Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 5, 2021, page 128-139

Modelling and Forecasting Wind Speed Data by Using ARIMA Fourier Model

[1]Nurul Amiera Salleh and [2]Siti Rohani Mohd Nor

[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia.

e-mail : [1] nurulamierasalleh@gmail.com, [2] sitirohani@utm.my

Abstract Wind is a renewable energy and wind speed is a key atmospheric characteristic that has a significant impact on the energy business. As the uses of wind speed are getting wide world, it has become a concern among administrators and also power grid dispatchers. Therefore, anaccurate measurement of wind speed prediction are needed in order to determine it future trend. Hence, the wind speed are modelled and forecast in this study. ARIMA Fourier model are used in forecast the wind speed data of Senai Station. A comparison between Autoregressive Integrated Moving Average ARIMA, and ARIMA Fourier has been made in determining which model are best used in forecasting wind speed data. The fit of the model are measures using Akaike's Information Criterion (AIC). Forecasting accuracy of both model are then quantify by Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Absolute Error (RMSE). Model with lowest error are selected as the best forecasting model. As a result, ARIMA Fourier model are chosen to forecast the wind speed data of Senai Station instead of ARIMA.

Keywords: Windspeed; Forecasting; Modelling; ARIMA; ARIMA Fourier.

1 Introduction

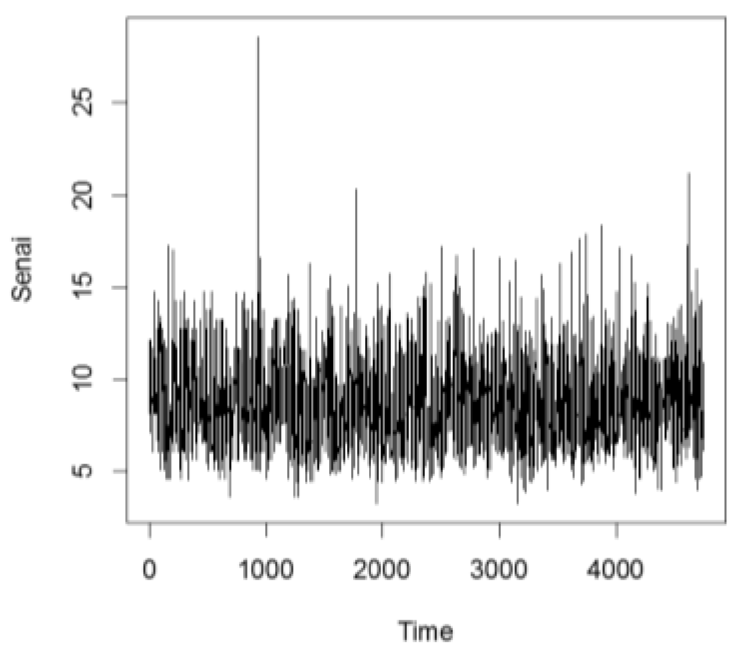
Wind is one of renewable green energy in the earth and wind speed is a key atmospheric characteristic that has a significant impact on the energy business (Jamaludin *et al*, 2016). There are some factors that can affect the wind speed which are temperature, air pressure, centripetal acceleration, earth's rotation and more. Time series, a collection of data recorded over a period of time discrete or continuous. As time series is applicable in various field, it provides tolls for selecting a model that can be used to forecast future circumstance. Forecasting is where the historical data are usedto determine the direction of future trends.

The data of wind speed are taken from Senai station. The data cover from 1st January 1985 until 31st December 2001. Forecasting wind speeds is critical for the safe and efficient functioning of wind turbines, beside it can help to reduce imbalance charges and penalties in global, competitive knowledge in real time stock trading, more efficient project construction, operations and maintenance planning (Lerner, 2009). Autoregressive integrated moving average (ARIMA) is a generalisation of an autoregressive moving average (ARMA) model. It is fitted to time series data either to best comprehend the data or to predict future points in the series (forecasting). ARIMA model are used to model and forecast the time series of the data but it is not adequate. Therefore, some regression are added to the ARIMA model and it then transform into ARIMA Fourier. The aim is to handle time series data more efficiently.

