Wind Resource Modelling using the Markov Transition Probability Matrix Method

Therese Chambers
University of Technology, Jamaica, Kingston, Jamaica, tchambers@utech.edu.jm

ABSTRACT

As obtaining measured wind data for the simulation and optimisation of stand alone PV-wind-diesel hybrid energy systems can be difficult in Latin American and Caribbean locations due to incomplete data, the aim of this work is to develop a suitable half hourly wind model that is representative of the resource at the location within the region. As wind resource is stochastic, to develop a suitable model, a stochastic modelling method has to be employed. The method used in this work is the The Markov transition probability matrix (MTPM) method for modelling. The available data was used with the MTPM method to produce four year long models with half hourly time steps. Validation of the models allowed for the selection of the model that best represented the measured data. The model with results most similar to the measured wind speed data can be used in the simulation and optimisation of PV-wind-diesel hybrid energy systems.

Keywords: Wind modelling, Markov transition probability matrix, Wind resource

1. INTRODUCTION

Half hourly wind data was required as an input to the simulation of a small rural remote PV-wind-diesel energy system. As obtaining measured wind data for the location of the energy system was difficult and the available data was incomplete, the aim of this work was to develop a suitable half hourly wind model that is representative of the resource at the location of the proposed rural remote PV-wind-diesel hybrid energy system. The Markov transition probability matrix (MTPM) method for modelling the wind resource was applied. The MTPM method calculates the probability of an event occurring in the next time step based on the event taking place at the present time (Grinstead and Snell, 2006). The wind speed resource models were generated based on the probabilities in the MTPM calculated from measured wind speeds.

To ensure that the wind speeds modelled were representative of the measured data, they were validated by comparison to the measured data. The model with results most similar to the measured wind speed data was selected for use in the simulation and optimisation of the PV-wind-diesel hybrid energy system. The wind resource models were validated to ensure that the models were representative of the measured data. The validation included the comparison of means and standard deviations (Masters et al., 2000, Shamshad et al., 2005, Poggi et al., 2003, Sfetsos, 2000, Amarakoon and Chen, 2001). After comparing the means and standard deviations, Amarakoon *et al* did not validate their wind resource model further as the percentage deviation between the mean and the measured data was less than 14% (Amarakoon and Chen, 2001). Other validation tests include the comparisons of the cumulative distributions and probability density functions (Masters et al., 2000, Shamshad et al., 2005, Poggi et al., 2003). The Kolmogorov-Smirnov statistical test on the cumulative distributions has been used to determine if the underlying distributions of the measured and modelled wind speed data were the same (Masters, 1999). Comparison of the auto-correlation coefficient at lag 1 has been used as a measure of the dependence of a wind resource value on a previous value, at set number of prior time steps (lags) (Shamshad et al., 2005, Ettoumi et al., 2003, Masters, 1999, Poggi et al., 2003, Sfetsos, 2000).

2. WIND SPEEDS AND STATES

The basis of the Markov technique relies on the determination of the MTPM. The events of the MTPM in wind modelling are the wind speeds. For each wind speed, the probability that the speed will transition from one wind speed state to the next was determined. The measured wind speeds vary over a wide range with non integer values therefore making the number of possible events large. The corresponding transition matrix would be large, with each transition probability being very small.

To overcome the problem of a large number of events, states were used to represent a range of wind speeds. Wind speeds that were within a 1m/s range were placed within the corresponding state. The resulting MTPM had twelve states. The probability of transitioning from one state to another was used to create the wind model.

The steps used to calculate the MTPM to create the model, were (Masters, 1999):

1. The transition from each wind speed state to other states was counted for the entire data set using Equation 1. This created a matrix that contained the number of transitions from one wind speed state to others.

$$W_{ij} = \sum_{j=1}^{k} T_{ij} \qquad \boldsymbol{i = 1, 2, \quad k} \qquad (1)$$

   Where:

   $\boldsymbol{W}$ is the number of times the wind speed transitioned from wind speed $\boldsymbol{i}$ to $\boldsymbol{j}$ for the full set of data.

