Wind Speed Forecasting using MRA based Adaptive Wavelet Neural Network

K. Bhaskar, *Student Member, IEEE,* S.N. Singh, *Senior Member, IEEE*

***Abstract*—This paper addresses the problem of predicting hourly wind speed using Adaptive Wavelet Neural Networks (AWNN). It employs wavelets as its activation function in the hidden layer. Due to time-frequency localiztion property and adapting the wavelet shape according to training data set instead of adapting the parameters of the fixed shape basis function, WNNs have better generalization properties in contrast to the classical Feed Forward Neural Network (FFNN). The wavelet based multiresolution analysis (MRA) is applied on wind series to decompose in to smooth and detail signals for better wind characterization and reliable forecasting. The transformed data of historical wind series were trained and tested over various periods of time using AWNNs. The forecasting results are better when compared with AWNN as a forecasitng model alone, FFNN with Back Propagation (BP) training algorithm as a forecasting model, and MRA based FFNN as a forecasting model.**

***Index Terms*—Adaptive wavelet neural network, multiresolution analysis, short-term wind forecasting.**

I. Introduction

GLOBAL warming resulting mainly due to fossil fuel combustion threatens our ecosystem, our health and our economy. According to the International Energy Agency's (IEA) findings published in June 2006 [1], the oil production is predicted to "peak" by 2014; gas by 2030, resulting aggressive energy price rises. The awareness of the limited availability of fossil fuels and the recognition of uncertainity in environmental conditions due to massive release of fossil CO_2 in the last two centuries acted as a driving mechanism for growth of non-conventional energy conversion systems. By the end of 2000, wind farm contributed, world wide, an estimated 60-GW aggregate capacity. Which rose to 160-GW by 2009. According to GWEC 2009 report [2], a 400-GW of wind generation capacity is estimated by 2014.

In contrast to conventional power plants, the electricity produced by wind forms almost entirely depends on meteorological conditions, perticularly the magnitude of the wind speed. The fluctuating or intermittent nature of the wind power production due to unforeseen variations of the wind conditions constitutes a new challenges for the system operators, perticularly in countries with high penetration levals of wind energy like Denmark, Spain and Germany. Because of the intermittent nature of the wind, utilities traditionally do not consider the contribution from wind forms in their economic scheduling and unit commitment problems. Also with the changing requirements, some of the Transmission System Operators (TSO) are introducing new grid codes, the guidelines for the connection of wind power in to the existing grid. They impose on the present day wind turbines that they should behave in a manner akin to conventional synchronous generators, for example, predictable and controllable real and reactive power output [3]. In order to mitigate the undesirable effects of wind fluctuations in growing wind penetration, an accurate wind power forecasting tool is fully justified.

K. Bhaskar and S.N. Singh are with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, India e-mail: (bkanna@iitk.ac.in; snsingh@iitk.ac.in).

Several techniques have been identified for wind forecasting. A literature survey of wind forecasting technology can be found in [4]–[6]. These techniques can be categorized in to physical systems and statistical systems. Physical systems are based on the results of Numerical Weather Prediction (NWP) systems [7]. The basic problem is to transform the wind speed given by the NWP on a coarse numerical grid to the on-site conditions at the location of the wind form. Physical systems use parameterizations based on a detailed physical description of lower atmosphere by considering several factors like surface roughness and its changes, scaling of the local wind speed within wind forms, wind form layouts and turbine power curves. The corrected wind speed is then plugged into the corresponding power curve of the wind turbine to determine the power output. Since NWP models are complex mathematical models, they are usually run on super computers, which limits the usefulness of NWP methods for on-line or very short-term operation of power system. Statistical models use explanatory variables and online measurements, usually employing recursive techniques, and artificial neural networks. Furthermore, physical models must and statistical models mey use NWP models.

Sideratos et al [8] applied neural networks and fuzzy logic techniques for estimation of wind form output. A fuzzy based self organized map classifies the input data to three classes depending on the magnitude of the wind speed. For each class, different RBF neural network are used as a prediction models.

Wavelet theory is emerging as an important tool in many applications in signal processing and numerical analysis due to time-frequency localization property [9]. Hunt and Nason [10] first applied wavelets to wind speed forecasting. They tried to model the response time-series in terms of a multiscalewavelet decomposition of the reference site data time series. Lei et al [11] and Diogo et al [12] applied wavelet based multiresolution analysis to decompose highly nonlinear wind speed time series in to several approximate time series. Then Auto Regressive Moving Average (ARMA) models are used for forecasting each approximate time series. In [13], Khan et al used spline smoothing method and linear curve fitting with least mean square error method for predicting the approximate and detail signals obtained from wavelet decomposition. Dong et al [14] presented wavelet decomposition based ANN as a forecasting model for wind power prediction.

A Wavelet Neural Network (WNN) was first developed by Zhang et al [15] as an alternative to the classical feed-forward neural network for approximating arbitrary nonlinnear functions. Due to the local properties of wavelets and the concept of adapting the wavelet shape according to training data set instead of adapting the parameters of the fixed shape basis function, WNNs are having better generalization property. The WNNs have been successfully applied in the field of function learning [16], nonlinear system identification [17], time series prediction [18]. Pindoriya et al [19] used adaptive wavelet neural network for energy price forecasting. However, to the best of the author's knowledge, AWNN has not yet used as a wind forecasting model.

