Tanzania Journal of Science 45(2): 209-215, 2019 ISSN 0856-1761, e-ISSN 2507-7961

Prediction of Wind Power Density for Electricity Generation at Makambako, Tanzania Using Auto-Regression Integrated Moving Average (ARIMA) Model

Josam Samson and Reuben R Kainkwa
Department of Physics, University of Dar es Salaam, P. O. Box 35063, Dar es Salaam, Tanzania
**Corresponding author e-mail address: rkainkwa@gmail.com*

Abstract

Wind speed data used in this study were from Makambako and were recorded at a height of 10 m. In the analysis, the data were extrapolated to 80 m above ground level and wind speed from July to November were found to have an average value of 17.58 m/s. November was found to be the windiest month for the whole duration while the calm month was found to be March. The wind power density was found to be highest in the month of November for the whole study duration. The months of July to November have higher wind power density ranging from 1,739.00 W/m^2/month to 12,244.35 W/m^2/month at 80 m above ground level while the remaining months have values below the mentioned range. The findings of this study have shown that among five predictions, three showed acceptable predictions of 90.40%, 96.38% and 73.42%, one showing a fair prediction of 65.90% while the fifth one giving unacceptable prediction of 28.83%. It is proposed that ARIMA model be used to predict wind speed for other months apart from March 2009 and November for the years 2009, 2010, 2013 and 2016 that were predicted in this study.

Keywords: ARIMA model, average wind speed, prediction, wind power density

Introduction

Wind energy is one of the potential renewable energy sources, which can be harnessed in a commercial way. Unlike other sources of energy, it is cheap, clean, inexhaustible and price stable (Mahyoub 2006). One of the main setbacks of wind energy is its variation in space and time. For optimal extraction of wind energy, detailed information of wind characteristics at the site of interest is essential before designing a wind farm project (Kumwenda 2011). The utilization of wind power depends on a good knowledge of wind characteristics at the site(s) of relevance. The knowledge about the continuity of wind energy supply in a year is more important than that for the total amount of energy in a year. Wind forecasting is a very important tool in giving the future wind speed characteristics at a site of interest. Using measured wind data at a site, it is possible to predict future characteristics of this meteorological variable by applying standard forecasting models (Chang 2014).

Gnana (2011) reported that Auto Regressive Moving Average (ARMA) Model is widely used in the fields of hydrology, dendrochronology and meteorology. It is further reported that time-series prediction model basing on historical data for short time predictions is also very common whereas persistence model forecasts wind speed under the assumption that wind speed is a meteorological variable that varies from time to time.

The acronym ARIMA stands for Auto-Regression Integrated Moving Averages. There is always a forecasting equation for the intended objective whereby Lags of the stationarized series are called “Auto-Regressive” (AR) terms, a series which needs to be differenced for stationarity is an “integrated” (I) series and Lags of the forecast errors are called “Moving Average” (MA) terms (Nau 2014). The author further reported that a non-seasonal ARIMA is classified as an "ARIMA *(p,d,q)*", where *p* is the number of autoregressive terms, *d* is the number of non-seasonal differences needed for stationarity and *q* is the number of lagged forecasts errors in the prediction equation.

Many studies have been done by different researchers on assessment of wind characteristics and wind energy potential in different sites of Tanzania (Kumwenda 2011, Rajabu 2014, Myinga 2016). The studies were involved with analysis of wind data but they did not forecast for future available wind power for electricity generation at the sites. The main objective of this study was to predict the wind power density for electricity generation at Makambako, Tanzania using Auto-Regression Integrated Moving Average (ARIMA) model.

Materials and Methods

The site where the data were collected is Makambako, a town situated in Njombe Region on the Southern Highlands of Tanzania. It is located roughly 40 km north of Njombe town at around latitude 8º 51' S and longitude 34º 50' E. Makambako was chosen because it is one of the areas in Tanzania with highest wind speeds as reported in previous studies such as those by Mwanyika (2004) and Mgwatu and Kainkwa (2012).

