International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Forecasting Wind Speed Data by Using a Combination of ARIMA Model with Single Exponential Smoothing

Nur A.B. Kamisan[1*], Muhammad H. Lee[1], Siti F. Hassan[2], Siti M. Norrulashikin[1], Maria E. Nor[3], Nur H.A. Rahman[4]

[1] Science Mathematical Department, Faculty of Science, Universiti Teknologi Malaysia, Johor 81310, Malaysia
[2] Centre of Science Foundation Study, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[3] Fakulti Sains Gunaan dan Teknologi, Universiti Tun Hussein Onn Malaysia, Johor 86400, Malaysia
[4] Science Mathematical Department, Universiti Putra Malaysia, Serdang 43400, Malaysia

Corresponding Author Email: nurarinabazilah@utm.my

ABSTRACT

Wind serves as natural resources as the solution to minimize global warming and has been commonly used to produce electricity. Because of their uncontrollable wind characteristics, wind speed forecasting is considered one of the best challenges in developing power generation. The Autoregressive Integrated Moving Average (ARIMA), Simple Exponential Smoothing (SES) and a hybrid model combination of ARIMA and SES will be used in this study to predict the wind speed. The mean absolute percentage error (MAPE) and the root mean square error (RMSE) are used as measurement of efficiency. The hybrid model provides a positive outcome for predicting wind speed compare to the single model of ARIMA and SES.

1. INTRODUCTION

Electricity is a type of energy source which is used in our everyday lives to power devices. Today, global warming, energy crisis and depletion of fossil fuels are the major challenges faced in the world. Producing renewable power supply is one of the solutions for industry to reduce greenhouse gas emissions from everyday activities. Wind energy is one of the sustainable and environmentally friendly among the other renewable sources of energy [1]. It has evolved rapidly to produce electricity in recent decades. At the moment, wind power is widely used not only in developed countries but also in developing countries.

The wind speed forecasts will reduce operating costs, increase power system strength and also boost the power generation capacity and efficiency of wind turbines [2, 3]. The unnatural characteristics of the wind, however, had made it difficult to predict wind speed with minimal error [2]. Compared with wind, conventional power is easier to predict. Wind power generation is unpredictable as it regulates the weather, wind speed direction, variations in seasonal temperature and humidity of the air. The unpredictable wind speeds will affect the amount of electricity produced and will therefore create challenges to balance the power between supply and demand [4].

One of the solutions in minimizing the deficiency of wind energy is by predicting the wind speed. There are several wind speed forecasting approaches that include physical, mathematical, and artificial intelligence (AI) approaches for the predicted development. All forecasting models can offer different results and varying degree of precision [1]. Statistical methods are common techniques of analysis of the data in time series [5].

Single exponential smoothing method has been created by Robert Goodell Brown in 1956 and developed by Charles C. Holt in 1957. There are three types of exponential smoothing model; the single exponential smoothing, double exponential smoohing and triple exponential smoothing. Single exponential smothing is for data with no trend and seasonality. Double exponential smoothing is for data with trend but has no seasonality. While, triple exponential smoothing is for data with trend and seasonality [6]. Since exponential smoothing is a simple and straightforward model, it is usually being used as a combination with other model [7, 8].

Many wind speed forecast studies have been done by using ARIMA models in energy sector. Grigonyte and Butkeiciute [9], used ARIMA models to forecast the wind speed in short term for Baltic region because ARIMA is a well-known model used to forecast short term wind speed. Based on their findings, ARIMA is appropriate for a short term prediction such as 6 to 8 hours period ahead compare to the 24 hours ahead prediction and they came to a conclusion that ARIMA forecast quality depends on consistency of the data. Palomares et al. [10] compared the performance of ARIMA and neural network for short term wind speed forecast and found that ARIMA is a better model than neural network since ARIMA could give similar results as neural network but the calculation number is lower compare to neural network. This shows that ARIMA is a good model in forecasting a short term wind speed.