   $\boldsymbol{T_{ij}}$ is a transition from wind speed state $\boldsymbol{i}$ to wind speed state $\boldsymbol{j}$ and is either $\boldsymbol{0}$ or $\boldsymbol{1}$.

2. The next step was to count transitions from each wind speed as shown in Equation 2.

$$C_i = \sum_{i=1}^{k} T_i \qquad (2)$$

   Where:

   $\boldsymbol{C_i}$ is the total number of transitions from wind speed state $\boldsymbol{i}$ to all other wind speeds states.

   $\boldsymbol{T_i}$ is a transition from wind speed state $\boldsymbol{i}$ and is either $\boldsymbol{0}$ or $\boldsymbol{1}$.

3. The transition probabilities of the MTPM were calculated using Equation 3. Each probability is the number of transitions from a wind speed state to another divided by the total number of transitions from the particular wind speed state.

$$P_{ij} = \frac{W_{ij}}{C_i} \qquad (3)$$

   Where:

   $\boldsymbol{P_{ij}}$ is the probability of transitioning from wind speed state $\boldsymbol{i}$ to wind speed state $\boldsymbol{j}$.

3. WIND DATA

The location of the measured wind data is situated at approximately latitude 18.5^{o}N, longitude -77^{o}W, Jamaica. The measurements were recorded at a height of 10m. Data was provided by the Climate Branch of the Meteorological Service of Jamaica.

The monthly average wind speed of the measured data varied from a low of 3.11m/s to a high of 5.84 m/s, as seen in Table 1. During the period June to the end of October, the region is susceptible to tropical storms (wind speeds of 17–33m/s (Kantha, 2006)) and hurricanes (winds speeds in excess of 33m/s) that travel in a west north westerly direction. These high wind speeds exceed wind turbine cut-out speeds and are therefore not useful for generating electricity. In other periods, wind behaviour in the Caribbean is dominated by the North Easterly trade winds. The wind speeds were measured in knots at 2 minute intervals and organised in days and months.

Table 1 Monthly average wind speed data for latitude 18.5°N, longitude -77°W

Month	Average wind speed (m/s)
January	3.95
February	3.51
March	3.75
April	4.25
May	5.27
June	5.84
July	5.36
August	4.2
September	4.59
October	3.53
November	3.11
December	3.71

3.1.1 MARKOV TRANSITION PROBABILITY MATRIX (MTPM)

The calculation of the Markov transition probability matrix was completed using a simple recursive code that created a square 12x12 matrix. Each state was 1m/s wide with the minimum wind speed state (state 1) including wind speeds between 0 and 1m/s and the maximum wind speed state (state 12) including wind speeds greater than or equal to 11m/s. Each transition from a particular wind speed state to another was counted. The total number of transitions from each wind speed state was used to calculate the probabilities in the MTPM.

3.1.2 WIND MODEL

The calculated MTPM was used to generate a sequence of state transitions for the year. The program used to develop the sequence of state transitions for the year is a modification of the simple Markov chain program created by Zirbel (Zirbel, 2006). The size of the MTPM used by Zirbel was increased from a 3x3 to a 12x12 matrix and file data input and output was added. The length of the wind speed state time series was 17520 time steps.

To generate the sequence of state transitions, random numbers were generated and used to determine the wind speed state by summing the probabilities of the row in the MTPM that corresponded to the state of the random number. The probabilities were added until the sum was greater that the random number.

