Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n1p198

The development of an optimization procedure in WRBNN for time series forecasting

By Rukun et al.

Published: 26 April 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 01, April 2016, 198-212
DOI: 10.1285/i20705948v9n1p198

The development of an optimization procedure in WRBNN for time series forecasting

Rukun Santoso*[a, b], Subanar[b], Dedi Rosadi[b], and Suhartono[c]

[a]*Statistics Department of Diponegoro University, Semarang, Indonesia*
[b]*Mathematics Department of Gadjah Mada University, Yogyakarta, Indonesia*
[c]*Statistics Department of Sepuluh Nopember Institute of Technology, Surabaya, Indonesia*

Published: 26 April 2016

A forecasting procedure based on a wavelet radial basis neural network is proposed in this paper. The MODWT result becomes an input of the model. The smooth part constructs the main pattern of forecasting model. Meanwhile the detail parts construct the fluctuation rhythm or disturbances. The model considers that each of the transformation level contribute to the forecasting result independently. The nonlinearity properties included in the MODWT result is controlled by radial basis functions. The LM test is used to explore the number of wavelet coefficient clusters in every transformation level. The membership of cluster is determined by the $k-$means method. The least square method (OLS or NLS) can be used to estimate the parameters of model.

keywords: DWT, MODWT, radial basis, time series, wavelet, WRBNN

1 Introduction

Economical activity, such as stock market, commodity market and currency market, can be performed as random variables. For examples, stock return rate, stock price indices, commodity production number, commodity price, commodity demand number, and currency exchange rate refer to random variables. The random variables observed on regular time periods will construct time series data. Theory and methods of time

*Corresponding author: rukunsantoso@undip.ac.id

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

series grow up continuously since Box and Jenlins' historical book, *Time Series Analysis: Forecasting and Control*. In general, the Box-Jenkins method, a linear class model, works well when the process fulfills stationery condition.

It is found that the economical data generally has nonlinear properties and the variance changes over time (heteroscedastic). In this condition, the Box-Jenkins method may provide a less satisfying solution. There are some proposed methods to address this problem of nonlinearity. Threshold Autoregressive (TAR) model is one such examples. The model assumes that the process jumps in some alternative functions which refer to the threshold domain value (Tong, 1990). Engel (1982) proposed a time series model which accommodates this heteroscedastic problem. The model is called as ARCH (autoregressive conditional heteroscedasticity) model refer to the assumption that the residual of process has a variance which is dependent on time. Therefore, the ARCH model consists of both the mean part model and the variance part model. The generalization of ARCH model is proposed by Bollerslev (1986) and called as GARCH (Generalized ARCH) model. All models mentioned in this paragraph are parametric models which have model parameter assumptions.

Non-parametric class models are sometimes preferred to parametric class model. Examples of recently non-parametric models are the neural network model (Haykin, 1999), wavelet model (Murtagh et al., 2004), fuzzy model (Popoola, 2007) and hybrids of these models. Usually, non-parametric methods are simpler in the analytical sense but involve more numerical computation. Fortunately, some computer software and hardware are available to support the computation. The first step to decide to whether to use a linear model is to identify whether there is a nonlinearity. Teräsvirta et al. (1993) proposed a test to determine nonlinearity with neural network approximation. Lee et al. (1993) use a neural network approximation with random weights to construct the test.

This paper proposes a development of procedure inspired by wavelet neuro model (Murtagh et al., 2004). The wavelet coefficients selected from each level are treated as univariate or multivariate random variables and become to inputs for radial basis nodes. The model is called wavelet radial basis neural network (WRBNN) model. The use of wavelet to build a function was proposed by Ogden (1997). The use of radial basis function with univariate inputs to approximate a nonlinear function in the neural network structure was proposed by Haykin (1999). The WRBNN model combines the of wavelet, radial basis function and neural network into one structure. $\mathbb{R}$ software (R Core Team, 2014) with the `wavelets` package (Aldrich, 2013) is used for computations needs.

2 Nonlinear Time Series Model

Box and Jenkins have been are pioneers in mathematical time series modeling. Usually, this model is called as Autoregressive Moving Average (ARMA). If the process fulfills the stationery condition, a solution can be reached using the ARMA model. Differentiation of data is common to obtain a stationary condition. The general form of the ARMA model described in Eq. (1) is included to linear model class.

$$Y_t = \mu + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{1}$$

where ϵ_t is a normal random variable with a mean of zero and a fixed variance σ^2.