Wind speed data collected at the height of 10 m AGL for the study site were obtained from Tanzania Meteorological Agency (TMA) and they covered a period of ten years from 2007 to 2016. The data for the year 2012 were collected on 10 minutes' intervals, that is, 00:00, 00:10, 00:20… 23:40, 23:50. The ten minutes data were used to evaluate the daily mean values, monthly and annual mean wind speeds.

The mean wind speed was evaluated using the standard equation for mean values. Vertical extrapolations of wind speed from 10 m to higher levels up to 80 m were accomplished through the relation (Gipe 1999):

$$\frac{V_x}{V_r} = \left(\frac{Z_x}{Z_r}\right)^{\alpha} \qquad (1)$$

where V_x and V_r are the mean wind speeds at new height Z_x and Z_r reference height, respectively, and α is the power law exponent. On the other hand the available wind power density P_{ao} was evaluated using the relation (Gipe 1999):

$$P_{ao} = \frac{1}{2}\rho V^3 \qquad (2)$$

where ρ is density of air which was taken as 1.225 kg/m^3 and V is the wind speed in m/s.

ARIMA model is an approach to time series forecasting in which there is a predicting equation which needs to be differenced for stationarity hence giving an "integrated" (**I**) series. Lags of the stationarized series are called "Auto-Regressive" (AR) terms and Lags of the forecast errors are called "Moving Average" (MA) terms. The wind patterns with different wind speeds at different times are defining what is termed as integrated (I) series since they are summed up. The differences from these series are defining what is called differenced series for stationarity. A non-seasonal ARIMA is classified as an "ARIMA *(p,d,q)",* where *p* is the number of autoregressive terms, *d* is the number of non-seasonal differences needed for stationarity, and *q* is the number of lagged forecasts errors in the prediction equation. The equation for the predicted value of *y* in period *t*, based on data observed up to the period *t*–1, can be expressed as (Nau 2014):

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} \qquad (3)$$

where μ is a constant, ϕ_p is the AR coefficient at *lagp*, θ_q is the MA coefficient at *lagp*, and $e_{t-p} = y_{t-p} - \hat{y}_{t-p}$ is the forecast error that was made at period *t-p* where y_{t-p} is the total value with fluctuations while $\hat{y}_{t-p}$ being the value without the fluctuations. Note that the MA terms in the model (the lags of the errors) are conventionally written with a negative sign rather than a positive sign. The constant term may or may not be assumed to be equal to zero.

Yaziz et al. (2013) considered the autoregressive integrated moving average (ARIMA) models to the statistical properties with accurate forecasting over a short period of time. The author eased the implementation of the study and be able to handle nonstationary data. Consequently, the case concluded with the hybrid model of ARIMA (1,1,1)–GARCH (0,0,2) for transformed data which was given by (Nau 2014):

$$y_t = 0.274 y_{t-1} + 0.726 y_{t-2} - 0.992 e_{t-1} \quad (4)$$

The empirical results indicated that the proposed hybrid model ARIMA-GARCH method had improved the estimation and forecasting accuracy where (Nau 2014):

$\phi_{t-1} = 0.274$, $\phi_{t-2} = 0.726$ and $\theta_{t-1} = -0.992$.

Since in this study the intention is to forecast power density, then the general formula for the prediction is given by:

$$P_t = \phi_1 P_{t-1} + ... + \phi_q P_{t-q} - \theta_1 e_{t-1} - ... - \theta_r e_{t-r} \quad (5)$$

where P_t stands for predicted wind power density from the previous wind power density while other symbols carry their usual meanings.

It is required to determine the values of *p* as autoregressive terms and *q* as number of lagged forecasting terms to be used in the equation for predicting the stationarized series *y;* by looking at plots of the outcome locations and partial autocorrelations *y*. The autocorrelation of y at *lagk* is the correlation between *y* and itself lagged by *k* periods; it is the correlation between y_t and y_{t-1}. In *regressIt* software terminology, this is the correlation between the series y and the series *y_LAGk*. In SPSS software terminology it is the correlation between y and *LAG(y,k)*. *Lagk*, refers to when the data are computed one period after another at lagging series as for example the average wind speed of one month is 8.5 m/s while that of next month being 8.42 m/s, then *lagk* = 0.08 m/s where *k* is the lag period between the two successive months where there is a one lag period.

ARIMA model equation was used in predicting the wind power density from the wind speed data. The data were statistically analyzed using SPSS software and Excel. After testing and analyzing the data through ARIMA model, the expression for ARIMA model equation for the wind power density prediction was adopted.

Results and Discussions

Wind speed

Tables 1 and 2 show that the month of November has the highest monthly average wind speed of 7.45 m/s s at 10 m equivalent to 17.58 m/s at 80 m AGL. The average wind speed for the whole study period at the 10 m and 80 m are 5.31 m/s and 12.53 m/s, respectively. March is the month with the lowest wind speed of 3.75 m/s and 8.85 m/s at 10 m and 80 m, correspondingly. The months with wind speed above the average values are July to November while the speed for the remaining months is below the average wind speed. The highest wind speed was observed in November 2009 and the value at the height of 80 m is 27.14 m/s. November was found to be the windiest month during the study period, and as such, further detailed analysis of the wind behavior was conducted. Table 3 shows that the annual average wind speeds for the month of November are 7.45 m/s and 17.58 m/s at 10 m and 80 m AGL, respectively for the whole study duration. The years with wind speed less than the average are 2014, 2015 and 2016 while the wind speeds of remaining years are above the average value for the whole period. It is interesting to note that the values of wind speed for the years 2014 to 2016 are far below those of the previous years. It was believed by the authors that there was something wrong with the data which is either due to climate change or errors in the measuring instruments. However, the officers from TMA informed the authors that the data were correct. But the abnormality for the year 2014 where the speed was 2.00 m/s for January at 10 m AGL which is about 36 % of the year 2017 may be due to climate change or errors in the measuring instruments.

Wind power density

Table 3 shows the available annual wind power densities for the years 2007 to 2016 at 10 m and 80 m AGL, respectively, together with average power for the month of November. The average annual available wind power densities for the study period at 10 m and 80 m AGL are 316.50 W/m^2/month and 4,160.17 W/m^2/month, correspondingly as depicted in the table. The year 2009 had the highest power densities which are 931.54 W/m^2/month and 12,244.35 W/m^2/month at 10 m and 80 m AGL, respectively. The years with power densities greater than the average are 2008, 2009 and 2010 whereas for the remaining years the values are less than the average.

Table 1: Monthly average wind speed (m/s) at the height of 10 m (2007–2016)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
2007	5.50	6.00	4.50	5.00	5.00	4.00	6.00	7.00	8.00	8.50	7.50	5.50	6.04
2008	5.50	5.50	4.50	5.00	6.00	5.50	6.00	6.50	8.50	9.00	8.50	5.00	6.29
2009	4.50	4.00	4.50	4.00	6.00	5.50	6.00	5.50	7.00	7.50	11.50	5.00	5.92
2010	4.50	4.50	4.00	6.50	5.00	6.00	7.00	6.00	8.00	7.00	9.50	7.00	6.25
2011	5.50	6.00	4.00	4.50	5.00	6.00	6.50	7.00	7.50	8.00	8.00	5.50	6.13
2012	5.50	5.00	4.50	5.00	6.00	5.50	6.50	7.00	8.00	8.50	7.50	6.00	6.25
2013	4.50	5.00	5.00	5.00	5.50	4.50	6.00	7.00	7.50	7.00	8.00	3.00	5.67
2014	2.00	2.50	2.00	2.50	3.50	3.50	3.50	3.50	4.50	4.50	4.50	3.50	3.33
2015	3.00	3.00	2.50	2.50	3.00	2.50	3.50	4.00	4.50	5.00	4.50	3.50	3.46
2016	3.00	2.50	2.00	2.50	2.50	5.00	4.00	4.50	5.00	5.50	5.00	3.50	3.75
Average	4.35	4.40	3.75	4.25	4.75	4.80	5.50	5.80	6.85	7.05	7.45	4.75	5.31