In this paper, wavelet based multiresolution analysis (MRA) is performed, to decompose the input wind time series in to detail and smooth signals upto the desired scaling level. Seperate AWNN model is used to forecast each of the detail and smooth signals. Which when added gives the actual wind series forecast. The data used in this paper is a hourly mean values of 10 minute time samples of wind speed from 17th January 2006 to 16th August 2006, taken from meteorological mast tower located at Thompson Island, Boston Harbor, USA [21]. To demonstrate the effectiveness of proposed technique, the results are compared with forecasting methods based on 1) AWNN forecsting method alone, 2) FFNN forecasting method with back propagation training algorithm and 3) MRA based FFNN forecasting method.

II. Wavelet Fundamentals

Wavelets are mathematical tools for analysis of time series and images. From time series analysis point of view, Continuous Wavelet Transform (CWT) are designed to work with time series defined over the entire real axis. And Discrete Wavelet Transform (DWT) deals with series defined essentially over a range of integers ($t = 0, 1, 2, \ldots, N-1$). A wavelet is a small wave grow and decays essentially in a limited time period. Fig. 1 depicts few basic wavelets. A wavelet can be defined as a real valued function $\psi(\cdot)$ defined over the real axis $(-\infty, \infty)$ and satisfying the two basic properties,

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

and the square of $\psi(\cdot)$ integrates to unity

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

Equation (2) imposes $\psi(\cdot)$ to make some excursions away from zero, and (1) imposes that any excursion it makes above zero must be cancelled out by excursion below zero, which implies $\psi(\cdot)$ must resemble a wave. To get wavelet a practical utility, it is necessary to impose conditions beyond (1) and (2) and admissible condition is one among them. A wavelet $\psi(\cdot)$ is said to be admissible if its Fourier transform

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(u) e^{-i2\pi f u} du, \text{ is such that}$$

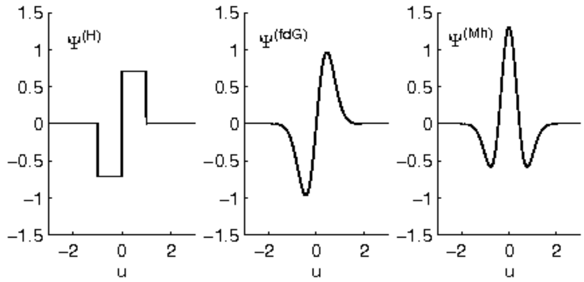

Fig. 1. Three wavelets. From left to right, Haar wavelet; first derivative of Gaussian function; and Mexican hat wavelet.

$$C_\psi \equiv \int_0^\infty \frac{|\ \Psi(f)\ |^2}{f} df \quad \text{satisfies } 0 < C_\psi < \infty. \tag{3}$$

This admissible condition allows the reconstruction of a function $x(\cdot)$ from its continuous wavelet transform [9]. The wavelets tell us how weighted averages of certain other function vary from one averaging period to the next. And this is achieved by shifting and scaling of wavelet function such as

$$\psi_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} \psi(\frac{u-t}{\lambda}) \;, \tag{4}$$

generally called as translated and dilated version of wavelet. The factor $1/\sqrt{\lambda}$ is multiplied to satisfy the condition given by (2). The CWT defined by

$$W(\lambda, t) = \int_{-\infty}^{\infty} \psi_{\lambda,t}(u).x(u)du \tag{5}$$

presents all the information in $x(\cdot)$. If $\psi(\cdot)$ satisfies the admissible condition (3) and if the signal $x(\cdot)$ satisfies $\int_{-\infty}^{\infty} x^2(t)dt < \infty$ then we can recover $x(\cdot)$ from its CWT as

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^{\infty} W(\lambda, u) \frac{1}{\sqrt{\lambda}} \psi\left(\frac{t-u}{\lambda}\right) du \right] \frac{d\lambda}{\lambda^2} \tag{6}$$

and we also have [22]

$$\int_{-\infty}^{\infty} x^2(t)dt = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^{\infty} W^2(\lambda, t)dt \right] \frac{d\lambda}{\lambda^2}. \tag{7}$$

The left hand side of (7) defines the energy in the signal $x(\cdot)$; and it says that $W^2(\lambda, t)/\lambda^2$ essentially defines an energy density function that decomposes the energy across different scales and times which leads to multiresolution analysis.

A. Discrete Wavelet Transform and MRA

As the two dimensional CWT depends on just a one dimensional signal, it is abvious that there is a lot of redundancy in CWT. Moreover it takes more computational time and large memory space. DWT an easy method, can be thought of as a jusicious subsampling of $W(\lambda, t)$, which can still help us in reconstructing the signal $x(\cdot)$. One specific version of DWT which deals with just "dyadic" scales $\lambda = 2^j$, where j is an integer indicating the scaling level. The corresponding wavelet as a function of dyadic scaling level j and position k is given by

$$\psi_{j,k}(u) = 2^{-\frac{j}{2}}\psi\left(2^{-j}u - k\right) \quad j,k \in Z. \tag{8}$$