In fact, there is no guarantee that the process can be modeled by linear model class, which provides motivation to discover nonlinear models. The first heteroscedastic model, the ARCH (*Autoregressive Conditional Heteroscedasticity*) model, was proposed by Engel (1982). This model assumes that the residual ϵ_t of Eq. (1) is nonlinear and dependent on the last residuals as described in Eq. (2).

$$\epsilon_t = \sigma_t v_t \tag{2}$$

where $v_t \sim N(0,1)$ and variance σ_t^2 depends on time, as described in Eq. (3).

$$\sigma_t^2 = \alpha_0 + \sum_{n=1}^{s} \alpha_n \epsilon_{t-n}^2 \tag{3}$$

Bollerslev (1986) has developed the ARCH model in Eq. (3) into GARCH (Generalized ARCH) model as described in Eq. (4).

$$\sigma_t^2 = \alpha_0 + \sum_{n=1}^{r} \beta_n \sigma_{t-n}^2 + \sum_{n=1}^{s} \alpha_n \epsilon_{t-n}^2 \tag{4}$$

The existences of heteroscedastic properties can be investigated using the LM test which was developed by Lee et al. (1993).

A nonlinear model can also consist of dependencies as described in Eq. (5).

$$Y_t = \mu + \sum_{i=1}^{p} \omega_i \Phi(Y_{t-i}) + \epsilon_t \tag{5}$$

where Φ is the nonlinear function such as logistic function, exponential function, high polynomial and radial basis function. The use of the radial basis function in nonlinear models can be found in Haykin (1999) with special notes can be found in Orr (1996; 1999).

3 Wavelet Based Approximation

Wavelet (mother wavelet) is a small wave function that can construct an orthonormal bases for $L_2(R)$ space (Daubechies, 1992). Every mother wavelet has a unique father wavelet or scaling function. Wavelet is usually symbolized by ψ and ϕ for father wavelet. Mother and father wavelets build a wavelet family through translation and dilatation functions as described in Eq. (6).

$$\begin{aligned} \psi_{j,k}(t) &= 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \\ \phi_{j,k}(t) &= 2^{-\frac{j}{2}} \phi(2^{-j}t - k) \end{aligned} \tag{6}$$

The construction of wavelet bases for $L_2(R)$ is inspired by the construction of Fourier bases for $L_2[-\pi, \pi]$ using sines and cosines functions (Ogden, 1997). The wavelet properties supporting the wavelet bases construction described in Eq. (7).

$$\begin{aligned}\int_{-\infty}^{\infty} \phi_{j,k}(t)\phi_{j,m}(t) &= \delta_{k,m} \\ \int_{-\infty}^{\infty} \phi_{j,k}(t)\psi_{l,m}(t) &= 0 \\ \int_{-\infty}^{\infty} \psi_{j,k}(t)\psi_{l,m}(t) &= \delta_{j,l}\delta_{k,m}\end{aligned} \tag{7}$$

where

$$\delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

Wavelet and scaling function combine to create a multiresolution space where every function $f \in L_2(R)$ can be described as a linear combination of dilation-translation form of wavelets. This formulation can be seen in Eq. (8).

$$f(t) = \sum_{k \in \mathbb{Z}} c_{J,k}\phi_{J,k}(t) + \sum_{j \leq J} \sum_{k \in \mathbb{Z}} d_{j,k}\psi_{j,k}(t) \tag{8}$$

Multiresolution space containing Eq. (8) can be described in Eq. (9)

$$L_2(R) \supseteq S_1 \oplus D_1 = S_2 \oplus D_2 \oplus D_1 = S_J \oplus D_J \oplus D_{J-1} \oplus \cdots \oplus D_1 \tag{9}$$

where $\oplus$ describes an orthogonal sum of two vector spaces. S_J describes the main pattern of function, meanwhile $D_j, j = 1, 2, \cdots, J$ describe the detail parts or residuals pattern of the function (Daubechies, 1992). The main pattern consists of a smooth function which usually can be approximated using a linear combination of scaling coefficients $c_{J,k}$. The disturbances of the original function are carried out in the detail pattern, and can be approximated using a linear combination of wavelet coefficients $d_{j,k}$. The coefficients $c_{J,k}$ and $d_{j,k}$ described in Eq. (8) can be computed by Eq. (10).

$$\begin{aligned}c_{j,k} &= \int_{-\infty}^{\infty} f(t)\phi_{j,k}(t)dt \\ d_{j,k} &= \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt\end{aligned} \tag{10}$$

