Time Series Analysis*, vol. 20, no. 1, 1999, pp. 87-127.
[5] J. Fan and Q. Yao, "Nonlinear Time Series: nonparametric and parametric methods", Springer-Verlag, New York, 2003.
[6] S. Handoyo, *et al*., "Implementation Of Particle Swarm Optimization (Pso) Algorithm For Estimating Parameter Of Arma Model Via Maximum Likelihood Method", *Far East Journal of Mathematical Sciences*, vol. 102, no. 7, 2017, pp. 1337-1363.
[7] C.K. Peng, *et al*., "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series", *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 5, no. 1, 1995, pp. 82-87.
[8] S. Poorani, and R. Murugan, "A Non-Linear Controller for Forecasting the Rising Demand for Electric Vehicles Applicable to Indian Road Conditions", *International Journal of Electrical and Computer Engineering,* vol. 6, no. 5, 2016, pp. 2274-2281.
[9] A. Bruce and H.Y. Gao, "Applied Wavelet Analysis with S-PLUS", Springer-Verlag, New York, USA, 1996.
[10] L. Debnath, "Wavelet Transform and Their Application", Birkhhauser, Boston, 2002.
[11] H. Sabrol, and S. Kumar, "Recognition of Tomato Late Blight by using DWT and Component Analysis", *International Journal of Electrical and Computer Engineering (IJECE),* vol. 7, no. 1, 2017, pp. 194-199.
[12] V.A. Kumar, C. Dharmaraj, and C.S. Rao, "A Hybrid Digital Watermarking Approach Using Wavelets and LSB", *International Journal of Electrical and Computer Engineering,* Vol. 7, No. 5, 2017, pp. 2483-2495.
[13] S. Soltani, "On the use of the wavelet decomposition for time series prediction", *Neurocomputing*, Vol. 48, No.1, 2002, pp. 267-277.
[14] O. Renaud, J.L. Stark, and F. Murtagh, *"Prediction Based on Multiscale Decomposition"*, *Int. Journal of Wavelets, Multiresolution and Information Processing,* vol. 1, no. 2, 2003, pp. 217-232.
[15] L. Fausett, *"Foundamentals of Neural Networks: Architecture, Algorithm, and Application"*, Florida Institute of Technology, Prentice Hall Inc. Englewoods Cliffs, New Jersey, 1994.

[16] G.P. Zhang, and M. Qi, "Neural network forecasting for seasonal and trend time series", *European journal of operational research*, vol. 160, no. 2, 2005, pp. 501-514.

[17] Kajitani, K.W.H. Yoshio, and A.I. McLeod, "Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian lynx data", *Journal of Forecasting*, vol. 24, no. 2, 2005, pp. 105-117.

[18] E.S. Chng, S. Chen, and B. Mulgrew, "Gradient radial basis function networks for nonlinear and nonstationary time series prediction", *IEEE transactions on neural networks*, vol. 7, no. 1, 1996, pp. 190-194.

[19] Y. Chen, B. Yang, and J. Dong, "Time-series prediction using a local linear wavelet neural network", *Neurocomputing,* vol. 69, no. 4, 2006, pp. 449-465.

[20] Subanar, and Suhartono, *"New Procedures for Model Building in Wavelet Neural Network for Forecasting Non Stationary Time Series"*, *Presented at 5th Asian Mathematical Conference*, Malaysia, 2009.

[21] F.F.M. El-Sousy, "Robust wavelet-neural-network sliding-mode control system for permanent magnet synchronous motor drive", *IET Electric Power Applications*, vol. 5, no.1, 2011, pp. 113-132.

[22] P. Bunnoon, "Electricity Peak Load Demand using De-noising Wavelet Transform integrated with Neural Network Methods", *International Journal of Electrical and Computer Engineering,* vol. 6, no. 1, 2016, pp. 12-20.

[23] S. Poorani, and R. Murugan, "A Non-Linear Controller for Forecasting the Rising Demand for Electric Vehicles Applicable to Indian Road Conditions", *International Journal of Electrical and Computer Engineering,* vol. 6, no. 5, 2016, pp. 2274-2281.

[24] S. Kamley, S. Jaloree, and R.S. Thakur, "Performance Forecasting of Share Market using Machine Learning Techniques: A Review", *International Journal of Electrical and Computer Engineering*, vol. 6, no. 6, 2016, pp. 3196-3204.

[25] N.R. Sari, *et al*., "Enabling External Factors for Inflation Rate Forecasting Using Fuzzy Neural System", *International Journal of Electrical and Computer Engineering*, vol. 7, no. 5, Oct. 2017, pp. 2746-2756.

[26] S. Handoyo, and A.S. Prasojo, "Applied Fuzzy System with R Software", UB press, Malang, Indonesia, 2017.

BIOGRAPHIES OF AUTHORS

Heni Kusdarwati is a lecturer and a researcher at Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang. She is a member of the Indonesian Mathematical Society (IndoMS). She received the B.S. degree in Computation Statistics from Institut Pertanian Bogor, Indonesia in 1985 and the M.S degree in Climatology from the same University in 1992. Her research interests include Time series Analysis, Forecasting tehnique, Operation Research, and Statistical Theory.

Samingun Handoyo is a lecturer and a researcher at Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang. He is the member both of the Indonesian Mathematical Society (IndoMS) and the International Association of Engineers (IAENG). He received the B.S. degree in Mathematics from Universitas Brawijaya, Malang, Indonesia in 1997 and the M.Cs degree in Computer Science from Universitas Gadjah Mada, Yogyakarta, Indonesia in 2010. His research interests include statistical computing, Neural Network, Fuzzy System, and Partial Least Square path modeling using R Software.