Wind Prediction in Malaysia

S. W. Lee, B. C. Kok, K. C. Goh and H. H. Goh

Abstract The wind speed prediction in Kudat, Malaysia had been done by using Mycielski-1 approach and K-mean clustering statistical method. There is some improvement in obtaining the random number of Mycielski-1. Besides, the comparison of K-means clustering with the optimal number of K is presented in this paper. The wind prediction is important to study a favorable site's wind potential. The prediction is based on 3 years history data provided by Meteorology Department of Malaysia and 1 year data as the reference to check the accuracy of both algorithms. The basic concept of Mycielski-1 algorithm is to predict the next value by looking to history data. Meanwhile, the K-means clustering can group the values with similar mean into the same group, and the prediction can be done by getting the probability of occurrence. The result shows the prediction of Mycielski-1 algorithm and K-means clustering are promising. The wind speed is predicted in order to obtain the mean power for energy planning.

Keywords Wind prediction · Mycielski-1 · K-means clustering · Wind power in Malaysia · Energy planning

1 Introduction

The growth of renewable energy around the world became significant since last decade. The existences of renewable energy enable mankind to harness and convert it to electricity to reduce the dependency on fuel. There are various kinds of energy resources. Wind energy is one of the renewable energies which has high

S. W. Lee · B. C. Kok · K. C. Goh · H. H. Goh (✉)
Department of Electrical Power Engineering, Faculty Electrical
and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja,
86400 Batu Pahat, Johor, Malaysia
e-mail: hhgoh@uthm.edu.my

possibility to be harnessed in Malaysia. Malaysia is a country without four seasons. However, Malaysia encounters two monsoon seasons. Southwest Monsoon starts from late May to September, meanwhile Northeast Monsoon starts from November to March [1]. Due to the Northeast Monsoon, there will be a heavy downpour in the east coast area of Peninsular Malaysia and West Sarawak. During the raining season, the wind speed in certain states will be relatively high.

According to the Ministry of Energy, Green Technology and Water Malaysia, it is targeted to install 350 MW of renewable energy in Malaysia which is 300 MW in West Peninsular Malaysia and 50 MW from Sabah under Ninth Malaysia Plan. Unfortunately, wind speed in Malaysia is not consistence. For the very beginning stage on developing the wind power plant in Malaysia, wind prediction is necessary in order to have a well energy planning in a potential site. Estimation of the wind power generated can be obtained when the wind speed is predicted.

One of the cities in Sabah was chosen for this research purpose. Kudat has a population of 33,378 people with the geographical location at latitude of 6.9°N and longitude of 116.84°E. Kudat is located at 131 km from Kota Kinabalu, the capital of Sabah. By using the 3TIER software, the wind speed at the seaside faced to South China Sea in Kudat is relatively high if compared with other states [2]. Previously, there was a study about wind energy potential in Kudat and Labuan [3]. Weibull distribution function is used to determine the wind profile. Yearly wind power density of the particular places were obtained. Besides, the Perhentian Island in peninsular of Malaysia also can be used as the location to harness the solar and wind energy [4]. The hybrid power generation had been also studied by Onar et al. [5].

In addition, there are a lot of studies regardless to wind turbine and energy utilisation [6, 7]. For instance, load frequency control can be fully utilised to a grid connected wind farm. Apart from that, the wind power system reliability is very important in the context of harnessing wind power [8, 9]. The system reliability should be known before the alternative energy could be utilised due to the big amount of expenses on wind farm maintenance. Hence, wind map or wind capacity of a location is important so that the investment and efforts put on it is worthy [10, 11]. In terms of energy planning, the output prediction of wind power generated can be predicted so that the proper wind site can be decided [12, 13].

