The Effect of Missing Wind Speed Data on Wind Power Estimation

Fatih Onur Hocaoğlu, Mehmet Kurban

Anadolu University, Dept. of Electrical and Electronics Eng., Eskisehir, Turkey
{fohocaoglu,mkurban} @ anadolu.edu.tr

Abstract. In this paper, the effect of possible missing data on wind power estimation is examined. One−month wind speed data obtained from wind and solar observation station which is constructed at Iki Eylul Campus of Anadolu University is used. A closed correlation is found between consecutive wind speed data that are collected for a period of 15 second. A very short time wind speed forecasting model is built by using two−input and one−output Adaptive Neuro Fuzzy Inference System (ANFIS). First, some randomly selected data from whole data are discarded. Second, 10%, 20% and 30% of all data which are randomly selected from a predefined interval(3−6 m/sec) are discarded and discarded data are forecasted. Finally, the data are fitted to Weibull distribution, Weibull distribution parameters are obtained and wind powers are estimated for all cases. The results show that the missing data has a significant effect on wind power estimation and must be taken into account in wind studies. Furthermore, it is concluded that ANFIS is a convenient tool for this kind of prediction.

1 Introduction

Wind is expected to be an important source of electric energy in the future in many regions. Many research groups in different countries have undertaken the development of commercial wind power plants. Wind speed is extremely important for electricity generation from wind turbine. The distribution of wind speeds is important for the design of wind farms, power generators. It is very important for the wind industry to be able to describe the variation of the wind speeds. The effective utilization of wind energy entails a detailed knowledge of the wind speed characteristics at a particular location. The characteristics of wind must be determined by using at least one year wind speed and wind direction data. The missing data should not exceed 10% according to the standards [1]. It is not possible to collect the data without any defect. There are a lot of studies for such numerical weather prediction problems in literature such as autoregressive moving average models (ARMA), Kalman filters [2], bilinear and smooth threshold autoregressive models, artificial intelligence techniques including the use of Multi Layered Perceptrons [3], Radial Basis Functions [4] and Recurrent [5] Neural Networks as well as Adaptive Neuro Fuzzy Systems. In [6] some of these models and different artificial intelligence based approaches are reviewed and compared in terms of Root Mean Square Error(RMSE) criteria for wind speed time series forecasting and better prediction results were obtained by ANFIS in most cases. The aim of this study is to minimize the effect of missing data on wind power estimation for a region which has not been taken into account in wind power estimation studies yet [7–9]. In this scope, one−month wind speed data with almost no defect that is measured and collected for a period of 15 seconds is studied. Some randomly selected data from whole data and from the interval (3−6)m/sec are discarded to represent missing data. According to high prediction capability at such studies an ANFIS model as described at Section 2 is built and the missing data are predicted for all cases. All data are fitted to the Weibull distribution mentioned in Section 3. The results are presented and discussed at Section 4 and 5 ,respectively.

2 ANFIS

ANFIS can incorporate fuzzy if−then rules and also, provide fine−tuning of the membership function according to a desired input−output data pair [10, 11]. The ANFIS structure that is used for this study is given in Fig. 1

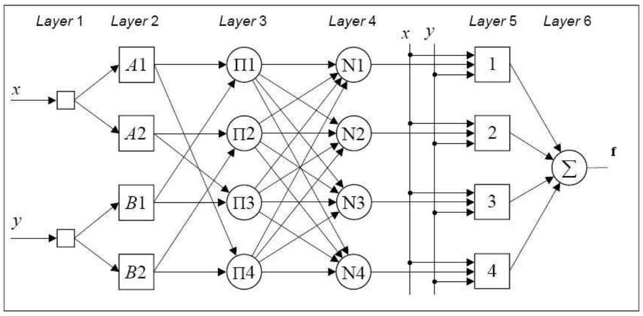

Fig. 1. The ANFIS structure

In Fig. 1; first layer is known as input layer, each neuron in the second layer corresponds to a linguistic label and the output equals the membership function of this linguistic label, each node in layer 3 estimates the firing strength of a rule, which is found from the multiplication of the incoming signals, each node in layer 4 estimates the ratio of the i th rule's firing strength to sum of the firing strength of all rules, j, the output of layer 5 is the product of the previously found relative firing strength of the i th rule, the final layer computes the overall output as the summation of all incoming signals from layer 4. In this study weighted average procedure is used for defuzzify operation, a back−propagation training method is employed to find the optimum value for the parameters of MF(membership functions) and a least squares procedure is employed for the linear parameters on the fuzzy rules, in such a way as to minimize the error between the input and the output pairs, both input1 and input2 are divided into four MFs. The shapes of the MF's before and after training procedure are given in Fig. 2.

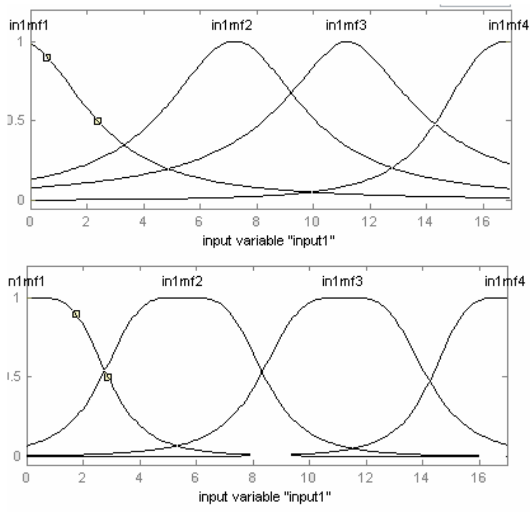

Fig. 2. The MFs of input1 before and after training, respectively

3 Weibull Distributions

To test the accuracy of the prediction operation the Weibull distribution which is the most popular due to its ability to fit most accurately the variety of wind speed data measured at different geographical locations in the world is used. Statistical estimation of unknown parameters from random sample is an important problem that can be solved by many establish methods, such as, the least square method (LSM), the weighted least square method (WLSM), the maximum likelihood method (MLM), the method of moments (MM), the method based on quantiles (QM) and a lot of modifications of these methods. The 2−parameter Weibull probability density function (p.d.f.) is given by equation. 1

$$f_w(v) = (\frac{k}{c})(\frac{v}{c})^{k-1} \exp(-(\frac{v}{c})^k) \tag{1}$$

where $f_w(v)$ is the probability of observing wind speed v, k is shape parameters of Weibull p.d.f., c is scale parameters of Weibull p.d.f. the k values range from 1.5 to 3. for most wind conditions. The cumulative distribution of the Weibull distribution is given as follows:

$$F_w(v) = 1 - \exp(-(\frac{v}{c})^k) \tag{2}$$

Hennessey [12], Justus at all [13], discussed a lot of estimation methods for Weibull distribution. In here, MLM is used. If $x_1, ..., x_n$ is random sample from Weibull distribution, then the log−likelihood function, $L(c)$can be written as in equation. 3 and equation. 4 for MLM

$$L(c,k) = \prod_{i=1}^{n} f_W(x_i, c) \tag{3}$$

$$L(c,k) = \prod_{i=1}^{n} (\frac{k}{c})(\frac{v}{c})^{k-1} \exp(-(\frac{v}{c})^k) \tag{4}$$

Taking the natural logarithm of likelihood function, we obtain equation,5

$$\ln L(c,k) = \sum_{i=1}^{n} \ln(\frac{k}{c}) + (k-1)\ln(\frac{k}{c}) - k\frac{v_i}{c} \tag{5}$$

any value c and k maximizing 5 is called maximum likelihood estimator (MLE), denoted by $\hat{k}_{MLE}$

$$\frac{d \ln L(c)}{dc} = \sum_{i=1}^{n} \frac{-1}{c} + (k-1)(-\frac{1}{c}) + k\frac{v_i}{c^2} = 0 \tag{6}$$

$$\frac{d \ln L(c)}{dk} = \sum_{i=1}^{n} \frac{1}{k} + \ln(\frac{k}{c}) + (k-1)(\frac{1}{k}) - \frac{v_i}{c} = 0 \tag{7}$$

$\hat{k}_{MLE}$, $\hat{c}_{MLE}$ which maximizes 5 can be obtained from the solution of 6 and 7.

4 Results

The correlation coefficient between consecutive wind speeds are obtained to be 0.9857. The distributions of all data are given in Fig.3

It is obtained that, when randomly selected data discarded from whole data the distributions did not change a lot and so power estimations from Weibull distribution did not change much more. From the interval 3−6 (m/s); 10%, 20% and 30% randomly selected data are discarded. The distributions of wind speed after discard are given in Fig. 4

It is obvious from Fig.5 that the distributions of the data are changed significantly. An ANFIS structure is built to predict the missing data. Two data sets, namely, the training and the testing set are employed for the ANFIS. Fig. 5 shows that the constructed ANFIS structure is convenient for the prediction operation.

The former data set is used during the identification process of the fuzzy model while the latter one is used to evaluate the forecast capabilities of the obtained model. The data sets contain patterns formulated from historical data. The prediction values are incorporated to the discarded data and distributions are obtained as given in Table 1

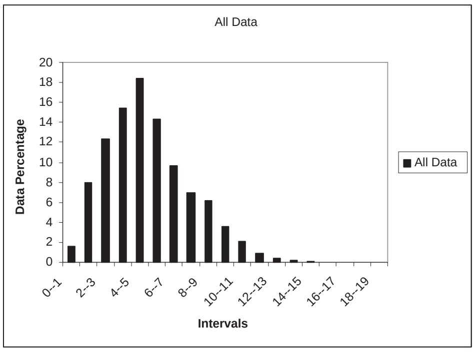

Fig. 3. The distributions of the data

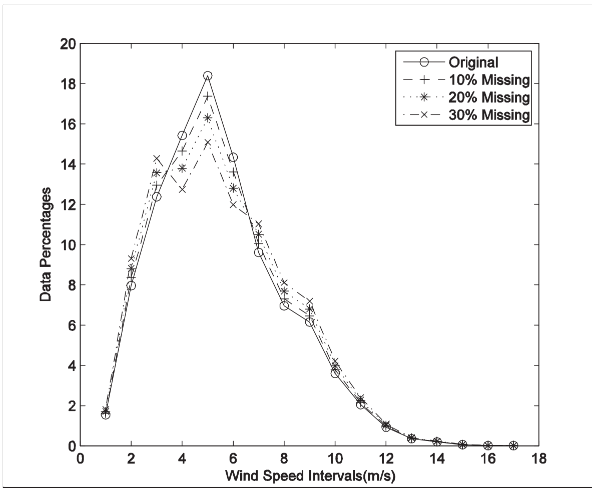

Fig. 4. The distributions after discarding

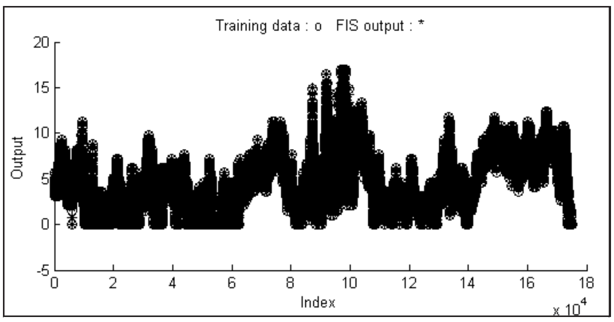

Fig. 5. After training the outputs that are found by ANFIS and actual data outputs.

Table 1. The distributions after the prediction procedure

Wind speed(m/s)	Original	P10	P20	P30
0-1	1.5466	1.5466	1.5467	1.5466
1-2	7.9644	7.9644	7.9645	7.9644
2-3	12.3792	12.4144	12.4406	12.4721
3-4	15.4269	15.4761	15.5339	15.5417
4-5	18.394	18.3764	18.3948	18.4196
5-6	14.3402	14.2984	14.2148	14.1905
6-7	9.6117	9.5911	9.5748	9.5438
7-8	6.9590	6.9559	6.9530	6.9457
8-9	6.1508	6.1495	6.1502	6.1484
9-10	3.6011	3.6011	3.6012	3.6012
10-11	2.0509	2.0509	2.0509	2.0509
11-12	0.9296	0.9296	0.9296	0.9296
12-13	0.3556	0.3556	0.3556	0.3556
13-14	0.2021	0.2021	0.2021	0.2021
14-15	0.0607	0.0607	0.0607	0.0607
15-16	0.0121	0.0121	0.0121	0.0121
16-17	0.0152	0.0152	0.0146	0.0152

Where P10, P20 and P30 represent the distributions after incorporation of the predict of discarded 10%, 20% and 30% data into remaining data respectively. The values of the estimated powers after data prediction and the value of actual power are given on Fig.6 For all cases, the data are fitted to Weibull distribution, parameters are obtained and powers are estimated. The results are given in Table 2

Table 2. Weibull distribution parameters and power estimations from Weibull distribution before and after data predictions

	$\hat{k}$	$\hat{c}$	$\hat{v}_m$	$\hat{\sigma}$	$\hat{v}_{mod}$	v_{max}
Original	5.5242	2.0331	4.8944	2.5209	3.9597	7.7373
10% Missing	5.5531	1.9985	4.9214	2.5742	3.9242	7.8567
After 10% pre.	5.5242	2.0335	4.8944	2.5205	3.9603	7.7365
20% Missing	5.5851	1.963	4.9516	2.6321	3.8857	7.9885
After 20% pre.	5.5095	2.0082	4.8824	2.5428	3.9093	7.7728
30% Missing	5.6211	1.9255	4.986	2.697	3.8422	8.1375
After 30% pre.	5.5237	2.0328	4.894	2.5211	3.9587	7.7373

Finally, estimated wind powers after data prediction are obtained and given in Fig.6

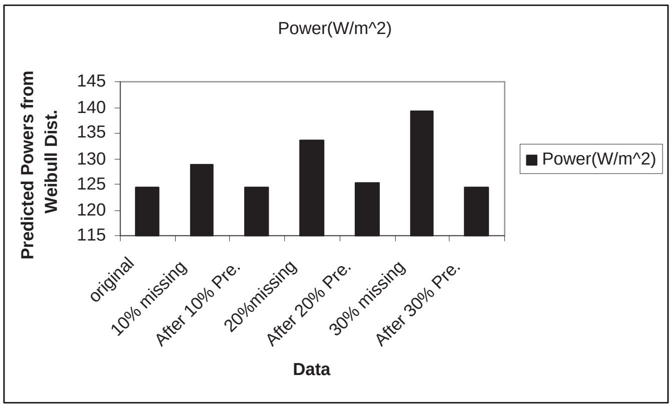

Fig. 6. The power values for actual data, missing data and forecasted data.

5 Discussions and Conclusions

In this paper, the effect of missing data on determining the wind power estimation for a region is examined. It is thought that there can occur any fault in the data collection unit at random short time intervals. In this scope firstly, some randomly selected data are discarded from whole data. It is seen that the distributions of data did not change a lot. Then from the interval (3−6)m/sec randomly selected 10%, 20% and 30% of all data are discarded and it is observed that the distributions are changed significantly. It is obtained that if there are randomly occured faults they can be tolarated but if the faults are occured at specificaly time intervals in which the wind speed regime is in a specific interval that times it is necessary to consider the missing data effect on wind power estimation. In this paper, to consider the effect of the missing data an ANFIS structure is built. The missing data are predicted. The distributions of data are obtained by incorparating the missing data to remaining. Considering the Root Mean Square(RMS) Energy of the distribution of original data to be 38.1035, the RMS Errors between actual distribution and P10, P20, P30 are obtained to be 0.0190, 0.0436 and 0.0541, respectively. Also to test the accuracy of the model, the data are fitted to Weibull distribution and wind power estimations are obtained for all situations. The MLM is used to find the parameters of the Weibull distribution. Actual wind power per square meter is obtained to be 124 where the powers per square meters are obtained for the missing 10%, 20%and 30% data 129, 133 and 139 respectively. After incorporating the missing data to remaining, wind powers per square meter are obtained to be 124.8, 126 and 125.2, respectively. According to litrature, it is tolarable for 10% missing but it is concluded that if the missing data is sourced from a specific interval, the estimated wind power from Weibull distribution is obtained with 4.03% error. In conclussion, it is obvious from the results that, missing data has a significant effect on wind power estimation and must be taken into account in wind power studies and ANFIS is a convernient tool for this kind of a study. Other known models such as latent variable modelling, bayesian estimation, bilinear and smooth threshold autoregressive models, kalman filters, ARMA models etc. may also be used to determine the effect of missing wind speed data together with ANFIS and best model may be established according to prediction results. Such studies can be regarded as a future work of this study.

References

1. Köse R., An evaluation of wind energy potential as a power generation source in Kütahya, Turkey, Energy Conversion and Management, vol. 45, (2004) 1631−1641.
2. Bossanyi E.A., Short-term wind prediction using Kalman filters, Wind Engineering vol.9 (1985) 1−8.
3. Alexiadis M., Dokopoulos P., Sahsamanoglou H., Manousaridis I., Short term forecasting of wind speed and related electrical power, Solar Energy vol.63 (1998) 61−68.
4. Beyer H.G, Denger T., Hausmann J., Hoffmann M., Rujan P., Short term prediction of wind speed and power output of a wind turbine with neural networks, EWEC 94, 5th European Wind Energy Association Conf., (1994) 349−352.
5. Kariniotakis G., Stavrakakis G., Nogaret E., Wind power forecasting using advanced neural network models, IEEE Trans Energy Conversion vol.11 (1996) 762−767.
6. Sfetsos A., A comparison of various forecasting techniques applied to mean hourly wind speed time series, Renewable Energy 21 (2000) 23-35.
7. Ahmed Shata A.S., Hanitsch R., The potential of electricity generation on the east coast of Red Sea in Egypt, Renewable Energy, vol. 31, (2006), 1597−1615.
8. Ozerdem B., Turkeli H.M., Wind energy potential estimation and micrositting on Izmir Institute of Technology Campus, Turkey, Renewable Energy, vol. 30, (2005) 1623−1633.
9. Weisser D., A wind energy analysis of Grenada: an estimation using the Weibull density function, Renewable Energy, vol. 28, (2003), 1803−1812.
10. Jang J. S., ANFIS: Adaptive−Network−based Fuzzy Inference System, IEEE Trans. Systems, Man and Cybernetics vol.23, (1993) 665−684.
11. Jang J. S. and Sun C., Predicting Chaotic Time Series With Fuzzy If−Then Rules, In Proc. IEEE Fuzzy Systems Conf., (1993) 1079−1084.
12. Hennesessey J., Some Aspects of Wind Power Statistics, J. Appl. Meteoral, vol.16, (1977) 119−128.
13. Justus C., Hargraves W. R., Mikhail A., and Graber D., Methods for estimating wind speed frequency distribution, J. Appl. Meteoral, vol.17, (1978) 350−353.