Although the studies above have shown that ARIMA is a good model for wind speed forecasting, nowadays, hybrid model also get a lot of attention in enhancing the wind speed forecasting as suggested by Grigonyte and Butkeiciute [9] in their paper where they mentioned, a combination of ARIMA with other model could improve the forecast of the wind speed. One of the finding on hybrid ARIMA model can be seen from the study of Cadenas and Rivera [11]. It shown that hybrid ARIMA and ANN has improved the forecasting of the wind

speed in Mexico compare to just using a single model of ARIMA or ANN only. Ye et al. [12] combined ARIMA with neural-fuzzy system to forecast the wind speed and found that this model outperform the single used of ARIMA and neural-fuzzy system models. There are many other studies not discussed in this paper have shown that hybrid model has improved the performance of forecasting. By combining two or more models together, it could cover one model limitations with other model advantages. Thus, in this paper a hybrid between ARIMA and SES is proposed.

2. METHODOLOGY

2.1 ARIMA model

The Autoregressive Integrated Moving Average (ARIMA) is widely used for short and long term data in time series analysis. This model was commonly used in a variety of sectors including energy, economics and finance. However, ARIMA has a limitation where it is only appropriate for stationary data.

The ARIMA model is the combination of three concepts which are autoregression, differencing and moving average [13]. The general form of the Box-Jenkins approach with non-seasonality is defined as:

$$\phi_p(B)\nabla^d y_t = \theta_q(B)a_t \tag{1}$$

where,

$\phi_p(B)$ is the nonseasonal autoregressive operator of order p,

$\theta_q(B)$ is the nonseasonal moving average operator of order q,

B is a backward shift operator,

$\nabla^d y_t$ is the degree of differencing,

a_t is a white noise process with a mean equal to 0 and a variance equal to σ_a^2,

$$\begin{aligned}
&\nabla = \nabla_1 = 1 - B, \\
&\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \\
&\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \\
&B^j y_t = y_{t-j}.
\end{aligned} \tag{2}$$

After data pre-processing, the next step is identified suitable model by testing the stationarity of the data. The stationary data has constant mean, variance and auto-covariance all the time. ACF and PACF graphs will be used to check the stationarity of data. Checking stationarity of data can also be done by the Augmented Dickey-Fuller (ADF) test. The null hypothesis is set as $\pi = 0$ while alternative hypothesis is set as $\pi < 0$. ADF test statistics can be formulated as:

$$DF_\tau = \frac{\hat{\pi}}{SE(\hat{\pi})} \tag{3}$$

where, $\hat{\pi}$ is the estimation of π and $SE(\hat{\pi})$ is standard error of coefficient.

The null hypothesis will be rejected from the ADF test statistics value if DF_τ at significance level is smaller than the critical value. Rejecting null hypothesis shows the data is stationary. The data is otherwise non-stationary. When the data is non-stationary, it takes transformation to make the data

stationary which is called differentiation. The number of differencing orders should be selected correctly so that the data from the time series fluctuate around the mean value. After obtaining stationary data, the selection of the appropriate order of autoregressive (p) and moving average (q) will be performed based on the ACF and PACF plots. The criteria for the ACF and PACF functions are shown in Table 1.

Table 1. ACF and PACF criteria

Model	ACFs	PACFs
MA (q)	Exponential decay	Cuts off after lag q
AR (p)	Cuts off after lag q	Exponential decay
MA (p, q)	Exponential decay	Exponential decay

After that, parameter estimation will be done to obtain the model's coefficients. The model of the ARIMA will be constructed based on the criteria in the table above. Then, the residual analysis is used for diagnostic testing to assess the adequacy of the model. If the diagnostic check is passed, the selected model will be used to begin a prediction. Otherwise, the model will be amended and repeated to estimate.

2.2 Single exponential smoothing

One of the simplest methods in time series is called Single Exponential Smoothing (SES). With slow change of means over time, SES will be applied to the stationary results. The general formula for SES is express as follows:

$$F_{t+1} = \beta y_t + (1 - \beta)F_t \tag{4}$$

where,

F_{t+1} is forecast value for the next period t,

F_t is forecast value in period t,

β is smoothing constant,

y_t is observed value in period t.

