Renewable and Sustainable Energy Reviews 54 (2016) 1389–1400

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Wind energy assessment considering wind speed correlation in Malaysia

H.H. Goh [a,*], S.W. Lee [a], Q.S. Chua [a], K.C. Goh [a], K.T.K. Teo [b]

[a] *Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400 Parit Raja, Johor, Malaysia*

[b] *School of Engineering and Information Technology, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*

A R T I C L E I N F O

Article history:
Received 23 July 2014
Received in revised form
16 April 2015
Accepted 23 October 2015
Available online 11 November 2015

Keywords:
Wind energy
Renewable energy
Mycielski
K-means clustering
Weibull

A B S T R A C T

Renewable energy is the current trend of energy sourcing. Numerous scientists, inventors, and engineers are working hard to harness renewable energy. The application of renewable energy is very wide; it can be as small as lighting an LED bulb or as large as generating the electricity of a town or even a country. Wind energy plays an important role in the context of electricity generation. Wind energy is highly dependent on the wind speed at a wind site. Wind prediction is necessary for a wind energy assessment of a potential wind farm. In this study, the wind energy assessment is based on wind prediction using the Mycielski algorithm and *K*-means clustering in Kudat, Malaysia. The predicted results are analysed using Weibull analysis to obtain the most probable wind speed. From the results of this study, *K*-means clustering is more accurate in prediction when compared with the Mycielski algorithm. The most probable wind in Kudat is sufficient to operate the wind turbines.

Contents

* Corresponding author. Tel.: +60 16 741 6824.
E-mail address: jonathanhhgoh@gmail.com (H.H. Goh).

http://dx.doi.org/10.1016/j.rser.2015.10.076
1364-0321/

Nomenclature

Name	Symbol	SI Unit	Description
1. Giga Watt	GW	GW	unit of power (electric).
2. Multi criteria decision making method	MCDM	–	abbreviation of a statistical method.
3. Number of cluster	k	–	to be determine by using K-means clustering algorithm.
4. Function	f	–	symbol of function.
5. Wind speed	v	m s^{-1}	speed of wind measure in meter per second.

1. Introduction

Renewable energy is the current trend of energy generation. Various types of energy resources are available. The selection of suitable energy resources is highly dependent on the potential site's topography. Wind power is suitable for sites with wind speed greater than the cut in speed. Meanwhile, hydropower is suitable for areas with high water capacity. The intensity of solar radiation is the keystone that determines the suitability of solar power in a particular area.

The trend of renewable energy's application is because of the saturation of carbon dioxide in the atmosphere. Although carbon dioxide might be absorbed by the plants during the daytime through photosynthesis, excessive carbon dioxide emission and illegal deforestation contributed to global warming. The effects of global warming have recently been seen and encountered by mankind. The increasing sea water level could sink some islands and countries in the coming decades.

Electricity is necessary sources in our life. However, conventional power generation using coal fire definitely produces carbon dioxide because of the fuel's combustion during the energy conversion process. According to the United States Energy Information Administration (U.S. EIA), 31,780.36 million metric tons of carbon dioxide were emitted from electric power generation in 2010, and it has increased 10% since 2006. Approximately, half of the global carbon dioxide emission was produced by coal [1,2].

To overcome climate issues, more countries are implementing renewable electricity generation. According to the Renewable 2012 Global Status Report [3], the investment in renewable energy increased from US $ 161 billion to US $ 257 billion. This investment increased significantly up to 60% within 3 years. Wind power is one of the fastest growing renewable energies. The total capacity of wind energy installed in the year 2009 was 159 GW and was gradually increased by 50% in 2011 [3,4]. The amount of wind power installed globally has steadily increased since 2010.

According to Bloomberg New Energy Finance (BNEF), a newly established renewable energy power plant is cheaper than a newly built coal and gas-fired power plant [5]. Wind energy is 14% cheaper than the new coal-fired power plant and 18% cheaper than a new gas-fired power plant, excluding the cost of emissions under the Australian carbon pricing scheme.

Numerous researchers from other countries are putting efforts into the carbon-reducing issues [6–10]. Although the types of renewable energy used might differ from each other, the main objective is to reduce carbon emission during electricity production. As discussed in the paper [6], the issue of biomass energy generation was discussed. A system dynamic study was done by using feedback loop to investigate the ways to achieve 30 GW biomass power generation in China. It is undeniable that government policies and acts are main players in renewable energy development. In the paper [7], government strategies, policies, acts, and planning were discussed for Queensland scenario. On the other hand, Turkey is an energy importing country. Potential of various type of renewable energies that could be utilised in Turkey was discussed [8]. Energy market will be very promising if Turkey's government can take consideration on renewable energy development. The efforts on promoting renewable energy can mitigate global warming issues [10].

The development of renewable energy is increasing gradually since last decades; education line should have been contributed a lot in this context. As technology become more advance, there will be new directions for academicians for knowledge transfer [9].

For this paper, the main objective is to investigate the viability of wind power implementation in Malaysia. In Malaysia, the renewable energy market, especially in wind power, is not as mature as other countries [11]. Hence, a preparation and assessment of wind power are needed. The assessment of wind speed prediction at the potential site will be analysed in this paper using the Mycielski algorithm, K-means clustering and a Weibull distribution. Related research works conducted earlier will be explained in the next chapter.

2. Literature review

In this section, a literature review of respective methods is presented. This review can be categorised into wind prediction, system prediction and wind site selection. In wind speed prediction, there are two methods used for the analysis: the Mycielski algorithm and K-means clustering. The results for both prediction methods are compared using a Weibull distribution. Scale and shape parameters need to be determined beforehand to achieve better analysis results.

2.1. Wind power in Malaysia

Wind speed is one of the key factors in determining the potential power that could be generated from a wind turbine. A wind resource map could be very helpful at the early stage of the development process. Throughout this research, the feasibility of wind power in Malaysia is emphasised. There are some researchers that contributed the assessments of wind energy potential in Malaysia. For instance, Sopian et al. from Universiti Kebangsaan Malaysia analysed the wind energy potential over 10 years for 10 different stations in 1995 [12]. The outcome from the research was the data bank of wind patterns in select wind measurement stations in the Peninsula of Malaysia, Sabah and Sarawak, by using a Weibull Distribution. Additionally, wind direction was tabulated in the paper. In 2011, a group of researchers from the University of Malaya used a Weibull distribution function to analyse the wind potential, especially in Kudat and Labuan, which is located in Sabah, Malaysia [13]. Islam et al. used the “WRPLOT” software to show wind direction. Graphical illustrations of the results provide clear presentations for readers. The researchers claim that Kudat and Labuan are suitable for a small-scale wind power plant. Additionally, research had been performed regarding wind farm allocation in Malaysia using a multi-criteria decision-making method (MCDM) [14]. In this particular research, two locations were chosen as the research target: Kota Bahru and Kudat. Literature studies in Malaysia are summarised in Table 1.

Table 1
The summary of literature studies for wind speed analysis in Malaysia.