2 Methodology

2.1 The Dataset

The data used in this study was obtained from Malaysia meteorological department. A wind speed data of Senai station are used in this observations. A total of 4248 observations cover from 1st January 1985 until 31st December 2001. In this study, the last 365 observations will be used to forecast the dataset by using the best fitted model.



2.2 ARIMA Model

The autoregressive integrated moving average ARIMA (p, d, q) model of the time series $\{y_t\}$, t = 1,2,… is a very popular time series modelling, defined as

$$\emptyset(B)\Delta^d y_t = \theta(B)\varepsilon_t \tag{1}$$

where

y_t = wind speed data

ε_t = random error terms at time t

B = backward shift operator

then, $\emptyset(B)$ and $\theta\ (B)$ are order of p and q that are defined as

$$\emptyset(B) = 1 - \emptyset_1 B - \ \emptyset_2 B^2 - ..-\emptyset_p B^p \tag{2}$$

$$(B) = 1 - \theta_1 B - \ \theta_2 B^2 - ..-\theta_q B^q \tag{3}$$

where $\emptyset_1$, $\emptyset_2$, $\emptyset_p$ are autoregressive coefficients that attempt to predict an output of a system basedon the previous outputs and θ_1, θ_2, θ_q are the moving averages coefficients.

The first stage in modelling ARIMA (p,d,q) processes is determining whether the time seires is stationary or non- stationary. Before modelling can be done, non-stationary data must be differencing using the right order of differencing d. the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series are used to determine the appropriate values of autoregressive order p and moving average q.

2.3 ARIMA Fourier Model

The wind speed data has been subjected to a Fourier transform in order to reveal the series as a composition sinusoidal function.

ARIMA Fourier need to be added to the ARIMA model in order to deal with multiple seasonality. The time series y_t are then being extract in term of frequency domainy_f to structure it into sinusoidal functions. Below are the equation needed.

$$yf = \sum_{t=-\infty}^{\infty} y(t)^{-12\pi ft}$$

where *f* represent the frequency when y_f is evaluated. It can also write as:

$$e^{-12\pi ft} = \cos(2\pi ft) - i\sin(2\pi ft)$$

The application of formula transform the original signal y_t onto a set of sinusoidal functions, each correspond to a particular frequency component.

$$y_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} yfe^{-12\pi ft}dt$$

The equation above used to show how much frequency component needed to synthesize the original signal y_t.

2.4 Ljung-Box Test

Ljung and Box proposed a *Q*-Test called Ljung–Box test which is commonly used in Autoregressive Integrated Moving Average (ARIMA) modelling. It is applied to the residuals of a fitted ARIMA model, and in such applications, the hypothesis actually being tested is that the residuals from the ARIMA model have no autocorrelation, which is based on the *Q*-statistic given as:

$$Q = (N + 2)\sum_{j=1}^{L} \frac{\hat{\rho}_j^2}{(N-j)}$$

where N is the sample size, L is the number of autocorrelation lags included in the statistic, and 2 is the square sample autocorrelation at lag *j*. the Q-test statistic is asymptotically x^2 distributed under H_0 (no serial correlation). The acceptance of the null hypothesis of model accuracy under 95% significant levels is indicated by *p*-values above 0.05.

2.5 Measurement Errors

The accuracy of models is measured by Mean Absolute Error (MAE), Mean

Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Average magnitude of the errors in a set of forecast without considering the direction are called MAE. Formula of MAE are shown below:

$$MAE = \frac{\sum_{t-1}^{n} (y_t - \hat{y})}{n}$$

MAPE measure the forecast accuracy in percentage terms. It is calculated as the average of the unsigned percentage errors as shown below:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{y_t - \hat{y}}{y_t} \times 100$$

RMSE is used to measure the differences between the values forecast by the model andvalue observed.

$$RMSE = \sqrt{\frac{\sum_{t-1}^{n}(y_t - y)^2}{n}}$$

yt = Original value in time period t
$\hat{y}$ = Prediction value in time period t
n = Number of observations

The larger the error value, the less accurate the dataset's forecast. As a result, the forecasting model with the fewest RMSE and MAPE is chosen as the best forecasting model.