Smoothing constant can be determined by trial and error experimentation to obtained optimal results.

2.3 Hybrid model

Hybrid model is a mixture of two statistical models, or a mixture of Artificial Intelligence (AI) or other forms. There are several ways to combine to predict wind speed. The first model will be used to forecast the out-sample results, while the second model will be used to forecast the first model's out-sample residuals. ARIMA will be combined with SES in this study to form hybrid method. The first model is ARIMA model and the second is model SES. Taking the idea from Zhang [14], he developed hybrid formula system.

$$y_t = L_t + N_t \tag{5}$$

where,

L_t is the first component,

N_t is the second component.

First component is estimated by ARIMA model and residuals, e_t obtained from the ARIMA model.

$$e_t = y_t - \hat{L}_t \tag{6}$$

where, $\hat{L}_t$ is the forecast value for ARIMA at time t.

Applied this theory in the study, a hybrid time series forecast formula can be expressed as

$$\hat{Y}_t = \hat{L}_t + \hat{E}_t \qquad (7)$$

where, $\hat{E}_t$ is the estimated residuals value for SES.

Hybrid method mainly involves the process of finding $\hat{L}_t$ and $\hat{E}_t$ before forecasting obtained. The steps-by-steps hybrid modelling process is illustrated below in Figure 1:
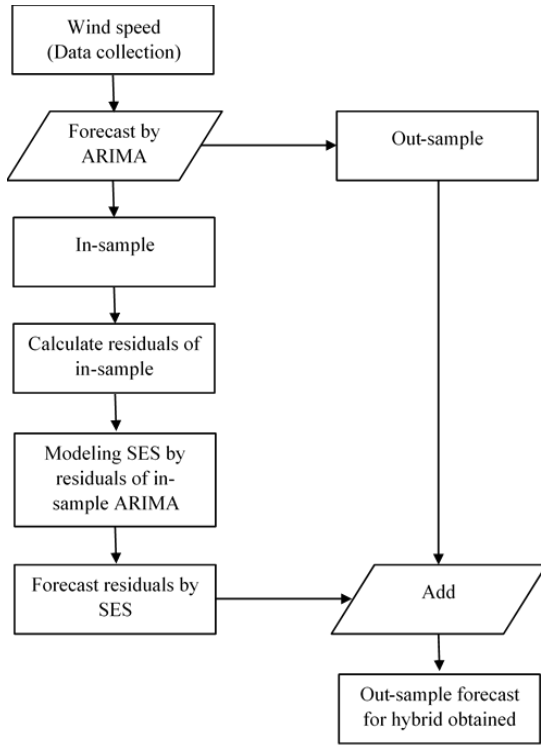

Figure 1. Step-by-step process of the hybrid model

Once hybrid modelling process is complete, the performance of hybrid method will be calculated and compare with the other models.

2.4 Performance indicator

The performance will be measured for each model by using MAPE and RMSE. The following shows MAPE and RMSE formula:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|(100) \qquad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2} \qquad (9)$$

3. RESULTS AND DISCUSSION

3.1 Single exponential smoothing

Time series plot of Alor Setar monthly average wind speed from January 2013 to June 2017 is plotted to observe whether the data has trend or seasonality. Below is the time series plot for Alor Setar.