Authors	**Contribution**	**Field**
K. Sopian M.H.Y. Othman A. Wirsat (1995)	Ten years wind speed analysis inclusive direction of wind blow in ten stations by using Weibull distribution.	1. Weibull distribution. 2. Wind speed analysis in Malaysia.
M.R. Islam R. Saidur N.A. Rahim (2011)	WRPLOT for wind speed and wind direction analysis in Kudat and Labuan. Recommend small scale wind power plant to establish in both sites.	1. Wind speed analysis in Malaysia. 2. Weibull analysis. 3. Wind farm site selection.
H.H. Goh S.W. Lee B.C. Kok S.L. Ng (2011)	Wind farm allocation based on multi-criteria decision making method on Kota Bahru and Kudat. The results shown that Kudat is more suitable for wind power plant installation.	1. Wind speed analysis in Malaysia. 2. Multi criteria decision making, AHP and Fuzzy AHP. 3. Wind farm site selection

2.2. Mycielski algorithm

In terms of wind speed analysis, Bivona et al. analysed the hourly wind speed in Sicily using a Weibull distribution function [15]. The authors used the meteorology department's wind data to study the characteristics of wind speed at nine locations in Sicily. In that particular research, the fitness of the Weibull distribution function to wind speed was clearly seen. As part of that work, Louka et. al. introduced modelling that is being applied in Greece [16]. For instance, the University of Athens used SKIRON modelling to forecast wind speed for up to 5 days ahead. At the same time, the Regional Atmospheric Modelling System (RAMS) was developed at Colorado State University and Mission Research Inc. ASTeR Division. RAMS can forecast wind up to 48 h later. In addition, an adaptive fuzzy neural network (F-NN) also applied wind power prediction for up to 120 h ahead. However, these methods cannot handle systematic errors that are caused by local adaptation problems. The authors proposed Kalman filtering to improve the performance of the aforementioned methods. Kalman filtering is one of the statistically optimal sequential estimation procedures for dynamic systems. The results from the research show that systematic errors can be eliminated using Kalman filtering.

For this research, the prediction of wind speed is performed using the Mycielski algorithm. The Mycielski approach is a data compression method that has been widely used in communications engineering. This method fully utilises historical data as the reference for the prediction value. The Mycielski method is actually the advanced version of the Limpel Ziv (LZ). Research on hourly wind speed prediction in Turkey was performed by Hacaoglu et al. using the Mycielski approach [17]. It is a new approach in wind power prediction. Researchers analysed and predicted the wind speed in Kayseri, Izmir and Antalya. The result of the prediction is promising and very close to the actual data. The comparison of data fitting for both actual and predicted data was performed using a Weibull distribution function. The comparison proved the accuracy of the predicted result. In 2011, the same group of researchers modified the algorithm to solve the looping problem by adding a random number into the predicted data [18]. The modified algorithm is called Mycielski-1 and Mycielski-2. In Mycielski-1, a random number between -0.4 to 0.4 is added to the predicted value. This random number can be changed according to the requirement of the research. Meanwhile, the historical data were rounded to the nearest integer number and divided into a few clusters in Mycielski-2. The prediction is performed by randomly selecting the historical data from a different cluster. In addition, the authors also made the comparison between the modified Mycielski approach and Markov chain.

2.3. K-means clustering

The application of *K*-means clustering can be traced to more than a half century ago [19,20]. The main idea of *K*-means clustering is to group sets of data into clusters. The data in each cluster can be analysed. Normally, this method is used in the application of image indexing [21,22]. Cao et al. proposed an additional algorithm to overcome the “zero value” dilemma which affects the efficiency of centroid selection in image recognition [21]. Additionally, it is also widely being used in system architectures [23,24]. An and Mattausch proposed the combination of a computer's hardware and the *K*-means software to shorten the time of the image segmentation process [23]. The results show that the time of execution is gradually reduced by using the proposed method. Meanwhile, Di Fatta et al. researched the application of the *K*-means algorithm in communications engineering. Clustering of the multimedia data could reduce the loss of messages [24]. Researchers are using the clustering to analyse protein interaction [25] and the natural environment [26].

K-mean clustering is a statistical data mining method. The method of choosing the number of k in the algorithm decisively leads to appropriate analysis [22,27]. It has become an argument of the k number selection for decades. The selection might present in different ways with different applications. For instance, the k cluster in the prediction of velocities on motorways that had been performed by Asamer and Din was by self selection [28]. In that particular paper, the cluster is divided into 4 centred clusters known as centroids. The quantisation error for 4 clusters is reduced dramatically when compared with 1–3 clusters.

Moreover, the *K*-means algorithm is applicable in the prediction of software faults [29]. The Quad Tree method was added into the *K*-means algorithm to find the initial cluster. It was claimed that the combination of the algorithms will make the prediction perform better. In the context of prediction, Kusiak and Li applied *K*-means clustering to predict wind power that could be produced by a wind turbine [30]. Five parameters have been investigated for power prediction. The power produced is categorised into a few clusters up to 1500 kW for different wind speed. The variation of wind speed varies the power produced. The performance of the prediction is then tested by using the mean absolute error (MAE), mean relative error (MRE), standard deviation of MAE and standard deviation of MRE [31].

Many improvements of *K*-means clustering have been performed by researchers to overcome the drawbacks of the algorithm [27,32–42]. Most modifications were based on the initialisation of the centroids. As stated in the research of Xu [27], the centroids of data would be inappropriate if the “out of boundary” data had never been updated by the algorithm. The term “dead unit” describes this circumstance. The time consumption for the algorithm is critical to pattern recognition technology. Hence, the improvement of the algorithm's processing time is a hot topic for researchers [23,38–40]. This is a great improvement in image and signal processing technology. By shortening the processing time, the process will be more efficient.

2.4. Weibull distribution

Weibull distribution is commonly used to describe the wind speed frequency distribution of a region. There are various types of Weibull distributions available. Different types of methods are applied to estimate Weibull parameters. The application of the Weibull distribution type is dependent on the research requirements. For decades, Weibull distribution was used in wind load

Table 2
The iteration of wind state.

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y_i (m s^{-1})	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0

studies by Davenport [43]. Another group of researchers found that the Weibull distribution was useful and appropriate for wind energy applications [44–46]. Justus et al. noted the advantages of five methods in Weibull distribution in wind energy applications [44]. The five methods mentioned in the research were as follows:

i. Least squares fit to observed distribution, also known as the graphic method [47].
ii. Median and quartile wind speeds.
iii. Mean wind speed and standard deviation, also known as the empirical method [48].
iv. Mean wind speed and fastest mile.
v. Trend of k vs $\overline{V}$.

As in the journal paper [44], Justus et al. commented that methods (iv) and (v) are simpler when compared with others. Additionally, these methods also provided reasonably accurate representations of the actual wind speed frequency distribution.

Decades later, comparisons of the Weibull parameter estimations were performed by several researchers [47,48]. Akdag and Dinler recommended a power density method as the parameter estimation and the performance is compared with conventional methods [47]. However, Chang analysed six different types of parameter estimation methods. From the result, Chang concluded that maximum likelihood, modified maximum likelihood and the moment method (empirical method) are more accurate when compared to others [48].

Few types of parameter estimations are applicable in the different fields of study being applied [49–53]. In the research of Seguro and Lambert [54], a two-parameter Weibull function was used to analyse wind energy. In that research, Seguro and Lambert commented that the maximum likelihood parameter estimation method is more suitable than the graphical method. In terms of wind speed analysis, the hourly wind speed in Sicily was analysed using a Weibull distribution function by the researchers [15]. Researchers used the meteorology department's wind data to study the characteristics of wind speed at nine locations in Sicily. In that particular research, the fitness of the Weibull distribution function to wind speed was clearly seen. Meanwhile, previous research [13,55,56] also preferred the two-parameter Weibull distribution as the wind frequency analyser. Islam et al. used the moment method to estimate the Weibull parameters for the wind energy assessment in Malaysia [13]. Despite the diversity of parameter estimation, the selection of estimation method should consider the availability of the data.

3. Methodology

The methods discussed in the previous chapter were commonly applied in various fields of study. To suit the corresponding methods for this study, some alterations are needed.

