Wind Speed Distribution: A case study of Mersing, Malaysia

Daniel Derome, Halim Razali, Ahmad Fazlizan, Sohif Mat, Mohd Azimin Elias
Solar Energy Research Institute
Universiti Kebangsaan Malaysia
43600 Bangi Selangor Malaysia
muhammaddanielderome@gmail.com

Abstract— **Mersing is one of the places that have the potential for wind power development in Malaysia. It is often suggested by researchers as an ideal place for generating electricity from wind power. But before a place is chosen, there are several factors need to be considered in order to get an accurate estimation of the actual situation. Additionally, it can also avoid resource wastage and maximize profitability to various parties. Among the factors that the previous researchers have suggested are the wind speed itself, the average amount of annual wind received, minimum wind speed required and more. For this study, the focus is to identify the shape of the wind speed distribution of the Mersing area. Mersing is located in southern part of Peninsular Malaysia and affects by the monsoon seasons. The data is recorded at a height of 20m above sea level. This process is performed for each wind speed data obtained for each area. This study is important because the wind data for any region has its own shape. This is evidenced by the findings of earlier researchers in which most places have different forms of wind speed data. There are also researchers who have obtained different forms of wind speed data distribution for the same area. Therefore, this study is very important to be implemented so that estimation of the wind distribution at the particular area will be more accurate. The data used is wind data from April 2017 to November 2017. The chosen method is to use the average wind speed for a different period as suggested by the literature. Thus, the wind speed data is averaged to 1 minute, 10 minutes and 60 minutes. This study is conducted to see whether the selection of the average type of wind speed data will affect the shape of the wind speed data distribution or otherwise. Data for the first 4 months is used as training while for the next 4 months is used for validation. The data were analyzed using the EasyFit 5.6 Profesional software. From the analysis, it can be seen that the shape of the wind speed distribution in Mersing varies according to the average type used. It can be concluded that the average type used will affect the shape of the wind speed distribution. Lastly, this will also affect the accuracy of estimation and decisions made.**

Keywords— *wind power; wind speed; wind speed distribution*

I. INTRODUCTION

Renewable energy is something familiar to us today. It has become famous due to several factors. Among the key factors that cause renewable energy to be popular are the world oil crisis and the pollution of gas dioxide gas. The sequence of uncontrolled and excessive carbon dioxide emissions reduces indirectly to another crisis in the world, which is global warming. Continued from this, many countries are very keen to develop renewable energy. Malaysia also participates in renewable energy use as it does not want to be left behind by other countries. According to [1], [2], Malaysia has embarked on a first step in the search and use of renewable energy from RMK8 (2001-2005). It cannot be denied that this renewable energy has some advantages. Among them is clean, virtually no pollution and cost-effective.

Nowadays, there is various renewable energy that has been identified. Among the renewable energy that researchers choose is solar energy and wind energy. According to [2], it can be seen through the speed in technology development for both the energy. However, wind power is more preferable choice for renewable energy [1]. This is probably because the wind is a source that is not fixed and constantly changing. So it requires more complex calculations than other renewable energy [3].

From the Asian countries, India and China are the pioneers in this wind energy field [3]. Given the success of India, Malaysia is also a country close to the equatorial line [4]. This allows Malaysia to become a country with some advantages and disadvantages. Among the advantages that Malaysia has earned is one that receives a large amount of wind every year. The wind has been used since ancient times, generally for sailing ships in the oceans and in the agricultural sector [3]. In addition, it can also generates electricity.

Wind is one of the renewable energies available in Malaysia. However, in the context of Malaysia, at present, wind power is still low in comparison with other Asian countries. This can be proved through a report issued by [5]. While from the world point of view, the use of wind power for the production of electricity is very high. According to a report submitted by International Renewable Energy Agency (IRENA) 2017, there has been a fourfold increase from 2007 to 2016. This indicates that wind energy has its own advantages in generating electricity.

Due to this, there are still many studies to be done to ensure the use of wind power to produce electricity into reality and consumer choice in Malaysia. Through literature studies, many factors need to be considered to enable such an area to become a reality as a power-generating area of wind power.