From the time series plot in Figure 2 above, the monthly average wind speed has no trend and no seasonality. To fulfil

the property of the data, exponential smoothing with one smoothing parameter will be used to analyze the data.
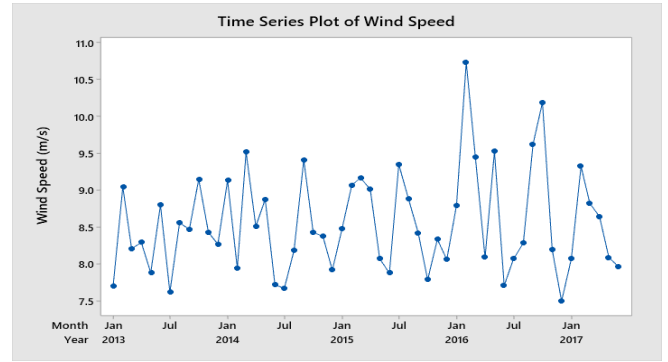

Figure 2. Time series plot of Alor Setar wind speed

The smoothing parameter can be determined by using trial and error experimentation to obtain optimal results. Minitab can assist in finding smoothing parameter to get optimal results. Based on the Minitab output, the smoothing parameter α for in-sample data is 0.0336. After optimal smoothing parameter is obtained, the equation for SES model can be expressed as:

$$F_{t+1} = 0.0336\, y_t + (1 - 0.0336\,)F_t \qquad (10)$$

For single exponential smoothing, the α had great influence on the F_{t+1}. When the best α value is chosen, the fitted value will also be closer to the actual value.

3.2 ARIMA model

Before we start modelling the data, the normality of the data is checked first by using Anderson-Darling test. The Anderson-Darling (AD) test p-value will determine whether the data meets normal distribution. In this study, the data do not follow normal distribution at level of significance 0.05, so it will perform power transformation with optimal power parameter λ=-3. Pre-processing of a data should be done before the modelling process. Box-cox transformation is a technique of pre-processing data primarily to transform raw data to a normal form of distribution.

Data stationarity is the major property of ARIMA. If the data are not stationary, the transformation of initial time series data into stationary data would require differentiation. The Augmented Dickey-Fuller (ADF) test can determine the stationarity of the data. Figure 3 is the result of the ADF test generated by the R software.

Figure 3. Result of ADF test generated by R

The hypothesis is set to $H_0: \pi = 0$ vs $H_1: \pi < 0$. From the ADF test obtained, p-value is equal to 0.01 which is smaller than critical value, 0.05. Since p-value is smaller than critical value, it has sufficient evidence to reject the null hypothesis. This means that the data is stationary at level of significance $\alpha = 0.05$.

Next, we need to determine the orders of the ARIMA model. The number of orders in ARIMA can be determined by observing the significant lags in ACF and PACF plot. Figure 4 is the ACF and PACF plots.
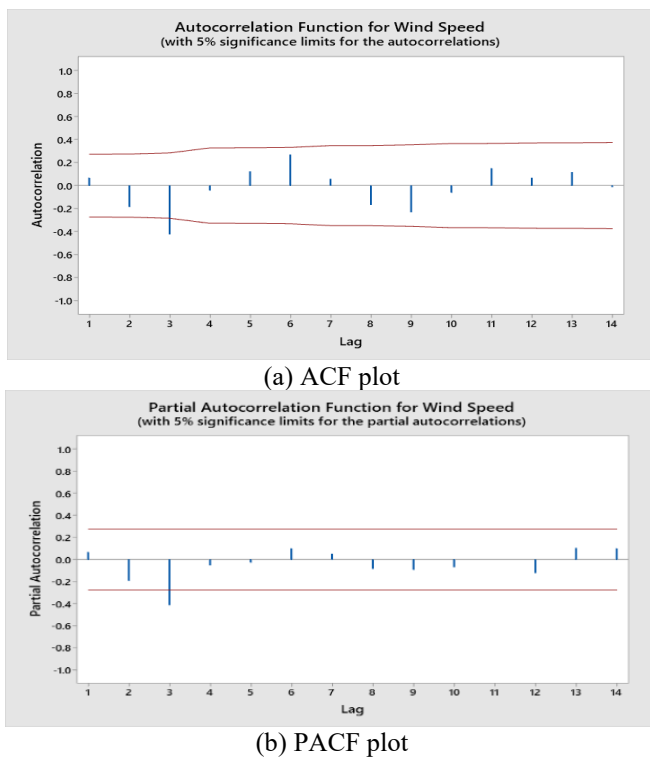

