# An Optimized ANN Measure-Correlate-Predict Method for Long-term Wind Prediction in Malaysia

Yong Kim Hwang[1], Mohd Zamri Ibrahim[1*], Ali Najah Ahmed[2,3], and Aliashim Albani[1]
[1]Eastern Corridor Renewable Energy Research Group, School of Ocean Engineering
Universiti Malaysia Terengganu, Terengganu, Malaysia
[2]Department of Civil Engineering, Universiti Tenaga Nasional, Selangor, Malaysia
[3]Institute for Energy Infrastruture, Universiti Tenaga Nasional, Selangor, Malaysia
zam@umt.edu.my

***Abstract*—The major issues on the wind measurement campaign are the data measured in a short period and the occurrence of missing data due to the failure of the measurement instrument. Meanwhile, Measure-Correlate-Predict (MCP) method had widely been used to predict the long-term condition and missing data at the measurement site based on nearest Malaysian Meteorological Department (MMD), Meteorological Aerodrome Report (METAR) and extended Climate Forecast System Reanalysis (ECFSR) data. In this research, the long-term wind data at selected potential sites in Malaysia were predicted by optimized Artificial Neural Networks (ANNs). The Genetic Algorithm (GA) was applied to optimize the ANN. Five different ANN MCP models had been designed based on different types of reference data and different temporal scales to predict wind data at three target sites. Weibull frequency distributions and RMSE examined predicted wind data. The prediction of ANN had been improved in between 20.562% to 113.573% by GA optimization. The best R-value obtained from optimization were affected the Weibull shape and scale of predicted data. At last, the result revealed that the optimized ANN model could predict the long-term data for the target site with better accuracy.**

***Index Terms*—Artificial Neural Network, Genetic Algorithm, Measure-Correlate-Predict, Meteorological data, Re-analysis data.**

## I. INTRODUCTION

The wind measurement campaign is essential to determine the wind quality of the potential site for virtous planning of any wind energy project. However, wind measurement campaigns generally conducted as minimum as a one-year duration due to the capital cost of wind project were mainly allocated for wind turbine generator (WTG). A few years wind data measured are inappropriate to estimate mean wind speed for a 20 years wind project's site due to the inter-annual variability of wind speed. Moreover, it is difficult to prevent a missing data encounter in a wind measurement campaign, which caused by device failure or disconnection wind mast. A statistical method, Measure-Correlate-Predict (MCP) technique, had been proposed in the 1940s to predict the long-term mean wind speed for a target site from a single reference site [1].

The wind energy industry had been developed MCP method as a standard tool in software applications for long-term wind speed forecast [1]. Established MCP models are mainly divided into two groups as a linear and nonlinear model. MCP algorithm such as linear fit, polynomial fit, quadratic fit, and matrix bins, those are commonly used linear MCP models as presented in [2]. Undeniable that not all reference sites have a linear nature with the target sites. In addition, the linear method is also sensitive to the outliers of the data [3].

Several new artificial intelligence (AI) approaches for wind speed, and power prediction has been established with the non-linear pattern in wind series. Indeed, the AI methods developed were included artificial neural network (ANN). ANN is a technique mainly used to map random input vector(s) to the corresponding random output vector without pre-assuming any fixed relationship between them. In other words, neural networks can learn from past data, recognise hidden patterns or relationships in historical observations and use them to forecast future values [4], [5]. As the results, it is not surprising that the neural network approach has attracted overwhelming attention in time series forecasting [6]. Additional advantages of the neural network approach over the conventional forecasting schemes include data error tolerance, ease in adaptability to online measurements, and lack of any excess information (other than time series history of wind speeds) [4], [5].

ANNs have been found to be better than various traditional time series models [6], [7]. It had documented that Lapedes and Farber [8] were the first utilised artificial neural networks to model the nonlinear time series. The neural networks seem to be performing well and extremely parsimonious in its requirements for data point from the time series. Consequently, various wind speed and power prediction or forecasts have been established with ANNs approaches.

More and Deo [4] employs two neural networks technique, back propagation neural network and recurrent neural network, to forecast daily, weekly and monthly wind speed. The network forecasting was reasonably close to the corresponding measurements than traditional stochastic time-series model of ARIMA. The forecasting interval reduced, from monthly to daily, forecasting accuracy decreased due to the overfitting or large training patterns.

Chang [9] proposed back propagation neural network for wind power forecasting based on 2400kW Wind Energy Conversion Systems (WECS) in Taichung coast. The established model was forecasting wind power every 10 min and verified by the historical power generation data. The numerical results show that the proposed forecasting method is accurate and reliable. However, the determination of some neurons in hidden layer was based on trial-and-error method.

Li and Shi [5] examined wind speed forecasting based on three types of conventional neural networks and evaluated by three different kinds of metrics. There are no single neural networks models met all evaluation metrics. In other words, the selection of a best-performing model for a dataset with different evaluation criteria is not appropriate.

Addision et al. [10] investigated the feasibility of using neural networks to make predictions of long-term energy at a target site. In the research, wind speed and direction from one reference station is using to study the effectiveness of neural networks in the prediction of wind speed at a target site. The accuracy of the prediction improved 5-12% by comparing with standard MCP algorithms, and the best results were obtained using multilayer perceptron networks with numerous hidden units.

Bechrakis et al. [11] present an ANN method to simulate time series of 1-year data at the target site with 1-month and 2-month concurrent data from target sites. Mean values, Weibull distribution parameter and cross-correlation coefficient had been executed; and the results indicated that ANN accomplished to estimate 1-year data at the target site with 1-month and 2-month concurrent data.