If $\mathbf{X}$ represents a real valued finite time series defined as $\mathbf{X} = [X_1, X_2, \ldots, X_N]^T$, where N is an integer multiple of 2^J, then the discrete wavelet transform of $\{\mathbf{X}\}$ is given by [23]

$$\mathbf{W} = \mathcal{W}\mathbf{X} \tag{9}$$

where $\mathbf{W}$ is a column vector of length $N = 2^J$ whose nth element is the nth DWT coefficient W_n, and $\mathcal{W}$ is an $N \times N$ real-valued matrix defining the DWT and satisfying orthonormality condition $\mathcal{W}^T\mathcal{W} = I_N$. The nth wavelet coefficient W_n is associated with a particular scale and with a particular set of times. As $\mathcal{W}$ is orthonormal, reconstruction of X is possible by premultiplying on both sides of (9) with $\mathcal{W}^T$

$$\mathcal{W}^T W = \mathcal{W}^T\mathcal{W}\mathbf{X} = \mathbf{X}. \tag{10}$$

Then vector $\mathbf{X}$ can be expressed as an addition of $J+1$ vectors of length N as $\mathbf{X} = \mathcal{W}^T\mathbf{W} =$

$$\left[\mathcal{W}_1^T, \mathcal{W}_2^T, \ldots, \mathcal{W}_J^T, \mathcal{V}_J^T\right]\begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_J \\ V_J \end{bmatrix}$$

$$= \sum_{j=1}^{J}\mathcal{W}_j^T W_j + \mathcal{V}_J^T V_J = \sum_{j=1}^{J}\mathcal{D}_j + \mathcal{S}_J \tag{11}$$

the first J vectors are associated with a perticular scale referred as detail signals denoted by $\mathcal{D}_j$ and the last vector has all its elements equal to the sample mean of X and referred as smooth signal $\mathcal{S}_J$, which leads to the multiresolution analysis. MRA, analyzes the signal at different frequencies with different resolution. MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. This approach makes sense with signals having high frequency components for short durations and low frequency components for long duration, which happens in all practical cases.

B. Maximum Overlap Discrete Wavelet Transform (MODWT)

In this paper, a modified version of the DWT called Maximum Overlap Discrete Wavelet Transform (MODWT) is used. It has been discussed in the wavelet literature under the name,'Wavelet frames' [24]. In contrast to orthogonal DWT, the MODWT for a time series $\mathbf{X}$ is a highly redundant non orthogonal transform yielding the column vectors of $\mathbf{W}$, each of dimension N. But still MRA of a signal can be performed like DWT. The choice of the MODWT over DWT is due the advantages that this modified version brings, the most important of which are: (i)MODWT of level j is well defined for any sample size N, where as DWT restricts the sample size to be an integer multiple of 2^J. (ii)As mentioned earlier, like DWT, the MODWT can also be used to form a MRA.

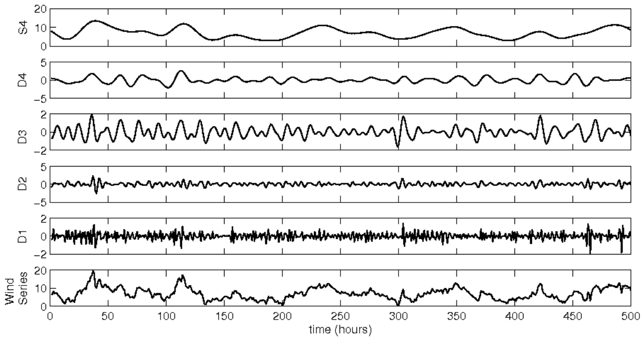

Fig. 2. MRA of level 4 applied to wind speed time seres using DB-4 wavelet filter.

However, in constrast to the DWT, the MODWT details and smooths are associated with zero phase filters, thus making it easy to line up features in an MRA with the original time series.

To perform the MRA on a given wind series using MODWT, 'Pyramid Algorithm' is used, which was introduced in the context of wavelets by Mallat [22]. It allows $\mathbf{W} = \mathcal{W}\mathbf{X}$ to be computed using only $O(N)$ multiplications, where as brute force technique involves N^2 multiplications. The Pyramid algorithm used here is based on linear filtering operations. The original signal is made to pass through high pass and low pass filters. The signal resulting from high pass filter is called detailed signal, and the other signal resulting from low pass filter represents smooth signal. Fig. 2 shows the MRA of wind time series using MODWT

III. AWNN AS A WIND FORECASTING MODEL

A. Architechture of Wavelet Neural Network

The general structure of a WNN model is depicted in Fig. 3. Like FFNN it comprises of an input layer, a hidden layer and a linear output layer. In this paper Mexican hat is choosen as a mother wavelet. It is the second derivative of gaussian function and defined by

$$\psi(x) = \left(1 - x^2\right)e^{-0.5x^2}. \tag{12}$$

Due to its symmetricity in shape, having explicit expression, and providing an exact time frequency analysis it has been considered to be the most appropriate mother wavelet function [19]. If n represents the dimension of the input vector $X = [x_1, x_2, \ldots, x_n]^T$ then the wavelet family can be generated over the entire input space by translating and dilating the mother wavelet given by

$$\begin{aligned}\psi_{a,b}(x_i) &= \left(1 - \left(\tfrac{x_i - b}{a}\right)^2\right)e^{-0.5\left(\frac{x_i - b}{a}\right)^2} \\ & i \in n; \quad a, b \in \Re; \quad a > 0. \end{aligned} \tag{13}$$

The input data in the input layer is directly transmitted to the wavelet layer. The n-dimensional wavelet basis function can be calculated by the tensor product of all one dimension wavelets [17]. Therefore the output z of the hidden layer neurons will be given by

$$z_j = \prod_{i=1}^{n}\psi_{a_{ij},b_{ij}}(x_i) \quad j \in m \tag{14}$$

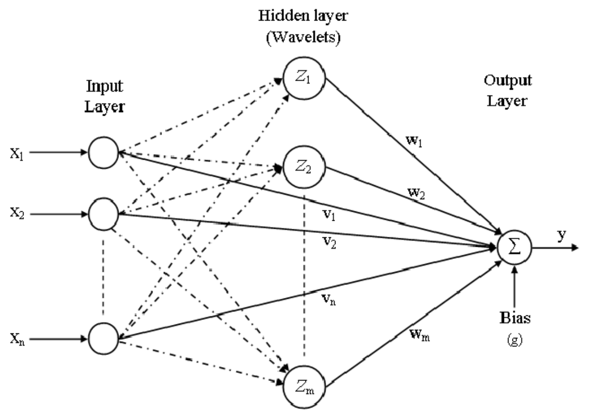

Fig. 3. General Structure of Wavelet Neural Network

in order to map the linear input-output relation, it is customary to have additional direct connections from input layer to output layer, as there is no point in using wavelets for reconstructing linear terms [15]. The output of the WNN can be computed as

$$y = \sum_{j=1}^{m} w_j z_j + \sum_{i=1}^{n} v_i x_i + g \tag{15}$$

where w_j is the layer weight between jth wavelon and output node, v_i is the input weight between ith input node and output node, and g is the bias at output node.

B. Training Algorithm

The standard back propagation gradient descent algorithm [25] is used for training the wavelet neural network, as the output function computed by the WNN is differentiable with respect to all unknown parameters, translation and dilation coefficients, weights and bias of the network. Training is based on minimisation of cost function also called as mean square error (MSE) given as

$$E = \frac{1}{2N}\sum_{p=1}^{N}[e(p)]^2 \quad \text{and} \quad e(p) = y^d(p) - y(p) \tag{16}$$

where $y(p)$ is the model output, and $y^d(p)$ is the desired output for a given pth input pattern. N represents the the total training patterns. And the updation of free parameter is given as

$$a(k+1) = a(k) + \eta\Delta a(k) + \alpha\Delta a(k-1) \tag{17}$$

where a represents the general unknown free variable, η is the learning rate and α is the momentum parameter. Then the changes in free parameters are found as

$$\Delta w_j = -\frac{\partial E}{\partial w_j} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial w_j} = e z_j \quad j \in m \tag{18}$$

$$\Delta v_i = -\frac{\partial E}{\partial v_i} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial v_i} = e x_i \quad i \in n \tag{19}$$

$$\Delta g = -\frac{\partial E}{\partial g} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial g} = e \tag{20}$$

$$\Delta a_{ij} = -\frac{\partial E}{\partial a_{ij}} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial z_j}\frac{\partial z_j}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial a_{ij}}$$

$$= e w_j z_j \left[\frac{1}{a_{ij}}\right]\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2 \left[3 - \left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2\right] e^{-0.5\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2} \tag{21}$$

$$\Delta b_{ij} = -\frac{\partial E}{\partial b_{ij}} = -\frac{\partial E}{\partial y}\frac{\partial y}{\partial z_j}\frac{\partial z_j}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial b_{ij}}$$

$$= e w_j z_j \left[\frac{1}{a_{ij}}\right]\left[\frac{x_i - b_{ij}}{a_{ij}}\right] \left[3 - \left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2\right] e^{-0.5\left[\frac{x_i - b_{ij}}{a_{ij}}\right]^2} \tag{22}$$

During training process the over fitting of training data can be stopped at right point using cross-validation. The available data set is divided in to training, validation and testing subsets. The training set is used to compute the gradients and update all the free parameters of the network. The error on the validation set is monitered during the training session. In early stopping criteria, when the validation error starts increasing after a number of predefined iterations, the training is stopped and the parameters at the minimum of the validation error are returned for the optimal network complexity.

C. An Adaptive Learning Rate

To make the learning process converge more rapidly than the conventional method, where in both learning rate and momentum parameters are kept contant during the learning process, an adaptive learning rate has been considered [19], [20]. The proposed adaption rule is as follows

$$\eta(n+1) = \begin{cases} 1.05\,\eta(n) & , \Delta E_n > 0 \\ 0.7\,\eta(n) & , otherwise \end{cases} \tag{23}$$

where $\eta(n)$ is learning rate at iteration n, and $\Delta E_n = E(n-1) - E(n)$ with $E(n)$ being the mean squared error (16) at the end of the nth iteration. Here, the basic idea is to increase η when ΔE_n is positive and decrease η when ΔE_n is negative. Note that for positive ΔE_n the error is decreasing, which implies that the network parameters are updated to the correct direction. If parameters move on to the wrong direction, causing error to increase, the direction is ignored in the next iteration by decreasing the learning rate.

IV. Results and Discussions

A. Correlation Analysis and Input Variable Selection

To evaluate the performance of MRA based AWNN, a three layer FFNN is also developed. Log sigmoidal activation function $1/(1+\exp^{-av})$ in the hidden layer and linear activation function in the output layer are used. Where positive constant a is the slope of the sigmoisal function and v is the weighted sum at neuron. That is the forecasting results of FFNN, AWNN and MRA based FFNN forecasting methods are compared with the forecasting results of proposed MRA based AWNN forecasting method. In FFNN and AWNN forecasting methods, the actual wind series is given as input signal. And for MRA based FFNN and AWNN forecasting methods the detailed and smooth signals are the input signals. To help, improve the training performance of the forecasting models the input data are normalized between $[-1, 1]$.