This research is sponsored by grant code AP-2015-003 Universiti Kebangsaan Malaysia

According to [1] among the factors that are the researcher's discussions are the wind speed at high altitude, wind direction, and the shape or nature of the wind itself. Besides that, there are researchers who are studying the selection of wind distribution forms for a particular area. This can be seen through the findings of past studies in which each area has a different shape of wind distribution. There are many wind data distributions obtained from previous research findings. Among them are Weibull, Rayleigh, Burr, Gamma, and others.

In addition, the average selection of wind speed to see the shape of the wind distribution is also often said. The findings of literature studies show that most researchers use the average wind data for each hour of their study. However, the recommendations from the previous study indicate that the data are analyzed using a smaller average data. According to [6], from the findings of the survey shows that the selection using the larger data average will cause the results being less accurate. This is due to the small amount of information obtained when the average data is enlarged. It coincides with the wind speed itself. It is known that the wind change at all times. It is a type of energy that is not fixed. For this reason, the study involves wind energy needs to be carried out until now.

The main purpose of this study is to see whether the average wind speed selection will influence the shape of the wind speed distribution of the Mersing area. In a study involving each data, the key point is to determine the form of data [7]. This is due to each data has its own form. It is distinguished by the nature of the data.

II. STUDY AREA AND DATA

Selection of Mersing as a suitable area for power generation from wind power began as early as 1995. The selection of the study area is based on several factors. Among the factors is the suggestion from the previous study which suggested Mersing as one of the ideal places for generating electricity from wind power [1], [2], [8], [9]. Among other factors is the location of the Mersing itself which is located at 43.6m at mean sea level [2]. In addition, the Mersing geographical area, which lies opposite the South China Sea, also allows Mersing to gain a large amount of wind throughout the year. This also makes Mersing vulnerable to the sea breeze and land breeze. The data were obtained from a weather station installed at the University Kebangsaan Malaysia (UKM) - Mersing Marine Ecosystem Center (EKOMAR), (Figure 1). It is recorded at a height of 20m above sea level. The data used is wind data from April 2017 to November 2017.

Fig. 1. Location of Mersing (Image from Google Maps)

III. METHODOLOGY

Wind data is divided into several types of average groups. Among them is 1 minute, 10 minutes and 60 minutes. Missing data will not be taken into account as analytics. The data were analyzed using the software, EasyFit 5.6 Professional. This software will be used to obtain the corresponding shape (distribution) of the wind data obtained. In addition, it is also used to obtain descriptive values for the data. Descriptive data is required as it is the basic information of a data. Among those to be analyzed is the average value of wind speed, standard deviation and ultimately data skewness.

A. Estimation distribution

For this study, it has been determined to use the Maximum Likelihood Method (MLM) to find the value of the parameter. This method needs to do a most numerical iteration to compute the value of the parameter and it was proposed by Steven & Smulders in 1979 [10]. Let $v_1, v_{2,....}v_n$ be a random sample size of n drawn from a probability density function $f(v_i, \theta)$, where θ is unknown parameter.
Likelihood function,

$$L = \prod_{i=1}^{n} f_{v_i}(v_i, \theta)$$

Maximum likelihood estimator (MLE) of θ is the value of θ that maximizes L or equivalently, the logarithm of L. According to [11], often but not always, the MLE of θ is a solution of

$$\frac{dLogL}{d\theta} = 0$$

The value of the parameters obtained will be a substitute in the distribution that is deemed appropriate to obtain the probability density function (PDF) value. Whereas the PDF will be plotted together with the histogram of the data. It aims to look and see the distribution that corresponds to the histogram.

B. *Statistical tools*

The selection of the best distribution will also involve some statistical tools. The reason to use statistical tool is due to as a tool to see the effectiveness of each selected distribution. Among them are Chi-Square, Kolmogorov-Smirnov, and Anderson-Darling.

Here is a formula for Chi-Square,

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed data and E_i is the expected data [12].

Formula for Kolmogorov-Smirnov,

$KOL = Max|F(v) - F_n(v)|$, where v is the identifies the set of the velocity to be considered, $F(v)$ is the cumulative probability distribution for specific distribution and $F_n(v)$ is the experimental histogram [13].