(a) ACF plot

(b) PACF plot

Figure 4. ACF and PACF plots for ARIMA model identification

ACF and PACF have no significant spikes except at lag 3 which suggesting the orders of the data is $p=3$ and $q=3$. While for the d, the orders is equal to zero due to no differencing is required. Thus, from the information above, the ARIMA model for this data is set as ARIMA(3,0,3). And by using Minitab software, the parameter for the ARIMA(3,0,3) is generated as the equation below:

$$
\begin{aligned}
y_t = {}& 1.9970y_{t-1} - 1.9900y_{t-2} + 0.9960y_{t-3} + a_t \\
& - 1.9742a_{t-1} + 1.7672a_{t-2} \quad\quad (11) \\
& - 0.8390a_{t-3}
\end{aligned}
$$

The next step is checking the adequacy of the model. The adequacy of the model can determine whether the model can fit the data. If the p-value of Ljung-Box test statistics is larger than critical value at significance level, we can conclude that the residuals of the model is not autocorrelated. Ljung-Box test is a statistical test to check whether there is an existance a serial of autocorrelation of residuals. Figure 5 below shows the Minitab output for Ljung-Box test for ARIMA (3,0,3). The p-value of the Ljung-Box test is larger than $\alpha=0.05$ which indicates the ARIMA (3,0,3) model fit in with the data very well.

Figure 5. Ljung-Box test result from R

3.3 Hybrid model

The hybrid method is continuing to proceed by taking in-sample residuals of ARIMA model discussed in 3.2 to forecast the out-sample residuals by using SES model. As in this hybrid method, SES is used to forecast the out-sample residuals instead of forecast the monthly average wind speed. To forecast the monthly average wind speed, the out-sample wind speed that has been obtained from ARIMA will be added to the forecasted out-sample residuals that will be obtained from the SES.
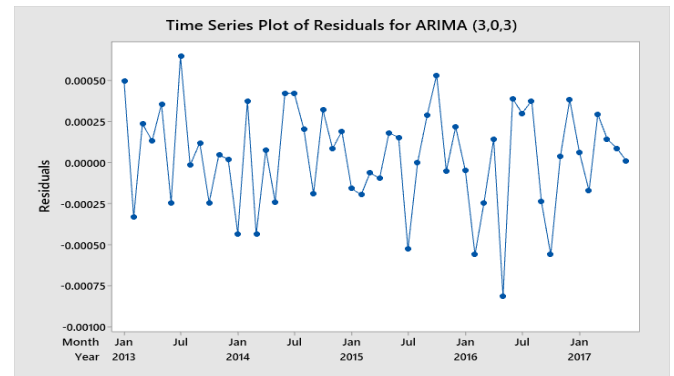

Figure 6. Time series plot of residuals of in sample from ARIMA model

As shown in Figure 6, the residuals of the in sample from ARIMA model does not have any trend or seasonal pattern although its fluctuate around 0.0000. due to this pattern, SES model is chosen to forecast these residuals. The smoothing parameter, β of the residuals is calculated from 0.1 to 0.9 to find the best β. From the MAPE and RMSE in Table 2 below, smoothing parameter β equal to 0.4 has the smallest value if MAPE and RMSE.

Table 2. MAPE and RMSE for smoothing parameter of residuals

β	MAPE	RMSE
0.1	6.904	0.628
0.2	6.478	0.578
0.3	6.285	0.553
0.4	**6.243**	**0.547**
0.5	6.289	0.553
0.6	6.379	0.564
0.7	6.481	0.578
0.8	6.579	0.591
0.9	6.661	0.603