The originality of this paper lies on the optimized ANN method for MCP algorithm for long-term data prediction for target site in Malaysia. Besides, different types of long-term data forms reference site which are meteorological data obtained Malaysian Meteorological Department (MMD), Meteorological Aerodrome Report (METAR) wind data, and extended Climate Forecast System Reanalysis (E-CFSR) data were examined for the prediction of data in target side. Weibull distribution, correlation coefficient, and root mean squared error (RMSE) were executed between both predicted and measured data as a tool to examine the predicted wind data.

## II. Case Study: Wind Assessment Project in Malaysia

Target sites in this study which referred to the measured sites under a wind assessment project conducted by Universiti Malaysia Terengganu. Five potential sites had been measured in this project were Kijal, Kuala Terengganu, Kudat, Langkawi, and Mersing. Wind speed were measured in ten minutes interval, and its average to hourly and daily for the suitable scale of concurrent data used. Long-term concurrent data used, MMD data, METAR data and E-CFSR data were described in the next subsection. Meanwhile, Kijal and Langkawi sites were not discussed in this paper due to lacking nearest concurrent data of MMD data and METAR data. The concurrent period of the data or measured period for the sites respectively as below:

- Kuala Terengganu: October 2011 until November 2013 (26 months)
- Kudat: October 2012 until May 2015 (32 months)
- Mersing: October 2012 until April 2015 (31 months)

### *A. Concurrent Long-term Data*

#### *1) MMD Data*

Existing local historical meteorological wind data always as the reference for wind energy studies [12], [13]. Hence, the meteorological data in this study were obtained from the Malaysian Meteorological Department (MMD). The concurrent MMD data were in the scale of daily data. Besides wind data, the air temperature and pressure data used in this research also obtained from MMD.

#### *2) METAR Data*

METAR is known as the international meteorological code for an aviation routine weather report or Meteorological Aerodrome Report. METAR observations usually are taken and disseminated on the hour [14]. METAR wind data are observed from surface observing (METAR) stations and frequently used for long-term reference data. METAR stations in Malaysia are located at airports and observations are usually collected at hourly increments at the height of 10 meters above ground level (m.a.g.l) [15]. The concurrent METAR data were in the scale of hourly and downloaded via WindPRO software.

#### *3) E-CFSR Data*

The NCEP Climate Forecast System Reanalysis (CFSR) was designed and executed as a global, high resolution, coupled atmosphere-ocean-land surface-sea ice system to provide the best estimate of the state of these combined domains over a 31-year period from 1979 to 2009 [16]. It was initially completed in January 2010 and has also been extended as an operational, real-time product into the future. The extended CFSR (E-CFSR) is 0.5 degree of spatial resolution, temporal resolution is 1 hour, and the period from 1979 until present. The CFSR-Extended dataset may suffer from inconsistencies if periods before and after the end of 2010 [16]. The concurrent E-CFSR data were in the scale of hourly and downloaded via WindPRO software.

## III. Methodology

### *A. Data Pre-processing and Data Normalisation*

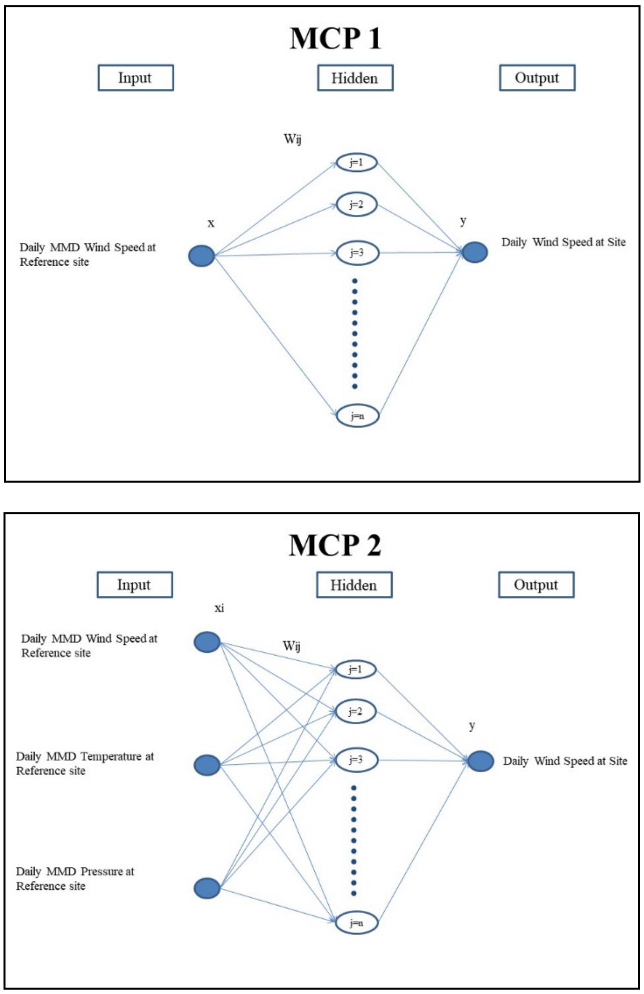
All data were normalized to produce the data within the range of 0 to 1 by min-max scaling technique. The input and output of the ANN were normalized to minimize the error which caused by the high range value of the dataset. Min-Max Normalization, which shown in (1).

$$x' = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

where $x_{max}$ is the maximum value of the data, $x_{min}$ is the minimum value of the data and $x_i$ is the data value need to be normalized.

The zero value which obtained from the prediction of ANN would affect the Weibull fitting. These zero values were substituted by the value 0.03 m/s which considered as calm or no wind condition.