3.1. Mycielski prediction

Mycielski is an algorithm that performs a prediction on time series data. The prediction is performed based on historical data. The wind speed prediction will be focused on the selected potential wind farm. The wind speed data will be modified to fit in the Mycielski algorithm. To utilise the Mycielski algorithm, the obtained wind data have to be rounded into the nearest wind states. This step purposely simplifies the searching process. Rounding is based on Eq. (1).

$$\text{if}: |x-y_i| \le 0.2;$$

$$\text{then}: x=y_i, \quad i \ge 0 \tag{1}$$

where x denotes the wind speed; y denotes the value of the wind state and i is positive integer that represents the iteration of wind state as shown in Table 2.

After the rounding process was completed, the Mycielski-3 algorithm can proceed further. The prediction starts with searching the latest historical data back to the earlier historical database. The algorithm will keep searching the exact same historical data until the pattern of searched data never appears in the history data set. The prediction value is the next value of the last best fit pattern.

For instance, let the data that need to be predicted be $\hat{x}[n+1]$, where n represents the number of the time series data sample. To predict the data precisely, the difference in the predicted and the actual value of should be minimal. The historical data, f_n, can be expressed as the function shown in Eq. (2).

$$f_n = (x[1], x[2], \ldots, x[n-1], x[n]) \tag{2}$$

The algorithm searching process starts from the latest data $x[n]$ back to $x[1]$. The main objective of this algorithm is to find the longest pattern of historical data that matches the pattern of the current data. Hence, the pattern searching is started from the nearest historical data; in this case, $x[n]$. If the value of $x[n]$ happened in the historical data, the algorithm will continue to search the pattern of $(x[n-1], x[n])$. The algorithm will make a prediction when no equivalent pattern can be found in the remaining database. For example, the pattern of $(x[n-1], x[n])$ happened in $(x[5], x[6])$ before, and no pattern of $(x[n-2], x[n-1], x[n])$ can be found in the historical data. Therefore, the algorithm will halt. The prediction will be accomplished by taking the value of $\hat{x}[n+1] = x[7]$, because the pattern after $(x[5], x[6])$ is $x[7]$. This can be explained by the main philosophy of this algorithm, which is the pattern of the historical data, and then the current data will be the same pattern. Referring to the previous works by Fidan et al. and Gerek [17,18], an equation was created to express the aforementioned prediction process. This particular equation is shown in Eq. (3).

$$m = \underset{L}{\operatorname{argmax}}\{x[k]=x[n], x[k-1]=x[n-1]\},$$
$$f_{n+1} = x[n+1] = x[m] \tag{3}$$

Although it was not written in the papers published by Mehmet Fidan et al., it is believed that the symbol L in the equation represents the location of the longest pattern found in the history. The prediction value will be the data after the longest pattern. Hence, the prediction value can be expressed as $\underset{L}{\operatorname{argmax}}[n-L+1]$. As mentioned earlier, the prediction value is $x[n+1]$, so the prediction value can be expressed as in Eq. (4).

$$x[n+1] = \underset{L}{\operatorname{argmax}}[n-L+1] \tag{4}$$

However, a looping error happened in the earlier version of the Mycielski approach. Hence, Mehmet Fidan et al. modified the approach by adding the random number range from -0.4 to $+0.4$

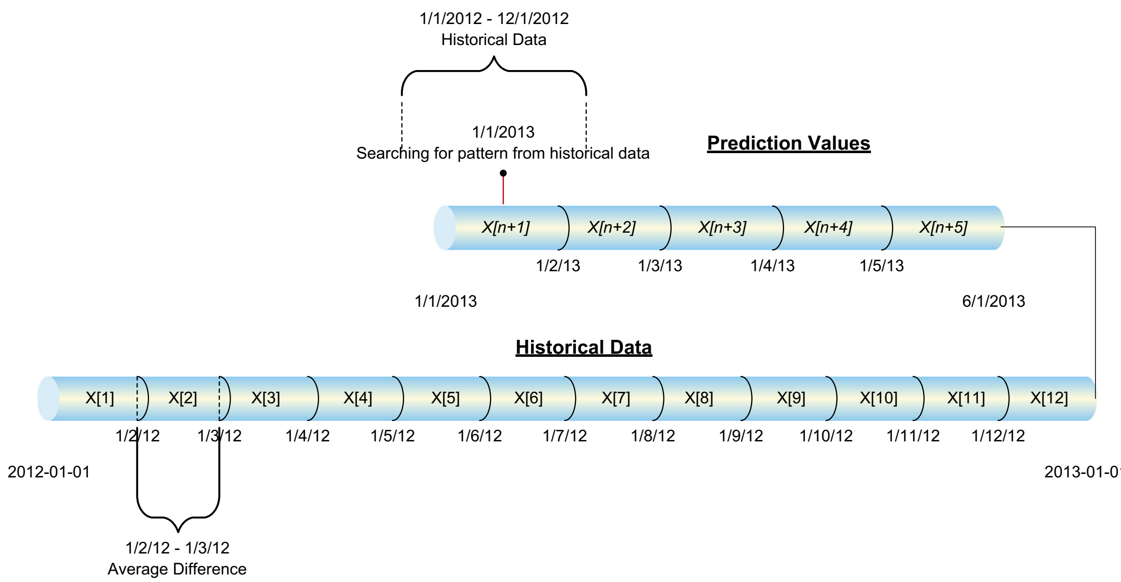

Fig. 1. The flow of Mycielski algorithm.