Therefore, the out-sample residuals of the wind speed will be using smoothing parameter of 0.4 for the SES model. After we generated the forecasted residuals, the forecast wind speed from ARIMA (3,0,3) will be added to the forecasted residuals in order to obtain a new forecast dataset of the wind speed.
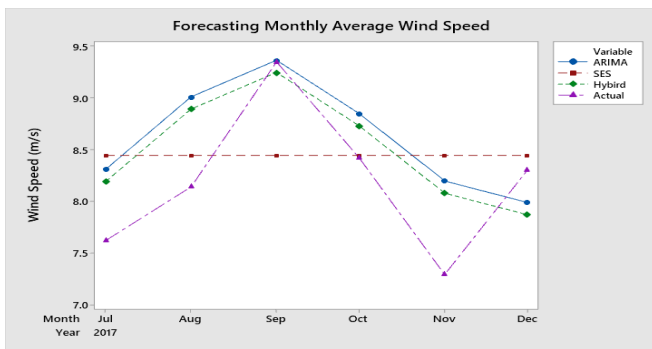
3.4 Models comparison

The performance of each model will be measured by using MAPE and RMSE. Performance of the three models will be compared to see the best model that could forecast the data accurately. Model with the smallest value of MAPE and RMSE will be selected as the best model.

Table 3. Model comparison by using MAPE and RMSE

Model	MAPE	RMSE
ARIMA(3,0,3)	6.85	0.62
SES	7.98	0.75
Hybrid	**6.23**	**0.55**

From the Table 3 above, it is clearly shown that the hybrid model has the smallest value of MAPE and RMSE. Although the different between all the three models are small, it is still an improvement in the forecasting. ARIMA and hybrid model has a quite small differences and this is understood since the hybrid model comes from ARIMA and SES model. The pattern is expected to follow the ARIMA model with a little adjustment of the residuals to improve the forecast and make it closer to the actual data. To help us see the forecast clearer, a time series plot is used to illustrate the forecast from each model with the real data.

Figure 7. Graphical plot of models comparison

The performance of the models can be seen in Figure 7. The forecasted wind speed by hybrid model is the closest to the actual wind speed and followed by ARIMA (3,0,3). As being mentioned in the paragraph above, the hybrid model is expected to follow the ARIMA pattern but closer to the actual data. However, for the SES model, if we refer to the result in Table 3, although the MAPE and RMSE values have a small difference with the other two models, but when it is illustrated into a plot, we could see that the SES forecast is actually a straight line. From the MAPE and RMSE we might think that SES is a good model but when being illustrate, SES is actually not a good model since it did not give any pattern of the forecast wind speed.

As for the hybrid model, by using the ARIMA model and SES model, it has shown an improvement of the forecast. The ARIMA model has the advantage of capturing the pattern of the data while the SES could model the residuals that has linear pattern. Due to the uncontrollable wind characteristics, simple model such as SES model is not enough to capture the fluctuation in the data. Therefore, the combination of few time series model should be considered in order to improve the forecasting. As the results, hybrid model with combination of ARIMA and SES have the highest accuracy in this study.

4. CONCLUSION

In conclusion, the hybrid model gives a promising result compare to ARIMA and SES model used as forecasting model alone. As mention earlier, hybrid model could give a better forecast since the model is a combination of two models which could complement each other. In this case, ARIMA is used to forecast the speed whilst the SES is used to forecast the residuals. By combining these two forecasts, a new forecast will give data which are closer to the real data. For further improvement, a combination with more advanced model such as neural network or fuzzy time series could be considered.

As for the performance presentation it is important to include the result in illustration too. It could help visualise the result of a finding. Even when performance test gives small value, it is still important to use graphical plot to see the real outcome.

REFERENCES

[1] Zhou, Q., Wang, C., Zhang, G. (2019). Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. Applied Energy, 250: 1559-1580. https://doi.org/10.1016/j.apenergy.2019.05.016

[2] Wang, Y., Gao, J., Xu, Z., Luo, J., Li, L. (2020). A prediction model for ultra-short-term output power of wind farms based on deep learning. International Journal of Computers Communications & Control, 15(4): 29-38. https://doi.org/10.15837/ijccc.2020.4.3901

[3] Zhao, J., Guo, Z.H., Su, Z.Y., Zhao, Z.Y., