EUROPEAN ACADEMIC RESEARCH, VOL. I, ISSUE 7/ OCTOBER 2013
ISSN 2286-4822, www.euacademic.org
IMPACT FACTOR: 0.485 (GIF)
DRJI VALUE: 5.9 (B+)

Forecasting Average Daily Wind Speed of Hyderabad (Sindh): an ARIMA Modelling Approach

MUZAFFAR UDDIN QAZI
Department of Physics
Federal Urdu University of Arts, Science and Technology
Karachi, Pakistan
ZOHAIB AZIZ
Department of Statistics
Federal Urdu University of Arts, Science and Technology
Karachi, Pakistan
SAIF UR REHMAN
Department of Physics
Federal Urdu University of Arts, Science and Technology
Karachi, Pakistan

Abstract:

For wind energy forecasting a number of techniques are available. Some of the latest physical based model techniques are computational learning system such as artificial neural networks (ANN), particle swam optimization (PSO) and vector support machines (VSM). On the other hand, statistical models are easy to model and cheaper to develop, compared to other models. In statistical models, we use past history of wind data to forecast the present over next few days. In this research article, autoregressive integrated moving average (ARIMA) model is used to forecast wind speed for next few days. ARIMA is a well-known time series statistical model. This model has shown good forecasting results. Wind energy forecasting plays a key role to address the challenges in wind energy as the wind energy is used for electric power generation.

Key words: ARIMA, Hyderabad, Sindh, wind speed forecasting

I. Introduction

The demand of energy in the developing countries has risen more over the past three decades and it will be increased rapidly in the future. In Pakistan shortage of electricity could cause major economic, political and social problems. For many developing countries like Pakistan, much of the additional energy needed is supplied by imported oil and this further burdens these countries already saddled with high oil import bills.

The increase in the use of fossil fuel for the electricity production in the developing countries like Pakistan increases the local and regional air pollution as well as atmospheric concentrations of green house gases. Pakistan is adversely affected due to climate change in the region especially between 2010 and 2011.

An economically and environmentally sound approach to energy development offers potentially large benefits for countries like Pakistan.

Pakistan is facing a serious challenge of energy deficit and wind energy resources can play an important role in bridging the deficit. Renewable energy can take electricity to remote rural areas where electricity transmission becomes too expensive.

II. Literature Review

Wind intermittency is the biggest task to implementing wind energy as a reliable autonomous source of electrical energy. Therefore on large scale wind penetration requires answers to a lot of problem, such as real time grid operations, competitive market designs, service requirements and costs, quality of power, capacity of transmission system, stability and reliability of power system and optimal reductions in green house gas emissions of entire power system [1][2]. Wind forecasting is an efficient tool to overcome many of these problems. For example, a correct forecast can help to develop well-functioning hour or day ahead markets [3]. Improving wind power forecast has significant economic and technical advantages. Fabbri et al. (2005) proposed a statistical method for calculating energy expenses associated with forecast errors for wind energy generation [4]. Milligan et al. (2004) described several statistical forecasting models known as Autoregressive Moving Average (ARMA) models to predict wind speed and wind power output in hour-ahead market [5]. Mirande et al. (2006) discussed short term forecasting of 1 hour-ahead using time scale in which authors present Bayesian approach to develop an Autoregressive model of the sixth order [6]. Milligan et al. (2003) examined ARMA with generation data for six hours-ahead predictions [7]. Sideratos et al. (2007) proposed a hybrid approach for 1 hour to 48 hours ahead forecast [8].

III. Case Study: Hyderabad (Sindh)

The study area, Hyderabad (Sindh), is the second largest city of the Sindh province. It is situated between 25.23^0N latitudes and 68.29^0E longitudes with an elevation of 13 meters (43 feet). According to 2010 census the population of the city is 1578,367. Hyderabad has a hot dessert climate with warm conditions round the year. During the period of April to June, winds that blow usually bring along clouds of dust, while the breeze that flows at night is more pleasant.

IV. Data

The sample data consist of average daily wind speed of Hyderabad (Sindh) from January 2008 to December 2011. The data are obtained from metrology department, Karachi (Sindh). Wind speed data have been measured at 0000 UTC (Knots). Wind speed is a ratio scale time series variable. For forecasting purpose, we shall hold last week of December 2011 such that comparison of observed and forecasted value could be possible.

V. Methodology

We examine the wind speed series through popular Box-Jenkins methodology [9], but technically known as ARIMA methodology. It is the most widely used methodology for the analysis of time series data. We shall use univariate ARIMA models, because we have a single series that is being analyzed through its past values. It is the fundamental theme of ARIMA methodology. This methodology is based on four silent steps.

1. Differencing the series to make a series stationary.
2. Identification of the tentative model.
3. Model estimation.
4. Diagnostic checking for the model adequacy.
5. Forecasting.

Consider the figure below, representing the flow chart of Box-Jenkins methodology [10].

VI. ARIMA Model

Consider the general form of ARIMA process. If d is a nonnegative integer represents the differencing order, then $\{X_t\}$ is an ARIMA (*p, d, q*) process if $Y_t = (1-L)^d X_t$ is a causal ARMA (*p, q)* process. So, $\{X_t\}$ satisfies a difference of the form

$$\phi_p(L)(1-L)^d X_t = \theta_q(L) Z_t \qquad (1)$$

Where, $\{Z_t\} \sim WN(0, \sigma^2)$, L is the back-shift operator and the stationary AR operator

$$\phi_p(L) = \left(1 - \phi_1 L - \cdots - \phi_p L^p\right) \qquad (2)$$

and the invertible MA operator

$$\theta_p(L) = \left(1 - \theta_1 L - \cdots - \theta_p L^p\right) \qquad (3)$$

share no common factors [11].

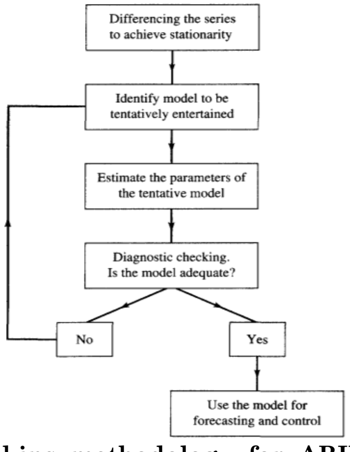

Fig. a: The Box-Jenkins methodology for ARIMA models. Source: Maddala (2001).

VII. Power Transformation

Box and Cox (1964) introduced the power transformation that is used for stabilizing the variance. It is expressed as

$$T(X_t) = \frac{X_t^{\lambda}-1}{\lambda} \qquad (4)$$

Consider some commonly used values of λ and their associated transformation [12].

Values of λ	Transformation
-1.0	$\frac{1}{X_t}$
-0.5	$\frac{1}{\sqrt{X_t}}$
0.0	$\ln X_t$
0.5	$\sqrt{X_t}$
1.0	X_t

Source: Wei (2006)

VIII. Information Criteria

We have two popular criteria for model selection Akaike's (1974) information criteria (AIC) and the Schwarz's (1978) Bayesian information criteria (SBIC). These statistics are symbolically written as

$$AIC = ln(\hat{\sigma}^2) + \frac{2K}{n} \qquad (5)$$

$$SBIC = ln(\hat{\sigma}^2) + \frac{K}{n} ln(n) \qquad (6)$$

Where K=p+q-1 and n=sample size.

IX. Empirical Analysis

If we look at the Fig. 1, it represents the time series plot of average daily wind speed (0000-UTC) of Hyderabad during January 2008 to December 2011. The variance of series is not looking at stabilize mode, but we have no authenticity to give comments on variance earlier. After applying the Box-Cox transformation technique the general view regarding the variance may clear. By taking bird's eye view on wind series plot it may suggest that the series is behaving like stochastic stationary, and later it may be confirmed by unit root testing. Further technical treatment will be followed step by step through Box-Jenkins methodology.

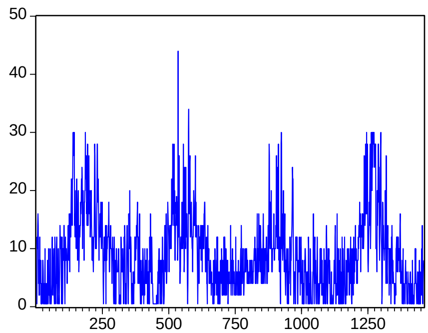

Fig. 1: Average daily wind speed [at 0000 UTC (Knots)] of Hyderabad (Sindh) from January 2008 to December 2011.

The first step of Box-Jenkins methodology is to stabilize the variance of the series. For this purpose, we need to apply Box-Cox/power transformation. Basically, power transformation is based on a particular value of λ that stabilizes the variance. But, the searching of the value of λ requires more practicing. The issue has been resolved through TSA package [13] in R [14] introduced by Chan (2012) [15].

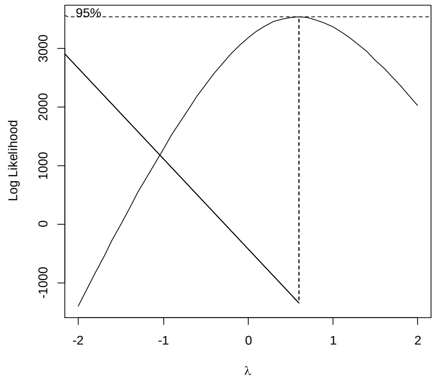

Fig. 2: Plot for determining the maximum likelihood estimate of λ required for Box-Cox transformation.

If we look at the Fig.2, that is showing the maximum likelihood estimate of λ required for stabilizing the variance of the series. If we see the vertical dotted line in the Fig. 2, it concluded that the maximum likelihood value of λ may be 0.54.

Of course, this particular value cannot exactly be determined from the Fig. 2. It has been found by the particular command of R for maximum likelihood estimate that may be inappropriate to code here.

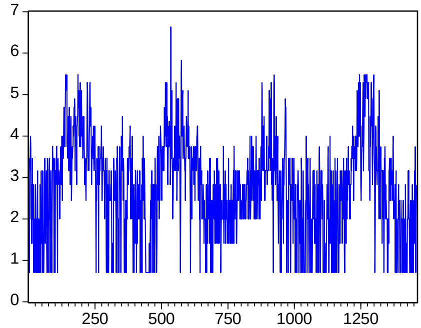

Fig. 3: Variance-Stabilized Average daily wind speed [at 0000 UTC (Knots)] of Hyderabad (Sindh) from January 2008 to December 2011.

Basically, we have two decimal place values. But for mathematical suitability we require the value to one decimal place. After round off the value may be 0.5. The amazing situation is that we get the value 0.5 that provides the well-known square root transformation. So, we have to take square root of given series so that we could apply the remaining steps on stabilized series. Fig. 3 is showing the variance-stabilized wind speed series.

Now, we have to find the integrated order of wind series, which is the second most prominent step of Box-Jenkins/ARIMA methodology. We need to use some tests for stationarity through which we could determine the integrated order of series. There are some well-known tests to check the stationarity namely ADF, PP and KPSS tests. Through these tests we can suggest the wind series stationary at level, on first difference or second difference.

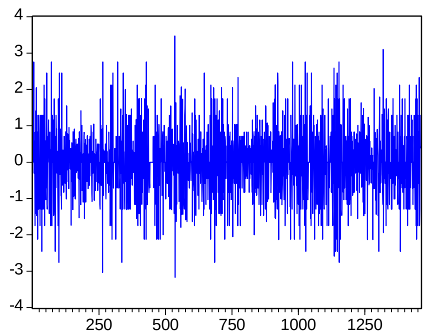

Fig. 4: First difference average daily wind speed [at 0000 UTC (Knots)] of Hyderabad (Sindh) from January 2008 to December 2011.

Table 1 reports Augmented Dickey-Fuller (ADF)[1], Phillips-Perron (PP)[2] and Kwaitkowski Phillips-Schmidt-shin (KPSS)[3] tests for stationarity of wind speed series of Hyderabad (Sindh). Further, it shows the variances of series at level, on 1st-difference and 2nd-difference. ADF, PP and KPSS tell us that series is stationary at level with 1 percent significance. If

[1] Null hypothesis is related to non-stationarity of series.
[2] It also has similar null hypothesis to ADF
[3] Null hypothesis is related to stationarity of series.

we go towards the 1st-difference column, all tests are proving the series is stationary at the 1 percent level. Fig. 4 is confirming the high degree of stationarity as compared to level series. Finally, if we look at 2nd-difference column, it also explains that the series is stationary at 1 percent level.

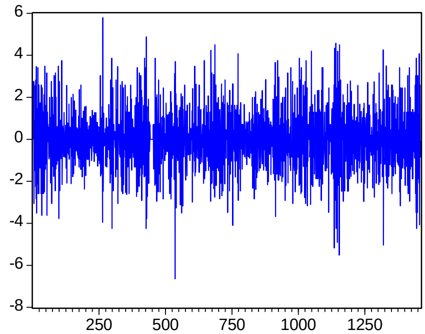

Fig. 5: Second difference average daily wind speed [at 0000 UTC (Knots)] of Hyderabad (Sindh) from January 2008 to December 2011.

TABLE 1
TESTS FOR STATIONARITY

	Level	1st-Difference	2nd-Diffrence
ADF Test	-6.3554*	-21.44185*	-16.3725*
PP Test	-24.7357*	-167.6100*	-640.1929*
KPSS Test	0.0777*	0.0653*	0.0179*
Trend & Intercept	✓		
Intercept		✓	
None			✓
Variance	1.374	0.944	2.457

*means significant at the 1 percent level

Fig. 5 is also confirming the high degree of stationarity among level and first difference series. However, we have sufficient evidence of variance inflate problem in second difference that is discussed in the last lines of current paragraph. We used trend and intercept, intercept and none option in each test at level, on 1st-difference and 2nd difference respectively. But the question is raised here if stationarity is achieved at level so why we are checking the stationarity on

1st-difference and 2nd difference. It may be the cause of over differencing. The answer may provide the variances of each situation. The difference of series, however, can be taken to get more extent of stationarity. Yet we have to keep in mind whether we are going towards the over differencing problem or not. The 1st-difference of series is providing the minimum variance among different situations. So, we shall use the first difference wind speed series. In short, we can conclude the wind speed series is integrated of order 1 i.e. I(1).

After determining the integrated order, difference required to make a series stationary, of wind series, we shall go to detect the orders of autoregressive and moving average that is the third and maybe more practicing step of used methodology. We shall divide this step into two parts. Part I may describe the orders of autoregressive and moving average polynomials instantly through extended autocorrelation function (EACF). Part II shall describe the tentative model searching technique through some information criteria namely AIC and BIC.

The ACF and PACF is not a prominent tool for determining the order of ARMA model. Tsay and Taio (1984) [11] propose EACF. The manual calculations of EACF are the tedious work. Chan (2012) provides a statistical package TSA in R-language. This package provides the simplified table of EACF. Instead of involving in actual calculation of EACF, we need to make a simplified EACF table for the series. The Table 2 shows the simplified sample EACF table for the daily wind speed of Hyderabad (Sindh) from January 2008 to December 2011. It shows a triangular pattern of O with its upper left vertex at the order (p,q)=(1,1). It may describe that the orders of both autoregressive and moving average polynomials are 1. Hence, we may suggest the ARIMA (1,1,1) for the average daily wind speed of Hyderabad (Sindh) from January 2008 to December 2011. But, latter it will be confirmed by using tentative model searching technique on the basis of AIC and BIC.

TABLE 2
SIMPLIFIED SAMPLE EACF TABLE FOR THE DAILY WIND SPEED OF HYDERABAD FROM JANUARY 2008 TO DECEMBER 2011

AR Order p	MA Order q 0	1	2	3	4	5	6	7	8	9
0	X	X	O	O	O	O	O	O	O	O
1	X	O	O	O	O	O	O	O	O	O
2	X	X	O	O	O	O	O	O	O	O
3	X	X	X	O	O	X	O	O	O	O
4	X	X	X	X	O	O	O	O	O	O
5	X	X	X	X	X	O	O	O	O	O
6	X	X	X	X	X	X	O	X	O	O
7	X	X	X	X	X	X	X	X	O	O

Note: X denotes significant (non zero) autocorrelation at the 1 percent level.
O denotes insignificant (zero) autocorrelation at the 1 percent level.

We envision Table 3 that expresses different ARIMA models with their autoregressive and moving average parameters' estimates, their corresponding t-statistics within the

TABLE 3
IDENTIFICATION OF TENTATIVE MODELS ON THE BASIS OF AIC & BIC

	AR-ESTIMATES			MA-ESTIMATES			AIC	BIC
ARIMA	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3		
(1,1,0)	-0.3012* [-12.0561]	-	-	-	-	-	2.6867	2.6904
(0,1,1)	-	-	-	-0.7105* [-38.7321]	-	-	2.5754	2.5791
(1,1,1)	0.3694* [12.5886]	-	-	12.5886* [-72.9120]	-	-	2.4838	2.4910
(1,1,2)	0.3412* [4.3539]	-	-	-0.8855* [-10.8409]	-0.0278 [-0.4045]	-	2.4841	2.4959
(2,1,1)	0.3682* [12.4168]	-0.0190 [-0.6668]	-	-0.9125* [-64.6413]	-	-	2.4841	2.4949
(2,1,2)	-0.5612* [-10.4182]	-10.4182* [9.8550]	-	0.0212 [0.4605]	-0.8608* [-20.3434]	-	2.4793	2.4938
(2,1,3)	0.5123 [0.7720]	-0.1058 [-0.4568]	-	-1.0574 [-1.5917]	0.1718 [0.2921]	-0.0335 [-0.4575]	2.4866	2.5047
(3,1,2)	-0.5601* [-11.6447]	0.3275* [8.1151]	-0.0131 [-0.4610]	0.0236 [0.5858]	-0.8574* [-23.0524]	-	2.4796	2.4978
(3,1,3)	-0.2867 [-0.9448]	-0.2320 [-0.8322]	0.1533 [1.2868]	-0.2564 [-0.8463]	-0.1367 [-0.4280]	-0.425500 [-1.5224]	2.4863	2.5081

Note: The numbers in brackets are t-statistics.
*means significant at the 1 percent level.

brackets, AIC and BIC. We consider without intercepting ARIMA models for wind speed data of Hyderabad due to highly insignificance of intercepted term. We checked nine tentative models. We shall select the model which has minimum AIC and/or BIC. AIC is minimum for ARIMA (2,1,2) but BIC is minimum for ARIMA (1,1,1). We select ARIMA (1,1,1) due to parsimonious assumption of the time series models. It may be helpful here, because we should avoid large number of parameters in the time series models. It might be dangerous for our forecasting results reliability. We see in the Table 3 the estimates ϕ_1 and θ_1 are significant at 1 percent level for ARIMA (1,1,1), while, there is no significant estimate of θ_1 for ARIMA (2,1,2) at 1, 5 even 10 percent level, whereas remaining estimates of ARIMA (2,1,2) are significant at 1 percent level. Remarkable information is that this searching technique of tentative ARIMA models also suggests ARIMA (1,1,1) on the basis of minimum BIC with consideration of parsimonious assumption. It confirms the detection of ARIMA order through EACF technique. If we want to search the rival or second best model of ARIMA (1,1,1) for the considered series then it may be ARIMA (2,1,2). But, it is not related to our objective while we have one best model.

As far as model adequacy is concerned, first, we want to get white noise residuals. In other words, if autocorrelations are individually and jointly insignificant these may suggest that residuals are white noise. Second, we require normality of residuals that may be checked through Jarque-Bera test.

Second column of Table 4 explores all given autocorrelations that are not individually significantly different from zero except at lag 4, 6 and 11. Autocorrelation at lag 4, however, has no high extent of departure to the upper class interval i.e. 0.0153. On the other hand, last column of Table 4 reveals that all given autocorrelations are jointly insignificant at the 5 percent level. Jarque-Bera statistics is significant at 5 percent level of significance that provides the evidence of normality regarding the residuals. Fig. 6 shows the histogram of residuals that is seemed normal. These results provide the evidence of white noise residuals. Long and short, ARIMA (1,1,1) is adequate for the given series.

TABLE 4
DERTERMINATION OF MODEL ADEQUACY

Lags	**Autocorrelation**[4]	**Q-statistic**[5]
1	0.015	0.3464***
2	-0.011	0.5178***
3	-0.036**	2.4132***
4	-0.005	2.4545***
5	0.006	2.5144***
6	0.018**	3.0124***
7	0.012	3.2317***
8	-0.001	3.2323***
9	0.000	3.7278***
10	-0.014	3.7281***
11	-0.048**	4.0354***
12	-0.006	7.4792***
13	0.050	7.5240***

SKEWNESS	-0.150861
KURTOSIS	3.040108
JARQUE - BERA	5.628144****

**means autocorrelation is individually significantly different from zero at the 5 percent level.

***means p-value of Q-statistic is less, than 5 percent level of significance.

****means errors are significantly normally distributed at the 5 percent level.

---

[4] If autocorrelations are within the interval of $\pm 1.96\frac{1}{\sqrt{n}}$ (=$\pm$ 0.015296) they are not individually significantly different from zero at the 5 percent level. Where n=number of observations.

[5] If the P-value of Q-statistic is less than equal to 0.05 then autocorrelations are not jointly significant different from zero at the 5 percent level.

Eventually, the last and objective step of Box-Jenkins methodology is forecasting through suggested model. That was our ultimate goal. Table 5 explores the forecasts values of last week, December 2011 from ARIMA (1,1,1) against the observed or hold values of studied series. To check the accuracy of forecasts, we report three measures of forecast error Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) into the Table 5. All these forecasts measures from ARIMA (1,1,1) are minimum among the rival models' measures. We did not report those measures due to inconvenience of their models.

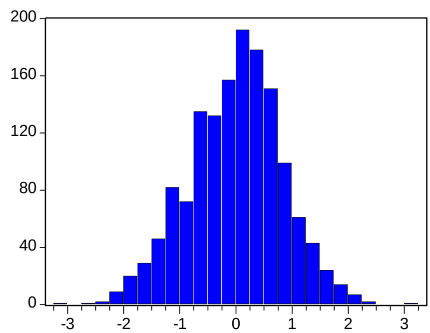

Fig. 6: Histogram of residuals from the fitted ARIMA (1,1,1).

TABLE 5

FORECASTING OF SUGGESTED MODEL

FORECAST TIME	FORECATS
	ARIMA(1,1,1)
December 25, 2011	5.5820
December 26, 2011	2.6855
December 27, 2011	7.1385

December 28, 2011		3.8930
December 29, 2011		1.9353
December 30, 2011		3.6822
December 31, 2011		4.5313
FORECAST ERRORS	**RMSE**	**4.5553**
	MAE	**3.4674**
	MAPE	**126.4284**

X. Conclusion

The average daily wind speed of Hyderabad (Sindh) is forecasted and compared using ARIMA model. This model is used as a prediction tool. The transformation, standardization, estimation and diagnostic checking processes are analyzed. According to the results shown in tables 1 – 5, we check the accuracy of forecast error report, three measures of forecast errors root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The o values of these forecast errors for RMSE, MAE and MAPE are 4.5553, 3.4674 and 126.4284 respectively. ARIMA model show better forecasting results.

BIBLIOGRAPHY

Box, G.E.P., G.W. Jenkins, and G.C. Reinsel. 2007. *Time Series Analysis: Forecasting and Control*. John Wiley. [9]

Brockwell, P.J., and R.A. Davis. 2002. *Introduction to Time Series and Forecasting*, 2nd ed. New York: Springer. 180-181. [11]

Chan, K.S., J. Cryer. 2008. *Time Series Analysis with Applications in R*, 2nd ed. Springer. [15]

Fabbri A., T.G.S. Roman, J.R. Abbad, V.H.M. an Quezada. 2005. "Assessment of the Cost Associated with Wind Generation Prediction Errors in a Liberalized Electricity Market." *IEEE Trans Power System* 20(3):1440-1446. [4]

Lund, H. 2005. "Large Scale Integration of Wind Power into Different Energy Systems." *Energy* 30(13): 2402-2412. [2]

Maddala, G.S. 2001. *Introduction to Econometrics*, 3rd ed. New York: Macmillan. 200-201. [10]

Milligan, M., M. Schwartz, and Y.H.H. Wan. 2004. "Statistical Wind Power Forecasting for U.S. Wind Farm." 17th Conference on Probability and Statistics in the Atmospheric Series/2004. American Metrological Society. Annual Meeting Seattle Washington. January 11-15, 2004. [5]

Milligan, M., M. Schwartz, and Y. Wan. 2003. "Statistical Wind Power Forecasting Models: Results for U.S. Wind Forms." Austin, TX. May 18-21, 2003. [7]

Mirande, M.S., and R.W. Dunn. 2006. "One-Hour Ahead Wind Speed Prediction Using a Bayesian Methodology." 1-6. IEEE Power Engineering Society General Meeting 2006. [6]

Negnevitsky, M., P. Johnson, and S. Santore. 2007. "Short Term Wind Power Forecasting Using Hybrid Intelligent Systems." 1-4; 24-28. *IEEE Power Engineering Society General Meeting*. June 2007. [3]

Sideratos, G., and N.D. Hatziargyrion. 2007. "An Advanced Statistical Method for Wind Power Forecasting." *IEEE Trans. Power System* 22(1): 258-265. [8]

Wei, W.W.S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley. [12]

Wu. Y.K., and J.S. Hong. 2007. "A Literature Review of Wind Energy Forecasting Technology in the World." 504-509. *IEEE Power Tech 2007*. Lansanne. 1-5 July 2007. [1]

http://cran.rproject.org/web/packages/TSA/index.html[13

http://www.r-project.org/ [14]