3 Result and Discussion

3.1 Data Description

Table 1 Descriptive statistics of daily wind speed of Senai station

Mean	Median	Minimum	Maximum	Standard Deviations	Skewness	Kurtosis
8.4590	8.2000	3.3000	28.6000	2.0608	1.0925	6.4409

The median, as seen in Table 1, indicates that the data's middle value is slightly lower than the mean. With 2.060 as the standard deviation, the data dispersion around the mean level is minimal. This indicates that the data was relatively stable. The data are considered to be right- skewed since the data have positive skewness and most of the data distribution is focused on the left. The dataset, however, was discovered to be non-normally distributed. As a result, two methods were utilized in this work to capture the variability and seasonality of wind speed data: the traditional method based on ARIMA model and the ARIMA Fourier model.

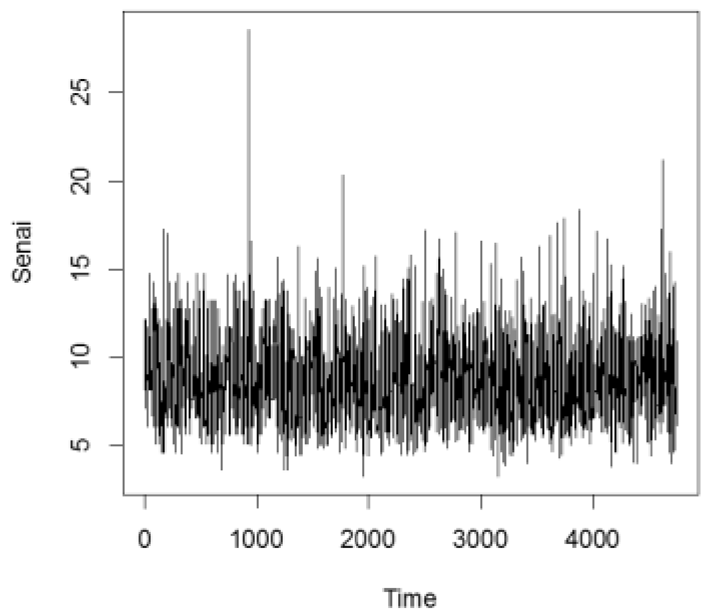


Figure 1 Times series plot of wind speed data

The total of 4748 data are then separated into two which is in- sample and out- sample data.In- sample data are range from 1st January 1985 until 31st December 2000 and are used to forecast the next 365 observations. The last 365 observations from 1st January 2001 until 31st December 2001 are used as out- sample data to be compare with the forecasted data. The trend of 4748 observations are recorded in Figure 1. From figure, it is shows that there are seasonality trend in theplot with the show of regular spikes in certain intervals. ARIMA Model and ARIMA Fourier Model