*B. Design of Artificial Neural Network (ANN)*



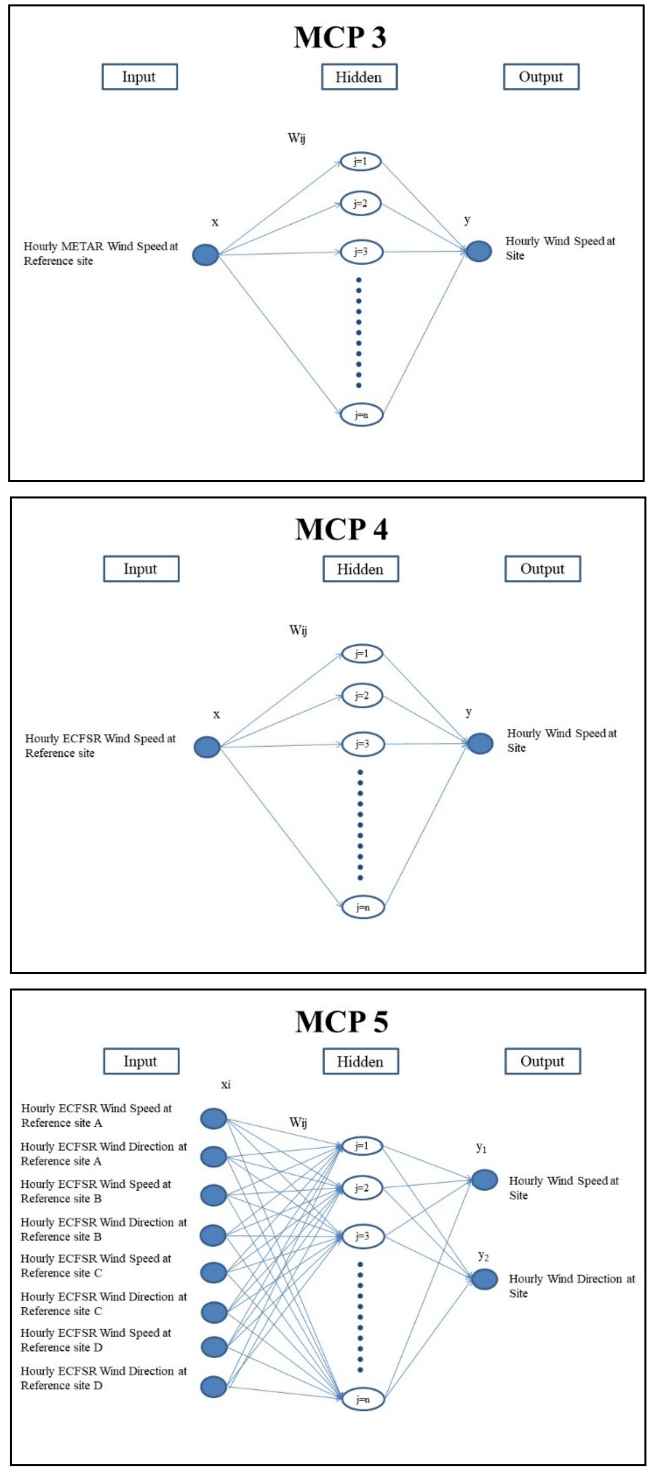




Figure 1. Architecture of Artificial Neural Network.

MCP models in this paper were designed based on Multilayer Perceptron (MLP) with a single hidden layer. A supervised MLP with input and output nodes also known as Feed-Forward Neural Networks (FFNN). Five different designs of the MCP were shown in Fig. 1 and the description as below respectively:

- MCP1: MCP model that concurrent data was MMD daily data. The input and output nodes were daily MMD wind speed at the reference site and daily wind speed at the target site.
- MCP2: MCP model that concurrent data was MMD daily data. The input nodes were daily MMD wind speed, air temperature, and pressure at a reference site; and output node was daily wind speed at the target site.
- MCP3: MCP model that concurrent data was METAR hourly data. The input and output nodes were hourly METAR wind speed at the reference site and hourly wind speed at the target site.
- MCP4: MCP model that concurrent data was E-CFSR hourly data. The input and output nodes were hourly E-CFSR wind speed at the reference site and hourly wind speed at the target site.
- MCP5: MCP model that concurrent data was E-CFSR hourly data. The input nodes were hourly E-CFSR wind speed and wind direction at four nearest reference sites, and output nodes were hourly wind speed and wind direction at the target site.

Levenberg-Marquardt backpropagation function was the training algorithm, and the gradient descent with momentum function was the learning function for the weight and bias. Meanwhile, the transfer function and the number of neurons in hidden layer were optimized by Genetic Algorithm (GA) optimization method. The combination of the transfer function for the hidden layer and output layer were selected between Linear, Tan-Sigmoid and Log-Sigmoid. The ranges of the number of neurons were calculated by the (2):

$$n=\sqrt{n_i+n_o}+\alpha \tag{2}$$

Where, $n$ is the number of neurons in the hidden layer, $n_i$ is the number of input neurons, $n_o$ is the number of output neurons, and α is a number between 1 and 10.

*C. Genetic Algorithm Optimisation of ANN*

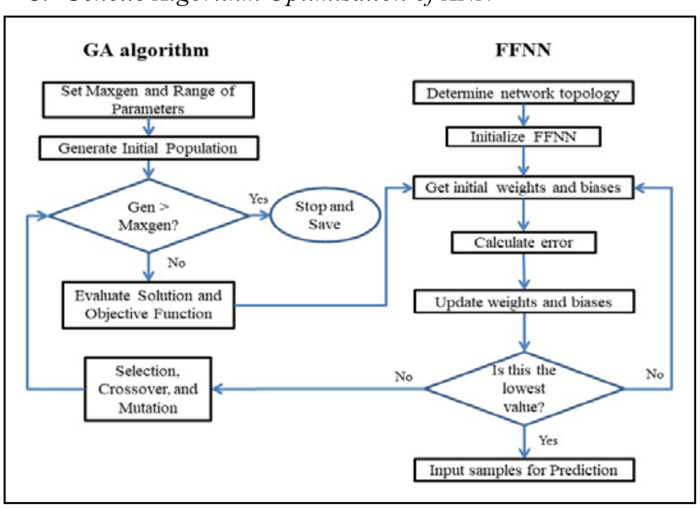



Figure 2. Flowchart of GA-FFNN.