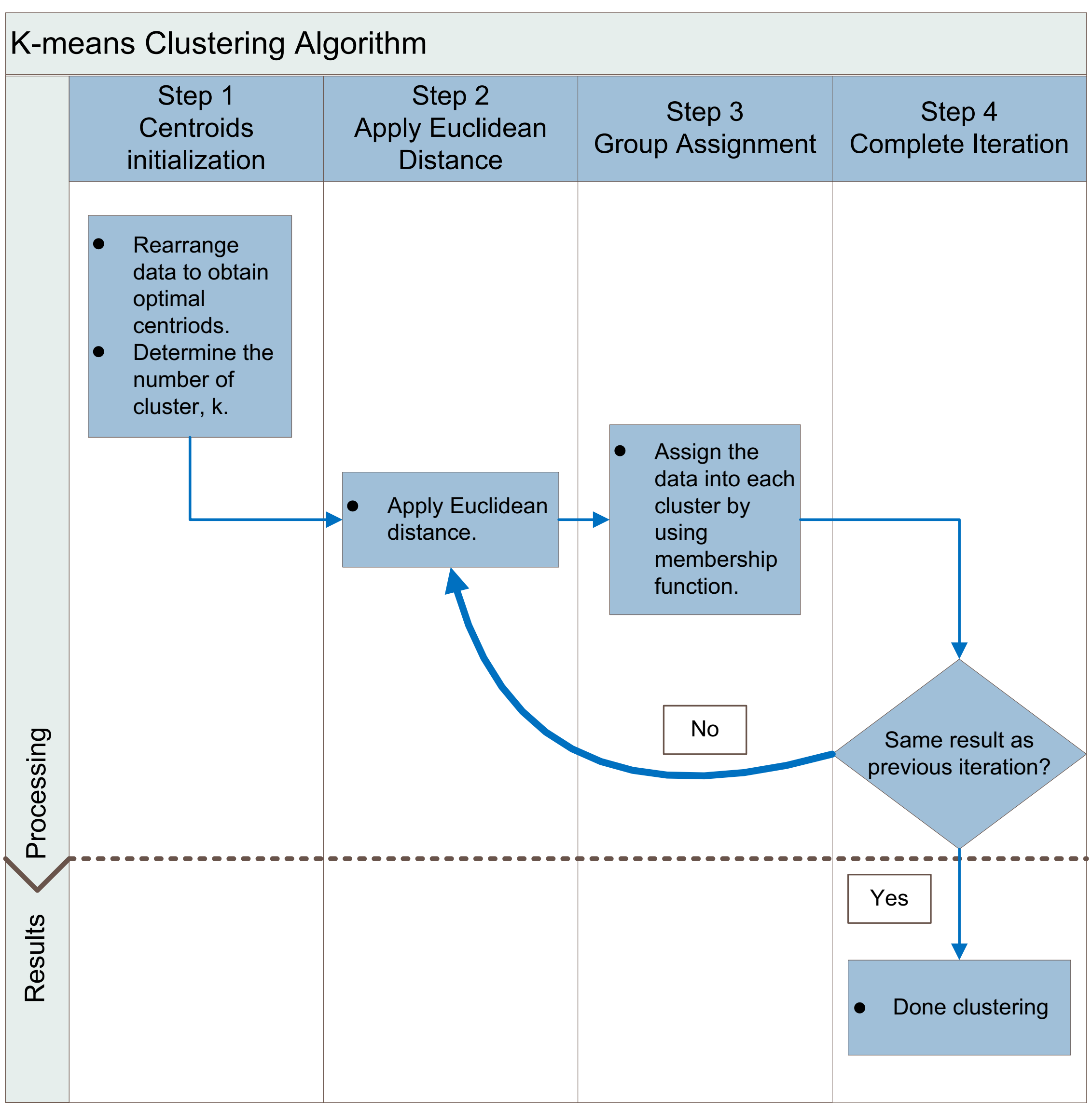

Fig. 2. The basic flow of *K*-means clustering algorithm.

to the predicted value to make the Mycielski-1 approach. In addition, the randomness of the number will cause the predicted result to become unreliable. As one of the contributions of this research, a novel approach to determining the random number is presented. The random number is found to be more reliable by obtaining the average difference, d_{avg} from the history. The principle of obtaining the d_{avg} is the same as the principle of Mycielski, which is the transitional behaviour of wind speed. Basically, the d_{avg} can be obtained by taking the difference between the months and finding the average for the past few years, depending on the database. The random number will be added into the predicted value. Once the prediction is completed, the predicted value will be updated in the historical data. The next prediction $\hat{x}[n+2]$ can be done based on the updated history data. This approach is named Mycielski-3. An illustration of the Mycielski-3 algorithm is shown in Fig. 1.

The random number will be added to the predicted value. Once the prediction is completed, the predicted value will be updated in the historical data. Last but not least, the next prediction $x[n+2]$ can be performed based on the updated historical data.

3.2. K-means clustering prediction

In this section, one of the innovative approaches in wind speed prediction is presented. *K*-means clustering is a technique that classifies cases into the most similar groups. However, it needs a long computation time because it uses manual calculations. There might be many calculation steps, or iterations, needed to find the suitable cluster. Therefore, the computation time for the iterations can be minimised with the aid of computer software. The statistical software, SPSS was used for the analysis of the clustering process. The software includes analysis methods for various types of required analysis. For the analysis of wind speed prediction, the first element to sort out is the number of clusters that could be generated from a series of 96 data.

The general idea of the *K*-means clustering algorithm is described in Fig. 2. The Euclidean distance is involved in the *K*-means process and is mathematically presented in Eq. (5).

$$d(x_t, c_r) = \sqrt{(x_t - c_r)^2} \tag{5}$$

where x_i denotes the data point, $x_t = x_1, x_2, x_3, \ldots, x_N$; and r denotes the centroid, .

The number of cluster, k, is initialised, and cluster assignment of the data can be represented by the group member function as stated in Eq. (5). The dataset ranges from x_1, x_2, x_3, … x_N and is represented in the general form of x_t. The x_t data are assigned into k clusters. The assignment of data into the jth cluster is presented in Eq. (6).

$$I(j|x_t) = \begin{cases} 1, \text{if} : j = d\min_{1 \le r \le k} \sqrt{(x_t - c_r)^2}; \\ 0, \text{otherwise} \end{cases} \tag{6}$$

As in Fig. 2, step 2 and step 3 repeat until the convergence of the centroids. When such a circumstance occurs, the centroid is considered stable and the optimal solution of the average distances is met. For implementation of *K*-means clustering in the prediction of wind speed, both manual and computer aid methods are utilised. SPSS is used to obtain the number of clusters, k, for 8 years of wind speed data in Kudat. Once the value for each cluster was defined, the computation of wind prediction could be achieved.

For the next step, the Euclidean distance in between the wind dataset and the cluster had to be obtained. This step purposely matches the wind data to a cluster. Before obtaining the second probability function, the frequency of each cluster that occurred in a particular month throughout 8 years is obtained. The number of clusters can be determined using a scree diagram. A scree diagram is a graph that consists of the latent root, which can judge the best number of factors from a factor analysis. The suitable cluster number is selected when there is a larger difference in the distance coefficient from one point to another and then subtracting from the number of total cases. Additionally, the numbers of the cluster can be identified using numerical observation of the distance coefficient.

3.3. Weibull estimation

The Weibull distribution is commonly known as a distribution that expresses wind speed frequency distribution well. More than five methods are being discussed by researchers around the world on the two Weibull parameters' estimation. There are advantages and disadvantages for the proposed methods. However, the most decisive criterion that determines the method to be used in the Weibull analysis is the provided dataset. Different time series of wind speed data in terms of hourly, daily or monthly sets might contribute to various analysis methods. For this research, two Weibull parameter's estimations are used to analyse wind speed data. The descriptions of the calculation steps for the parameters' estimation are presented in this section.

Under normal circumstances, there are two Weibull parameters: scale parameter, c (m s^{-1}) and shape parameter, k (dimensionless). Generally, the probability distribution function of the Weibull distribution is defined as in Eq. (7).

$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{7}$$

Normally, the scale parameter and shape parameter can be altered to the desirable height. However, it is stated in [46] that: "*the alterations are valid provided the terrain is relatively flat* …". Therefore, the surface roughness for Kudat might not be suitable for the alteration using a modification of the scale parameter and shape parameter to the corresponding height.

The relationship between the mean wind speed, $\overline{V}$, standard deviation, σ and k is defined in Eq. (8).

$$k = \left(\frac{\sigma}{\overline{V}}\right)^{-1.086} \tag{8}$$

The wind speed data are in monthly mean values. Therefore, the fifth method, as discussed in Section 2.4, is used. According to Justus et al. [44,46], a constant that relates mean wind speed and the shape parameter k. The variability of the wind speed in the lower stage (10th percentile), average and higher stage (90th percentile) could be expressed as in Eq. (9).

$$k = \begin{cases} 1.05\sqrt{\overline{V}}, \text{low} \\ 0.94\sqrt{\overline{V}}, \text{average} \\ 0.83\sqrt{\overline{V}}, \text{high} \end{cases} \tag{9}$$

Once the value of k is obtained, the value of the scale parameter c can be computed. The equation for c is defined in Eq. (10).

$$\overline{V} = c\Gamma\left(1 + \frac{1}{k}\right) \tag{10}$$

$\Gamma(1+\frac{1}{k})$ is the gamma function and can be solved by using the Stirling approximation. Let $\Gamma(1+\frac{1}{k})$ be $\Gamma(1+x)$. In the gamma function, $\Gamma(1+x)$ will be analysed as $x\Gamma(x)$. Therefore, the Stirling approximation for $\Gamma(1+x)$ is defined in Eq. (11). Alternatively, the expanded version of Eq. (11) as in Eq. (12) can also be applied.

$$\Gamma(x) = \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \left[1 + \frac{1}{2(6x)} + \frac{1}{2^3(6x)^2} - \frac{139}{2^3(2.3.5)(6x)^3} + \dots\right] \quad (11)$$

$$\Gamma(x) = \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \left[1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51,840x^3} + \dots\right] \quad (12)$$

4. Mycielski-3 wind prediction result

Based on wind potential analysis, Kudat was selected as the potential wind farm site. Therefore, the wind data of Kudat will be used as the source for wind speed prediction in this section. Freeware called FirstLook, was provided by 3TIER for the wind rank assessment. By using FirstLook, the wind rank for a particular place around the world can be obtained. This enables researchers to analyse the corresponding wind speed at a preliminary stage of wind power development. Basically, Malaysia has two areas highlighted in light green colour that possessed a higher potential to harness wind energy. The highest wind rank is near the Kudat coast, as shown in Fig. 3.