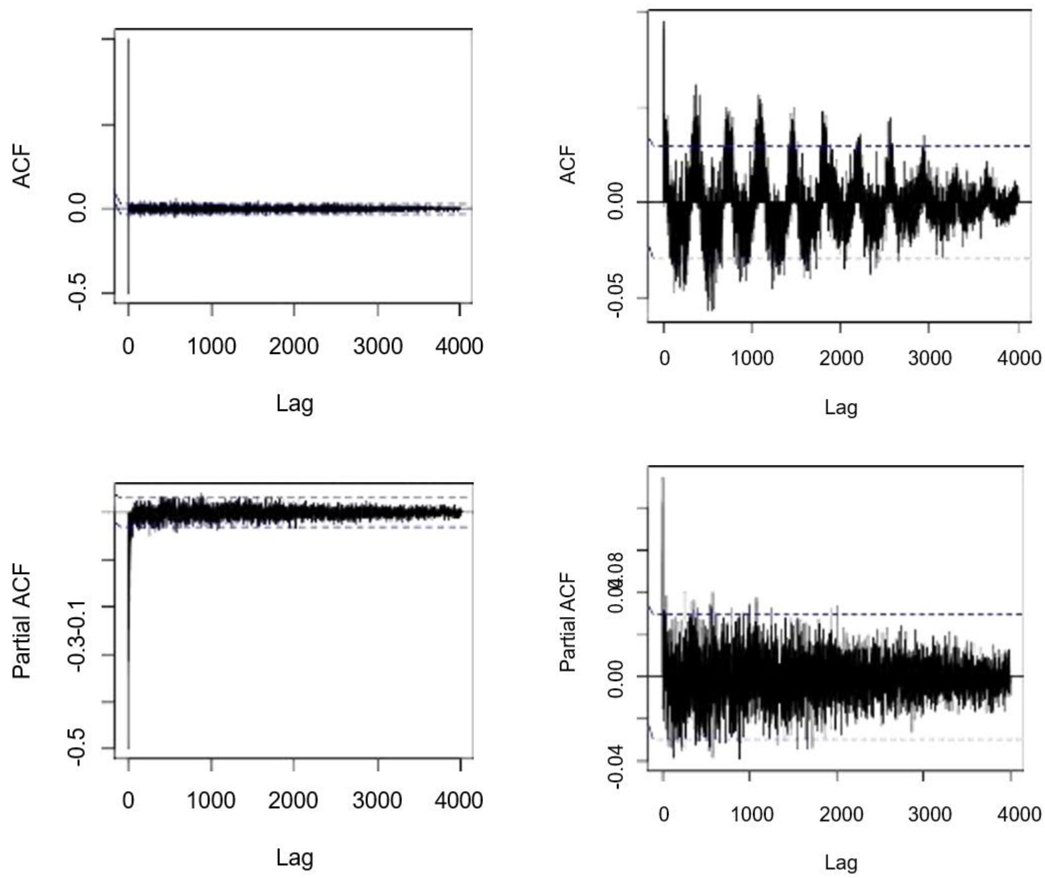
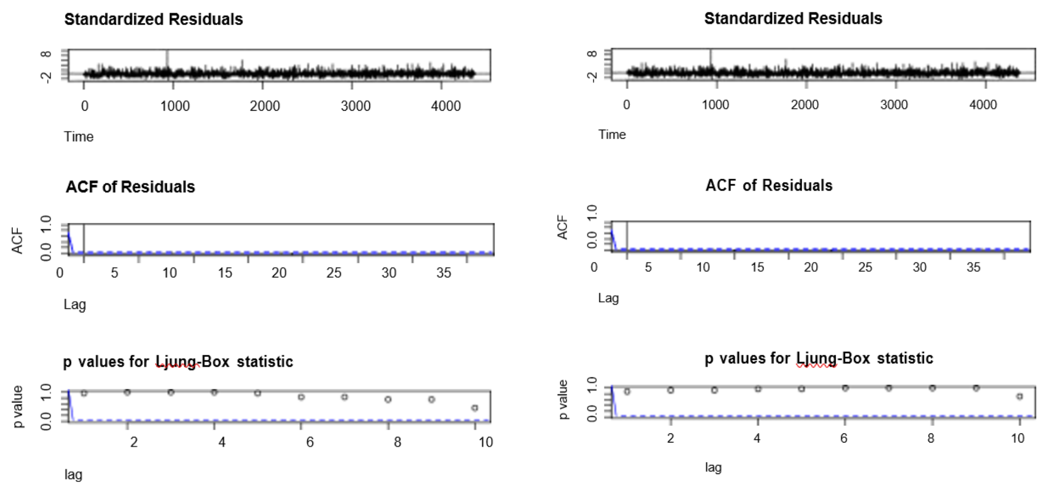


Figure 2 ACF and PACF plot of ARIMA (4,1,1) and ARIMA Fourier (1,0,1)

The Autocorrelation Function (ACF) and Partial Autocorrelation (PACF) must be displayed in order to get the model's equation. Moving Averagee (MA) order is denoted by ACF, while Autoregressive Order (AR) is denoted by PACF. The order of the autoregressive term (*p)* and moving average *(q)* must be determind in order to identify the best- fitting model. The Ljung-Box Test, which examines residuals from a time series model that resemble white noise, must be used to assess all feasible fitted models.