2 Background of Study

Wind prediction is important before any wind power plant project can be started. Potential benefit of a wind site can be retrieved by manufacturers even there was no local data for a particular site. Estimation of that particular site can be done by obtaining the data from the nearest wind measuring station. Beccali et al. [14] Serporta had used a neural network method, which called Generalized Mapping Regressor (GMR) to generate the wind velocity through the training to the neurons for Sicily. So, the wind speed behaviour must be obtained in order to utilise this approach. Authors used Matlab to generate coding for the estimation [14].

In terms of wind speed analysis, Bivona et al. [15] analyzed hourly wind speed in Sicily by using Weibull distribution function. They used the meteorology department's wind data to study on the characteristics of wind speed at nine locations in Sicily. In that particular research, the fitness of the Weibull distribution function to the wind speed was clearly be seen.

Apart of that, Louka et al. [16] introduced some modelling that applied in Greece. For instance, University of Athens used SKIRON modeling to forecast the wind speed for 5 days ahead. At the same time, Regional Atmospheric Modeling System (RAMS) was developed at Colorado State University and Mission Research Inc. ASTeR Division. RAMS can forecast the wind in 48 h later. In addition, adaptive fuzzy neural network (F-NN) also applied for the wind power prediction for 120 h ahead. However, these methods cannot handle the systematic errors which are caused by the local adaptation problems. The authors proposed Kalman filtering to improve the performance of aforementioned methods. It is one of the statistical optimal sequential estimation procedures for dynamic systems. Results show that the systematic errors can be eliminated by using Kalman filtering.

For this paper, the prediction of the wind speed is done by using Mycielski algorithm and K-mean clustering statistical method. Mycielski approach is a data compression method which has been widely used in communication engineering. This method fully utilises the history data as the reference for the prediction value. Mycielski method is actually the advance version of the Limpel Ziv (LZ). The research on hourly wind speed prediction in Turkey had been done by Hocaoglu et al. by using Mycielski approach [17]. It is a new approach in wind power prediction. Authors had analysed and predicted the wind speed in Kayseri, Izmir and Antalya. The result of prediction is promising and very close to the actual data. The comparison of data fitting for both actual and predicted data had been done by using Weibull distribution function. The comparison had proven the accuracy of the predicted result.

In 2011, same group of researchers had modified the algorithm in order to solve the looping problem by adding the random number into the predicted data [18]. The modified algorithm called Mycielski-1 and Mycielski-2. In Mycielski-1, the random number in between −0.4 and 0.4 is added into the predicted value. This random number can be changed according to the requirement of the research. Meanwhile, the history data were rounded to the nearest integer number and divided into a few cluster in Mycielski-2. The prediction is done by randomly select the history data from different cluster. In addition, authors also made the comparison between the modified Mycielski approach and Markov chain.

On the other hand, the K-mean clustering is a statistical method that arranges all data into few clusters. The number of cluster is determined by the user. However, there was a research shown that the optimal number of clusters is 4. This is due to the quantization error for 4 number of clusters is reduced dramatically if compare with 1–3 clusters [19]. J. Asamer and K. Din used K-mean clustering to predict the velocities on motorways. In the particular paper, the cluster is divided into

4 cluster centre. Besides, there are some researchers like Bishnu and Bhattacherjee [20] used the K-means clustering together with neural network algorithm to predict the software fault.

Once the prediction of the wind speed had been done, the suitable wind turbine for Malaysia can be chosen by referring to the international wind turbine specification. So, government can seek for the professional opinion from the wind turbine providers regarding the suitable wind turbine. According to the international wind turbine specification, there are IECI, II, and III which means high, medium and low wind respectively.

3 Methodology

According to the meteorological data obtained from the meteorology department Malaysia, the anemometer at Kudat was placed at latitude 6°55′N and longitude 116°50′E. It is also equivalent to latitude 6.916N and longitude 116.83E. By using FirstLook software provided by 3TIER, the actual place of the location of the anemometer can be found. The location of the anemometer in Kudat is shown in Table 1. The global wind rank calculated by FirstLook software is 66 %. 3TIER claims that 80 % of the wind project can be done if the wind rank is higher than 65 %. In addition, most of the wind projects can be found at the area which has the wind rank from 80 to 90 % (Fig. 1).