3.1 Discrete Wavelet Transform

In any wavelet, finite even points can be chosen to fulfill certain properties called a wavelet filter. It is usually symbolized by Eq. (11) and must fulfill the properties described in Eq. (12) (Percival and Walden, 2000).

$$\mathbf{h} = [h_0, h_1, \cdots, h_{L-1}] \tag{11}$$

which fulfill the following properties

$$\begin{aligned}\sum_{i=0}^{L-1} h_i &= 0 \\ \sum_{i=0}^{L-1} h_i^2 &= 1 \\ \sum_{i=0}^{L-1} h_i h_{i+2n} &= 0, \quad n \in \mathbb{Z}\end{aligned} \tag{12}$$

The scaling filter, which is symbolized by $\mathbf{g} = [g_0, g_1, \cdots, g_{L-1}]$, can be generated from wavelet filter. The relationship between $\mathbf{h}$ and $\mathbf{g}$ can be described in Eq. (13).

$$g_l = (-1)^{l+1} h_{L-1-l} \tag{13}$$

The filter described in Eq. (11) is the first level filter. Therefore, it is symbolized by $\mathbf{h}^{(1)}$. The up-sampled form of $\mathbf{h}^{(1)}$ is defined by inserting 0 (zero) between filter values not equal to 0. Therefore, the up-sampled form of Eq. (11) can be described in Eq. (14)

$$\mathbf{h}_{up}^{(1)} = [h_0, 0, h_1, 0, \cdots, 0, h_{L-1}, 0, h_{L-1}] \tag{14}$$

The $2^n d$ level filter is constructed by Eq. (15).

$$\mathbf{h}^{(2)} = \mathbf{h}_{up}^{(1)} * \mathbf{g} \tag{15}$$

where $*$ is a convolution operator. In general, wavelet and scaling filters are constructed by Eq. (16).

$$\begin{aligned}\mathbf{h}^{(j)} &= \mathbf{h}_{up}^{(j-1)} * \mathbf{g} \\ \mathbf{g}^{(j)} &= \mathbf{g}_{up}^{(j-1)} * \mathbf{g}\end{aligned} \tag{16}$$

The collaboration of wavelet and scaling filters builds discrete wavelet transformations (DWT). Consequently, every discrete realization of a function $f \in L_2(R)$ with fixed time increment can be decomposed into smooth part (S) and detail parts (D). Let $\mathbf{Y} = \{Y_t\}_{t=1}^{N}$ describes a discrete realization of function $f \in L_2(R)$ where $N > L$ and $N = 2^J$ for certain integer J. DWT at level j can be shown in Eq. (17)

$$\underset{N\times 1}{\mathbf{D}} = \underset{N\times N}{\mathbf{H}_j} \underset{N\times 1}{\mathbf{Y}} \tag{17}$$

where $\mathbf{H}_j$ is a transformation matrix at level j and $\mathbf{D}$ is a transformation result or wavelet coefficients matrix.

The transformation matrix at level $j = 1$ can be written as $\mathbf{H}_1 = [\mathscr{H}_1, \mathscr{G}_1]^T$. The first row until $(\frac{N}{2})^{th}$ of $\mathbf{H}_1, \quad j = 1, 2, \cdots, J$ constitutes a two-step translations of $\mathbf{h}^{(1)}$ as described in Eq. (18)

$$\mathscr{H}_1 = \begin{bmatrix} h_1 & h_0 & 0 & 0 & \cdots & 0 & 0 & h_{L-1} & h_{L-2} & \cdots & h_3 & h_2 \\ h_3 & h_2 & h_1 & h_0 & 0 & \cdots & 0 & 0 & h_{L-1} & h_{L-2} & \cdots & h_4 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & & \cdots & & 0 & h_{L-1} & h_{L-2} & & \cdots & h_1 & h_0 \end{bmatrix} \tag{18}$$

The $(\frac{N}{2}+1)^{th}$ row until N^{th} row of $\mathbf{H}_1$ constitutes a two-step translations of $\mathbf{g}$ as described in Eq. (19).

$$\mathscr{G}_1 = \begin{bmatrix} g_1 & g_0 & 0 & 0 & \cdots & 0 & 0 & g_{L-1} & g_{L-2} & \cdots & g_3 & g_2 \\ g_3 & g_2 & g_1 & g_0 & 0 & \cdots & 0 & 0 & g_{L-1} & g_{L-2} & \cdots & g_4 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & & \cdots & & 0 & g_{L-1} & g_{L-2} & & \cdots & g_1 & g_0 \end{bmatrix} \tag{19}$$