The core idea of GA-FFNN is to obtain the initial weights and biases of FFNN by optimising the transfer function and the number of neurons for the hidden layer to increase the accuracy of the prediction. The flow chart of the proposed GA-FFNN is shown in Fig. 2, and the calculation steps GA-FFNN are as follows:

***Step 1***. Determine the neuron number of the input layer, the output layer, the learning and the training function of FFNN respectively as described in the previous section.

***Step 2***. Set the population and generation size of GA optimization of FFNN were 100 and 50 respectively.

***Step 3***. Randomly set the lower and upper bound of the number of hidden layers and the combination of transfer functions and randomly generate the value or combination within the range.

***Step 4***. Initial position of the weight and biases of FFNN. Input training samples for calculating.

***Step 5***. Calculate the output of FFNN, subsequently, calculate the fitness evaluation of GA which set to minimize the uncorrelated data, (1-R).

***Step 6***. Selection of the best individual in the population, cross over, then mutated. The best and mean optimized results and plotted in the GA optimization best plot.

***Step 7***. Examine whether the maximum generation of GA has been reached, if it is not reached, loop back to step 3

***Step 8***. Long-term prediction of FFNN and saved results

*D. Weibull Distribution*

For statistical analysis of wind data, a probability distribution is a term that describes the likelihood that certain values of random wind speed will regime. In this case, Weibull distribution became a tool to examine the predicted wind data by comparing the k and c Weibull parameters for both predicted and measured data. The probability density function F(v), indicates the fraction of probability for which the wind is at a given velocity v. It is given by:

$$\mathrm{F(v)}=\left(\frac{\mathrm{k}}{\mathrm{c}}\right)\left(\frac{\mathrm{v}}{\mathrm{c}}\right)^{\mathrm{k}-1}\exp\left[-\left(\frac{\mathrm{v}}{\mathrm{c}}\right)^{\mathrm{k}}\right] \tag{3}$$

he Weibull factor k and c can also be estimated from the mean and standard deviation of wind data [17].

$$\mathrm{k}=\left(\frac{\sigma_{\mathrm{V}}}{\mathrm{V_m}}\right)^{-1.086} \tag{4}$$

$$\mathrm{c}=\frac{\mathrm{V_m}}{\Gamma\left(1+\frac{1}{\mathrm{k}}\right)} \tag{5}$$

Where the $\Gamma(x)$ = gamma function, $\Gamma(x)=\int_0^{\infty} t^{x-1}e^{-t}dt$

The mean and standard deviation were calculated by the equation below:

$$\mathrm{V_m}=\frac{\sum_{i=1}^{n}\mathrm{v_i}}{\mathrm{n}} \tag{6}$$

$$\sigma_V = \sqrt{\frac{\sum_{i=1}^{n}(v_i - V_m)^2}{n}} \quad (7)$$

*E. Root Mean Squared Error (RMSE)*

RMSE is a quadratic scoring rule which measured the average magnitude of the error. The RMSE gives a relatively high weight to large errors because of the errors is squared before averaged and the was shown its relevant mathematical expressions [5].

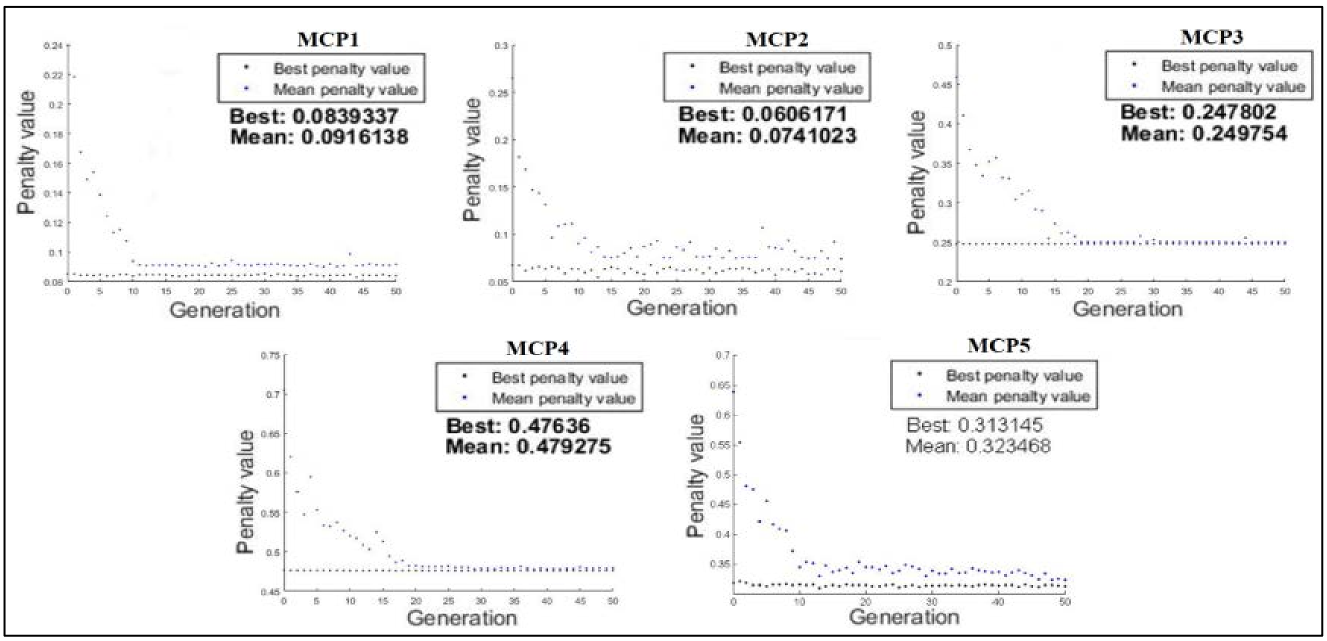
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}[x_i - y_i]^2}{n-1}} \quad (8)$$

where $x_i$, is the measured wind speed; $y_i$, is the forecasted wind speed and $n$, is the number of samples.