FirstLook only provided the wind rank but not the wind speed. Therefore, a complete analysis of the potential site essentially needs both wind data and wind rank. From data obtained from the Malaysian Meteorological Department, the actual place of the anemometer in Kudat is 6°55′N, 116°50′E, and it is equivalent to *lat*. 6.916, *long*. 116.83 in decimal. The global wind rank calculated by the FirstLook software is 66%, as shown in Fig. 4. For the project assessment, 3TIER claims that 80% of the wind project can be done if the wind rank is higher than 65%.

The location of the anemometer has a high wind rank, faces the Sulu Sea and is located at an airport. However, an airport is not a suitable location to install wind turbines, and therefore, the power plant should be installed away from the airport. It is found that the wind speed at the coast facing the South China Sea is higher than other locations. The global wind rank at the coastline facing the South China Sea is 80%. Because of changes in location, some amendments are needed for the mean wind speed provided by the Malaysian Meteorological Department to fulfil site conditions. Fig. 5 shows the proposed site's global wind rank, which was obtained using the Firstlook software. The distance from the anemometer station to the proposed site is approximately 25–30 km if using road access. However, the topography is quite different between the two sites. Therefore, the wind shear exponent could describe the wind regime in such a circumstance.

The initial wind rank at the anemometer station is 66% instead of 80%. Hence, the observed wind speed data in Kudat will be multiplied by 114% as defined by a 14% increment of the original values.

According to the methodology described in Section 3.1, the obtained wind data have to be rounded into the nearest wind

Fig. 4. Location of anemometer in Kudat.

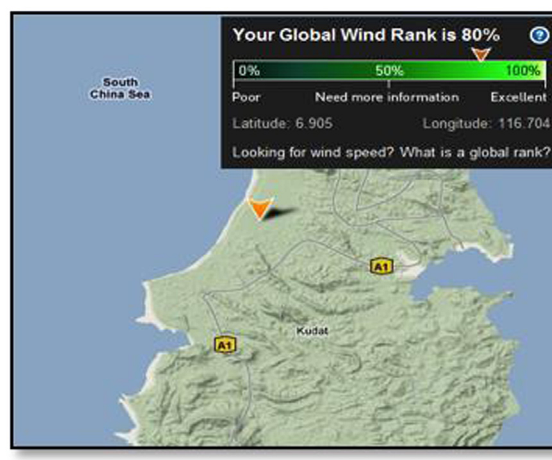

Fig. 5. Proposed location of wind turbine.

Fig. 3. The overview of wind rank in Malaysia.

Table 3
The wind speed data (m s^{-1}) for Kudat after rounding process.

Month year	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2002	*11.0*	*8.0*	*6.5*	*4.5*	*3.*	*3.5*	*5.5*	*6.5*	*4.0*	*4.0*	*3.0*	*7.0*
2003	*9.0*	*10.5*	*7.5*	*5.0*	*4.5*	*3.5*	*5.5*	*5.0*	*6.0*	*6.5*	*4.0*	*9.0*
2004	*11.0*	*7.5*	*5.5*	*5.0*	*5.0*	*7.5*	*4.5*	*6.5*	*3.0*	*6.5*	*3.5*	*4.5*
2005	*5.0*	*6.0*	*5.5*	*6.5*	*5.5*	*5.0*	*6.0*	*6.0*	*6.5*	*6.0*	*6.0*	*4.0*
2006	*6.5*	*9.0*	*6.0*	*5.5*	*5.5*	*5.5*	*5.5*	*6.5*	*6.0*	*7.0*	*5.5*	*6.0*
2007	*6.5*	*8.0*	*6.5*	*6.0*	*5.0*	*5.0*	*6.0*	*5.5*	*7.5*	*5.5*	*7.5*	*5.5*
2008	*6.5*	*6.5*	*5.5*	*5.0*	*6.5*	*5.0*	*5.5*	*4.5*	*6.5*	*5.0*	*4.5*	*5.0*
2009	*6.5*	*6.0*	*5.5*	*6.5*	*5.5*	*5.5*	*6.5*	*8.0*	*8.5*	*7.5*	*5.5*	*7.5*
2010	*8.0*	*9.0*	*8.0*	*6.5*	*5.5*	*5.0*	*5.0*	*6.0*	*5.0*	*6.0*	*5.0*	*5.0*

Table 4
The average difference of each month (m s^{-1}).

Month	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	–1.7	0.1	1.6	0.6	0.4	0	–0.6	–0.4	0.1	0	1.1	–1.1

states before proceeding to the Mycielski-3 prediction. The rounding process is based on Eq. (1). For the prediction requirement of the prediction, there are 16 wind states, as shown in Table 2. The purpose of the rounding process is to ensure the validity of the data during the prediction state. Too many wind states might cause recursion of the predicted data. The rounded modified mean wind speeds of Kudat from Year 2002 to Year 2009 are tabulated in Table 3 and will be used as the prediction database. The wind data for 2010 will be used as the verification dataset of the prediction efficiency. For easier assessment, the wind speed data are assigned to a sequence number.

To avoid the looping error, an average difference for each month is computed. By subtracting the earlier month from the later month, a value of either a positive or negative integer will be obtained. The average difference of each month can be calculated by averaging each month's summation for 8 years. The average difference is tabulated in Table 4 and will add to the predicted wind speed to create a random result.

By adding the average difference for January to the predicted value, the prediction result for January 2010 was obtained successfully. For data consistency, the predicted result will go through the rounding process using Eq. (1). The same prediction procedure applies to other months. The prediction result for 2010 is tabulated in Table 5.

Table 5
The predicted wind speed for 2010 (m s^{-1}).

Month	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	4	6.5	7.0	6.0	5.5	7.5	5.0	5.5	6.5	9.0	7.0	5.0

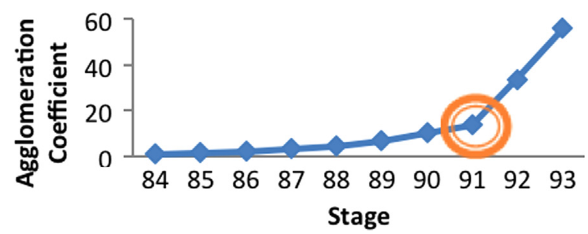

Fig. 6. The scree diagram of wind speed data in Kudat over eight years.

Table 6
The initial and final clusters' centre for the wind speed in Kudat.

	Cluster				
	1	2	3	4	5
Initial	10.9	8.8	6.7	4.7	2.9
Final	10.7	7.9	6.4	5.2	3.5

5. *K*-means clustering wind prediction result

A brief explanation of the *K*-means method was presented in Section 3.2. *K*-means implementation in wind speed prediction will be fully described in this section. To standardise the analysis, the wind speed database from the Mycielski-3 algorithm prediction was adapted in the *K*-means clustering.