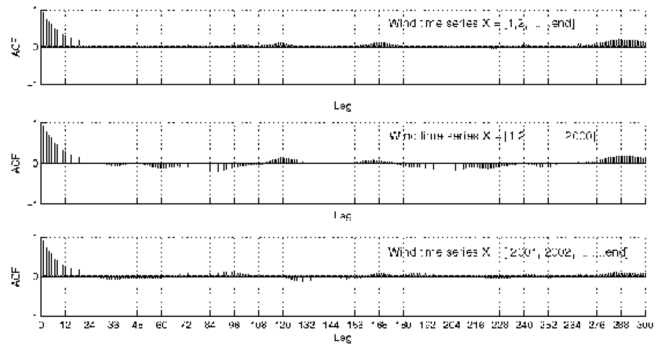

Fig. 4. ACF of Wind Time Series

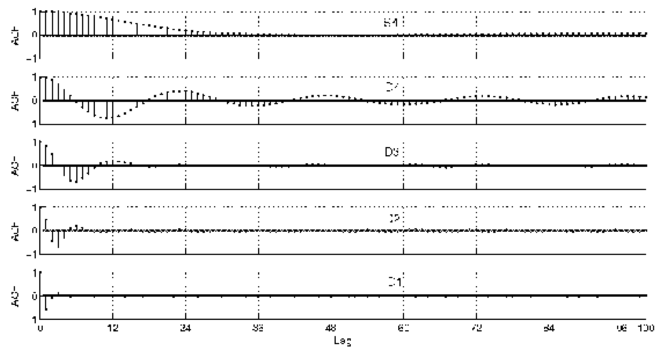

Fig. 5. ACF of Detailed and Smooth Time Series

The correlation test is performed prior to the model construction. The analysis is helpful to evaluate the relationship between the output and each influencing factor. ACF analysis on wind series is carried out by considering the entire series and also subsets of series at different intervals. Fig. 4 shows that in all cases there is no seasonal pattern exists and correlation factor decreases rapidly as the lag times increases. With considerations of these factors the input variables selected for FFNN and AWNN forecasting methods for predicting wind at hour h are

$$X_{h-1}, X_{h-2}, \ldots, X_{h-12} \tag{24}$$

Fig. 5 depicts ACF carried out on detailed and smooth signal of the decomposed wind series. It shows that ACF for S_4 decreases gradually with high correlation factors up to 12 lag hours. The ACF for $D_4, D_3, D_2$and D_1 are damped in nature with positive and negative correlations. In case of D_4, lag hours $1, 2, 3$ and $22, 23, 24$ are positively correlated and lag hours $11, 12, 13$ are negatively correlated. In case of D_3 lag hours $1, 2$ are positively correlated and lag hours $4, 5, 6, 7$ are negatively correlated. The ACFs for D_2 and D_1 are highly damped in nature. With these considerations the input variables selected for MRA based FFNN and AWNN forecasting methods for forecasting D_1 to D_4 and S_4 are given in Table I.

In [19], Pindoriya et al showed that with the increase in wavelons in the hidden layer of a AWNN, the Mean Absolute Percentage Error (MAPE) given by (25) for day ahead hourly load forecast also increases. In this paper, 2 wavelons in the hidden layer of AWNN are used. For FFNN, based on network pruning technique [25] the number of neurons in the hidden layer are selected. The network configuration used in this work for different forecasting models is given in Table II.

TABLE I
INPUT VARIABLES SELECTED FOR DIFFERENT FORECASTING MODELS

Input Series	*Forecasting Model*	
	MRA based FFNN	*MRA based AWNN*
S_4	$[1, 2, \ldots, 9]$	$[1, 2, \ldots, 9]$
D_4	$[1, 2, 3, 11, 12, 13, 24]$	$[1, 2, 3, 11, 12, 13, 24]$
D_3	$[1, 2, 3, 4, 5, 6, 7]$	$[1, 2, 3, 4, 5, 6, 7]$
D_2	$[1, 2, \ldots 6]$	$[1, 2, \ldots, 6]$
D_1	$[1, 2, \ldots, 6]$	$[1, 2, \ldots, 6]$

TABLE II
NETWORK ARCHITECHTURE AND SLOPE A SELECTED FOR DIFFERENT FORECASTING MODELS

Input Series	*Forecasting Model*		*slope a in*
	FFNN	*AWNN*	*FFNN*
Wind Series	$12-6-1$	$12-2-1$	20
S_4	$9-4-1$	$9-2-1$	6
D_4	$7-3-1$	$7-2-1$	16
D_3	$7-3-1$	$7-2-1$	25
D_2	$6-2-1$	$6-2-1$	33
D_1	$6-2-1$	$6-2-1$	40

B. Forecasting Results

Day ahead hourly forecsting and week ahead hourly forecasting are carried out for comparision of forecasting models. In both cases for network training five hundred hystorical data points prior to the test set data are considered as training set. And last 30% of training set is taken for validation. The measure of error between the actual and predicted wind speed is obtained using Mean Absolute Percentage Error (MAPE) as the prediction error as follows:

$$MAPE = \frac{1}{h} \sum_{i=1}^{h} \frac{\mid A_i - P_i \mid}{A_i} \times 100 \tag{25}$$

where A_i and P_i are the actual and predicted wind speed respectively at ith hour. h is the number of hours.

To avoid the adverse effect of hourly wind speed values close to zero, the mean error e can be computed as follows:

$$e(\%) = \frac{1}{h} \sum_{i=1}^{h} \frac{\mid A_i - P_i \mid}{\overline{A}} \times 100 \tag{26}$$

where

$$\overline{A} = \frac{1}{h} \sum_{i=1}^{h} A_i. \tag{27}$$

In this paper, after thorough investigations, the following system parameters are employed. The initial learning rate and momemtum parameters for AWNN forecasting model are set to 0.1 and 0.5 respectively and for FFNN forecasting model both are set to 0.5. The parameter a in FFNN for forecasting different time series are given in Table II. As a stopping criterion for all forecasting models, the mean square error

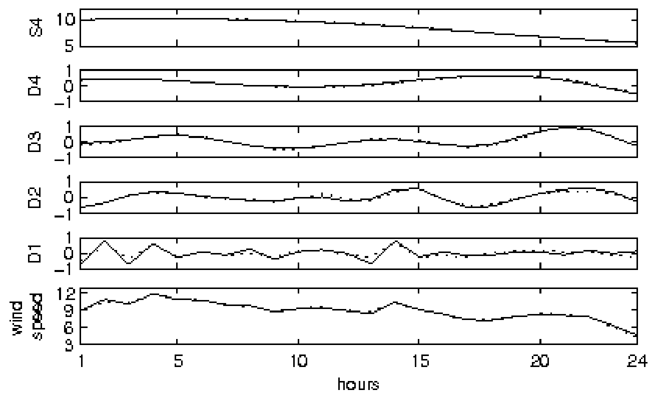

Fig. 6. Day ahead hourly forecasting using MRA based AWNN forecasting technique. Solid line indicates actual and dotted line indicates forecasted wind.

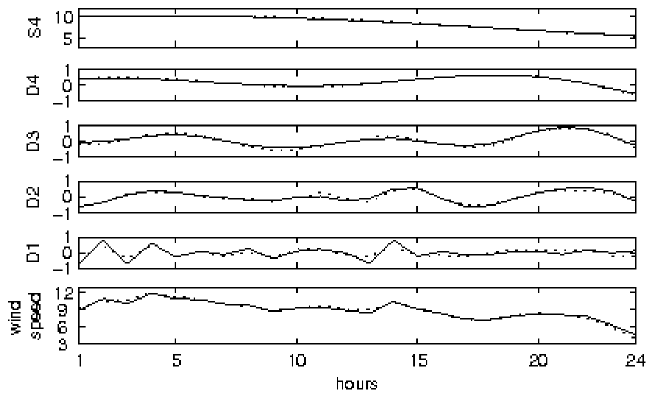

Fig. 7. Day ahead hourly forecasting using MRA based FFNN forecasting technique. Solid line indicates actual and dotted line indicates forecasted wind.

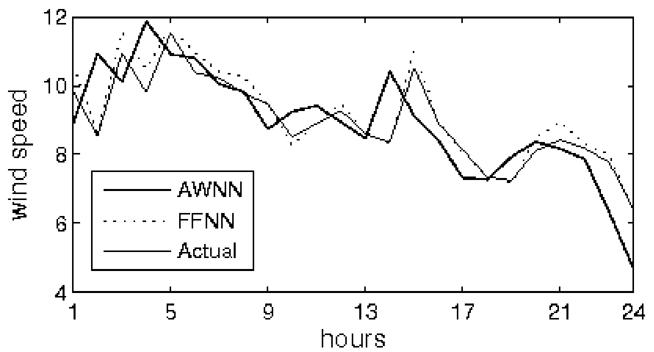

Fig. 8. Day ahead hourly forecasting using AWNN and FFNN forecasting techniques

(MSE) goal set is 0.0001 and maximum iteration of 500. In case of validation as an early stopping criterion the maximum fails are set to 100.

The day ahead hourly forecast using MRA based AWNN is depicted in Fig. 6. It shows the forecasting of smooth and detail signals along with the combined forecasting. Figs. 7 and 8 indicate the day ahead hourly forecasting by using other forecasting models. The hourly absolute percentage error (APE), daily mean absolute percentage error (MAPE) and daily mean error $e_{day}(\%)$ for all forecasting models are given in Table III. It shows that MAPE values are low for MRA based forecasting methods. The ACF of highly intermittent, non-stationary wind series has no consistant daily patterns and seasonal trends. Which made difficulty in selecting the input variables for FFNN and AWNN forecasting methods with out MRA. This problem is subdued using multiresolution analysis applied on input wind series. After MRA on wind series, the obtained smooth and detail signals are forecated seperately using FFNNs and AWNNs. Figs. 6 and 7 show that in all cases MRA based AWNN gave the better forecasting than MRA based FFNN. Perticularly forecasting the high frequency detail signal D_1 using AWNN is far superior than forecasting using FFNN. As neural networks are not optimistic in forecasting high frequency signals [26].

TABLE III
APE (%),MAPE (%), AND DAILY MEAN ERROR $e_{day}(\%)$ FOR DIFFERENT FORECASTING MODELS

Hour	Absolute Percentage Error (APE%)			
	MRA_AWNN	MRA_FFNN	AWNN	FFNN
1	2.8505	4.9867	10.1089	18.4444
2	3.8975	3.6412	21.7556	20.9300
3	3.7692	5.4134	8.1172	14.2327
4	0.4565	0.5904	17.2858	11.5924
5	1.2388	2.9309	5.7491	7.2626
6	0.3795	1.8251	3.9269	1.4854
7	0.8340	0.4269	1.7774	3.6305
8	1.6807	1.9370	0.4789	4.4138
9	1.6810	0.3545	8.2264	8.3934
10	0.6561	0.6896	7.9067	10.6391
11	2.3575	3.5416	5.4871	5.1101
12	1.4226	2.0484	3.7281	5.8905
13	3.9159	5.4256	1.4378	2.2363
14	2.0825	1.4952	19.8551	20.3647
15	1.2882	2.3340	15.5114	21.6190
16	2.1003	2.3703	6.0595	6.3449
17	0.4969	1.2614	11.0580	9.8553
18	2.7768	0.7485	0.8597	1.1879
19	1.0302	0.9685	8.6008	9.5856
20	1.1176	1.5904	3.0101	1.4424
21	2.1796	0.1996	3.2057	9.4420
22	2.2565	4.3390	3.8997	5.6457
23	3.1056	4.9569	22.7094	26.0498
24	7.3858	15.9204	37.0337	37.0060
MAPE	**2.1233**	2.9165	9.4912	10.9502
$e_{day}(\%)$	**1.9858**	2.6264	9.0845	10.4268

In case of week ahead hourly forecasting using MRA based AWNN and FFNN all network parameters are same as that of day ahead hourly forecasting eccept the learning rate in MRA based FFNN is changed from 0.5 to 0.1. The Fig. 9 shows the week ahead hourly forecasting using MRA based AWNN and FFNN respectively. Table IV gives the weekly MAPE and mean error. It shows that weekly MAPE 5.8468 and mean error 2.9155 obtained using MRA based AWNN forecasting method is far better than the weekly MAPE 15.5567 and mean error 5.5234 obtained using MRA based FFNN. This is due to limitations in capturing sudden changes in trends

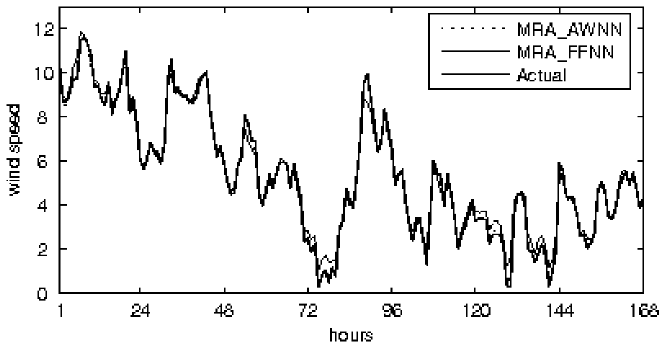

Fig. 9. Week ahead hourly forecasting using MRA based AWNN and FFNN forecasting techniques.

TABLE IV
WEEKLY MAPE (%), AND MEAN ERROR $e_{week}(\%)$ FOR DIFFERENT FORECASTING MODELS

	MRA based	
	AWNN	FFNN
MAPE	5.8468	15.5567
$e_{week}(\%)$	2.9155	5.5234

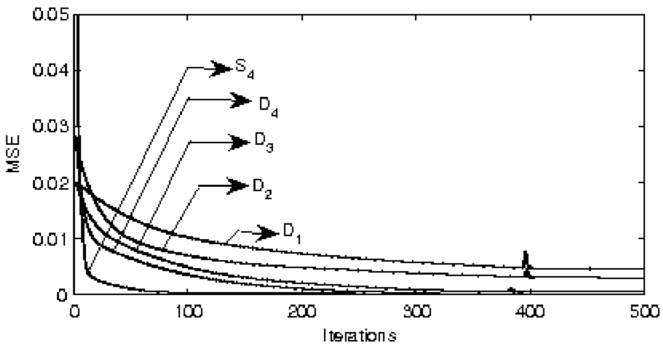

Fig. 10. MSE for forecasting smooth and detail signals using AWNN as a forecasting model.

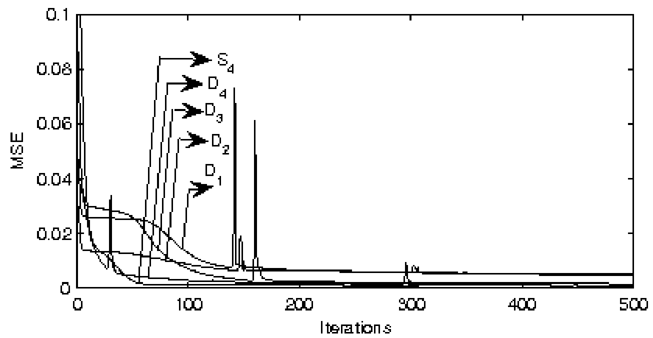

Fig. 11. MSE for forecasting smooth and detail signals using FFNN as a forecasting model.

by FFNN resulting large error values in forecasting. Figs. 10 and 11 shows the training performance in forecasting the smooth and detail signal using AWNN and FFNN forecasting models respectively. Table V gives the MSE values at the end of the 500 iterations for forecasting smooth and detail signals using AWNN and FFNN. It shows that the performance goal is met in forecasting S_4 and D_4 signal using AWNN forecasting method. Also at the end of maximum iterations the training performance of AWNN forecasting technique is much better for other high frequency signal when compared to the training performance of FFNN forecasting method.

TABLE V
TRAINING PERFORMANCE AWNN AND FFNN FORECASTING MODELS

	MSE value after 500 epochs	
	AWNN	FFNN
S_4	0.0001	0.0007
D_4	0.0001	0.001
D_3	0.0006	0.0014
D_2	0.003	0.0051
D_1	0.0046	0.0047

V. CONCLUSION

In this paper, a new method of wind prediction based on MRA and AWNN is presented. MRA decompose wind series into different resolution of smooth and detail signals. Then, AWNNs are used to forecast each decomposed signal to get a combined wind forecasting. Day ahead and week ahead hourly forecasting are carried out. The test results are presented and compared with MRA based FFNN as a forecasting model, AWNN and FFNN with out MRA as a forecasting models. It is observed that MRA based AWNN converges with higher rate and requires few training set as compared other methods. The forecasting results obtained using MRA based AWNN outperformed other methods.

VI. ACKNOWLEDGEMENT

The authors sincerely acknowledge the financial support provided by CPRI Bangalore under RSOP Scheme (Project no. CPRI/EE/20100038) to carry out this work

REFERENCES

[1] IEA website. [online]. Available:http://www.iea.org/publications/index.asp
[2] GWEC website. [online]. Available:http://www.gwec.net/index.php?id=8
[3] M. Tsili and S. Papathanassiou, *A review of grid code technical requirements for wind farms*, IET Renew. Power Gener., 2009, Vol. 3, Iss. 3, pp. 308-332.
[4] L. Landberg, G. Giebel, H.A. Nielsen, T. Nielsen, and H. Madsen, *Short-term prediction-An overview*, Wind Energy (Special Review Issue on Advances in Wind Energy), Vol. 6, no. 3, pp. 273-280, Jan. 2003.
[5] G. Giebel, L. Landberg, G. Kariniotakis, and R. Brownsword, *State-of-the-Art on Methods and Software Tools for Short-Term Prediction of Wind Energy Production*, European Wind Energy Conference, Spain, 2003.
[6] M. Lange and U. Focken, *New Developments in Wind Energy Forecasting*, IEEE Power and Energy Society General Meeting, 2008.
[7] M. Lange and U. Focken, *Physical Approach to Short-Term Wind Power Prediction*, Springer Verlag, Heidelberg, 2005. ISDN 3-540-25662-8
[8] G. Sideratos and N.D. Hatziargyriou, *An Advanced Statistical Method for Wind Power Forecasting*, IEEE Tran. Power Systems, Vol. 22, No. 1, Feb. 2007.
[9] I. Daubechies, *The wavelet transform, time-frequency localization and signal analysis*, IEEE Tran. Information Theory, Vol. 36, 1990, pp. 961-1005.
[10] K. Hunt and G.P. Nason, *Wind speed modelling and short term prediction using wavelets*, WInd Engineering, Vol. 25, No. 1, pp. 55- 61, 2001.

[11] CAO Lei and LI Ran, *Short-Term Wind Speed Forecating Model for Wind Farm Based on Wavelet Decomposition*, International Conf. on Electric Utility Deregulation and Restructing and Power Technologies, Nanjing, China, 2008.
[12] D.L. Faria, R. Castro, C. Philippart, and A. Gusmao, *Wavelet Pre-Filtering in Wind Speed Prediction*, in *Proc. POWERENG*, Lisbon, Portugal, 2009.
[13] A.A. Khan and M. Shahidehpour, *One Day Ahead Wind Speed Forecasting using Wavelets*, IEEE/PES power system Conference and Exposition, 2009.
[14] D. Lei, W. Lijie, L. Xiaozhong, G. Yang, L. Yilli, and W. Zhiwei, *Prediction of Wind Power Generation based on Time Series Wavelet Transform for Large Wind Farm*, in *Proc. PESA*, 2009.
[15] J. Zhang and A. Benveniste, *Wavelet Networks*, IEEE Tran. Neural Networks, Vol. 3, no.3, pp. 889-898, Feb. 2007.
[16] J. Zhang and G.G. Walter, Y. Miao, and W.N.W. Lee, *Wavelet neural networks for function learning*, IEEE Tran. Signal Processing, Vol. 43, no. 6, pp. 1485-1497, Jun. 1995.
[17] S.A. Billings and H.L. Wei, *Anew class of wavelet networks for nonlinear system identification*, IEEE Tran. Neural Networks, Vol. 16, no. 4, pp. 862-874, Jul. 2005.
[18] S.S Iyenger, E.C. Cho, and V.V. Phoha, *Foundation of Wavelet Networks and Applications*, Boca Raton, FL: Chapman & Hall.CRC, 2002, pp. 178-179.
[19] N.M. Pindoriya and S.N. Singh, *An Adaptive Wavelet Neural Network-Based Energy price Forecasting in Electricity Markets*, IEEE Tran. Power Systems, Vol. 23, no. 3, pp. 1423-1432, Aug. 2008.
[20] K.L. Ho, Y.Y. Hsu and C.C. Yang, *Short Term Load Forecasting using a Multilayer Neural Network with an Adaptive Learning Algorithm*, IEEE Tran. Power Systems, Vol. 7, No. 1, pp. 141-149, Feb 1992.
[21] CEERC Website. [Online]. Available: http://www.ceere.org/index.html
[22] S.G. Mallat, *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Tran. Pattern Analysis and Machine Intelligence, Vol. 11, no. 7, pp. 674-693, 1989.
[23] D.B. Parcival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, pp. 57-60, 2000. ISBN 0-521-64068-7
[24] M. Unser, *Texture Classification and Segmentation using Wavelet Frames*, IEEE Tran Image Processing, Vol. 4, pp. 1560-1595, 1995.
[25] S. Haykin, *Neural Netwoks a Comprehensive Foundation*, Prentice Hall International, New Jersey, pp. 161-172, 1999. ISBN 0-13-908385-5
[26] Z.A. Bashir and M.E. El-Hawary, *Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks*, IEEE Tran. Power Systems, Vol. 24, no. 1, pp. 20-27, Feb 2009.

K. Bhaskar (S'10) received the B.E. degree in Electrical Engineering from Andhra University, in 1999 and M.Tech. degree in Power Systems from NIT Surathkal, in 2001. Since then he is working as Assistant Professor in Electrical and Electronics Engineering Department, JNTUH College of Engineering, Hyderabad. He is currently pursuing his Ph.D. degree in the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, India.

S. N. Singh (SM'02) received the M.Tech. and Ph.D. degrees from Indian Institute of Technology Kanpur, Kanpur, India, in 1989 and 1995, respectively.

Presently, he is working as Professor in the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur. His research interests include Power System Restructuring, Power System Optimization and Control, Voltage Security and Stability Analysis, Power System Planning, and ANN application to Power System problems.

Prof. S.N. Singh is a Fellow of IE (India), a Fellow of IETE (India), and senior member IEEE.