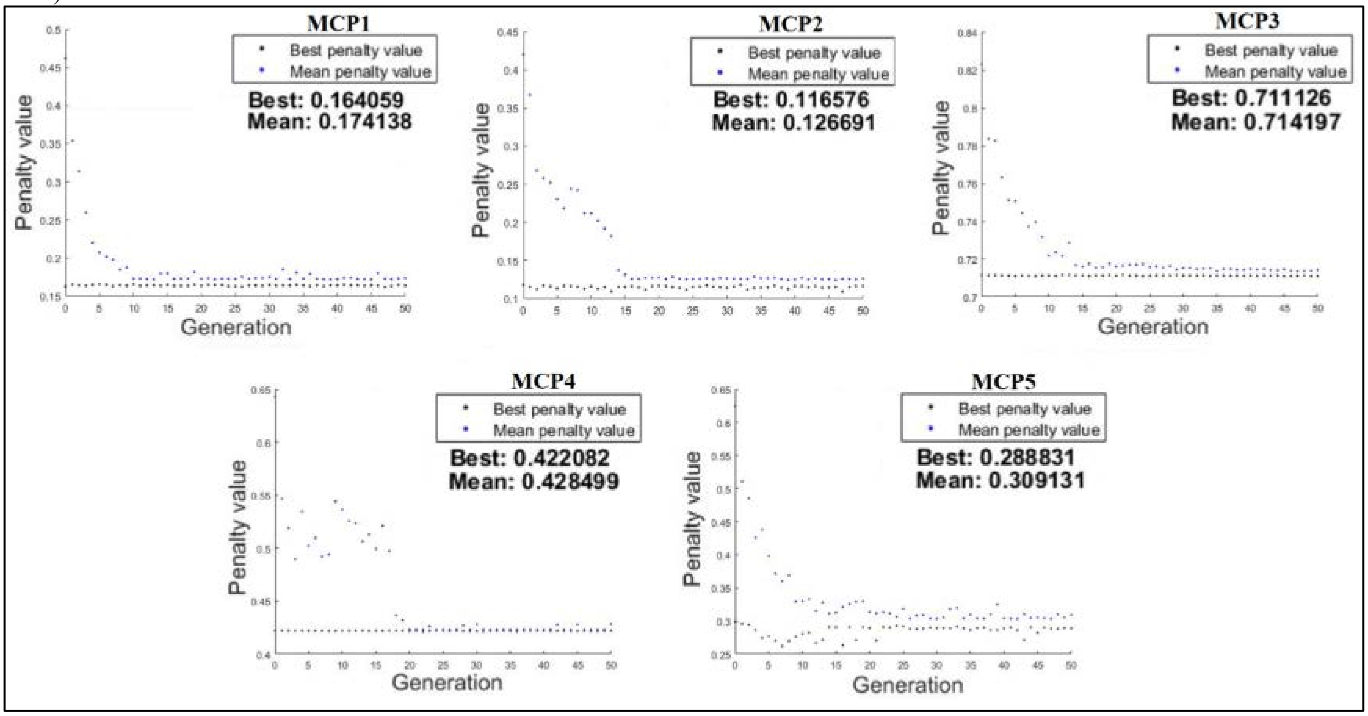
## IV. Results and Discussion

The objective function of GA is to minimize the uncorrelated data and increase the accuracy of prediction by controlling the number of neurons in hidden layer and combination of the transfer function in ANN. The optimization of GA achieved once the mean value of the fitness evaluation met the best fitness evaluation value. Results obtained that maximum generation set was sufficient for GA optimization, which the mean fitness evaluation value met the best fitness evaluation value within 50 generations, Fig. 3. In addition, the mean fitness evaluation values remained constant once met the best fitness evaluation values. The percentages of optimizations were in the range of 20.562 – 113.573 %, Tab. 1.