Diagnostic result of ARIMA (4,1,1) Diagnostic result of ARIMA Fourier (1,0,1)

Figure 3 Diagnostic result of ARIMA (4,1,1) and ARIMA Fourier (1,0,1)

The wind speed data are found to be adequate if the residuals is normal with zero mean, constant in variance, independent and are normally distributed. However, the plot of the residuals of ARIMA (4,1,1) shows several spikes which indicates the model could not capture some information from the data very well. From diagnostic test of ARIMA Fourier (1,0,1), it is found that ARIMA Fourier model is adequate to this series of data as the residuals of the dataset fluctuatearound mean level. ACF

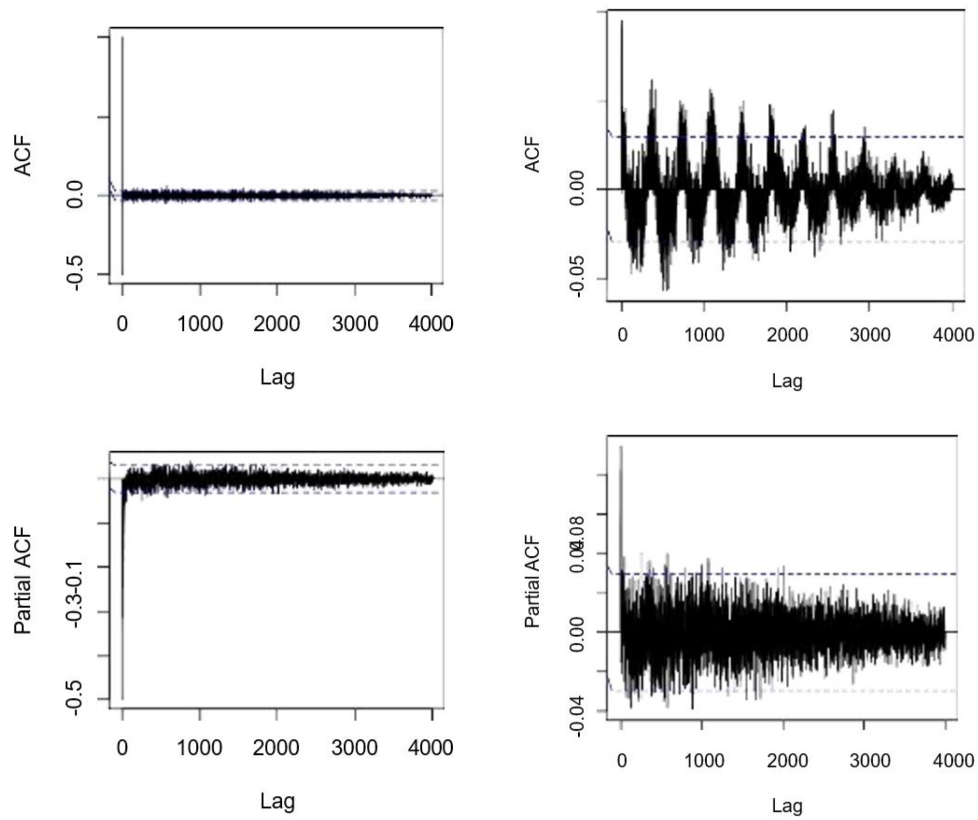


Figure 2 ACF and PACF plot of ARIMA (4,1,1) and ARIMA Fourier (1,0,1)

plot of the residuals also exhibit no relations between data. It is then reconfirm with all the value of *p*-value which yield above 0.05, which indicate that the hypothesis accept H0 and the residuals was normal. Hence, ARIMA Fourier model is suggested to use instead of ARIMA model.

4 Forecasting

The data for future wind speeds was forecasted over the next 365 days. The forecasting data was represented as a blue line, as shown in the diagram below.