While the formula for Anderson-Darling (AD) is

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))]$$

Where n =the sample size, $F(v)$ is the cumulative probability distribution the specific distribution and i=the ith sample, calculated when the data is sorted in ascending order [14].

IV. RESULT AND DISCUSSION

Among the results obtained from the analysis were descriptive data. For the August, the average wind speed was the highest in 2.9431m / s while the lowest was in October of 1.3947m / s. For standard deviation, the whole month is at a relatively high level. This means that the mean value of the data obtained is not suitable for assessment. This is because the high standard deviation value will cause the mean value to be unstable. From the findings, we can also see that the entire wind speed data is the skew to the right. This is based on positive values for the entire data. This has been summarized and can be seen from the Table 1.

TABLE I. THE DESCRIPTIVE STATISTIC FOR EACH MONTH AND TYPE OF AVERAGE

Month/type of average		Mean	Standard deviation	Skewness
August	1 minute	2.9431	1.2105	0.2864
	10 minutes	2.0289	1.1795	0.5817
	60 minutes	2.0189	1.1103	0.5014
September	1 minute	2.3462	1.4396	0.4059
	10 minutes	1.7675	1.0345	0.7651
	60 minutes	1.7702	0.9767	0.7158
October	1 minute	1.3974	1.0364	1.9821
	10 minutes	1.4005	0.7925	1.3750
	60 minutes	1.3947	0.6938	0.7628
November	1 minute	1.3437	0.7533	1.0001
	10 minutes	1.6399	0.9243	1.1815
	60 minutes	1.6377	0.8402	1.0484

A. *Statistical tool – Chi-Square*

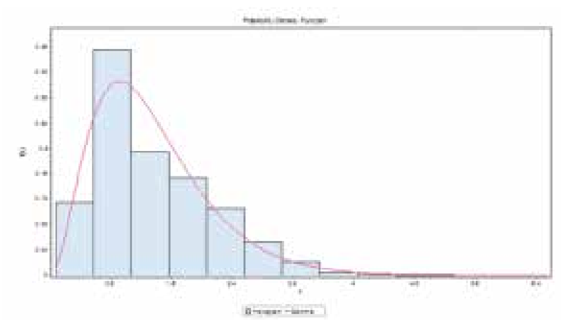
The initial conclusion that can be made is that there is no single data that resembles the shape of the Weibull distribution as is usual. From the Table 2, for August it is noteworthy that the average selection of either 1 minute, 10 minutes or 60 minutes does not give any effect to the distribution. But for September data, it can be seen that the average selection for 1 minute and 60 minutes can provide the same distribution which is Burr distribution. While from October data it can be concluded that distribution will change according to the average type of data taken. Last but not least, for November it was patent as in October since it was different from the average type used.

TABLE II. BEST FIT DISTRIBUTION BY CHI-SQUARE

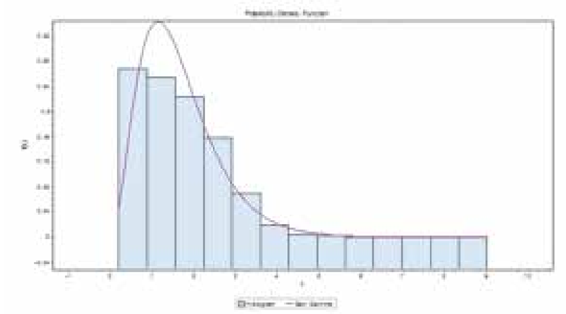
Month/type of average		Best fit distribution
August	1 minute	Burr
	10 minutes	Burr
	60 minutes	Burr
September	1 minute	Burr
	10 minutes	Gen. Gamma
	60 minutes	Burr
October	1 minute	Burr
	10 minutes	Gen. Gamma
	60 minutes	Gamma
November	1 minute	Gamma
	10 minutes	Gen. Gamma
	60 minutes	Rayleigh

Figure 2 (a-c) shows an example of histogram and probability density function for the selected month (November). The selected month based on which one can show the variability of the distribution wind speed data.

a.



b.



c.