The wind speed time series values, ***W***, were calculated using the means and standard deviations of each of the generated states in the time series (Equation 4).

$$\boldsymbol{W}(\boldsymbol{t}) = \boldsymbol{\mu_S}(\boldsymbol{t}) + \boldsymbol{\sigma_S}(\boldsymbol{t})\boldsymbol{R}(\boldsymbol{t}) \qquad (4)$$

Where:

W is the wind speed (m/s),

R is a random number between 0 and 1,

$\boldsymbol{\mu_S}$ and $\boldsymbol{\sigma_S}$ are the mean and standard deviation of the measured data for each wind speed state ***S***,

t is the time step.

3.1.3 WIND MODEL DEVELOPED USING TWO MARKOV TRANSITION PROBABILITY MATRICES

The measured wind data showed a daily wind speed change. The daily peak wind speed was observed to occur between 1200 and 1600 hrs each day with the daily low occurring between 0100 and 0400 hrs.

A model of the daily behaviour of the wind was produced using two probability matrices, one MTPM for night-time that displayed low wind speeds and one for daytime when the wind speeds were higher. The steps to generate the wind resource model were repeated, using the daytime and night-time probability matrices.

4. RESULTS

The output of the program included:

1. Markov transition probability matrices.
2. The wind resource models;
    a. Two obtained using the MTPM obtained from the measured data.
    b. One obtained using the MTPM's representing the diurnal wind speed behaviour.

5. VALIDATION

5.1.1 MEANS AND STANDARD DEVIATIONS

The values in Table 2 show that the mean wind speeds and standard deviations of the Models 1 and 2 are similar to that of the measured data. The mean and standard deviations of Model 2 is most similar to the measured data while two MTPM model is the least similar.

Table 2 Comparisons of mean and standard deviation of results

Wind speed data	Mean	Standard deviation	Difference in the means (%)
Measured	4.2738	3.0095	-
Model 1	4.0241	2.9275	-5.84
Model 2	4.2692	2.9719	-0.1
2 MTPM model	5.9435	4.0403	39

5.1.2 WIND SPEED FREQUENCY HISTOGRAM AND CUMULATIVE DISTRIBUTION FUNCTION (CDF)

A plot of the wind speed histogram of the models and the measured wind resource shows the frequency at which wind speeds fall into corresponding states (Figure 1). The histograms were generated using the 12 wind speed states that were used to calculate the MTPMs. The general shapes of the wind speed histograms of Models 1 and 2 are similar to that of the measured data. The 2 MTPM model had a high frequency of wind speeds greater than 11m/s. However, the visual comparison of the frequency histograms is not adequate to make a judgement on the validity of the models.

The wind speed cdf is a measure of the probability that wind speed values in the measured or modelled data are less than or equal to a particular wind speed. From visual observation, the cdf of Model 2, Figure 2, is the most similar to the measured data. The Kolmogorov-Smirnov test was used to compare the cdfs of the modelled wind data to that of the measured data.