Furthermore, $\mathscr{G}_1$ will be decomposed to $\mathscr{H}_2$ and $\mathscr{G}_2$ whenever $\mathbf{H}_2$ is carried out. $\mathscr{H}_2$ and $\mathscr{G}_2$ constitutes a four-step translations of wavelet filter and scaling filter in 2^{nd} level, repectively. The process can be continued to obtain $\mathbf{H}_j = [\mathscr{H}_j, \mathscr{G}_j]^T$ where $\mathscr{H}_j$ and $\mathscr{G}_j$ constitute 2^j periodical step of wavelet filter and scaling filter in j^{th} level, respectively. Furthermore, Eq. (17) can be written as Eq. (20).

$$\begin{aligned} \mathbf{D} &= \Big[\mathscr{H}_1\ \mathscr{G}_1\Big]^T \mathbf{Y} = \Big[\mathscr{H}_1\ \mathscr{H}_2\ \mathscr{G}_2\Big]^T \mathbf{Y} = \cdots = \Big[\mathscr{H}_1\ \mathscr{H}_2\ \cdots\ \mathscr{H}_J\ \mathscr{G}_J\Big]^T \mathbf{Y} \\ &= \Big[D_1\ S_1\Big]^T = \Big[D_1\ D_2\ S_2\Big]^T = \cdots = \Big[D_1\ D_2\ \cdots\ D_J\ S_J\Big]^T \end{aligned} \tag{20}$$

3.2 The Maximal Overlapping Discrete Wavelet Transform

The Maximal Overlapping Discrete Wavelet Transform (MODWT) has been judged to possess some additional worth compared to the DWT in time series analysis. For instance, MODWT does not be subsampled by two, and is well defined for any sample size. The number of coefficients in every level is equal to the sample size (Percival and Walden, 2000; Serroukh, 2012).

Let $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{g}}$ refer to MODWT wavelet filter and scaling filter, respectively. In every transformation level, there is a relationship between DWT filter and MODWT filter as described in Eq. (21)

$$\tilde{\mathbf{h}} = \frac{\mathbf{h}}{\sqrt{2}} \qquad \text{and} \qquad \tilde{\mathbf{g}} = \frac{\mathbf{g}}{\sqrt{2}} \tag{21}$$

The formulation of MODWT is described by

$$\tilde{\mathbf{D}} = \tilde{\mathbf{H}}_{\mathbf{j}}\mathbf{Y} \tag{22}$$

In the $j-$th transformation level, the dimensions of $\tilde{\mathbf{H}}_{\mathbf{j}}$ are $(j+1)N \times N$. This transformation matrix can be partitioned into $j+1$ submatrices, refering to each transformation level, so that the MODWT descibed in Eq. (22) can be written as

$$\begin{aligned}\tilde{\mathbf{D}} &= \left[\tilde{\mathscr{H}_1}\ \tilde{\mathscr{G}_1}\right]^T \mathbf{Y} = \left[\tilde{\mathscr{H}_1}\ \tilde{\mathscr{H}_2}\ \tilde{\mathscr{G}_2}\right]^T \mathbf{Y} = \cdots = \left[\tilde{\mathscr{H}_1}\ \tilde{\mathscr{H}_2}\ \cdots\ \tilde{\mathscr{H}_J}\ \tilde{\mathscr{G}_J}\right]^T \mathbf{Y} \\ &= \left[\tilde{D}_1\ \tilde{S}_1\right]^T = \left[\tilde{D}_1\ \tilde{D}_2\ \tilde{S}_2\right]^T = \cdots = \left[\tilde{D}_1\ \tilde{D}_2\ \cdots\ \tilde{D}_J\ \tilde{S}_J\right]^T\end{aligned} \tag{23}$$

For each index i, the submatrces $\tilde{\mathscr{H}_i}$ and $\tilde{\mathscr{G}_i}$ constitutes one periodic step of the wavelet filter and the scaling filter at the $i-$th level, respectively. For instance, the submatrix $\tilde{\mathscr{H}_1}$ can be described as follows:

$$\tilde{\mathscr{H}_1} = \begin{bmatrix} \tilde{h}_0 & 0 & 0 & 0 & \cdots & 0 & 0 & \tilde{h}_{L-1} & \tilde{h}_{L-2} & \cdots & \tilde{h}_2 & \tilde{h}_1 \\ \tilde{h}_1 & \tilde{h}_0 & 0 & 0 & 0 & \cdots & 0 & 0 & \tilde{h}_{L-1} & \tilde{h}_{L-2} & \cdots & \tilde{h}_2 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & \tilde{h}_{L-1} & \tilde{h}_{L-2} & \tilde{h}_{L-3} & \cdots & \tilde{h}_1 & \tilde{h}_0 \end{bmatrix} \tag{24}$$