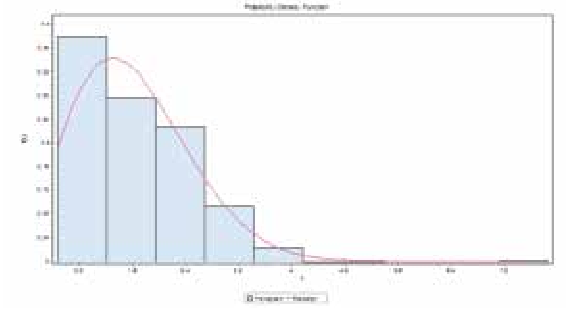


Fig. 2. Example of best-fit distribution for wind speed in month November by Chi-Square (a) Gamma (b) Gen. Gamma (c) Rayleigh

B. Statistical tool – Kolmogorov-Smirnov

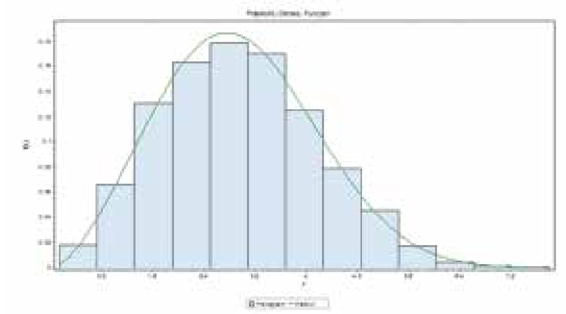
From the Table 3, the general conclusions that can be made are none of the data can get the same distributions for 1 minute and 60 minutes interval. For August and October, there was a slight difference compared to the other months. This is because, from the findings, there are non-similar distributions for different types of data. In addition, there is only one Weibull distribution option from 12 analysis conducted. It happens to Erlang and Rayleigh distribution which respectively get one nomination.

TABLE III. BEST FIT DISTRIBUTION BY KOLMOGOROV-SMIRNOV

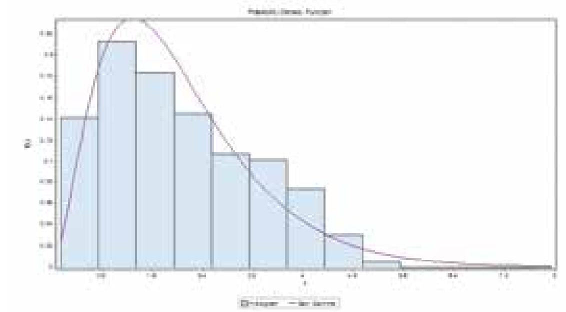
Month/type of average		Best fit distribution
August	1 minute	Weibull
	10 minutes	Gen. Gamma
	60 minutes	Burr
September	1 minute	Gen. Gamma
	10 minutes	Gen. Gamma
	60 minutes	Burr
October	1 minute	Burr
	10 minutes	Gamma
	60 minutes	Erlang
November	1 minute	Gamma
	10 minutes	Gamma
	60 minutes	Rayleigh

Figure 3 (a-c) shows an example of histogram and probability density function for the selected month (August). The selected month based on which one can show the variability of the distribution wind speed data.

a.



b.



c.

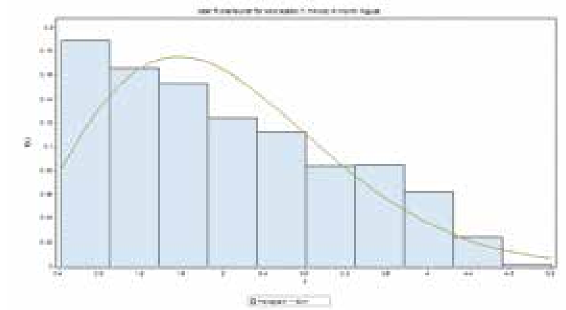


Fig. 3. Example of best-fit distribution for wind speed in month August by Kolmogorov-Smirnov (a) Weibull (b) Gen. Gamma (c) Burr

C. Statistical tool – Anderson-Darling

Based on Table 4, data for October indicate a slight difference from other months. This is due to the type of distribution obtained is the different for each average type of data. It seems to be impressed with the selection of data types used. Whereas from Table 4, the results for August, September, and November show that the same pattern, where the average options of 1 minute and 60 minutes will not give the same result. For this statistical tool, there are two Weibull distributions that have been accepted as the best fit distribution.

TABLE IV. BEST FIT DISTRIBUTION BY ANDERSON-DARLING

Month/type of average		Best fit distribution
August	1 minute	Weibull
	10 minutes	Burr
	60 minutes	Burr
September	1 minute	Weibull
	10 minutes	Gen. Gamma
	60 minutes	Gen. Gamma

October	1 minute	Burr
	10 minutes	Gamma
	60 minutes	Erlang
November	1 minute	Gamma
	10 minutes	Gamma
	60 minutes	Rayleigh

Figure 4 (a-c) shows an example of histogram and probability density function for the selected month (October). The selected month based on which one can show the variability of the distribution wind speed data.

a. b. c.

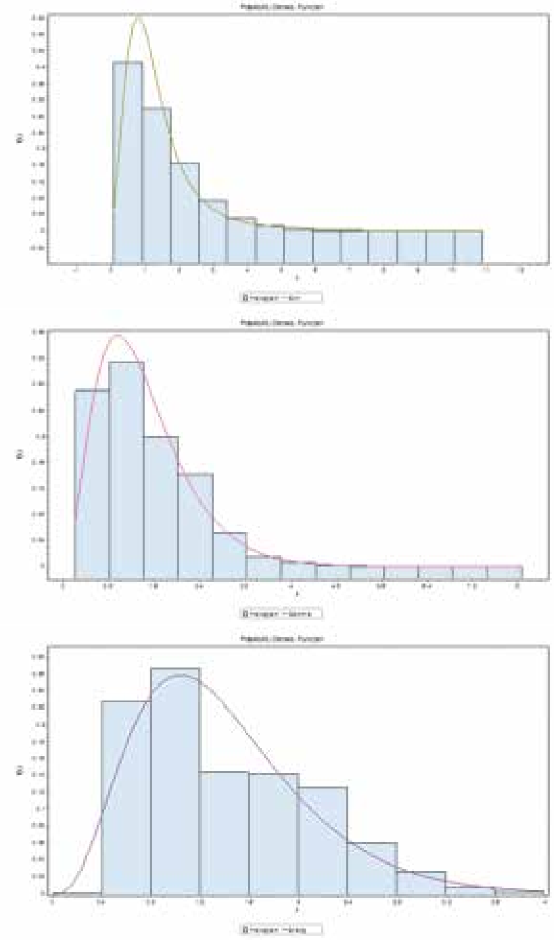


Fig. 4. Example of best-fit distribution for wind speed in month October by Anderson-Darling (a) Burr (b) Gamma (c) Erlang

Based on Figure 5, it can be concluded that Burr distribution is the highest choice. It has gained 42%. This means that the Burr distribution corresponds to the data for the average data per minute. While for Gamma and Weibull get 25% each. Gen. Gamma gets 8% to occupy the last place.

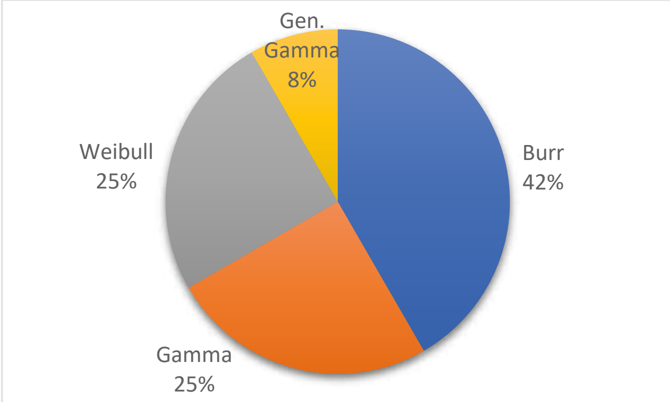


Fig. 5. List of distribution that best fits the 1-minute data (Mersing)

There are only three distribution for 10 minutes average data. Gen. Gamma distribution is the highest choice distribution for 10 minutes average data. It has gained 50%. This means that the Gen. Gamma distribution corresponds to the data for the average data per 10 minutes. While for Gamma and Burr get 33% and 17% each based on Figure 6. If seen from the previous data it is a continuation that all the distributions available for the average of 1 minute will also be on average 10 minutes