a) Kuala Terengganu



b)Kudat




c) Mersing

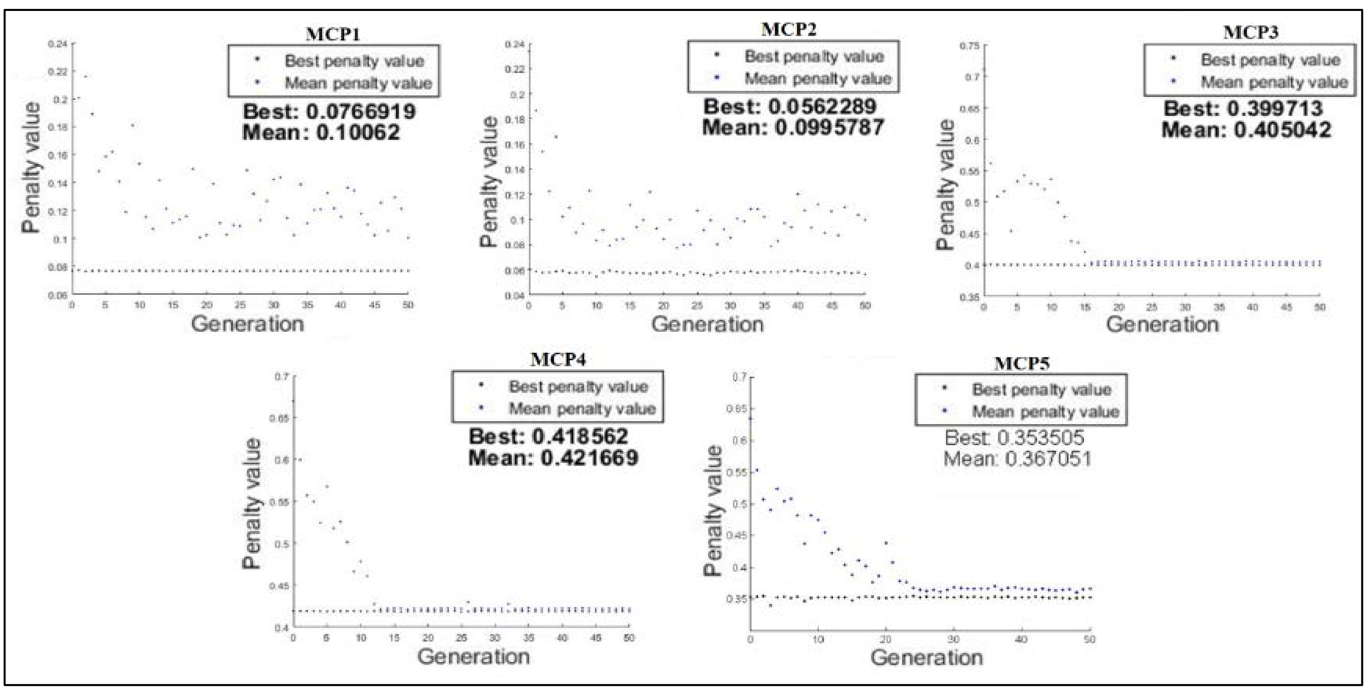
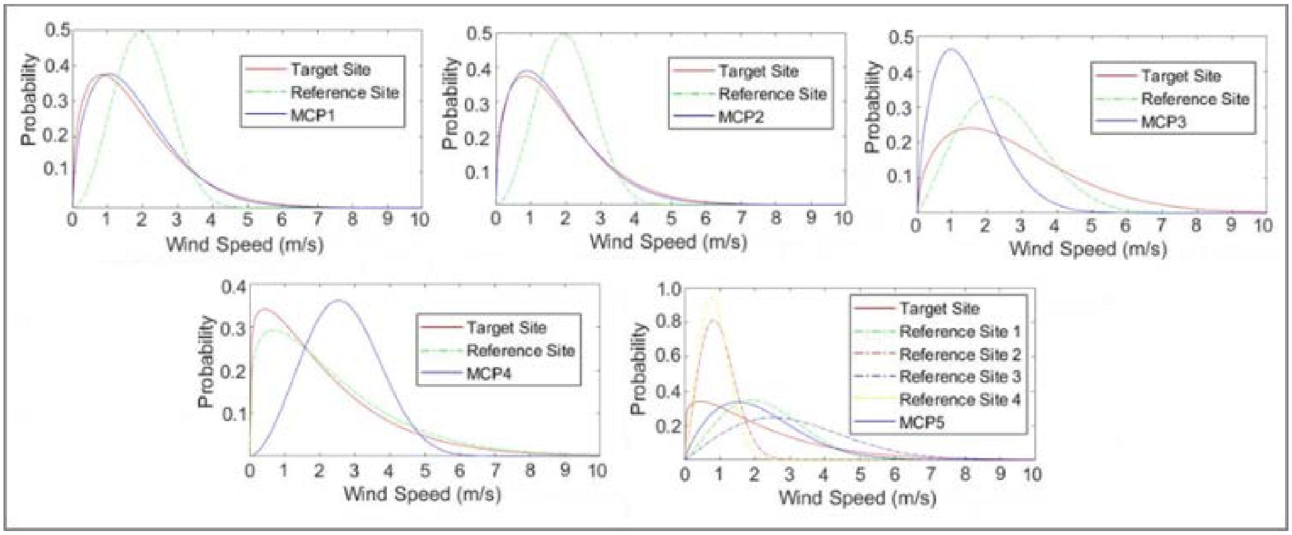
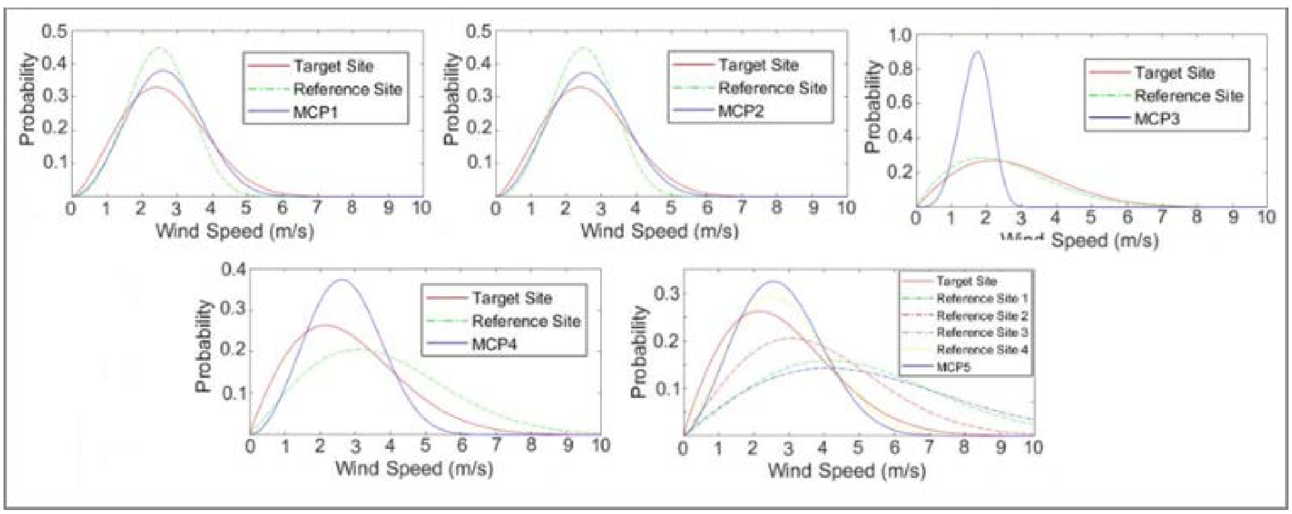