The location of anemometer is facing the Sulu Sea. However, the wind speed is higher at the coast face to South China Sea. The global wind rank at the coast facing South China Sea is 80 %. Hence, some amendments are needed to be done on the mean wind speed provided by the Meteorology Department Malaysia. Table 2 shows the higher global wind rank obtained by using Firstlook software (Fig. 2).

According to the wind profile power law, the wind speed listed in Table 1 can be further adjusted to height of 100 m. This is because the wind speed is higher at higher altitude. The adjustment is based on the formula:

$$\frac{u_x}{u_r} = \left(\frac{z_x}{z_r}\right)^{\alpha} \tag{1}$$

Table 1 The adjusted wind speed (ms^{-1}) to 80 % global wind rank over 4 years

Mth	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
Year													
2007	2.9	3.5	3.0	2.6	2.2	2.3	2.6	2.5	3.2	2.5	3.3	2.4	2.7
2008	2.9	3.0	2.4	2.3	2.9	2.2	2.5	2.1	2.9	2.2	2.1	2.3	2.5
2009	3.0	2.7	2.4	3.0	2.5	2.5	2.9	3.4	3.6	3.2	2.4	3.3	3.0
2010	3.5	3.9	3.4	2.9	2.5	2.3	2.2	2.7	2.3	2.7	2.2	2.3	2.7

Fig. 1 The location of anemometer in Kudat
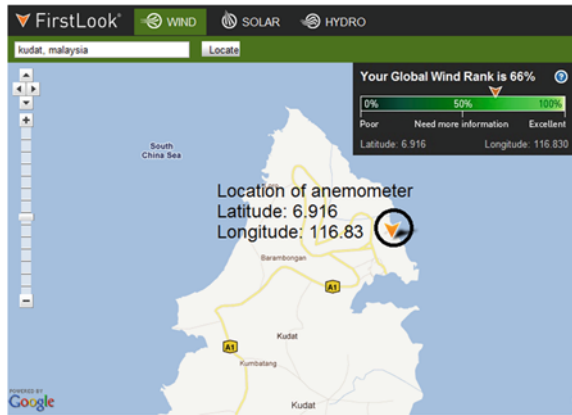
where u_x is the wind speed (ms^{-1}) at height of z_x (m). u_r denotes the reference wind speed (ms^{-1}) at reference height z_r (m). α denotes the constant of wind assessment during neutral stability which is approximately $\frac{1}{7}$ or 0.143.

3.1 Mycielski Algorithm

First of all, the recorded wind data has to be rounded into nearest wind states before proceeds to the Mycielski algorithm. This step is purposely making the searching process simpler. The rounding is based on the Eq. (2).

$$|x - y_i| \le 0.2; x = y_i, i \ge 0 \tag{2}$$

where x denotes the wind speed; y denotes the value of wind state i is positive integer represents the iteration of wind state.

For the requirement of this research, there are 12 wind states in total. The wind state can be tabulated in Table 2.

After the rounding process had done, Mycielski algorithm can be preceded. Mycielski is an algorithm which performs the prediction on the time series data. By using the history data, the prediction is done based on the data. The prediction starts with searching the latest history data back to the earlier history data. The algorithm will keep searching the exact same history data until the pattern of searched data never appear in the history data set. Hence, the prediction is done by taking the data before the algorithm stop. For instance, let the data that need to be predicted $\hat{x}[n + 1]$, where n represents the number of time series data sample.