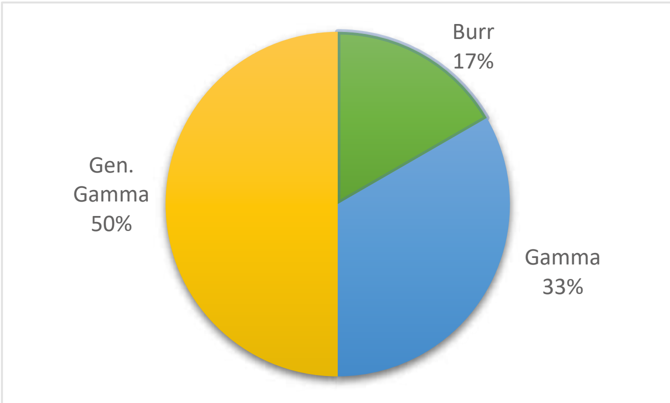


Fig. 6. List of distribution that best fit the 10 minutes data (Mersing)

According to Figure 7, it only has five distributions selected as the best fit distribution for 60 minutes data. The highest distribution is Burr distribution. It has gained 42%. This means that the Burr distribution corresponds to the data for the average data per 60 minutes. The results look same as average data per minute, but it is can't be a conclusion. It still needs more work to do before generalized it.

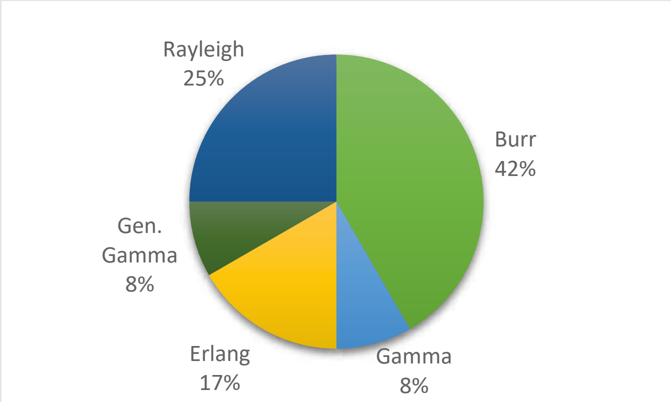


Fig. 7. List of distribution that best fit the 60 minutes data (Mersing)

As a result of selecting the best fit distribution for each type of average wind speed, it has been identified that the most distribution represents the wind speed in Mersing as in Figure 5. The intended distribution is Burr distribution, where it gets 33.3%. Then it was followed by Gamma and Gen. Gamma distributions which got 22.2% each. While for Rayleigh distribution it has been chosen as 3 times, which is equivalent to Weibull distribution. Last but not least, the last place is Erlang distribution where it gets only 5.5 %.

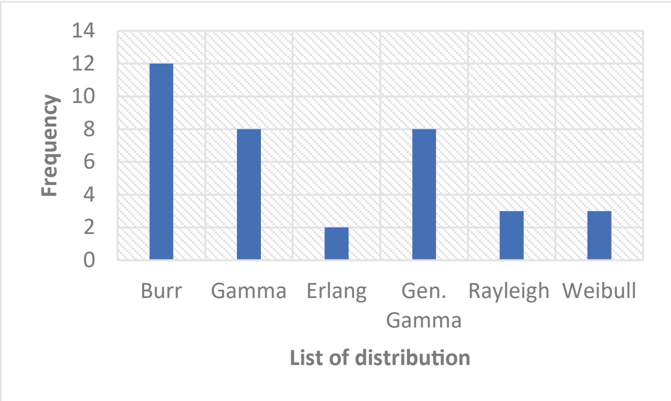


Fig. 8. List of distribution that best fit the data (Mersing)

V. CONCLUSION

This paper discusses the distribution of wind speed for Mersing. From the findings, it can be concluded that the selection of the average type of wind speed affects the distribution. This can at least deny that any data involving wind speeds would be justified to Weibull distribution. This is due to each data has its own form. Furthermore, making smaller averages will make a good result. This is because feature characteristics of the data being less affected.