Figure 3. Optimization results.

a) Kuala Terengganu




b) Kudat




c) Mersing

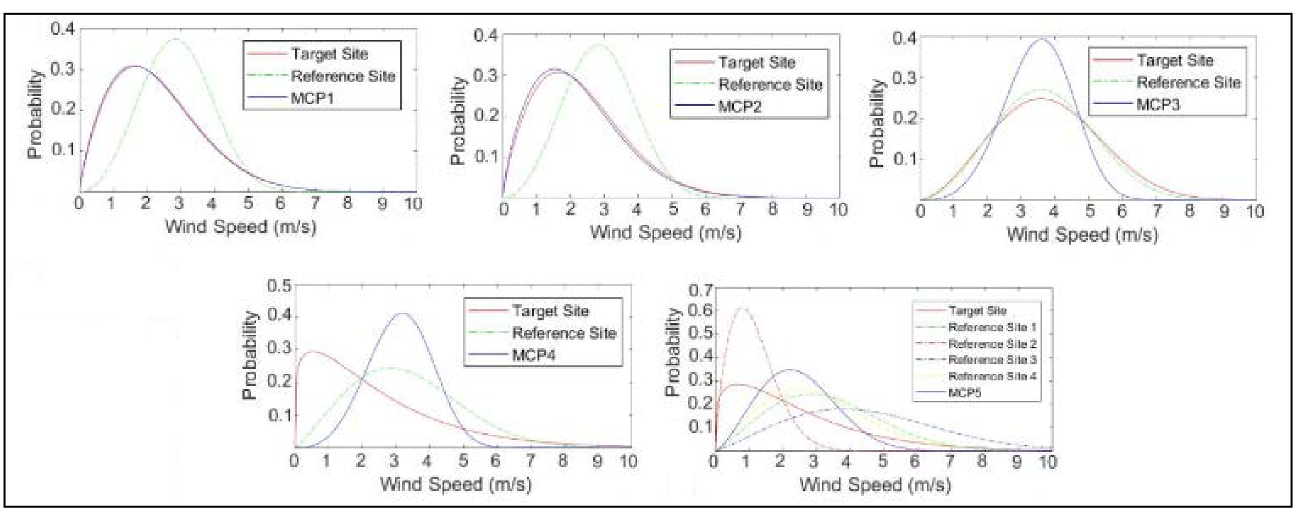



Figure 4. Weibull distribution plot.

TABLE I. CORRELATION COEFFICIENT, WEIBULL SCALE AND SHAPE PARAMETER, AND RMSE

| Site | Model | Optimization | | | Weibull Target Site data | | Weibull ANN Simulated data | | RMSE |
|---|---|---|---|---|---|---|---|---|---|
| | | Initial R | Best R obtained | Percentage Optimized (%) | Scale Parameter (m/s) | Shape Parameter | Scale Parameter (m/s) | Shape Parameter | |
| **Kuala Terengganu** | MCP1 | 0.760 | 0.916 | 20.562 | 1.968 | 1.425 | 2.003 | 1.562 | 0.577 |
| | MCP2 | 0.734 | 0.925 | 26.104 | 1.968 | 1.425 | 1.906 | 1.480 | 0.550 |
| | MCP3 | 0.540 | 0.751 | 38.933 | 3.109 | 1.517 | 1.676 | 1.671 | 1.959 |
| | MCP4 | 0.290 | 0.522 | 79.782 | 2.237 | 1.169 | 3.002 | 2.732 | 1.663 |
| | MCP5 | 0.351 | 0.682 | 94.000 | 2.237 | 1.169 | 2.385 | 1.794 | 1.353 |
| **Kudat** | MCP1 | 0.538 | 0.832 | 54.615 | 2.998 | 2.433 | 3.006 | 2.894 | 0.628 |
| | MCP2 | 0.564 | 0.878 | 55.632 | 2.998 | 2.433 | 2.992 | 2.822 | 0.541 |
| | MCP3 | 0.177 | 0.287 | 61.986 | 3.124 | 1.962 | 1.852 | 4.409 | 1.775 |
| | MCP4 | 0.352 | 0.576 | 63.961 | 3.175 | 1.914 | 3.036 | 2.877 | 1.272 |
| | MCP5 | 0.375 | 0.695 | 85.595 | 3.175 | 1.914 | 3.123 | 2.523 | 1.110 |
| **Mersing** | MCP1 | 0.760 | 0.922 | 21.284 | 2.614 | 1.789 | 2.581 | 1.768 | 0.545 |
| | MCP2 | 0.766 | 0.938 | 22.598 | 2.614 | 1.789 | 2.518 | 1.745 | 0.532 |
| | MCP3 | 0.290 | 0.618 | 113.573 | 4.288 | 2.653 | 3.878 | 3.997 | 1.288 |
| | MCP4 | 0.331 | 0.581 | 75.773 | 2.581 | 1.182 | 3.465 | 3.727 | 1.659 |
| | MCP5 | 0.358 | 0.643 | 79.664 | 2.610 | 1.230 | 2.791 | 2.379 | 1.316 |

Weibull distribution, Fig. 4, shown that the predicted data for MCP1 and MCP2 had similar distribution pattern with the target sites data. However, MCP3, MCP4 and MCP5 had different distribution pattern with the target sites data. Table 1 summarized the Weibull shape and scale parameters for target site data and predicted data as well as the RMSE.