In order to predict the data precisely, the difference of predicted $\hat{x}[n + 1]$ and the actual value of $\hat{x}[n + 1]$ should be minimum. The history data, f_n can be expressed as the function as shown in Eq. (3)

Table 2 The iteration of wind state

i	0	1	2	3	4	5	6	7	8	9	10	11	12
y_i (ms^{-1})	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0

Fig. 2 The proposed location of wind turbine

$$f_n = (x[1], x[2], \ldots, x[n-1], x[n]) \tag{3}$$

The searching process of the algorithm is started from the latest data $x[n]$ back to $x[1]$. The main objective of this algorithm is to find the longest pattern of the history data which match the pattern of current data. Hence, the pattern searching is started from the shortest history data, in this case will be $x[n]$. If the value of $x[n]$ was happened in the history data, the algorithm will continue to search the pattern of $(x[n-1], x[n])$. The algorithm will make a prediction when there is no same pattern can be found in the remaining history data sample. For example, the pattern of $(x[n-1], x[n])$ was happened in $(x[5], x[6])$ before, and there is no pattern of $(x[n-2], x[n-1], x[n])$ can be found in the history data. Hence, the algorithm will halt. The prediction will be done by taking the value of $x[7]$, since the pattern after $(x[5], x[6])$ is $x[7]$. This can be explained by the main philosophy of this algorithm, which is how the pattern of the data in history is, and then the current data will be the same pattern. Referring to the previous works which had done by Fidan et al. [17, 18], they had created an equation to express the afore-mentioned prediction process. The particular equation is shown in Eq. (4):

$$m = \arg\max_L \{x[k] = x[n], x[k-1] = x[n-1],$$
$$f_{n+1} = \hat{x}[n+1] = x[m] \tag{4}$$

Although it was not written in the papers published, it is believed that the symbol L in the equation represents the location of the longest pattern found in the history. Due to the prediction value will be the data after the longest pattern. Hence, the prediction value can be expressed as $\arg\max_{L}[n - L + 1]$. As mentioned earlier, the prediction value is $\hat{x}[n + 1]$, so the prediction value can be expressed as in Eq. (5).

$$\hat{x}[n + 1] = \arg\max_{L}[n - L + 1] \tag{5}$$

There was a cyclic fault happened in the earlier version of Mycielski approach. Hence, Mehmet Fidan, Fatih Onur Hocaoğlu and Ömer N. Gerek had modified the approach by adding the random number range from -0.4 to $+0.4$ to the predicted value and make it Mycielski-1 approach. However, the randomness of the number will cause the predicted result become unreliable. Due to this reason, authors had figured out another solution which is obtaining the average difference, d_{avg} from the history data in this paper. The principle of obtaining the d_{avg} is 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 find the average for past few years depends on the available data. The result shown this kind of random number finding is more reliable than the original Mycielski-1 approach. The details of findings will show in next section.

Then, the random number will be added into the predicted value. Once the prediction had done, the predicted value will be updated in the history data. Then the next prediction $\hat{x}[n + 2]$ can be done based on the updated history data.

3.2 K-Means Clustering

K-means clustering is one of the data mining methods by grouping all the data with the similar mean into a cluster. There may be a lot of clusters in a certain analysis. However, the optimal cluster number with the least error is 4 [19]. For this research, the optimal cluster been found from the hierarchical analysis by using Ward's method is 2. Due to the comparison purpose, 4 clusters analysis will be presented in this paper also.

In order to compare the efficiency of Mycielski and K-means clustering, the same data is used for the wind speed prediction computation. The wind speed data ranges from 2007 to 2009 are used for algorithm training. Meanwhile, the wind speed in 2010 will be used as the tester.

First of all, the optimal clusters solution must be identified before the computation started. The number of cluster might affect the accuracy at the end of the prediction. Once the number of cluster had been determined, the initial center of cluster of the data should be obtained. For this paper, the initial centers of clusters were chosen based on the largest and the lowest values. The classification of the

initial center of cluster might be varied for different research. The most of important part of the K-means clustering is the distance between the data and the center of cluster. The algorithm will stop when the shortest distance between the data and the center of cluster is achieved. The distance is calculated by using Euclidean distance. The Euclidean distance is formulated as in Eq. (6).

$$D_n = \sqrt{(x_n - x_i)^2 + (y_n - y_i)^2} \tag{6}$$

where D_n represents the distance between nth data and the center of cluster;

x_i and y_i denote the coordinate of the ith center of cluster;
x_n and y_n denote the coordinate of the nth center of cluster.