This study is an initial study of the shape (distribution) of wind speed in Mersing. This study needs to be continued so that we can determine the average wind standard that is suitable for wind speed assessment purposes. Selection of distribution for each data is a key requirement before proceeding to further analysis. If the selection of a data is done with a lenient, it will cause the overall finding to be inaccurate. So the good distribution selection process will help researchers find the approximate answer. This will avoid wasting of time, energy and money.

ACKNOWLEDGMENT

The authors would like to thank Universiti Kebangsaan Malaysia (UKM) for the research grant allocation under the Grad Challenge Research Project (AP-2015-003).

REFERENCES

[1] M. Z. Ibrahim, K. H. Yong, M. Ismail, A. Albani, and A. M. Muzathik, "Wind characteristics and gis-based spatial wind mapping study in Malaysia," *J. Sustain. Sci. Manag.*, vol. 9, no. 2, pp. 1–20, 2014.

[2] A. Belhamadia, M. Mansor, and M. A. Younis, "A study on wind and solar energy potentials in Malaysia," *Int. J. Renew. energy Res.*, vol. 4, no. 4, pp. 1042–1048, 2014.

[3] A. Zaharim, S. K. Najid, a M. Razali, and K. Sopian, "The Suitability of Statistical Distribution in Fitting Wind Speed Data," *Math. Methods, Comput. Tech. Non-Linear Syst. Intell. Syst.*, vol. 7, no. 12, pp. 386–389, 2008.

[4] A. Albani, M. Ibrahim, and K. Yong, "Wind Energy Investigation in Northern Part of Kudat, Malaysia," *Int. J. Eng.*, vol. 2, no. 2, pp. 14–22, 2013.

[5] IRENA, "Data and Statistics - IRENA REsource (Capacity and Generation)," *International Renewable Energy Agency*, 2017. [Online]. Available: http://resourceirena.irena.org/gateway/dashboard/?topic=4&subTopic=16.

[6] S. Ahmad, W. H. W. M. a, M. a Bawadi, and M. S. S. a, "Analysis of Wind Speed Variations and Estimation of Weibull Parameters for Wind Power Generation in Malavsia," *October*, vol. 18, no. 4, pp. 476–482, 2006.

[7] N. . Masseran, A. M. . b Razali, K. . b Ibrahim, W. Z. W. . Zin, and A. . Zaharim, "On spatial analysis of wind energy potential in Malaysia," *WSEAS Trans. Math.*, vol. 11, no. 6, pp. 467–477, 2012.

[8] K. Sopian, M. Y. Othman, and A. Wirsat, "The wind energy potential of Malaysia," *Renew. Energy*, vol. 6, no. 8, pp. 1005–1016, 1995.

[9] N. Sanusi, A. Zaharim, and S. Mat, "Wind energy potential: A case study of mersing, Malaysia," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 12, pp. 7712–7716, 2016.

[10] A. A. Junaid K. Khan, Feroz Ahmed, Zaheer Uddin, S. Tanweer Iqbal, Saif Uddin Jilani, Afaq Ahmed Siddiqui, "Determination of Weibull Parameter by Four Numerical Methods and Prediction of Wind Speed in Jiwani (Balochistan)," *J. Basic Appl. Sci.*, vol. 11, pp. 62–68, 2015.

[11] P. Bhattacharya and R. Bhattacharjee, "A Study on Weibull Distribution for Estimating the Parameters," *Wind Eng.*, vol. 33, no. 5, pp. 469–476, Oct. 2009.

[12] A. K. Azad, M. G. Rasul, and T. Yusaf, "Statistical diagnosis of the best weibull methods for wind power assessment for agricultural applications," *Energies*, vol. 7, no. 5, pp. 3056–3085, 2014.

[13] D. K. Kidmo, R. Danwe, S. Y. Doka, and N. Djongyang, "Statistical analysis of wind speed distribution based on six Weibull Methods for wind power evaluation in Garoua , Cameroon," *Rev. des Energies Renouveables*, vol. 18, pp. 105–125, 2015.

[14] Stephanie, "Anderson-Darling Test & Statistic: Definition, Examples," 2017. [Online]. Available: http://www.statisticshowto.com/anderson-darling-test/. [Accessed: 07-Mar-2018].