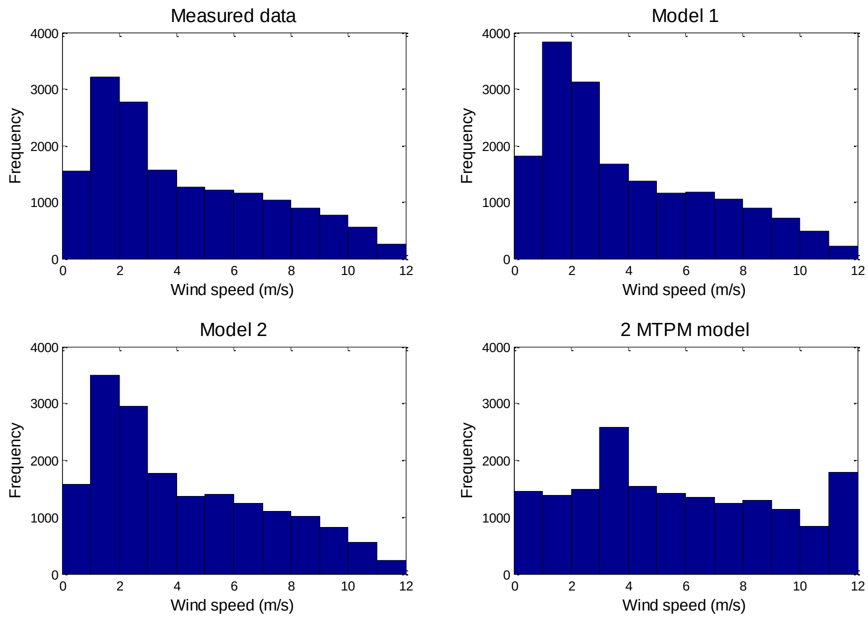

Figure 1 Measured and modelled wind speed frequency histograms

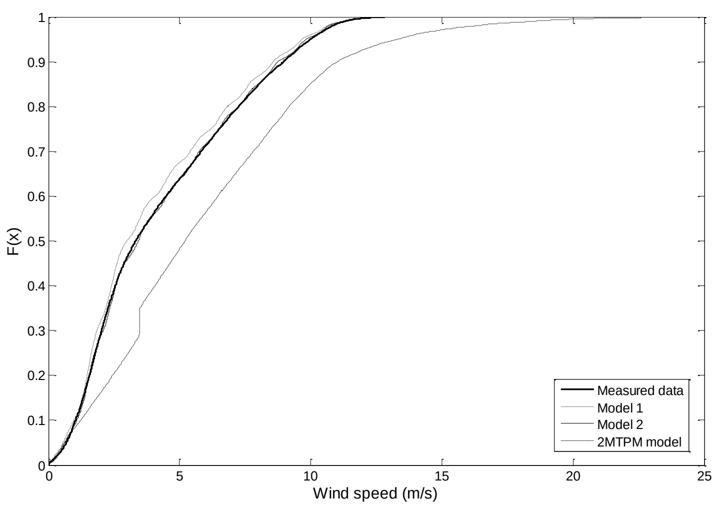

Figure 2 Annual measured and modelled wind cumulative distribution function plots

5.1.3 KOLMOGOROV-SMIRNOV TESTS (K-S)

The K-S test was used to compare the cumulative distribution functions of the measured and modelled wind resource time series to determine the similarity of the distributions. If the maximum difference between the distributions was less than the critical value at the 5% significance level, the hypothesis that the distributions were similar could not be rejected. The critical value, $\boldsymbol{KS_{crit}}$, is determined by Equation 5 (Masters, 1999, Gregory, 1963).

$$KS_{crit} = 1.36\sqrt{\frac{n_1 + n_2}{n_1 \times n_2}} \tag{5}$$

Where:

n_1 is the length of the measured wind speed time series

n_2 is the length of the modelled wind speed time series.

The K-S test was applied to the measured data and each of the models using the ***kstest2*** function for two sample tests in MATLAB. The maximum differences between the distributions of the models and the measured data were higher than the critical value. The negative result of the K-S test may be due to the effect of the missing data in the measured wind speed. The wind model was developed using an MTPM for the year, therefore, the missing data would impact on the transition probabilities of all months.

As the measured data for January was a full set of data, to show that the MTPM method is suitable for modelling wind speeds, the method was used to develop a January model. The K-S test was then used to determine the similarity between the January model and the corresponding measured wind resource. The MTPM for the January wind speed was calculated using the steps outlined in Section 3.1.1. A wind model for January was developed using the method in Section 3.1.1. The results of the K-S on the January wind model are given in Table 3. The critical value, KS_{crit}, for the January measured and modelled wind resource time series was 0.0499, given that the length of both time series was 1488 half hours. The cumulative distributions of the January modelled and measured wind speed is shown in Figure 3. The January data from the models developed using the one year MTPM are included in Figure 3 and Table 3 to illustrate the differences that occurred.