The process of finding the shortest distance will stop when the distances of all data for two consequences are met.

The wind speed prediction is done by obtained the probability of the occurrence of wind speed in that particular time. The wind speed data is first to be clustered by using K-means clustering. Then, the trend of the wind speed is studied. Last but not least, the probability of the wind speed either go upper cluster or lower cluster or maintain in the same cluster is obtained. In order to have a better prediction result, the K-means clustering of the wind speed for each month in different years is obtained. Finally, the prediction is done by using the probability tree method.

4 Results and Analysis

Table 3 shows the wind speed obtained from meteorology department Malaysia after apply wind profile power law and rounded up process. The wind data in year 2010 is purposely shown for reference. After round up process, the Mycielski approach is applied in order to obtain the prediction for year 2010. The random number for each month is shown in Table 4.

Mycielski approach is applied in order to make the prediction for the wind speed in year 2010. The random number as shown in Table 5 is added into the predicted wind speed.

The graph of obtained wind speed and predicted wind speed can be illustrated in Fig. 3. As shown in the graph, the shape of the graph is look alike for obtained

Table 3 The round up wind speed in Kudat

Mth	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Year
2007	3.5	4.5	3.5	3.5	2.5	3.0	3.5	3.0	4.0	3.0	4.0	3.0
2008	3.5	3.5	3.0	3.0	3.5	2.5	3.0	3.5	2.5	2.5	2.5	3.0
2009	3.5	3.5	3.0	3.5	3.0	3.0	3.5	4.5	4.5	4.0	3.0	4.0
2010	4.5	5.0	4.5	3.5	3.0	3.0	2.5	3.5	3.0	3.5	2.5	3.0

Table 4 The average difference (d_{avg}) for each month from 2007 to 2009

Mth	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
d_{avg}	0.3	0.3	−0.7	0.2	−0.3	−0.2	0.5	0	0.7	−0.8	0	0.2

Table 5 The rounded up prediction wind speed in 2010 after add in d_{avg}

Mth	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Prediction (ms^{-1})	3.5	5.0	4.5	4.0	2.5	3.0	4.0	3.5	5.5	4.5	4.0	2.5

Fig. 3 The graph of *obtained* and *predicted wind speed* for 2010 using Mycielski-1
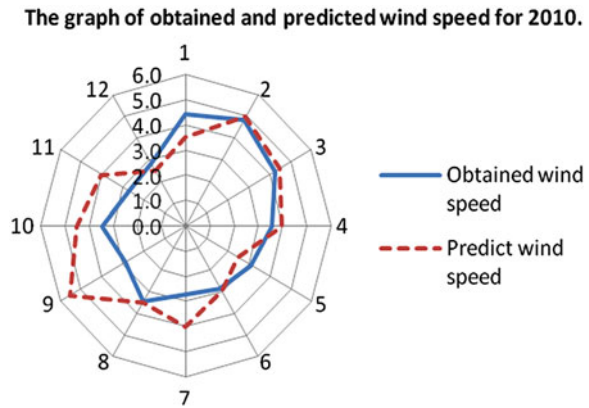
and predicted wind speed. However, there is much difference in September. This will be considered as the prediction error.

For K-means clustering, the wind speed data for 2007–2009 were sorting in ascending order. As in Fig. 4, the minimum wind speed is 2.1 ms^{-1} in August 2008 whereas the highest wind speed is 3.6 ms^{-1} in September 2009.

By using the wind speed data in Table 1, the center of clusters for 2 clusters and 4 clusters methods are shown in Tables 6 and 7 respectively. Meanwhile, the wind prediction by using K-means clustering is shown in Table 8. Result as shown in Table 8 can be illustrated in Fig. 5 (Table 8).