According to calculation steps as presented in Section 3.2, the wind speed data in Table 3 were tabulated in SPSS. The agglomeration schedule for the data was obtained. These data are purposely used to plot the scree diagram to determine the number of clusters involved in the wind speed prediction analysis. By using a graph plotter function in Microsoft Excel, the scree diagram, with respect to the stage number is plotted in Fig. 6. The orange-coloured circle in Fig. 6 indicates the numbers of the stages that started to become a steeper graph because a larger number of cases are joined together into significant clusters. Therefore, the ideal number of clusters for the analysis is five. There are 96 cases in total. By subtracting 91 from the total cases, there are five clusters.

The *K*-means clustering analysis for the wind speed in Kudat was then executed. The number of clusters was set to five. The initial cluster and final cluster of the wind speed in Kudat were obtained and shown in Table 6. From observation, there are minor changes in the clusters' centre in between the initial process and final result. This was because the *K*-means clustering algorithm was finding the best mean value that could represent the entire cluster. For the prediction of wind speed in Kudat, the result for the final clusters' centre will be used as the best mean values for those five clusters. For 8 years of data, there will be a total of 96 numbers of wind speed data recorded.

To obtain the probability occurrence for a particular cluster, the number of cases in each cluster should be obtained. As in Table 7, the total number of cases is $n=96$. The probability of occurrence, $P(E_1)$, can be easily computed. The number of total cases is divided by the number of cases for each cluster and results in the probability of occurrence in Table 8.

The K-means step in finding the Euclidean distance is a complicated calculation step. It needs to compute 96 data cases to sort out the distance between the corresponding wind speeds to the cluster centre. Because of this, the programming in Microsoft Excel was utilised to handle the bulk calculation. Relevant data were entered into the programme's user interface.

The probability of the significant cluster, $P(E_2)$, can be calculated using the aforementioned data. This probability indicates the frequency with which the same wind speed occurred in the respective cluster, in particular, in a month throughout 8 years. The most probable wind speed is obtained by multiplying both $P(E_1)$ and $P(E_2)$ as tabulated in Table 9. The highest value of probability will be treated as the predicted value by adding the random number that was obtained from the average difference. The predicted wind speed for Kudat in 2010 by using K-means clustering is tabulated in Table 10.

6. Weibull distribution assessment

The estimation of wind speed for 2010 in Kudat using a Weibull distribution will be conducted in this section. The estimation for wind speed data in 2010 required observed wind speed data to compute the Weibull parameters. The wind speed data in Kudat from 2002 to 2009 are categorised into a few stages. The classification of the categories is shown in Tables 11 and 12. To obtain the shape parameter k, Eq. (9) is used. Because the shape parameters are obtained in range, the scale parameters for Kudat are also in range. The scale parameters are computed corresponding to the wind speed of the particular month with respect to the shape parameters of the month. The data for both parameters are tabulated in Table 13.

Table 7
Number of cases for cluster 1 to cluster 5.

Cluster	Number of cases, x_i	Range (m s^{-1})
1	$x_1=3$	10.4–10.9
2	$x_2=14$	7.3–9.1
3	$x_3=32$	6.0–7.0
4	$x_4=37$	4.7–5.7
5	$x_5=10$	2.9–4.1

Table 8
The probability of occurrence for five clusters.

Cluster	$P(E_1)$
1	0.0313
2	0.1458
3	0.3333
4	0.3854
5	0.1042

The wind speed data are rearranged to extract useful information and plot the Weibull probability density function graph. By using Eq. (7), the Weibull probability density function can be obtained. The wind speed data are computed to achieve statistical requirements and are tabulated in Table 14. The graph of the observed and Weibull wind speed frequencies of Kudat in 2010 are plotted in Fig. 7. As in the figure, the low percentile of the Weibull parameters (purple-coloured line) represents the wind speed frequency in Kudat better than the average Weibull (green-coloured line) and high percentile Weibull (red-coloured line).

7. Discussion

The aforementioned algorithms were using spreadsheet to generate the result. The preset formula will interlink to each other when the data is keyed in the data insertion area. The computation time is highly depending on the insertion time by user. However, K-means clustering algorithm required extra step to sort the scree diagram. Currently, the average computation time for Mycielski algorithm is about 20 s inclusive of data insertion; meanwhile, average computation time for K-means clustering is about 1 min due to extra step involved. The comparison is visualised in Table 15. Nevertheless, a one-stop-computation software is currently developing by the authors. By using Visual Basic Advance (VBA) language, all related functions can be incorporated into only one software and the computation time can reduce to 1 s or even lesser.

The results from both the Mycielski-3 and K-means clustering are shown in Table 16. In terms of graphical analysis, the results for prediction are interpreted as shown in Fig. 8. From the figure, the plot of the line graph for K-means clustering describes quite well the wind speed for Kudat in 2010. The red line shows the highest frequency at 6 m s^{-1}. However, it over-predicted the majority of the site's wind speed. Meanwhile, the graph of the Mycielski-3 prediction was over-predicted at 8 m s^{-1}. Conversely, Weibull estimation by using the low percentile tended to be lower frequencies of the majority wind speed. The frequency in the analysis indicated the number of times that the wind blew at such a speed.

For this research, Weibull analysis is used as the comparison method for the proposed methodologies. Weibull can represent the wind profile but not the exact wind speed. Therefore, the prediction is based on the method that obtained higher accuracy from either Mycielski-3 or K-means clustering.

In the context of wind energy that could be harnessed as electrical energy, the power of the observed and predicted wind speeds is analysed and compared. The differences in wind speed are computed using RMSE. The RMSE formula is defined in Eq. (13). The x_i in the equation denotes the observed wind speed, whereas $\hat{x}_i$ denotes the predicted wind speed. The total number of data is expressed in N. The results of the comparison are shown in

Table 9
The percentage of the probability for predicted wind speed cluster 2010.

Cluster	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	0.78	0.39	0	0	0	0	0	0	0	0	0	0
2	1.82	7.29	1.82	0	0	1.82	0	1.82	3.65	1.82	1.82	3.65
3	16.67	12.50	12.50	12.50	4.17	0	12.50	16.67	16.67	16.67	4.17	8.33
4	4.82	0	19.27	24.09	28.91	24.09	24.09	14.45	0	9.64	14.45	14.45
5	0	0	0	0	1.30	2.61	0	0	2.61	1.30	3.91	1.30

Table 10
Result of *K*-means clustering wind speed prediction for Kudat in 2010.

Cluster	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Centroid	6.4	6.4	5.2	5.2	5.2	5.2	5.2	6.4	6.4	6.4	5.2	5.2
Random Numbers	−1.7	0.1	1.7	0.5	0.6	0	0.6	0.5	2	0	1.0	1.1
Prediction	4.7	6.5	6.9	5.7	5.8	5.2	4.6	5.9	6.6	6.4	6.2	4.1

Table 11
Classification of the wind speed categories.

i	1	2	3	4	5	6	7	8	9
Interval	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11

Table 12
Statistical summary of the wind speed categories.

i	Frequency	Percent	Cumulative percent
1	1	1.0	1.0
2	7	7.3	8.3
3	15	15.6	24.0
4	24	25.0	49.0
5	30	31.2	80.2
6	10	10.4	90.6
7	5	5.2	95.8
8	1	1.0	96.9
9	3	3.1	100.0
Total	96	100.0	

Table 13
The Weibull two parameters for Kudat from 2002 to 2010.