The results reveal that prediction of wind data with MMD data, MCP1 and MCP2, given greater accuracy, which the best R and RMSE were in the range 0.832 – 0.938 and 0.532 – 0.628 respectively. The high accuracy of R-value obtained compare to MCP3, MCP4, and MCP5 were due to the scale of the data which MCP1 and MCP2 were daily data. Meteorological data, air temperature and pressure added into the model, MCP2, had been increased the accuracy of the prediction, were R-value for MCP1 were in the range of 0.832 – 0.922 compared to MCP2 were 0.878 – 0.938.

MCP model which based on hourly data, MCP3, MP4 and MCP5, ECFSR data shown better correlation compared to METAR data except for sites Kuala Terengganu. These are due to the METAR reference site for Kuala Terengganu were nearest to the target site compared to three other target sites. Model with four ECFSR reference sites, MCP5, had improved the accuracy of the prediction compared to MCP4 with one ECFSR reference site. The best R and RMSE for MCP4 and MCP5 were respectively in the range 0.522 – 0581 and 0.634 – 0.695; 1.272 – 1.663 and 1.110 – 1.353.

Indeed, the best R - value obtained from optimization were affected the Weibull shape and scale of predicted data. These had been clearly shown by Weibull shape and scale parameter for MCP1 and MCP2 were closer to target sites data than MCP3, MCP4 and MCP5. The Weibull scale and shape parameter for MCP1 and MCP2 simulated data were in the range 2.003 – 3.006 and 1.562 – 2.894; 1.902 – 2.992 and 1.480 – 2.822 respectively, while the target sites data were 1.968 – 2.998 and 1.425 – 2.433. On the other hand, the lowest R best value, MCP3 at Kudat site, made the Weibull shape parameter for simulated data extremely higher than others, 4.409.

## V. Conclusion

GA optimization achieved within 50 generations, the maximum generations; and the percentage of optimizations were in the range of 20.562 – 113.573 %. Prediction of wind data with daily MMD data, MCP1 and MCP2, given greater accuracy than hourly METAR and ECFSR data, which the best R and RMSE were in the range 0.776 – 0.938. The best R-value obtained from optimization were affected the Weibull shape and scale of predicted data. As the conclusion, the optimized ANN model could predict the long-term data for the target site with better accuracy.

## Acknowledgment

The authors would like to thank the Ministry of Sciences, Technology and Innovation of Malaysia (MOSTI); Ministry of Education (MOE) Malaysia; and School of Ocean Engineering, Universiti Malaysia Terengganu (UMT) for providing supervision and financial support under VOT 59440.

## References

[1] J. A. Carta, S. Velázquez, and P. Cabrera, "A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 362–400, Nov. 2013.

[2] J. A. Carta, P. Cabrera, J. M. Matías, and F. Castellano, "Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study," *Appl. Energy*, vol. 158, pp. 490–507, 2015.

[3] A. L. Rogers, J. W. Rogers, and J. F. Manwell, "Comparison of the performance of four measure-correlate-predict algorithms," *J. Wind Eng. Ind. Aerodyn.*, vol. 93, no. 3, pp. 243–264, 2005.

[4] A. More and M. C. Deo, "Forecasting wind with neural networks," *Mar. Struct.*, vol. 16, no. 1, pp. 35–49, 2003.

[5] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, Jul. 2010.

[6] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479–489, 2010.

[7] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," *Appl. Soft Comput. J.*, vol. 7, no. 2, pp. 585–592, 2007.

[8] A. S. Lapedes and R. Farber, "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling," *Proc. 1. IEEE Int. Conf. neural networks*, 1987.

[9] W. Chang, "Application of Back Propagation Neural Network for Wind Power Generation Forecasting," *Int. J. Digit. Content Technol. its Appl.*, vol. 7, no. 4, pp. 502–509, 2013.

[10] J. F. D. Addison, A. Hunter, J. Bass, and M. Rebbeck, "A neural network version of the measure correlate predict algorithm for estimating wind energy yield," *Proc. 13th Int. Congr. Cond. Monit. Diagnostic Eng.*, pp. 917–922, 2000.

[11] D. A. Bechrakis, J. P. Deane, and E. J. McKeogh, "Wind resource assessment of an area using short term data correlated to a long term data set," *Sol. Energy*, vol. 76, no. 6, pp. 725–732, 2004.

[12] S. Ahmed, *Wind Energy: Theory and Practice*, Second. Delhi: PHI, 2011.

[13] I. AWS Scientific, "Wind Resource Assessment Handbook," 1997.

[14] "Aviation Routine Weather Report (METAR)," 1993. [Online]. Available: http://meteocentre.com/doc/metar.html. [Accessed: 14-Jul-2018].

[15] A. Albani, M. Ibrahim, and M. Hamzah, "Assessment of Wind Energy Potential based on METAR data in Malaysia," *Energy*, vol. 36, no. 1, pp. 985–992, 2013.

[16] S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H. Chuang, H.-M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg, "The NCEP Climate Forecast System Reanalysis," *Bull. Am. Meteorol. Soc.*, vol. 91, no. 8, pp. 1015–1058, Aug. 2010.

[17] J. . Manwell, J. . McGowan, and A. . Rogers, *Wind Energy Explained Theory, Design and Application*, Second Edi. John Wiley & Sons, 2009.