3.3 The MODWT Forecasting Method for Time Series

Generally, forecasting is the main purpose of modeling a process. There are various ways to make forecasts, ranging from naive models to complicated models. This paper is mainly concerned with the use of MODWT for time series forecasting as described in the following equation (25) (Murtagh et al., 2004; Renaud et al., 2003).

$$Y_{t+1} = \sum_{j=1}^{J} \sum_{k=1}^{|A_j|} \hat{a}_{j,k} d_{j,t-2^j(k-1)} + \sum_{k=1}^{|A_{J+1}|} \hat{a}_{J+1,k} c_{J,t-2^J(k-1)} + \epsilon_t \tag{25}$$

The highest transformation level is denoted by J. The coefficient set chosen at level j is denoted by A_j. For instance, Eq. (25) will become Eq. (26) when $J = 4$ and $|A_j| = 2$ for $j = 1, 2, 3, 4, 5$.

$$\begin{aligned} Y_{t+1} &= a_{1,1}d_{1,t} + a_{1,2}d_{1,t-2} + a_{2,1}d_{2,t} + a_{2,2}d_{2,t-4} + \\ & \quad a_{3,1}d_{3,t} + a_{3,2}d_{3,t-8} + a_{4,1}d_{4,t} + a_{4,2}d_{4,t-16} + \\ & \quad a_{5,1}c_{4,t} + a_{5,2}c_{4,t-16} + \epsilon_t \end{aligned} \tag{26}$$

The parameter estimation for Eq. (26) can be calculated by the method of least squares (LSE) or the Maximum Likelihood method when the distribution of ϵ_t is known. Rukun et al. (2003) showed that the data with higher autocorrelation tends to have a lower sum of squares error when forecast by the MODWT model. Table (1) shows a summary of the statistics for this wavelet based model.

Table 1: Staistics for the wavelet based model

Variable	Coefficient	Std. Error	t value	Prob
X1	1.15732	0.11383	10.167	0.000000
X2	-0.01144	0.12492	-0.092	0.92707
X3	1.14253	0.10770	10.608	0.000000
X4	-0.34138	0.11453	-2.981	0.00304
X5	1.02910	0.08725	11.794	0.000000
X6	-0.43924	0.07592	-5.785	0.000000
X7	1.15439	0.05657	20.405	0.000000
X8	-0.12914	0.03307	-3.905	0.00011
X9	1.01673	0.01704	59.653	0.000000
X10	-0.04169	0.01626	-2.564	0.01068
Residual standard error: 1.019 on 426 degrees of freedom				
Multiple R-squared: 0.9687, Adjusted R-squared: 0.968				
F-statistic: 1321 on 10 and 426 DF, p-value: ¡ 2.2e-16				

4 Wavelet Radial Basis Neural Network Model

The role of a radial basis as an activation function in a radial basis neural network model has been discussed in some publications(e.g. Haykin, 1999; Samarasinghe, 2006). An input nearer to a radial center will result a bigger response. So, the radial basis functions in the model play the role of an input classifier into homogeneous groups depending on each radial center. Let **X** describe a random variable which will be processed by a radial basis function. Usually, it is transformed to standard form:

$$r = \frac{x-\mu}{\sigma}, \quad \sigma > 0, \quad x, \mu \in \mathbb{R} \tag{27}$$

Some kinds of radial basis functions can be seen in Eqs. (28), (29), and (30).

Gaussian function:

$$\Phi(r) \;\; = \;\; exp\left(-\frac{r^2}{2}\right) \tag{28}$$

Multiquadrics function:

$$\Phi(r) \;\; = \;\; \sqrt{1+r^2} \tag{29}$$

Multiquadrics inverse function:

$$\Phi(r) = \frac{1}{\sqrt{1+r^2}} \tag{30}$$

If $\mathbf{X}$ represents a $p-$variat random variable, then $\boldsymbol{\mu}$ represents its mean vector, and $\sigma^2 = \Sigma$ represents its variance-covariance matrix. Furthermore, r denote the Mahalanobis distance as defined in Eq. (31).

$$r_x^2 = (\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}), \qquad \mathbf{X},\boldsymbol{\mu} \in \mathbb{R}^p \tag{31}$$

The wavelet based forecasting descibed in Eq. (26) is included in the class of linear models. Sometimes it does not adequately approximate the main part and/or the detailed part of the original function in the linear sense. This is to sufficient reason to develope Eq. (25) into a nonlinear form consisting of a main part and a detailed part. The model is called the *wavelet radial basis neural network* (WRBNN) model. This refers to the use of wavelets as a pre-processing tool, radial basis functions as nonlinear transfer functions, and neural network rule for optimizing the parameter estimation. Without loss of generality, there will be developed a WRBNN model with inputs from the results of MODWT at the transformation level of $J = 4$ and $A_j = 2$ for all j. For the sake of simplificity, the variables in Eq. (26) will be redefined to refer to the transformation level as described in Eq. (32).