Year	*Shape parameter, k (Dimensionless)*			*Scale parameter, c* (ms^{-1})		
	Low percentile	*Average*	*High percentile*	*Low percentile*	*Average*	*High percentile*
2002	2.43	2.17	1.92	6.36	6.33	6.29
2003	2.61	2.33	2.06	7.26	7.23	7.19
2004	2.50	2.24	1.98	6.71	6.68	6.64
2005	2.51	2.25	1.99	6.57	6.55	6.52
2006	2.61	2.34	2.07	7.12	7.10	7.07
2007	2.62	2.34	2.07	7.15	7.12	7.10
2008	2.47	2.21	1.96	6.37	6.35	6.32
2009	2.69	2.41	2.13	7.58	7.55	7.52
2010	2.61	2.33	2.06	7.13	7.10	7.07

Table 14
The time series data with respect to Weibull probability density function for Kudat in 2010.

i	Frequency	Percent	Cumulative percent	Weibull *pdf* (*low*)	Weibull *pdf* (*avg*)	Weibull *pdf* (*high*)
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	2	16.7	16.7	0.138	0.132	0.124
4	4	33.3	50.0	0.143	0.134	0.124
5	3	25.0	75.0	0.145	0.131	0.117
6	1	8.3	83.3	0.120	0.108	0.095
7	2	16.7	100.0	0.102	0.093	0.084
8	0	0	100.0	0	0	0
9	0	0	100.0	0	0	0
Total	12	100.0				

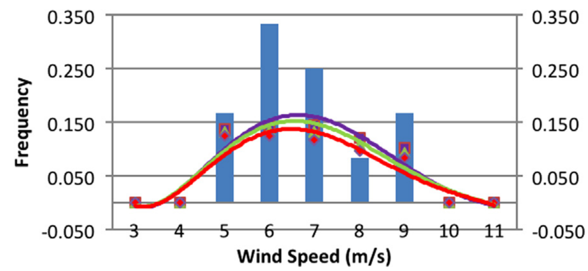

Fig. 7. Lines plot for Weibull frequencies in Kudat for year 2010.

Table 15
Computation time of algorithm.

	Mycielski	***K*-means Clustering**
Computation time, s	20	60

Table 16
Results of predicted wind speed for Kudat 2010.

Month	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mycielski	4	6.5	7.0	6.0	5.5	7.5	5.0	5.5	6.5	9.0	7.0	5.0
K-means	4.7	6.5	6.9	5.7	5.8	5.2	4.6	5.9	6.6	6.4	6.2	4.1
Observed	8.0	8.8	7.8	6.5	5.7	5.2	4.9	6.2	5.2	6.2	4.9	5.2

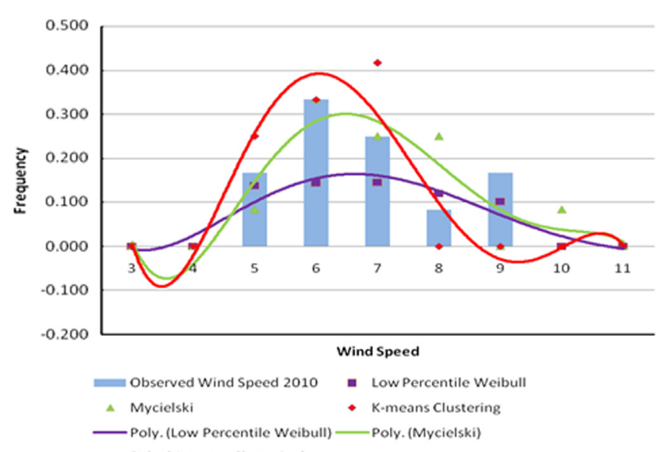

Fig. 8. The frequencies distribution of observed wind speed and predicted wind speed for Kudat in 2010.

Table 17
The comparison of the observed data and prediction.

Methods	*Average Wind Speed* ($m\,s^{-1}$)	*RMSE*	*Power* ($kWh\,m^{-2}\,year^{-1}$)
Observed	6.2	NA	1481.13
Mycielski	6.2	1.875	1479.54
K-means	5.7	1.391	1075.79

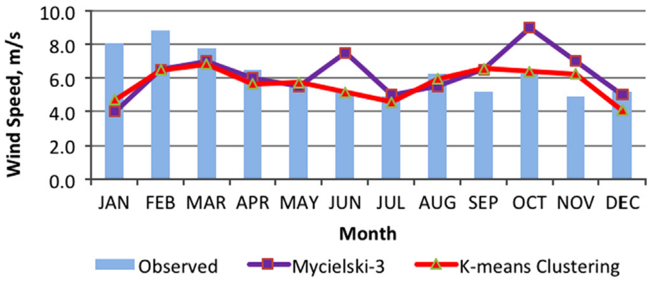

Fig. 9. The combine graph analysis for observed and prediction wind speed.

Table 17.

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)}{N}} \tag{13}$$

From Table 17, the average wind speed for the observed and Mycielski-3 algorithm matched perfectly. However, the line graph shown in Fig. 9 indicates that the Mycielski-3 prediction over predicted the wind speed in June and October. The larger gap in the prediction of the Mycielski-3 of these 2 months compensated for the under-predicted wind speed in January. Meanwhile, *K*-means clustering does not consist of any over-predictions that could compensate for the error in the first month. Although *K*-means clustering does not match the average wind speed, the RMSE value is significantly smaller than the Mycielski-3 prediction. Hence, *K*-means clustering prediction is preferable for this research as the RMSE of the method is the smallest. Additionally, the wind profile of *K*-means clustering is more similar to the observed wind speed when compared with the Mycielski-3 algorithm.

The predicted wind speed data using *K*-means clustering can be used as the data input for wind power system analysis. In such circumstances, a wind power system that includes overall system analysis should be performed beforehand. In this research, the Weibull distribution is not as suitable as the *K*-means clustering because the data sampling obtained lacked some statistical data. Therefore, the *K*-means clustering approach is more suitable when compared with others.

8. Conclusion

To summarise, the prediction of wind speed is necessary to initiate a new wind power plant in a new potential area. For this study, Kudat is a potential area that can be fully utilised in constructing a new wind power plant in Malaysia. To build a wind power plant, there might be some crucial criteria that need to be determined before commencement of the project. The criteria to be considered in a project included potential electricity that could be generated, selection of a wind turbine, cash flow analysis of the project and project management. Each criterion is highly related with the others.

The wind speed prediction performed in this research can be used as a decision maker's preliminary planning to convince investors to provide funding for relevant assessment at a potential site. There were two innovative approaches introduced: Mycielski-3 and *K*-means clustering. On the one hand, the random number approach in the Mycielski-3 developed successfully and increased the reliability of creating randomness of wind behaviour. On the other hand, *K*-means clustering that categorised wind speed into a few clusters gave a better prediction result compared with Mycielski-3 in terms of RMSE analysis.

Acknowledgement

The authors would like to thank the Ministry of Higher Education Malaysia (MOHE) for giving the scholarship to the author to accomplish this research. Furthermore, the authors would like to thank the Ministry of Education, Malaysia (MOE), the Ministry of Science, Technology and Innovation Malaysia (MOSTI) and The Office for Research, Innovation, Commercialization, Consultancy Management (ORICC), UTHM for financially supporting this research under the Fundamental Research Grant Scheme (FRGS) Grant no. 1417, EScience Fund no. S023 and IGSP Vot. U242 in funding this research.