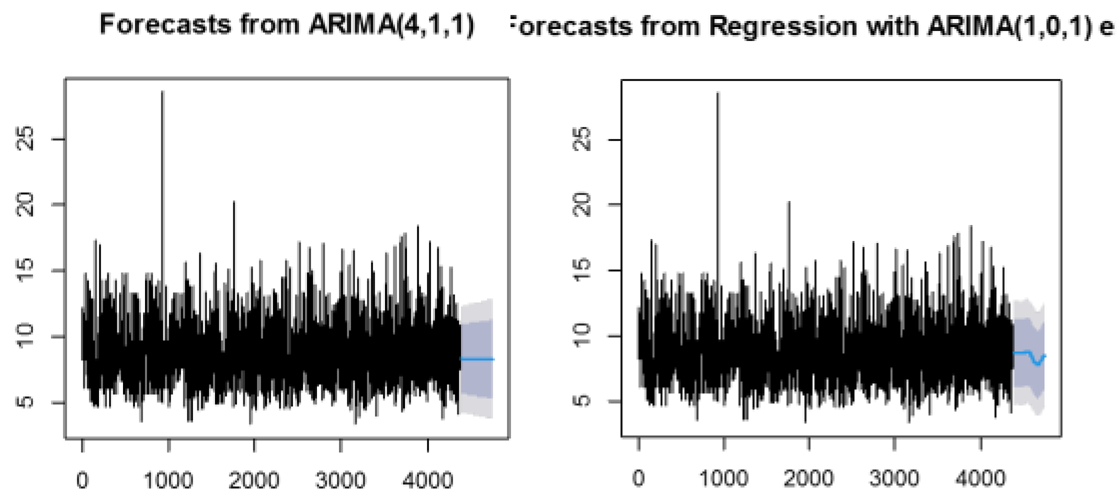


Figure 4 Forecast plot of ARIMA (4,1,1) and ARIMA Fourier (1,0,1)

To confirm that ARIMA Fourier (1,0,1) is better than ARIMA (4,1,1), the accuracy of both ARIMA and ARIMA Fourier models are compared by using measurement errors in terms of RMSE and MAPE values. Model that obtain lowest error values are dominated as the best model. The value of measurement errors for both models are shown in Table 2 below.

Table 2 The measurement errors of ARIMA (4,1,1) and ARIMA Fourier (1,0,1)

	In- sample		Out- sample	
	RMSE	MAPE	RMSE	MAPE
ARIMA (4,1,1)	2.0371	19.0490	2.0790	0.1778
ARIMA Fourier (1,0,1)	2.0209	18.8179	2.0774	0.1734

5 Conclusion

As a conclusion, ARIMA Fourier (1,0,1) model perform well with approximate value of MAPE and RMSE which are 18.8179 and 2.0209 for in-sample data and 0.1734 and 2.0774 for out-sample data when compare with ARIMA (4,1,1) model, with larger value of MAPE and RMSE approximately 19.0490 and 2.0371 for in- sample data and 0.1778 and 2.0790 for out- sample data respectively. Therefore, ARIMA Fourier (1,0,1) outperformed ARIMA (4,1,1) since it has the lowest measurement errors in terms of RMSE and MAPE. Hence, ARIMA Fourier (1,0,1) arebetter in capturing the wind speed data of Senai station.

References

[1] Akorede, M. F., Rashid, M. I. M., Sulaiman, M. H., Mohamed, N. B. and Ab Ghani, S.

B. (2013). Appraising the Viability of Wind Energy Conversion System in the Peninsular Malaysia.*Energy conversion and management,* 76**,** 801-810.

[2] Alsharif, M. H., Younes, M. K. and Kim, J. (2019). Time Series Arima Model for Predictionof Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. *Symmetry,* 11**,** 240.

[3] Fu, T. and Wang, C. (2018). A Hybrid Wind Speed Forecasting Method and Wind Energy Resource Analysis Based on a Swarm Intelligence Optimization Algorithm and an Artificial Intelligence Model. *Sustainability,* 10**,** 3913.

[4] Iwok, I. and Udoh, G. (2016). A Comparative Study between the Arima-Fourier Modeland the Wavelet Model. *American Journal of Scientific and Industrial Research,* 7**,** 137-144. [5]Jamaludin, A. R., Yusof, F., Kane, I. L. and Norrulasikin, S. M. A Comparative Study between Conventional Arma and Fourier Arma in Modeling and Forecasting Wind Speed Data. AIP Conference Proceedings, 2016. AIP Publishing LLC, 060022.

[6] Li, G. and Shi, J. (2010). On Comparing Three Artificial Neural Networks for Wind SpeedForecasting. *Applied Energy,* 87**,** 2313-2320.

[7] Niu, D., Liang, Y. and Hong, W.-C. (2017). Wind Speed Forecasting Based on Emd andGrnnOptimized by Foa. *Energies,* 10**,** 2001.

[8] Pearre, N. S. and Swan, L. G. (2018). Statistical Approach for Improved Wind Speed Forecasting for Wind Power Production. *Sustainable Energy Technologies and Assessments,* 27**,**180-191.

[9] Tian, C., Hao, Y. and Hu, J. (2018). A Novel Wind Speed Forecasting System Based onHybrid Data Preprocessing and Multi-Objective Optimization. *Applied Energy,* 231**,** 301-31