New symbol	Old synbol	New symbol	Old synbol
Y_i	Y_{t+1}		
X_1	$d_{1,t}$	X_2	$d_{1,t-2}$
X_3	$d_{2,t}$	X_4	$d_{2,t-4}$
X_5	$d_{3,t}$	X_6	$d_{3,t-8}$
X_7	$d_{4,t}$	X_8	$d_{4,t-16}$
X_9	$c_{4,t}$	X_{10}	$c_{4,t-16}$

(32)

The architecture of the proposed model can be seen in Figure 1. The input variables of the WRBNN model are similar to those of the wavelet based model described in Eq. (26) (see Murtagh et al. (2004)). There are two layers in the model. The first layer carries out a nonlinear process performed by the radial basis functions. The number of radial basis nodes in every input line is equivalent to the number of clusters that have occured. The mean and variance clusters are estimated refering to the input membership. The second layer carries out a linear process performed by a linear summation function. However, a nonlinear function can be used if needed. The mathematical form of the WRBNN estimation model can be seen in Eq. (33). The parameter estimation can be calculated by the least squares method.

$$\hat{Y}_i = \sum_{j=1}^{10}\sum_{k=1}^{q_j} \hat{a}_{j,k}\Phi_{j,k}(X_j) \tag{33}$$

The steps of the model building will be expressed chronologically as follows.

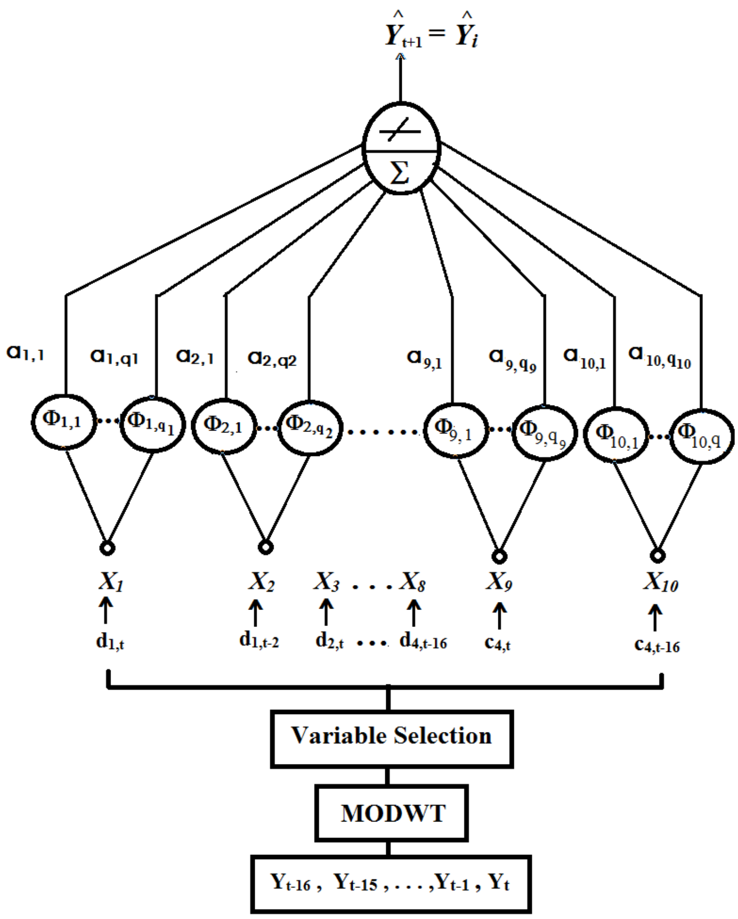

Figure 1: Model Architectur of WRBNN

1. The main pattern of the process usually dominates the model structure. This is an adequate reason for initiating the model by including X_{10} as descibed in Eq. (34) or Eq. (35). Equation (34) is used for the nonlinear structure and Eq. (35) for the linear. The choice of linear or nonlinear structure depends on its contribution as indicated by the R-squared value. This determination prevails for the other predictor variables.

$$Y_i = \sum_{k=1}^{q_{10}} a_{10,k}\Phi_{10,k}(X_{10}) + \epsilon_i \tag{34}$$

$$Y_i = a_{10}X_{10} + \epsilon_i \tag{35}$$

The hierarchical cluster method (see Johnson and Wichern (1982)) can be used to assist in deciding on the number of radial basis nodes in each input line. The mean of the radial basis nodes can be calculated by the $k-$means precedure. Next, the variance of the radial basis nodes can be calculated. The parameter estimation can be performed by the method of least squares. The significance test of the model parameters relies on the assumption of the normality of the errors.

