2008 IEEE Electrical Power & Energy Conference

A New ANN-Based Methodology for Very Short-Term Wind Speed Prediction Using Markov Chain Approach

S. A. Pourmousavi Kani, *Student Member, IEEE*, and G. H. Riahy

***Abstract*--since year 2000, the increase of the installed wind energy capacity all over the world (mainly in Europe and United States) attracted the attention of electricity companies, wind farm promoters and researchers towards the short term prediction, mainly motivated by the necessity of integration into the grid of an increasing 'unknown' (fluctuating) amount of wind power. Besides, in a deregulated system, the ability to trade efficiently, make the best use of transmission line capability and address concerns with system frequency, accurate very short-term forecasts are motivated more than ever. In this study, very short term wind speed forecasting is developed utilizing Artificial Neural Networks (ANN) in conjunction with Markov chain approach. Artificial neural networks predict short term values and the results are modified according to the long term patterns due to applying Markov chain. For verification purposes, the integrated proposed method is compared with ANN. The results show the effectiveness of the integrated method.**

***Index Terms*—Artificial Neural Networks, Wind Speed, Markov Chain, very short-term prediction, Mean Absolute Percentage Error, Correlation factor.**

I. Introduction

Wind energy is one of world's fastest-growing energy technologies. Both the global and U.S. wind energy (US as the biggest wind energy market) experienced a record year in 2005. According to the figures released by the Global Wind Energy Council (GWEC), in 2005 there was a 43.4% increase in annual additions to the global market and a 564.2% increase in the annual addition to the U.S. wind power generation capacity. 2006 was just one of the record years. In the past 11 years, the global wind energy capacity has increased more than 17 times - from 3.5 GW in 1994 to almost 60 GW by the end of 2005. In the United States, wind power capacity has increased more than 3 times in only 5 years (from 2554 MW in 2000 to 9149 MW by the end of 2005), also. The future prospects of the global wind industry are also very promising even based on a conservative projection that the total global wind power installed could reach 160 GW by 2012 [1].

S. A. Pourmouavi Kani is with the Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran (ali.pourmousavi @gmail.com).

G. H. Riahy is with the Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran (e-mail: gholam@aut.ac.ir).

978-1-4244-2895-3/08/$25.00 ©2008 IEEE

On the other hand, due to increasing wind energy penetration to the grid, it may cause a serious reduction in operating economy and reliability, especially for utilities with large wind power penetration. So, it may increase the cost of wind energy produced, whereas wind industry needs to compete against conventional energy resources. In this situation, wind power prediction permits scheduling the connection or disconnection of wind turbines or conventional generators, thus achieving low spinning reserve and optimal operating cost. Even in this case, accurate prediction of wind power is beneficial for power grid management, matching demand and supply, and stabilization of power market.

Since wind power is a function of wind speed, power forecasts are generally derived through wind speed forecasts. In this area, Very short term prediction is a subclass of the wind power time prediction (in opposition to the wind power spatial prediction). The time scales concerning very short term prediction are in the order of some minutes (for the forecast horizon) and from seconds to minutes (for the time step). Very short term prediction is mainly oriented to the spot market, system management and scheduling of some maintenance tasks, being of interest to system operators, electricity companies and wind farm promoters. Also, control system of a wind turbine, requires prediction times in the range of seconds ahead of our interest, because the major problem with the control of wind turbines is the delays associated with the wind turbine system. These delays affect the response of the system in respect to controller action [2]. In the literatures [2-3], it is shown that the wind speed prediction is an important issue for wind energy conversion systems. Short term wind speed prediction can be used for dynamic control of a wind turbine, due to importance of short-term decisions. The short-term decisions could be classified as connection of a load, changing the pitch of the blades and/or any other control action which involves delays [2].

Since year 2000, more attention to very short term wind and power prediction have been concentrated on the non-linear models [4]. Hoppmann et al [5] developed a system for short term wind speed prediction on the basis of continuous wind measurements at distinct locations of a high-speed railway line. In this method, every 20 minutes the average and standard deviation of wind speed of the last 10 minutes is calculated. Then, the maximum expected wind speed is calculated in four steps. The model has been independently verified by Deutscher Wetterdienst (German Weather Service). Pinson et al [6] used on-line SCADA measurements, as well as numerical weather prediction as input, to predict the power production of wind parks 48 hours ahead. Adaptive fuzzy-neural networks are applied either for short term (1-10 hours) or long term (1-48 hours) forecasting. One year evaluation results are presented on the case study of Ireland.

