6th Global Conference on Power Control and Optimization, Las Vegas, USA, 6-8 August 2012

WIND PREDICTION IN MALAYSIA USING MYCIELSKI-1 APPROACH

S. W. LEE, B. C. KOK, K. C. GOH, H. H. GOH
Department of Electrical Power Engineering, Faculty Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.
hhgoh@uthm.edu.my

Abstract
In this paper, the wind speed prediction in Kudat, Malaysia had been done by using Mycielski-1 approach. There is some improvement in obtaining the random number of Mycielski-1. The wind prediction is important to study a favourable 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 this algorithm. The basic concept of this algorithm is to predict the next value by looking to history data. The result shows the prediction of Mycielski-1 algorithm is promising. The wind speed is predicted in order to obtain the mean power for energy planning.

Keywords: Wind prediction, Mycielski

1. Introduction
The growth of renewable energy around the world is 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 have high possibility to be harnessed in Malaysia. Malaysia is a country without four seasons. However, Malaysia encounters two monsoon seasons. The Southwest Monsoon starts from late May to September, meanwhile the Northeast Monsoon starts from November to March [1]. Due to the Northeast Monsoon, there will be 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 350MW of renewable energy in Malaysia which is 300MW in West Peninsular Malaysia and 50MW from Sabah under Ninth Malaysia Plan. Unfortunately, the 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 potential site. Estimation of the wind power generated can be obtained when the wind speed is predicted.
For this paper, one of the cities in Sabah had been chosen for 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 131km from Kota Kinabalu, the capital city 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 others states [2]. Previously, there was a group of researchers had studied about the wind energy potential in Kudat and Labuan [3]. The Weibull distribution function is used to determine the wind profile. The yearly wind power density of the particular places had been obtained.

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 the manufacturer even there was no local data for a particular site. The estimation of that particular site can be done by obtaining the data from the nearest wind measuring station. M. Beccali, G. Cirrincione, A. Marvuglia, and C. 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 [4]. So, the wind speed behaviour must be obtained in order to utilise this approach. Authors used Matlab to generate the coding for the estimation [4].
In terms of wind speed analysis, S. Bivona, R. Burlon, and C. Leone had analysed the hourly wind speed in Sicily by using Weibull distribution function [5]. 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, P. Louka *et. al.* introduced some modelling which are being applied in Greece [6]. 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) had been developed at Colorado State University and Mission Research Inc. /ASTeR Division. RAMS can forecast the wind in 48 hours later. In addition, adaptive fuzzy neural network (F-NN) also applied for the wind power prediction for 120 hours 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. 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 F.O. Hacaoglu, M. Fidan, and O.N. Gerek by using Mycielski approach [7]. 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 [8]. The modified algorithm called Mycielski-1 and Mycielski-2. In Mycielski-1, the random number in between -0.4 to 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.

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 Fig.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%.

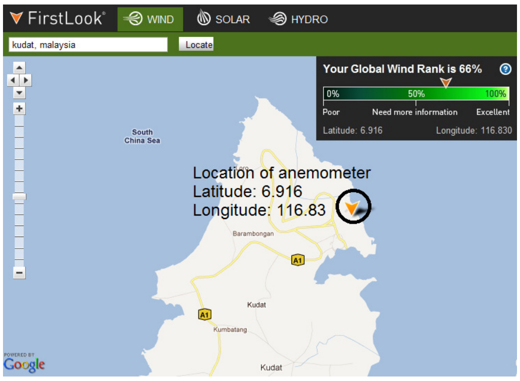

Fig.1. The location of anemometer in Kudat.

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. Fig.2 shows the higher global wind rank obtained by using Firstlook software.

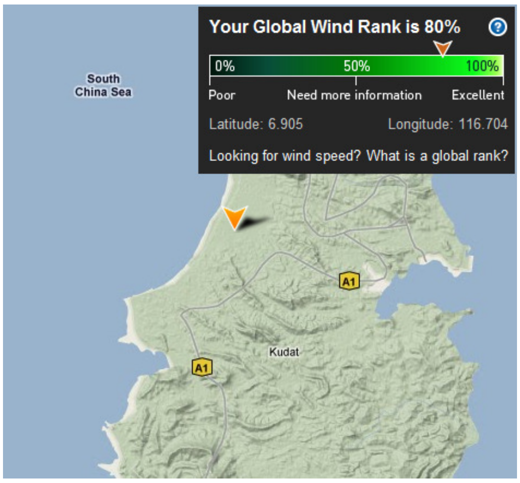

Fig.2. The proposed location of wind turbine.

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

Month Year	JAN	FEB	MAC	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Annual
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

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

According to the wind profile power law, the wind speed listed in

Table **1** can be further adjusted to height of 100m. This is because the wind speed is higher at higher altitude. The adjustment is based on the formula:

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

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.

4. 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 equation (2).

$$|x - y_i| \leq 0.2; x = y_i, i \geq 0 \quad (2)$$

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 equation (3)

$$f_n = (x[1], x[2], ..., x[n-1], x[n]) \quad (3)$$

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 Mehmet Fidan, Fatih Onur Hocao□lu and Ömer N. Gerek [7, 8], they had created an equation to express the aforementioned prediction process. The particular equation is shown in equation (4):

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

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 equation (5).

$$\hat{x}[n+1] = \arg\max_L [n-L+1] \quad (5)$$

There was a cyclic fault happened in the earlier version of Mycielski approach. Hence, Mehmet Fidan, Fatih Onur Hocao□lu 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.