5 Discussion

The Mycielski algorithm delivers a good prediction for the mean wind speed in Kudat, Malaysia. By using Mycielski-1 approach, the wind speed for year 2010 is predicted. From the result in Fig. 6, the lowest accuracy is about 55 %. In terms of the monthly comparison, the prediction results for February, March, June, and August are perfectly matched with the obtained data. The smallest percentage of difference is in April whereas the largest difference is in September. However, by looking into the history data for September in 2007, 2008 and 2009 respectively,
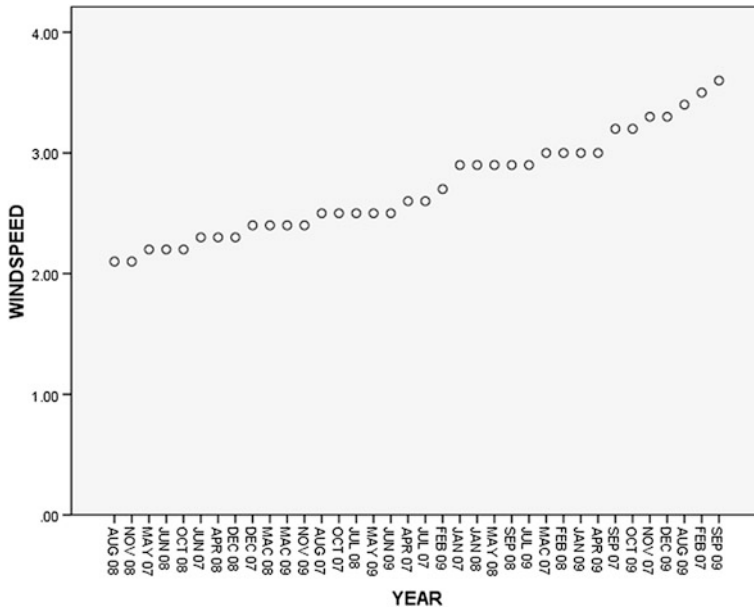
Fig. 4 The wind speed from 2007 to 2009 in ascending order

Table 6 The K-means clustering using 2 clusters

Cluster	1	2
Center of cluster (ms^{-1})	2.39	3.12

Table 7 The K-means clustering using 4 clusters

Cluster	1	2	3	4
Center of cluster (ms^{-1})	2.21	2.5	2.99	3.42

Fig. 5 The comparison between *actual* and *predicted* wind speed
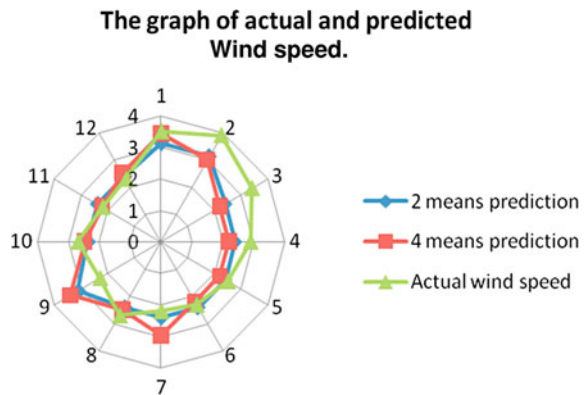
Table 8 The predicted wind speed by using K-means clustering