Table 3 Maximum difference between the January measured and modelled wind speed time series cumulative distribution functions

January time series	Mean (m/s)	Maximum difference from measured time series
Measured	3.9531	-
Year MTPM Model 1	3.6911	0.1458
Year MTPM Model 2	4.1588	0.0773
January MTPM	3.9583	0.0410

The results of the K-S test indicate that the cumulative distribution of the January model developed with the January MTPM is similar to the January measured data. The mean wind speed is also the most similar to that of the January measured data.

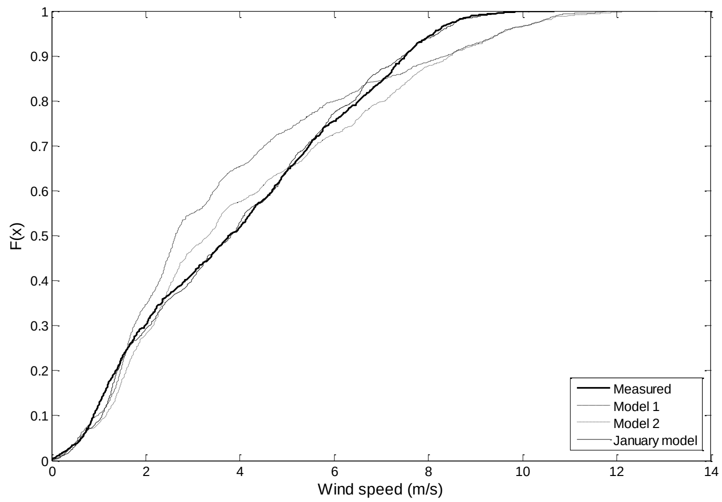

Figure 3 January measured and modelled wind cdf plots

5.1.4 PROBABILITY DENSITY FUNCTION (PDF)

The shape factor of the Weibull pdf of the wind speed data gives an indication of the suitability of the site for wind turbine use to generate electrical energy. The comparison of this parameter for the measured and modelled data was used in the model validation.

The probability distribution function of a Weibull distribution is given by Equation 6 (Masters, 2004).

$$F(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} e^{\left[-\left(\frac{v}{c}\right)^{k}\right]} \tag{6}$$

Where

k is the shape factor

c is the scale factor

v is the wind speed

The values of ***k*** and ***c*** were determined by Equations 7 and 8.

$$k = \left(\frac{\sigma}{\mu}\right)^{-1.086} \tag{7}$$

Where:

μ is the mean of the time series

σ is the standard deviation of the time series

$$c = \frac{\mu}{\Gamma\left(1 + \frac{1}{k}\right)} \tag{8}$$

Where:

Γ is the gamma function.

The shape and scale factors of the measured and modelled data were calculated and are given in Table 4.

Table 4 **Weibull parameters of measured and modelled wind data**

Data	Shape factor k	Scale factor c
Measured	1.43	4.72
Model 1	1.375	4.4
Model 2	1.432	4.7
2 MTPM model	1.437	6.51

The Weibull pdf plots for the measured and modelled data are shown in Figure 4. The shape factors ***k*** show a low wind speed regime and the suitability of the site for wind turbine use could be questioned. The value of ***k*** gives an indication of the variation of the wind speeds about the mean. A high ***k*** value, > 2, indicates a site where the variation of hourly mean wind speed about the annual mean is small (Burton et. al., 2001) while a value less than 2 indicates greater variability.

The low values of ***k*** indicate that, in the measured and modelled wind speed time series, there are wide variations in the half hourly average wind speeds about the annual means. The measured wind speed data includes wind speeds in excess of 16m/s and wind speeds of 0m/s. This indicates wide variations about the annual mean of 4.3m/s.

The scale factor ***c*** indicates the presence of higher wind speeds (Masters, 2004). Figure 4 shows that the 2 MTPM times series model has more time steps with high winds speeds than the other models and the measured data. This is also evident in the histograms in Figure 1 where the frequency of wind speeds in excess of 4m/s is higher than that of the other models and the measured data.

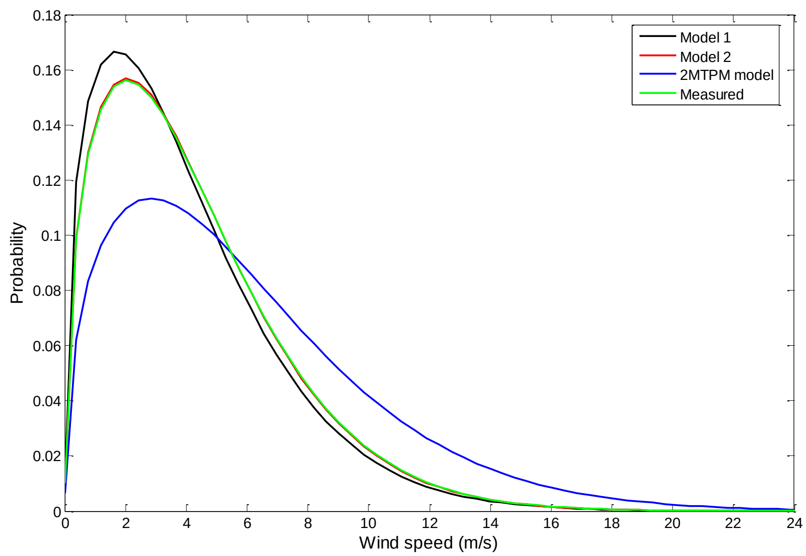

Figure 4 Weibull pdf plots of measured and modelled wind speed time series

For a wind regime where the value of ***k*** is 2, the scale factor ***c*** is directly proportional to the mean of the wind resource time series (Masters, 2004). Locations with a high average wind speed will have a high scale factor. The scale factor of the measured wind data and that of Model 2 are almost the same as seen in Figure 3 and Table 4.

The results in Table 4 and Figure 3 indicate that Model 2 is the most similar to the measured data in regard to scale and shape factor and the diurnal model the most dissimilar.

5.1.5 AUTOCORRELATION

Autocorrelation of the time series is a measure of the relationship between values of the series at different points in time (NIST/SEMATECH, 2006) and falls within the range of -1 to 1. Equation 9 (NIST/SEMATECH, 2006) is the equation for the calculation of the autocorrelation coefficients.

$$A = \frac{\sum_{i=1}^{N-k} (v_i - \mu)(v_{i+k} - \mu)}{\sum_{i=1}^{N} (v_i - \mu)^2} \tag{9}$$

Where:

A is the autocorrelation coefficient at (time) lag ***k***.

i is the time step

N is the maximum number of time steps

μ is the mean of the time series

v is the wind speed at time step ***i***.

When the autocorrelation is used to detect non-randomness, the first (lag 1) autocorrelation is that of interest (NIST/SEMATECH, 2006). The lag 1 autocorrelation indicates the dependence of a value in the time series on the value that occurred in the previous time step. A lag 2 autocorrelation would indicate the dependence on the value two time steps prior.

Equation 9 was applied to the measured and modelled wind resource for lag 1. The results are given in Table 5. From these results, the level of correlation between the values of the modelled wind resource at lag 1 was similar to that of the measured wind resource time series.

Table 5 Wind speed autocorrelation coefficient at lag 1

Model	Autocorrelation coefficient
Measured	0.9423
Model 1	0.9260
Model 2	0.9239
2 MTPM model	0.3062

6. CONCLUSION

The Markov technique for generating discrete time series was used to generate three wind resource models. Two models were generated using a transition probability matrix representing the yearly data. The third model was developed based on two transition probability matrices that reflected daytime and night-time variations.

The means and standard deviations of Models 1 and 2 were similar to those of the measured data.The results of the Kolmogorov-Smirnov test on the cumulative distribution functions show that the underlying distributions of the models and the measured data were not the same. The differences in the distributions may have occurred due to the missing data in the measured wind speed time series. The result of the K-S test on a January model created using a January MTPM calculated from a complete measured January wind resource dataset show that the model was similar to the measured resource. The auto correlation coefficient at lag 1 showed that the wind resource at any one time had a high dependence on the wind resource in the previous time step for the measured data and Models 1 and 2. Modelling with Markov transition probability matrices dictates that the value generated in a time step is dependent on the previous time step and not on those prior. Therefore a high autocorrelation at lag 1 was expected. The lag 1 autocorrelation coefficients of the measured data and Models 1 and 2 are similar.

While the results of the K-S test were negative, the comparison of the Weibull parameters and the autocorrelation coefficient of the measured and modelled wind speeds show that Models 1 and 2 developed using the single year MTPM are similar. The validation results indicate that Model 2 is a suitable representation of the half hourly wind speed. Model 2 was selected to represent the wind resource in a small PV-wind-diesel hybrid energy system simulation and optimisation. For a more accurate wind resource model, monthly MTPMs should be calculated and used to generate the model, as demonstrated with the January wind resource model.

REFERENCES

Amarakoon, A.M.D., Chen, A.A., (2001) "Modelling of wind speed and power across Jamaica". *Jamaica Journal of Science and Technology,*. Vol. 12. p. 48-64.

Burton, T., Sharpe, D., Jenkins, N., and Bossanyi, E., (2001) "Wind Energy Handbook", ed. John Wiley & Sons..

Ettoumi, F.Y., Sauvageot, H., and Adane, A.E., (2003) "Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution." *Renewable Energy,*. Vol. 28. Iss. 11: p. 1787-1802.

Gregory, S., (1963) "Statistical methods and the geographer". 2nd. ed. Longmans.

Grinstead, C.M. and Snell, J.L. (2006) "Introduction to probability." *American Mathematical Society* [Available from: http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/pdf.html]

Kantha, L., (2006) "Time to replace the Saffir-Simpson hurricane scale?" *EOS Transactions American Geophysical Union,*. Vol. 87. Iss. 1.

Masters, C.L., (1999) "The integration of wind farms on to distribution systems, in Department of Electrical Engineering and Electronics." *Thesis,*. University of Manchester Institute of Science and Technology.

Masters, C.L., Mutale, J., Strbac, G., Curcic, S., and Jenkins, N., (2000) "Statistical evaluation of voltages in distribution systems with embedded wind generation." *IEE Proceedings Generation Transmission and Distribution,*. Vol. 147. Iss. 4: p. 207-212.

Masters, G.M., (2004) "Renewable and efficient electric power systems." John Wiley & Sons.

NIST/SEMATECH. *NIST/SEMATECH (*2006) "e-Handbook of Statistical Methods". [Accessed. Available from: http://www.itl.nist.gov/div898/handbook/.

Poggi, P., Muselli, M., Notton,G., Cristofari, C., Louche A., (2003) "Forecasting and simulating wind speed in Corsica by using an autoregressive model." *Energy Conversion and Management,*. Vol. 44. Iss. 20: p. 3177-3196.

Sahin, A.D. and Sen, Z., (2001) "First-order Markov chain approach to wind speed modelling." *Journal of Wind Engineering and Industrial Aerodynamics,*. Vol. 89. Iss. 3-4: p. 263-269.

Sfetsos, A., (2000) "A comparison of various forecasting techniques applied to mean hourly wind speed time series." *Renewable Energy,*. Vol. 21. Iss. 1: p. 23-35.

Shamshad, A., Bawadi, M.A., Wan Hussin, W.M., Majid, T.A., and Sanusi, S.A., (2005) "First and second order Markov chain models for synthetic generation of wind speed time series." *Renewable Energy,*. Vol. 30. Iss. 5: p. 693-708.

Zirbel, Z., (2006) "MATLAB programs for applied probability: Simulation of Markov chains.", Bowling Green State University.

Authorization and Disclaimer