5. Result 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 and predicted wind speed. However, there is much difference in September. This will be considered as the prediction error.

6. 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.4, 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, 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.

7. Conclusion

In conclusion, the Mycielski 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%. Hence, the result shown the prediction is reliable. 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. 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. 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 utilisation of wind speed prediction can lead to a good development in green energy Malaysia.

Acknowledgment

Authors would like to thank the Ministry of Higher Education Malaysia (MOHE) for giving the scholarship to the author in order to accomplish this research. Furthermore, the authors would like to thank Ministry of Energy, Green Technology and Water Malaysia, Ministry of Higher Education, Malaysia (MOHE) 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.0905 in funding this research.

References

[1] MOSTI, "Monsoon." vol. 2011: Misnistry of Science, Technology and Innovation Malaysia.

[2] FirstLook, "Renewable Energy Risk Analysis: Wind, Solar, Hydro Data," 3TIER, 2011.

[3] M. R. Islam, R. Saidur, and N. A. Rahim, "Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function," *Energy,* vol. 36, pp. 985-992.

[4] M. Beccali, G. Cirrincione, A. Marvuglia, and C. Serporta, "Estimation of wind velocity over a complex terrain using the Generalized Mapping Regressor," *Applied Energy,* vol. 87, pp. 884-893

Table 3. The round up wind speed in Kudat

Month / Year	JAN	FEB	MAC	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
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}...

Month	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

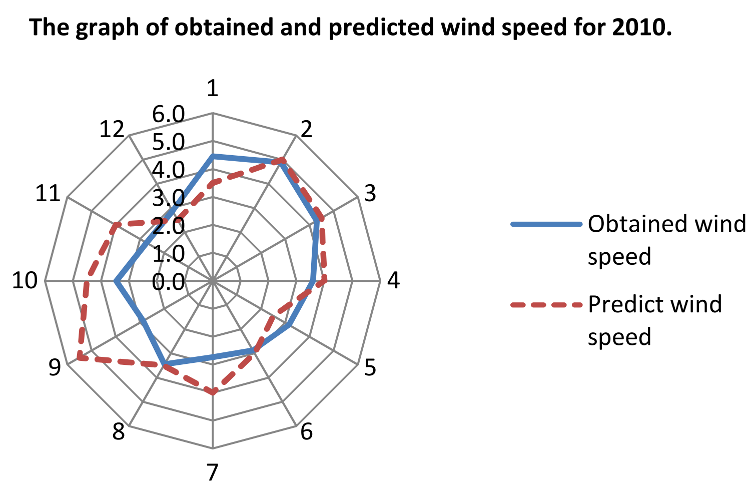

Fig.3. The graph of obtained and predicted wind speed for 2010.

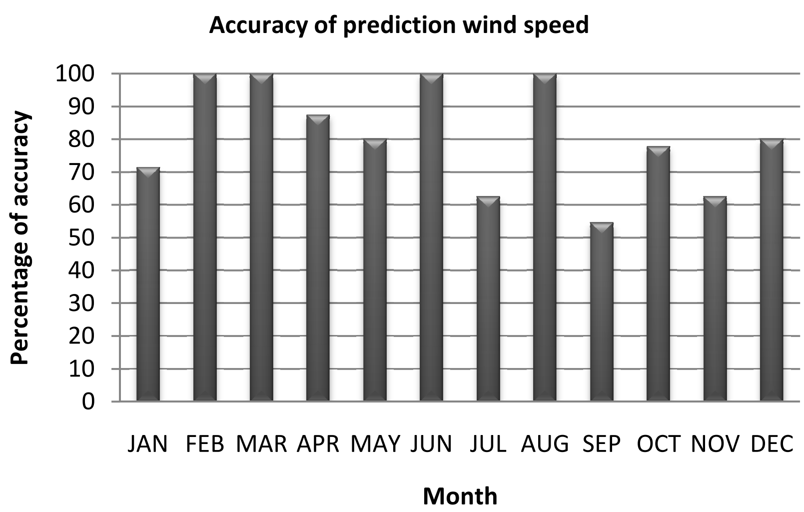

Fig.4. The graph of accuracy for wind speed prediction in 2010.

[5] S. Bivona, R. Burlon, and C. Leone, "Hourly wind speed analysis in Sicily," *Renewable Energy,* vol. 28, pp. 1371-1385, 2003
[6] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *Journal of Wind Engineering and Industrial Aerodynamics,* vol. 96, pp. 2348-2362, 2008.
[7] F. O. Hocaoglu, M. Fidan, and O. N. Gerek, "Mycielski approach for wind speed prediction," *Energy Conversion and Management,* vol. 50, pp. 1436-1443, 2009.
[8] M. Fidan, F. O. Hocaoğlu, and Ö. N. Gerek, "Improved synthetic wind speed generation using modified Mycielski approach," *International Journal of Energy Research,* pp. n/a-n/a