Mth (ms^{-1})	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2 Clusters	3.12	3.12	2.39	2.39	2.39	2.39	2.39	2.39	3.12	2.39	2.39	2.39
4 Clusters	3.42	2.99	2.21	2.21	2.21	2.21	2.99	2.5	3.42	2.5	2.21	2.5
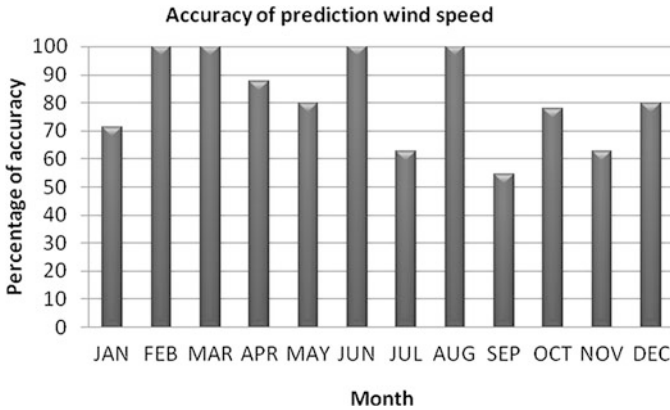
Fig. 6 The graph of accuracy for wind prediction using Mycielski-1 in 2010
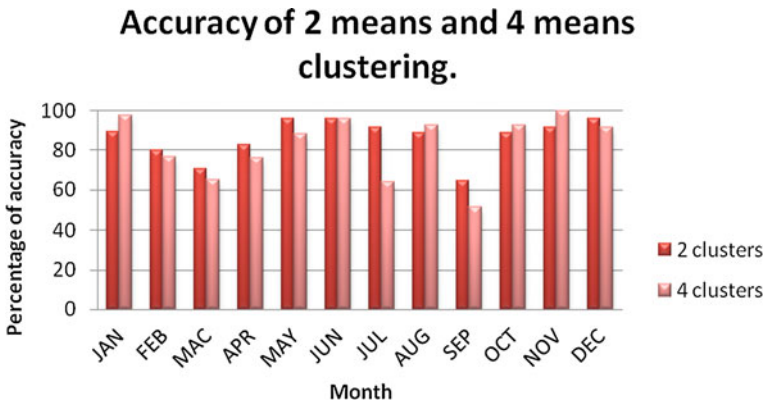
Fig. 7 The accuracy of wind prediction using 2 and 4 clusters

the wind speed is relatively high in this month. Mycielski approach assumes that 2010 is having higher wind speed in September. Same goes to July and November. The large difference of the obtained and predicted wind speed is considered as the computing error. However, the overall result shows that Mycielski-1 approach is successful to predict the wind speed in 2010. This prediction algorithm is suitable for hourly, monthly, and yearly wind speed prediction. For the future work, it is recommended to consider the capacity factor of the wind site as well.

On the other hand, in terms of accuracy of the prediction, the comparison of 2-means clustering and 4-means clustering is done as shown in Fig. 7.

From Fig. 7, the accuracy of each month is more than 70 % except in September 2010. The same reason applied in this error. By looking into the comparison of accuracy between 2-means and 4-means clustering, the 2-means clustering has higher overall accuracy if compared with the latter. In 2-means clustering, the average accuracy for 2010 wind speed prediction is 86.14 % whereas 4-means clustering is 3.51 % less than that, which is 82.63 %. Although the different of accuracy is not significant, the efficiency of determining the number of means is determined.

6 Conclusion

In conclusion, the Mycielski algorithm has average accuracy of 81.35 %. It means that this algorithm can predict well on the wind speed of Kudat. The wind speed for year 2010 is predicted and there are 9 out of 12 months have accuracy more than 70 %. On the other hand, for the 2-means clustering and 4-means clustering, the average accuracy is 86.14 % and 82.63 % respectively. In terms of prediction value, the 2-means clustering has 92 % of predicted wind speed is more than 70 %. Meanwhile, 4-means clustering has same amount of month that predicted the wind speed which accuracy achieved more than 70 %. Hence, the result shown that the prediction methods presented in this paper are reliable especially K-means clustering with the number of cluster obtained by using Ward's cluster analysis. In order to get more reliable result, it is better to have a huge history data as the database. The algorithm can be more precise due to the search process has more data to be compared. In the Mycielski-1 algorithm, the random number is added to the prediction value purposely to eliminate the recurrence error in predicting the wind speed. The proposed algorithm in finding the random number gives promising result. In order to analyze using K-means clustering, the computation tools are necessary in order to calculate a large number of Euclidean distances. Last but not least, the wind speed in Kudat shows that Malaysia can develop the wind energy as the mean wind speed is beyond the cut in speed of most of the wind turbines. However, the selection of the wind turbine is actually the key of success of this planning. In a nutshell, the utilization of wind speed prediction can lead to a good development in green energy Malaysia.