Table 2: Monthly average wind speed (m/s) at 80 m AGL (2007–2016)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2007	12.98	14.16	10.62	11.80	11.80	9.44	14.16	16.52	18.88	20.06	17.70	12.98
2008	12.98	12.98	10.62	11.80	14.16	12.98	14.16	15.34	20.06	21.24	20.06	11.80
2009	10.62	9.44	10.62	9.44	14.16	12.98	14.16	12.98	16.52	17.70	27.14	11.80
2010	10.62	10.62	9.44	15.34	11.80	14.16	16.52	14.16	18.88	16.52	22.42	16.52
2011	12.98	14.16	9.44	10.62	11.80	14.16	15.34	16.52	17.70	18.88	18.88	12.98
2012	12.98	11.80	10.62	11.80	14.16	12.98	15.34	16.52	18.88	20.06	17.70	14.16
2013	10.62	11.80	11.80	11.80	12.98	10.62	14.16	16.52	17.70	16.52	18.88	7.08
2014	4.72	5.90	4.72	5.90	8.26	8.26	8.26	8.26	10.62	10.62	10.62	8.26
2015	7.08	7.08	5.90	5.90	7.08	5.90	8.26	9.44	10.62	11.80	10.62	8.26
2016	7.08	5.90	4.72	5.90	5.90	11.80	9.44	10.62	11.80	12.98	11.80	8.26
Average	10.27	10.38	8.85	10.03	11.21	11.33	12.98	13.69	16.17	16.64	17.58	11.21

Table 3: The average wind power density (Wm^{-2}) at 10 m and 80 m AGL from 2007 to 2016 for the month of November

	Height (m)	
Year	10	80
2007	258.4	3,396.50
2008	376.2	4,944.20
2009	931.54	12,244.35
2010	525.1	6,902.60
2011	313.6	4,122.00
2012	258.4	3,396.50
2013	313.6	4,122.00
2014	55.8	733.6
2015	55.8	733.6
2016	76.6	1,006.40
Average	316.5	4,160.17

Prediction of wind power density

In the predictions the height at which the wind power density is considered is 80 m. Table 4 shows the autocorrelations for the month of November 2009 which were used to predict wind power density. The corresponding relation is:

$$P_t = 0.629P_{t-1} + 0.274P_{t-2} - 0.330e_{t-1} \quad (5)$$

where lags 1, 2 and 3 have been used.

Equation (5) was used to predict wind power density of the year 2009 in the month of November from those of 2007 and 2008 and compare with that calculated from the raw data. The equation now takes the form:

$$P_{2009} = P_{2008} + P_{2007} - 0.330e_{2008}$$
$$P_{2009} = 0.629(4{,}944.20) + 0.274(3{,}396.50) - 0.330(1{,}547.70)$$
$$P_{2009} = 3{,}529.80 \quad W/m^2/month \quad (6)$$

From Equation (6) the predicted value of the wind power density is 3,529.80 W/m^2/month while the one from the raw wind speed data given in Table 3 is 12,244.35 W/m^2/month. The ratio of the predicted value to the raw data gives 28.83% which indicates that the forecasted value is far below the raw data. The poor percentage is probably due to an abrupt change in wind speed data from 17.70 m/s in 2007 and 20.06 m/s in 2008 to 27.14 m/s in 2009.

Table 4: Autocorrelations of wind power density for the month of November 2009

Lags	Autocorrelations
1	0.629
2	0.274
3	–0.330
4	–0.367
5	–0.402
6	–0.415

From Table 5 of the autocorrelations the prediction equation for the wind power density for the month of November in the year 2010 is given as:

$$P_t = 0.629P_{t-1} + 0.275P_{t-2} - 0.330e_{t-1} \quad (7)$$

where lags 1, 2 and 3 have been used.

Equation (7) was used to predict wind power density of the year 2010 in the month of November from those of 2008 and 2009 and compare with that calculated from the raw data. The equation now takes the form:

$$P_{2010} = P_{2009} + P_{2008} - 0.330e_{2009}$$
$$P_{2010} = 0.629(12{,}244.35) + 0.275(4{,}944.50) - 0.330(7{,}299.85)$$
$$P_{2010} = 6{,}652.48 \quad W/m^2/month \quad (8)$$

Table 5: Autocorrelations of wind power density for the month of November 2010

Lags	Autocorrelations
1	0.629
2	0.275
3	–0.330
4	–0.367
5	–0.402
6	–0.415

From Equation (8) the predicted value is 6,652.48 W/m^2/month while the one from the raw data is 6,902.60 W/m^2/month as depicted in Table 3. The ratio of the predicted value to the raw data gives 96.38% which indicates that the forecasted value is very close to the raw data.

The autocorrelations from Table 6 under lags 1, 2 and 3 suggest the prediction equation for the year 2013 be given as:

$$P_t = 0.629P_{t-1} + 0.274P_{t-2} - 0.330e_{t-1} \quad (9)$$

Equation (9) was used to predict the power density for the month of November 2013 from those of 2011 and 2012 giving the following results:

$$P_{2013} = 0.629P_{2012} + 0.275P_{2011} - 0.330e_{2012}$$
$$P_{2013} = 0.629(3{,}396.50) + 0.274(4{,}122.00) - 0.330(725.50)$$
$$P_{2013} = 2{,}136.3985 + 1{,}129.428 - 239.415$$
$$P_{2013} = 3{,}026.41 \quad W/m^2/month \tag{10}$$

Table 6: Autocorrelations of wind power density for the month of November 2013

Lags	Autocorrelations
1	0.629
2	0.274
3	–0.330
4	–0.367
5	–0.402
6	–0.415

Equation (10) shows that the predicted value is 3,026.41 W/m^2/month while the one from the raw data being 4,122.00 W/m^2/month as shown in Table 3. The predicted value compared to the raw data gives the ratio of 73.42% which can be taken as an acceptable prediction.

The autocorrelations from Table 7 suggest that the prediction equation with lags 1, 2 and 3 for the year 2016 be given by:

$$P_t = 0.629P_{t-1} + 0.275P_{t-2} - 0.330e_{t-1} \tag{11}$$

Prediction for the year 2016 using Equation (11) is accomplished by imposing the data as follows:

$$P_{2016} = 0.629P_{2015} + 0.275P_{2014} - 0.33e_{2015}$$
$$P_{2016} = 0.629(733.60) + 0.275(733.60) - 0.330(0.00)$$
$$P_{2016} = 461.43 + 201.74 - 0.00$$
$$P_{2016} = 663.17 \quad W/m^2/month \tag{12}$$

On the other hand Equation (12) gives the predicted value for the year 2016 as 663.17 W/m^2/month while the one from the raw data being 1006.40 W/m^2/month as portrayed in Table 3. The ratio of the two values for year 2016 gives 65.90% which may be taken as a fair prediction.

Table 7: Autocorrelations of wind power density for the month of November 2016

Lags	Autocorrelations
1	0.629
2	0.275
3	–0.330
4	–0.367
5	–0.402
6	–0.415

Table 8 shows the autocorrelations under lags 1, 2 and 3 which suggest the prediction for the calm month of March 2009 to be given as:

$$P_t = 0.629P_{t-1} + 0.275P_{t-2} - 0.330e_{t-1} \tag{13}$$

Prediction for the month of March in the year 2009 using Equation (9) is evaluated as follows:

$$P_{2009} = 0.629P_{2008} + 0.275P_{2007}e_{2008}$$
$$P_{2009} = 0.629(733.60) + 0.275(733.60) - 0.330(0)$$
$$P_{2009} = 461.43 + 201.74 - 0.00$$
$$P_{2009} = 663.17W/m^2/month \tag{14}$$