Wang et al [7] developed an artificial neural network based predictor to predict wind speed. The whole process is divided into two parts: artificial neural networks predict short term value and the results are modified according to the long term pattern. In comparison with linear regression approach, the proposed method shows improvement in short term prediction as well as long term prediction.

In literature [8], comparison of two techniques for wind speed forecasting in the South Coast of the state of Oaxaca, Mexico is done. The Autoregressive Integrated Moving Average (ARIMA) and the Artificial Neural Networks (ANN) methods are applied to a time series conformed by 7 years of wind speed measurements. It concluded that seasonal ARIMA models present a better sensitivity to the adjustment and prediction of the wind speed for that case in particular.

Potter et al [9] described an adaptive neuro-fuzzy inference system (ANFIS) to forecasting very short term wind speed prediction. To provide comparison, a persistence model was developed using the same data. The ANFIS shows a huge reduction in mean absolute percentage error for the same data.

Oztopal [10] developed weighting factors of surrounding stations necessary for the prediction of a pivot station by an ANN technique. Trigonometric point cumulative semivariogram (TPCSV) approach results are compared with the ANN estimations for the same data by considering the correlation coefficient. The ANN approach presented better prediction than TPCSV.

Riahy et al [2] proposed a new method, based on linear prediction for very short term wind speed prediction. The method utilizes the linear prediction method in conjunction with filtering of the wind speed waveform. Real wind speed data based on experimental results is applied for verification. The results show significant performance of the proposed method.

Safavieh et al [11] proposed a new integrated method utilizing Wavelet-based networks and Particle Swarm Optimization (PSO) forecasting very short term wind speed prediction. PSO algorithm is applied for training a Wavelet networks. The proposed approach is compared to multi layer perceptron networks with Back Propagation training algorithm. Results show that the new approach improved Mean Absolute Percentage Error (MAPE) and Maximum error of prediction.

Pourmousavi et al [12] developed a new model for very short term wind speed prediction utilizing ANN, Markov Chain and linear regression. In this method, ANN is used to predict primary values. Then, second-order Markov Chain is applied to calculate transition probability matrix for predicted values. Finally, a linear regression among ANN estimated values and Markov Chain calculated probabilities are used for final prediction. The results in comparison with ANN show slightly decreasing in prediction errors.

In this study, ANN and Markov Chain are applied to predict future values. The whole process is also divided into three parts according to Fig 1: first, very short term values are predicted by artificial neural networks according to time step; second, third-order Markov Chain transition matrix is formed with experimental data. Finally, another artificial neural network is applied among estimated values from first ANN and probabilities of these values calculated by Markov process.

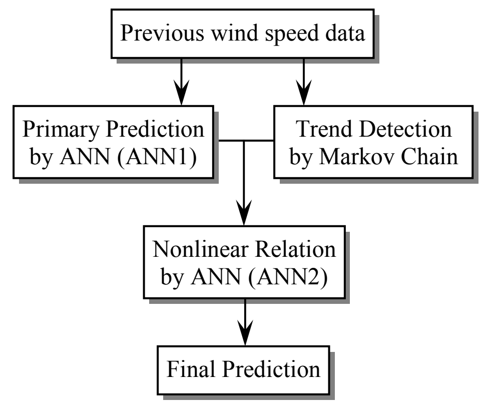

Fig. 1. flowchart of the proposed method.

The rest of the paper is organized as follows. In section 2, the new algorithm is described. The characteristics of the both ANNs are presented in section 3. Section 4 summarizes Markov chain approach used in this study. Section 5 illustrates the results of proposed method and its comparison with ANN method.

II. THE PROPOSED ANN-BASED METHODOLOGY.

The proposed methodology takes both short term and long term patterns into consideration, each of them are processed separately. The whole algorithm is composed of three modules. A schematic flowchart is shown in Fig. 2. In the first module, ANNs handles short term prediction. Too many inputs may cause over estimation due to ANN architecture complication and hence decreasing predictor performance. Therefore, only several past values are utilized as inputs. In the second module, the trend information is provided by Markov Chain, which performs the second-order Markov chain transition matrix using more than 600 most recently data values. Since Markov chain takes account of too many previous data to form transition matrix, the trend information in the long term is considered. The relation between predicted values by ANN and probabilities for these values is found utilizing another ANNs. Since there is no distinct relation between primary predicted values and their probabilities, a non-linear approach is applied using ANN to find this relation accurately. Therefore, final prediction for wind speed data is done by second ANN, when primary prediction by first ANN and probability calculation for these predicted values by Markov chain approach have been done. Fig. 2, shows the whole process clearly in three different sections. Number of preceding data used in each section is reported for better clarification of the process.

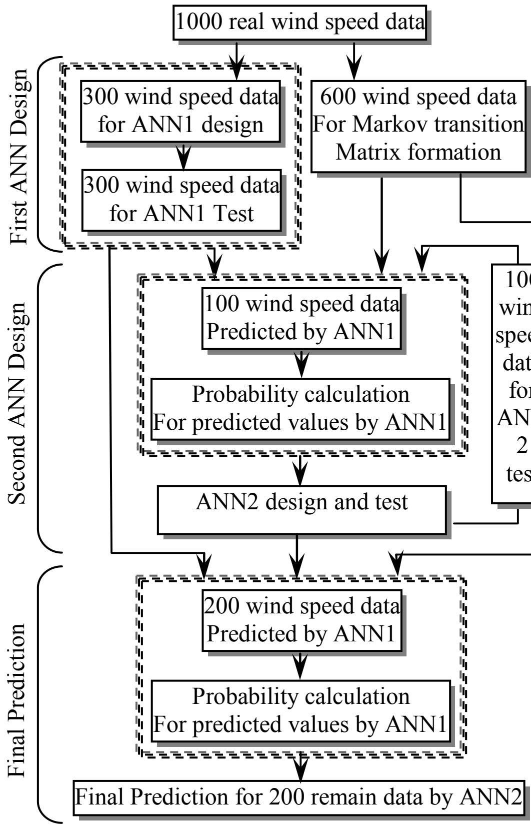

Fig. 2. The whole process in three different sections.

III. A Neural Network with Back Propagation Learning Algorithm.

The MLP neural network consists of simple processing elements (artificial neurons) arranged in layers: an input layer receiving the input variables, one or more hidden layers performing the required non linear input–output mappings, and an output layer producing the network outputs. Each neuron receives weighted inputs from all neurons in the preceding layer. Let W_{ij} be the weight associated with the link from neuron i in one layer to neuron j in the next downstream layer. The neuron sums all weighted inputs and, with reference to a threshold value, uses a non-liner activation function to determine its output. The modeling problem is solved by training on a set of solved examples in the form of input–output records. Training attempts to minimize the error between known and calculated network outputs over all training examples through optimizing the network weights. The mean square error (MSE) criterion is given by:

$$E = \frac{1}{2}\left[\sum_p \sum_p \left| t_{kp} - O_{kp} \right|^2\right] \tag{1}$$

Where t_{kp} and O_{kp} are the true and observed outputs, respectively, for neuron k in the output layer when input vector x_p corresponding to the p_{th} training record is applied to the network. Training with the back propagation algorithm involves iterative application of the training records, determining observed output errors for neurons in the output layer, back propagating these errors to all previous layers, and adjusting the weights so as to minimize the error criterion. The output from neuron j in a given layer (other than the input layer) is calculated as:

$$O_j = f\left(\sum_i W_{ij} . O_i\right) \tag{2}$$

Where i indicates a neuron in the preceding layer and f is the activation function for neuron j. The activation function is often a sigmoid function of the form:

$$f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

With the gradient descent approach to error minimization, weights are changed in proportion to the error gradient, i.e.

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} \tag{4}$$

Where η is a constant that determines the learning rate. To improve convergence characteristics, weight changes are also related to changes introduced in the previous iteration. At the nth iteration, the change in W_{ij} for the link from neuron i to neuron j is given by [11]:

$$\Delta W_{ij}(n) = \varepsilon \delta_j O_i + \alpha \Delta W_{ij}(n-1) \tag{5}$$

Where ε is the learning rate, α is the momentum factor, and δ_j is the error signal for the destination neuron j. When neuron j is in the output layer, δ_j is given by:

$$\delta_j = \left(t_j - O_j\right) O_j . \left(1 - O_j\right) \tag{6}$$

When neuron j is in a hidden layer, δ_j is given by:

$$\delta_i = O_j . \left(1 - O_j\right) \sum_k \delta_k W_{jk} \tag{7}$$

Where k indicates neurons in the succeeding layer next to that containing neuron j.

The learning rate and the momentum factor influence the speed and stability of network training. The process continues until the error criterion on the training set is reduced below a specified limit. To improve generalization on new out-of-sample data, early stopping criteria are often employed where a separate test data set is used to validate the resulting model and training is stopped when error on that data set starts to increase indicating the start of over estimation.

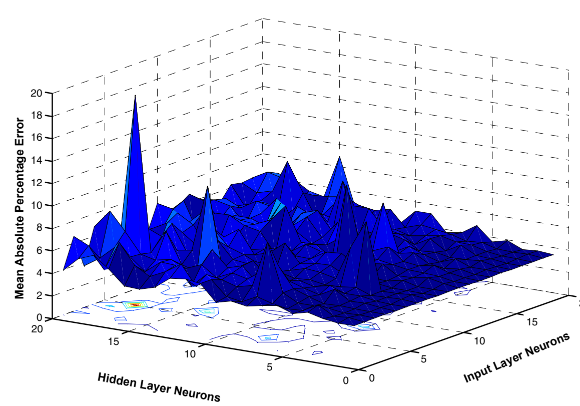

Fig. 3. Sensitivity analysis for number of input and hidden layers neurons in the primary ANN.

Sensitivity analysis is done to find the best structure of the both ANNs. In this study, one step ahead is predicted; therefore, first ANN as well as the ANN used in final prediction, has one neuron in output layer. So, number of output layer neuron never has been changed. Sensitivity analysis for input layer and hidden layer is illustrated in Fig. 3.

The best structure for this network is the following:

- Number of input neurons: 9
- Hidden layer: 2
- Number of output neurons: 1
- Number of training vectors: 30
- Value of the learning rate: (0.01-0.08)

Since the second ANN for final prediction has 2 inputs and 1 output parameters, number of input and hidden layers' neurons should be in the range of inputs. So, the best configuration for this network is the following:

- Number of input neurons: 2
- Hidden layer: 0
- Number of output neurons: 1
- Number of training vectors: 30
- Value of the learning rate: (0.01-0.05)

Hidden layer does not exist here. All analysis has been repeated 10 times to find the best answers.

IV. Markov Chain Approach

All Markov chain models are based on the transitional probability matrices of various time steps. Most often, a first-order Markov chain implies preservation of statistical parameters and especially the first-order autocorrelation coefficient in the synthetic sequences. In order to calculate the Markov chain transitional probabilities, initially the wind speed variation domain is divided into many states. Such a state categorization may be rather arbitrary depending on the purpose, but herein, it is determined according to the average and standard deviation of the available wind speed time series. [13]

For the Markov process, the probability of the given condition in the given moment may be deduced from information about the preceding conditions. A Markov chain represents a system of elements moving from one state to another over time. The order of the chain gives the number of time steps in the past influencing the probability distribution of the present state, which can be greater than one. Many natural processes are considered as Markov processes [13]. In fact, the probability transition matrix is a tool for describing the Markov chains' behavior. Each element of the matrix represents probability of passage from a specific condition to a next state. [14]

Let X(t) be a stochastic process, possessing discrete states space S={1,2,...,K}. In general, for a given sequence of time points $t_1 < t_2 < \ldots < t_{n-1} < t_n$ the conditional probabilities should be [14]:

$$\Pr\{X(t_n)=i_n \mid X(t_1)=i_1,\cdots,X(t_{n-1})=i_{n-1}\} = \Pr\{X(t_n)=i_n \mid X(t_{n-1})=i_{n-1}\} \tag{8}$$

The conditional probabilities $\Pr\{X(t)=j \mid X(s)=i\}=P_{ij}(s,t)$ are called transition probabilities of order *r=t-s* from state *i* to state *j* for all indices $0 \le s < t, with 1 \le i \, and \, j \le k$ .They are denoted as the transition matrix $P_{transition}$. For *k* states, the first order transition matrix *P* has a size of $k \times k$ and takes the form [14]:

$$P_{transition} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,k} \end{bmatrix} \tag{9}$$

The state probabilities at time *t* can be estimated from the relative frequencies of the *k* states. If n_{ij} is the number of transitions from state *i* to state *j* in the sequence of speed data, the maximum likelihood estimates of the transition probabilities is:

$$p_{ij} = n_{ij} \Big/ \sum_j n_{ij} \tag{10}$$

A second order transition probability matrix for *k* state can be shown symbolically as below:

$$P_{transition} = \begin{bmatrix} p_{1.1,1} & p_{1.1,2} & \cdots & p_{1.1,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{1.k,1} & p_{1.k,2} & \cdots & p_{1.k,k} \\ p_{2.1,1} & p_{2.1,2} & \cdots & p_{2.1,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{2.k,1} & p_{2.k,2} & \cdots & p_{2.k,k} \\ p_{3.1,1} & p_{3.1,2} & \cdots & p_{3.1,k} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k.k,1} & p_{k.k,2} & \cdots & p_{k.k,k} \end{bmatrix} \tag{11}$$

In this matrix the probability $p_{j.k,1}$ is the probability of the next wind speed state l if the current wind speed state is *k* and the previous wind speed state were *j*. It has a size of $k_2 \times k$. This is how the probability of making a transition depends on the current state and on the preceding states [13]. The following properties of the transition matrix are valid by definition. Any state probability varies between zero and one. Notationally,

$$0 < p_{j.k,l} < 1.0 \tag{12}$$

On the other hand, the row summation in the transition matrix is Equal to 1 and hence notationally,

$$\sum_{l=1}^{n} p_{j.k,l} = 1.0 \tag{13}$$

According to aforementioned mechanisms, transition matrix is formed by 600 preceding wind speed data. First, the state of each predicted wind speed by primary ANN is calculated for one step ahead. Then, according to transition and probability matrices, the probability of predicted value in the next step is calculated. This is the process which is done for every predicted wind speed data. Predicted values are produced in the previous step by primary ANN.

V. Results and Discussion

In this section, the results of proposed method in comparison with primary ANN are presented. According to

Fig. 2, 200 wind speed data is used for final prediction and proposed method evaluation. Both ANNs are designed using the other sort of data. Also, Markov chain state and transition probability matrices are formed with previous data which never used in this section.

To show the efficiency of the proposed approach (ANN-Markov), the method is compared with ANN trained by Back Propagation. Results indicate that proposed method can predict the short term wind speed better than conventional methods. Prediction is done for one step ahead, 2.5 seconds.

TABLE I
PREDICTED VALUES OF PRICE, NUMBER OF CUSTOMERS AND VALUE ADDED UP TO YEAR 2010.

#I	ANN-Markov			ANN		
	E_1	E_2	C	E_1	E_2	C
1	4.025	12.15	96.07	4.055	12.46	96.12
2	3.755	10.91	97.46	3.761	11.48	97.43
3	3.605	11.64	97.41	3.958	128.46	86.96
4	4.029	13.29	96.91	4.027	15.28	96.92
5	3.517	10.94	95.65	3.546	11.70	95.67
6	3.330	10.26	97.80	3.356	11.09	97.81
7	4.520	19.16	95.97	4.581	20.12	95.97
8	3.684	15.25	97.49	3.706	16.33	97.49
9	3.457	9.48	97.73	3.492	10.67	97.76
10	3.538	10.31	97.67	3.557	10.72	97.69

Table I, represents the ability of proposed method in predicting of wind speed. Considering the mean absolute percentage error (E_1), maximum error of prediction (E_2) and correlation coefficient (C) as criterions of efficiency, the best prediction is occurred using proposed method. Both MAPE and maximum error of prediction criterions are improved slightly in the proposed method. In wind turbine applications, particularly for control purposes, maximum error of wind speed prediction is very important. Also, wind parks with a large amount of electricity production are connected to the grid today, so, even a little improvement in MAPE refer to significant accurate wind power prediction. Prediction for both methods has been repeated 10 times to show the higher performance of the proposed method.

Fig 4, illustrates comparison between proposed method and ANN approach. For better comparison, just 20 predicted wind speed data are illustrated. It is clear that the proposed method followed the real wind speed data better than ANN approach. Also, the errors generated by both methods are depicted in Fig 5. In the range of data set shown in Fig 5, the maximum error of prediction is occurred by ANN approach. Also, as shown in Fig 5, the average of absolute errors produced with ANN approach is reduced applying ANN-Markov approach.

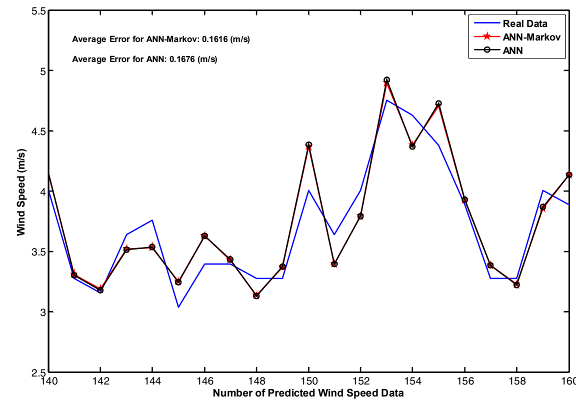

Fig. 4. wind speed prediction by ANN-Markov method and ANN.

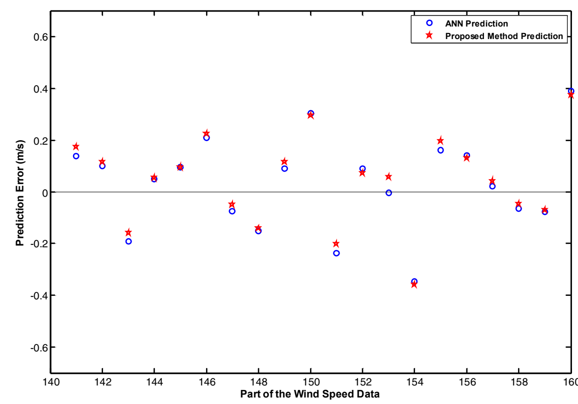

Fig. 5. errors generated by ANN-Markov method and ANN.

VI. CONCLUSION

In this study, a new integrated approach based on ANN method is applied to predict very short term wind speed prediction. A data set with 2.5 seconds resolution is used to evaluate proposed method. Considering some criterions, the proposed method is compared with ANN approach.

In this study, artificial neural networks predict short term value for one step ahead and the results are modified according to the long term patterns due to applying Markov chains.

Results show improvement in mean absolute percentage error in the proposed method. Also, Maximum errors of prediction are improved significantly. It is very important at wind turbine control applications in the range of seconds.

VII. REFERENCES

[1] Global Power Source, Global Wind Energy Council, http://www.gwec.net/

[2] G.H. Riahy, M.Abedi,"Short term wind speed forecasting for wind turbine applications using linear prediction method", Renewable Energy, 3(1),(2007) 35-41.

[3] Riahy G, Freere P, Dynamic wind turbine energy extraction showing the benefits of short term wind prediction. In: Proceedings of the Solar'96 conference. Darwin, Australia, OCT. 1996

[4] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, E. Feitosa, A Review on the young history of the wind power short-term prediction, Renewal and Sustainable Energy Reviews,.12(6) (2008) 1725-1744

[5] U. Hoppmann, S. Koenig, T. Tielkes, G. Matschke, A short-term strong wind prediction model for railway application: design and verification, Wind Engineering and Industrial Aerodynamics 90 (200) 1127-1134.

[6] P. Pinson, G. N. Kariniotakis, Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment, IEEE Bologna PowerTech Conference, June 23-26, Bologna, Italy, (2003).

[7] X. Wang, G. Sideratos, N. Hatziargyriou, L. H. Tsoukalas, Wind speed forecasting for power system operational planning, 8th International Conference on Probabilistic Methods Applied to Power Systems, Iowa State University, Ames, Iowa, September 12-16, 2004.

[8] E. Cadenas, W. Rivera, Wind speed forecasting in the south coast of Oaxaca Mexico, Renewable Energy 32 (2007) 2116-2128.

[9] C. W. Potter, M. Negnevitsky, Very short-term wind forecasting for Tasmanian power generation, IEEE Transactions on Power Systems 21 (2) (2006) 965-972.

[10] A. Oztopal, Artificial neural network approach to spatial estimation of wind velocity data, Energy Conversion and Management 47 (2006) 395-406.

[11]E. Safavieh, A. Jahanbani Ardakani, A. Kashefi Laviani, S. A. Pourmousavi, S. H. Hosseinian, M. Abedi, A new integrated approach for very short-term wind speed prediction using wavelet networks and PSO, Proceeding on International Conference on Power Systems, 12-14th December, 2007 Bangalore, India.

[12] S. A. Pourmousavi Kani, S. M. Mousavi, A. Kashefi Kaviani, G. H. Riahy, A new integrated approach for very short-term wind speed prediction using linear regression among ANN and Markov Chain, Proceeding on International Conference on Power System Analysis, Control and Optimization, 13-15th March 2008, Andhra Pradesh, India.

[13].A. D. Sahin, Z. Sen, First-Order Markov chain approach to wind speed modeling, Wind Engineering and Industrial Aerodynamics 89 (2001) 263-269.

[14] A. Shamshad, M. A. Bawadi, W. M. A. Wan Hussin, T. A. Majid, S. A. M. Sanusi, First and second order Markov chain models for synthetic generation of wind speed time series, Energy 30 (2005) 693-708.