References

[1] EIA, International Energy Statistics. Independent statistic & analysis. U.S. Energy Information Administration, EIA; 2012.
[2] Clean energy: air emissions. Available from: 〈http://www.epa.gov/cleanenergy/energy-and-you/affect/air-emissions.html〉; 5th July 2013.
[3] REN21. Renewable energy policy network for the 21st century. Renewables 2012 global status report. France; 2012.
[4] GWEC. Global wind statistics. Global Wind Energy Council; 2012.
[5] Bloomberg. Renewable energy now cheaper than new fossil fuels in Australia. Available from: 〈http://about.bnef.com/2013/02/07/renewable-energy-now-cheaper-than-new-fossil-fuels-in-australia/〉 [cited 20 February 2013].
[6] Xingang Z, Zhongfu T, Pingkuo L. Development goal of 30 GW for China's biomass power generation: will it be achieved? Renew Sustain Energy Rev 2013;25:310–7.
[7] Martin NJ, Rice JL. Developing renewable energy supply in Queensland, Australia: a study of the barriers, targets, policies and actions. Renew Energy 2012;44(0):119–27.
[8] Capik M, Yilmaz AO, Cavusoglu I. Present situation and potential role of renewable energy in Turkey. Renew Energy 2012;46(0):1–13.
[9] Jennings P. New directions in renewable energy education. Renew Energy 2009;34(2):435–9.
[10] Sims REH, Rogner H-H, Gregory K. Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. Energy Policy 2003;31(13):1315–26.
[11] KeTTHA. National renewable energy policy and action plan, Malaysia; 2008.
[12] Sopian K, Othman MYH, Wirsat A. The wind energy potential of Malaysia. Renew Energy 1995;6(8):1005–16.
[13] Islam MR, Saidur R, Rahim NA. Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function. Energy 2011;36(2):985–92.
[14] Goh HH, et al. Wind farm allocation in Malaysia based on multi-criteria decision making method. In: Proceedings of the national postgraduate conference (NPC); 2011.
[15] Bivona S, Burlon R, Leone C. Hourly wind speed analysis in Sicily. Renew Energy 2003;28(9):1371–85.
[16] Louka P, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerodyn 2008;96 (12):2348–62.
[17] Hocaoglu FO, Fidan M, Gerek ON. Mycielski approach for wind speed prediction. Energy Convers Manag 2009;50(6):1436–43.
[18] Fidan M, Hocaoğlu FO, Gerek ÖN. Improved synthetic wind speed generation using modified Mycielski approach. Int J Energy Res 2011.
[19] MacQueen J. Some methods for classification and analysis of multivariate observations. InL Proceedings of 5th Berkeley symposium on mathematical statistics and probability. University of California Press; 1967.
[20] Lloyd S. Least square quantization in PCM. IEEE Trans Inf Theory 1982;28 (2):129–37.
[21] Cao J, et al. Towards information-theoretic K-means clustering for image indexing. Signal Process 2012;93(7):2026–37.
[22] Xu L. How many clusters?: a Ying-Yang machine based theory for a classical open problem in pattern recognition. In: Proceedings of IEEE international conference on neural networks; 1996.
[23] An F, Mattausch HJ. K-means clustering algorithm for multimedia applications with flexible HW/SW co-design. J Syst Archit 2012:155–64.
[24] Di Fatta G, et al. Fault tolerant decentralised-means clustering for asynchronous large-scale networks. J Parallel Distrib Comput 2012;73(3):317–29.
[25] Pingping S, et al. Application of improved K-mean clustering in predicting protein–protein interactions. In: Proceedings of the international conference on biomedical engineering and informatics; 2008.
[26] Su X, Suo J. Prediction of sea clutter based on chaos theory with RBF and K-mean clustering. In: Proceedings of the international conference on radar; 2006.
[27] Xu L, Krzyzak A, Oja E. Rival penalized competitive learning for clustering analysis, RBF Net, and curve detection. IEEE Trans Neural Netw 1993;4 (4):636–49.

[28] Asamer J, Din K. Prediction of velocities on motorways by k-means clustering. In: Proceedings of the seventh Mexican international conference on artificial intelligence; 2008.
[29] Bishnu PS, Bhattacherjee V. Software fault prediction using quad tree-based K-means clustering algorithm. IEEE Trans Knowl Data Eng 2011;24(6):1146–50.
[30] Kusiak A, Li W. Short-term prediction of wind power with a clustering approach. Renew Energy 2010;35(10):2362–9.
[31] Kusiak A, Haiyang Z, Zhe S. Short-term prediction of wind farm power: a data mining approach. IEEE Trans Energy Convers 2009;24(1):125–36.
[32] Cheung Y-M. k-Means: a new generalized k-means clustering algorithm. Pattern Recognit Lett 2003;24(15):2883–93.
[33] Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. Pattern Recognit 2003;36(2):451–61.
[34] Budura G, Botoca C, Miclau N. Competitive learning algorithm for data clustering. Facta Universitatis Series; 2006. p. 261–9.
[35] Bagirov AM. Modified global -means algorithm for minimum sum-of-squares clustering problems. Pattern Recognit 2008;41(10):3192–9.
[36] Zalik KR. An efficient k-means clustering algorithm. Pattern Recognit Lett 2008;29(9):1385–91.
[37] Zhang Z, Zhang J, Xue H. Improved K-means clustering algorithm. In: Proceedings of Congress on Image and Signal Processing; 2008.
[38] Lai JZC, Huang T-J. Fast global k-means clustering using cluster membership and inequality. Pattern Recognit 2009;43(5):1954–63.
[39] Lai JZC, Huang T-J, Liaw Y-C. A fast -means clustering algorithm using cluster center displacement. Pattern Recognit 2009;42(11):2551–6.
[40] Chiang M-C, Tsai C-W, Yang C-S. A time-efficient pattern reduction algorithm for k-means clustering. Inf Sci 2010;181(4):716–31.
[41] Juntao, W, Xiaolong S. An improved K-Means clustering algorithm. In: Proceedings of the IEEE 3rd international conference on communication software and networks (ICCSN); 2011.
[42] Wang Y, Li, W, Gao R. An improved k-means clustering algorithm. In: Proceedings of world automation congress (WAC); 2012.
[43] Davenport AG. The relationship of wind structure to wind loading. In: Proceedings conference on wind effects on structures. London, England: National Physical Laboratory; 1963 p. 19–82.
[44] Justus CG, et al. Methods for estimating wind speed frequency distributions. J Appl Meteorol 1977;17:350–3.
[45] Hennessey JPJ. Some aspects of wind power statistics. J Appl Meteorol 1977;16:119–28.
[46] Justus CG, Hargraves WR, Mikhail A. Reference wind speed distributions and height profiles for wind turbine design and performance evaluation applications; 1976.
[47] Akdag SA, Dinler A. A new method to estimate Weibull parameters for wind energy applications. Energy Convers Manag 2009;50(7):1761–6.
[48] Chang TP. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. Appl Energy 2010;88(1):272–82.
[49] Carta JA, Ramirez P. Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. Renew Energy 2007;32(3):518–31.
[50] Touw AE. Bayesian estimation of mixed Weibull distributions. Reliab Eng Syst Saf 2009;94(2):463–73.
[51] Mahdi S, Ashkar F. Exploring generalized probability weighted moments, generalized moments and maximum likelihood estimating methods in two-parameter Weibull model. J Hydrol 2004;285(1):62–75.
[52] Watkins AJ. On maximum likelihood estimation for the two parameter Weibull distribution. Microelectron Reliab 1996;36(5):595–603.
[53] Safari B, Gasore J. A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda. Renew Energy 2010;35(12):2874–80.
[54] Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J Wind Eng Ind Aerodyn 2000;85(1):75–84.
[55] Celik AN. Energy output estimation for small scale wind power generators using Weibull representative wind data. J Wind Eng Ind Aerodyn 2003;91(5):693–707.
[56] Toure S. Investigations on the Eigen-coordinates method for the 2-parameter weibull distribution of wind speed. Renew Energy 2005;30(4):511–21.