Table 8: Autocorrelations of wind power density for the month of March 2009

Lags	Autocorrelations
1	0.629
2	0.275
3	–0.330
4	–0.367
5	–0.402
6	–0.415

Equation (14) shows that the predicted value is 663.17 W/m^2/month while that one from the raw data being 733.60 W/m^2/month as portrayed in Table 3. The predicted value compared to the raw data gives the ratio of 90.40% which is an acceptable prediction.

From the results obtained the predictions give the percentages of 28.83%, 96.38%, 73.42% and 65.90% for the years 2009, 2010, 2013 and 2016, respectively for the month of November. The prediction done for the calm month of March gave 90.40%. The prediction for November 2009 is not right as it gives 28.83%. The reason for failure may be due to the fact that the wind power density of the raw data changed abnormally from 3,396.50 W/m^2/month in the year 2007 and 4,944.20 W/m^2/month in the year 2008 to 12,244.35 W/m^2/month in the year 2009.

Conclusions

The wind power density is found to be highest in the month of November for the whole study duration. The months of July to November have higher wind power density ranging from 1,739.00 W/m^2/month to 12,244.35 W/m^2/month at 80 m AGL while the remaining months have values below the mentioned range. In this study among the five predictions three have shown acceptable predictions by 90.40% in March 2009, 96.38% and 73.42% in November 2010 and 2013 respectively. One gave a fair prediction of 65.90% in 2016 while one gave unacceptable results by 28.83% in 2009. The abnormality appearing in the year 2009 might have been due to abrupt change of wind speed or erroneous data in the measuring instruments as reported earlier. It is recommended that analysis of the wind speed characteristics on shorter time intervals be done. It is also proposed that ARIMA model be used to predict wind speeds for other months apart from March 2009 and November for the years 2009, 2010, 2013 and 2016 that were predicted in this study.

Acknowledgement

The authors would like to thank the staff of the Tanzania Meteorological Agency (TMA) for providing them with wind speed data of the study site.

References

Chang WY 2014 A literature review of wind forecasting methods. *J. Power Energy Engin.* 2: 161-168.

Gipe P 1999 Wind Energy Basics*:* A Guide to Small and Micro Wind System. Chelsea Green Publishing Company, White River Junction, England.

Gnana SK 2011 Computing Models for Wind Speed Prediction in Renewable Energy Systems.

Kumwenda JM 2011 *Assessment of wind power potential for electricity generation at Kititimo-Singida Region, Tanzania*. MSc Dissertation, University of Dar es Salaam.

Mahyoub HA 2006 A Statistical Analysis of Wind Speed Data and an Assessment of Wind Energy Potential in Taiz-Yemen.

Mgwatu MI and Kainkwa RR 2012 Establishing a probabilistic model of extrapolated wind speed data for wind energy prediction. *World Academy of Science, Engineering and Technology, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering* 6(10): 2055-2060.

Mwanyika HH 2004 *The potential of wind based-energy for electricity generation at Makambako in Iringa, Tanzania*. MSc Thesis, University of Dar es Salaam.

Myinga GM 2016 *Assessment of wind energy potential for electricity generation around Dodoma Airport, Tanzania*. MSc Dissertation, University of Dar es Salaam.

Nau RF 2014 Statistical Forecasting & Notes on non-seasonal ARIMA models. *Duke University–Fuqua School of Management.*

Rajabu J 2014 *Assessment of wind characteristics and available wind power at Kilimanjaro International Airport (KIA), Tanzania*. MSc Dissertation, University of Dar es Salaam.

Yaziz SR Azizan NA Zakaria R and Ahmad MH 2013 The Performance of Hybrid ARIMA-GARCH Modeling in Forecasting Gold Prices.