References

1. MOSTI: Monsoon. [cited 2011 28 December]; Available from: http://www.met.gov.my/index.php?option=com_content&task=view&id=69&Itemid=160&lang=english
2. FirstLook: Renewable energy risk analysis: Wind, solar, hydro data, 3TIER (2011)
3. Islam, M.R., Saidur, R., Rahim, N.A.: Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function. Energy **36**(2), 985–992 (2011)
4. Zuhairuse, M.D.: The development of hybrid renewable energy system (wind and solar) for sustainable living at Perhentian Island Malaysia. Eur. J. Soc. Sci. **9**(4), 557–563 (2009)
5. Onar, O.C., Uzunoglu, M., Alam, M.S.: Dynamic modeling, design and simulation of a wind/fuel cell/ultra-capacitor-based hybrid power generation system. J. Power Sources **161**(1), 707–722 (2006)
6. Badmasti, B., Bevrani, H., Naghshbandy, A.H.: Impacts of high wind power penetration on the frequency response considering wind power reserve. Int. J. Energy Optimization Eng. (IJEOE) **1**(3), 32–47 (2012). doi:10.4018/ijeoe.2012070102
7. Fahim, A.A., Khalil, E.E.: Energy efficient operation of university hostel buildings under indoor environmental quality requirements. Int. J. Energy Optimization Eng. (IJEOE) **1**(2), 96–109 (2012). doi:10.4018/ijeoe.2012040106
8. Hahn, B., Durstewitz, M., and Rohrig, K., Wind Energy (Reliability of Wind Turbines Experiences of 15 Years with 1,500 W), ed J Peinke et al, Springer, Berlin, pp. 329–332, 2007
9. Karki, R.: Renewable energy credit driven wind power growth for system reliability. Electr. Power Syst. Res. **77**(7), 797–803 (2007)
10. Jewer, P., Iqbal, M.T., Khan, M.J.: Wind energy resource map of labrador. Renewable Energy **30**(7), 989–1004 (2005)
11. Khan, M.J., Iqbal, M.T., Mahboob, S.: A wind map of Bangladesh. Renewable Energy **29**(5), 643–660 (2004)
12. Focken, U., et al.: Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. J. Wind Eng. Ind. Aerodyn. **90**(3), 231–246 (2002)
13. Landberg, L.: Short-term prediction of the power production from wind farms. J. Wind Eng. Ind. Aerodyn. **80**(1–2), 207–220 (1999)
14. Beccali, M., et al.: Estimation of wind velocity over a complex terrain using the generalized mapping regressor. Appl. Energy **87**(3), 884–893 (2010)
15. Bivona, S., Burlon, R., Leone, C.: Hourly wind speed analysis in Sicily. Renewable Energy **28**(9), 1371–1385 (2003)
16. Louka, P., et al.: Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J. Wind Eng. Ind. Aerodyn. **96**(12), 2348–2362 (2008)
17. Hocaoglu, F.O., Fidan, M., Gerek, O.N.: Mycielski approach for wind speed prediction. Energy Convers. Manage. **50**(6), 1436–1443 (2009)
18. Fidan, M., Hocaoğlu, F.O., Gerek, Ö.N.: Improved synthetic wind speed generation using modified Mycielski approach. Int. J. Energy Res. **36**(13), 1226–1237 (2012)
19. Asamer, J., Din, K.: Prediction of velocities on motorways by k-means clustering. In: 7th Mexican International Conference on Artificial Intelligence MICAI'08 (2008)
20. Bishnu, P.S., Bhattacherjee, V.: Software fault prediction using quad tree-based k-means clustering algorithm. IEEE Trans. Knowl. Data Eng. **24**(6), 1146–1150 (2012)