2. The next step is investigating the appropriateness of X_9 for being included in the model. It is chosen for the reason that nearest to the previous variable in the multiresolution structure (see Eq. (23)). The test is begun by calculating the error of the previous model as described by

$$\epsilon_i = Y_i - \sum_{k=1}^{q_{10}} \hat{a}_{10,k}\Phi_{10,k}(X_{10}) \tag{36}$$

The regression model consists of ϵ as calculated in Eq. (36) using as the dependent variables, the radial basis values of X_{10} and X_9, is performed as follows

$$\epsilon_i = \sum_{k=1}^{q_{10}} a_{10,k}\Phi_{10,k}(X_{10}) + \sum_{l=1}^{q_9} a_{9,l}\Phi_{9,l}(X_9) + \epsilon_i^* \tag{37}$$

The LM test procedure is performed to measure the contribution of X_9 to the model. As mentioned in (Lee et al., 1993), nR^2 is asymtoticaly distributed as $\chi^2(q_9)$ where n is the sample size, R^2 is the determination coefficient, and q_9 is the number of degrees of freedom, which is equal to the number of variables to be added. This means that X_9 is appropriate for being included in the model when $nR^2 > \chi^2_{\alpha,q_9}$.

3. If X_9 is appropriate for being included in the model then the initial model described in Eq. (34) is updated to Eq. (38). Oherwise, X_9 is rejected from the model and step 2 is repeated for X_8.

$$Y_i = \sum_{k=1}^{q_{10}} a_{10,k}\Phi_{10,k}(X_{10}) + \sum_{l=1}^{q_9} a_{9,l}\Phi_{9,l}(X_9) + \epsilon_i \tag{38}$$

4. The task expressed in steps 2 and 3 are repeated for the other variables described in Eq. (32). The significance test of the model parameters in every step is performed assuming the normality of the errors.

5 Result and Discussion

Data with nonlinear properties is needed to support the comprehensiveness of these observation. Fortunately, the $\mathbb{R}$ software (R Core Team, 2014) and its supporting packages usually have relevant examples of such data. In Addition, the need for relevant data can be met by generating simulation data. The data of the index of industrial production in the United States (`IIPUs`) in the `tsDyn` package will be used as an example in this paper (Hansen, 1999). `tsDyn` is a package of dynamical time series modeling included in the class of nonlinear models (see Stigler 2010 and Di Narzo et al. 2009).

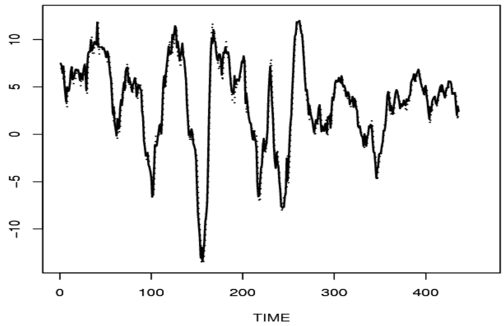

Figure 2: Plot of IIPUs data (solid line) and its predction (dots line) in wavelet based model

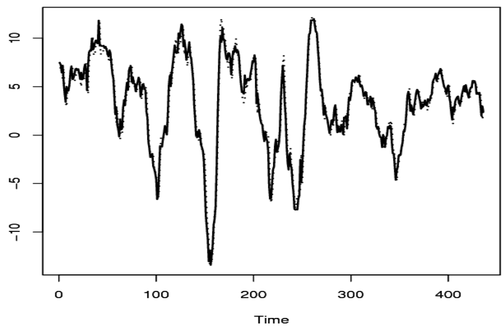

Figure 3: Plot of IIPUs data (solid line) and its predction (dots line) in WRBNN model

Table (1) shows that the variable X_2 does not make a significant contribution to the model in the linear sense. A careful investigation was performed to revise the model. The statistical summary of the WRBNN model can be seen Table (2). The results show that all variables make a significant contribution to the WRBNN model. Furthermore, the R squared of the new model is greater, which indicates that the new model is better.

6 Conclusion

The choice of initial mean cluster in the *k-means* procedure plays a role in the goodness of the final model. Computer packages may make aprovisional choice as was done in this paper. Some trials were performed and then the best result was choosen. However, expertise may need to be sharpened in order to chose the best value. The type of radial basis function also needs to be matched to the properties of the data. No one perfect method for all types of data exists. Although the IIPUs data can be approximated well by the WRBNN model, but it still needs to be compared to other models. Finally, an advanced WRBNN model is still open for development and testing.

Table 2: Staistics for the WRBNN model

RBF	Variable	Coefficient	Std. Error	t value	Prob.
-	X_{10}	-0.05846	0.01718	-3.403	0.000728
-	X_9	0.99749	0.01822	54.741	0.000000
-	X_8	-0.11615	0.03310	-3.510	0.000497
-	X_7	1.14460	0.05616	20.380	0.000000
-	X_6	-0.42897	0.07531	-5.696	0.000000
-	X_5	1.01633	0.07865	12.922	0.000000
-	X_4	-0.33759	0.11109	-3.039	0.002521
-	X_3	1.11906	0.10358	10.804	0.000000
Mq*)	X_2	0.14237	0.05062	2.813	0.005140
-	X_1	1.14656	0.11027	10.397	0.000000

Residual standard error: 1.01 on 426 degrees of freedom
Multiple R-squared: 0.9693, Adjusted R-squared: 0.9686
F-statistic: 1346 on 10 and 426 DF, p-value: < 2.2e-16

*) Multiquadrics radial basis function

Acknowledgements

This research was supported by the Direktorat Riset dan Pengabdian kepada Masyarakat Kementerian Riset, Teknologi, dan Pendidikan Tinggi, Republic of Indonesia

References

Aldrich, E. (2013). *Wavelets: A Package of Funtions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses.* http://CRAN.R-project.org/package=wavelets. R package version 0.3-0.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. of Econometrics*, 31:307–327.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Di Narzo, A. F., Aznarte, J. L., and Stigler, M. (2009). *tsDyn: Time series analysis based on dynamical systems theory.* R package version 0.7, http://stat.ethz.ch/CRAN/web/packages/tsDyn/ vignettes/tsDyn.pdf.

Engel, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *J. of Econometrica*, 50:987–1008.

Hansen (1999). Testing for linearity. *J. of Economic Survey*, 13(5):551–576.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation.* Prentice-Hall.

Johnson, R. A. and Wichern, D. W. (1982). *Applied Multivariate Statistical Analysis.* Prentice-Hall.

Lee, T. H., White, H., and Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *J. of Econometrics*, 56:269–290.

Murtagh, F., Starck, J. L., and Renaud, O. (2004). On neuro wavelet modeling. *Decision Support System*, 37:475–490.

Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhauser, Berlin.

Orr, M. J. L. (1996). *Introduction to Radial Basis Function Networks.* Centre for Cognitive Science, University of Edinburgh.

Orr, M. J. L. (1999). *Recent Advances in Radial Basis Function Networks.* Centre for Cognitive Science, University of Edinburgh.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge Univ. Press.

Popoola, A. O. (2007). *Fuzzy-Wavelet Method for Time Series Analysis.* PhD thesis, Surey University.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, http://www.R-project.org/, Vienna, Austria.

Renaud, O., Starck, J. L., and Murtagh, F. (2003). Prediction based on a multiscale decomposition. *Int. J. of Wavelets Multiresolution and Information Processing*, 1(2):217–232.

Rukun, S., Subanar, Rosadi, D., and Suhartono (2003). The adequateness of wavelet based model for time series. *J. Phys.: Conf. Ser.*, 423.

Samarasinghe, S. (2006). *Neural Network for Applied Science and Engineering*. Auerbach Pub., New York.

Serroukh, A. (2012). Wavelet coefficients cross-correlation analysis of time series. *Electronic Journal of Applied Statistical Analysis*, 5(2):289–296.

Stigler, M. (2010). *Threshold cointegration: Overview and implementation in R*. R package version 0.7-2, http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/ThCointOverview.pdf.

Teräsvirta, T., Lin, C. F., and Granger, C. W. J. (1993). Power of the neural network linearity test. *J. of Time Series Analysis*, 14(2):209–220.

Tong, H. (1990). *Nonlinear Time Series: A Dynamic System Approach*. Clarendon Press, Oxford.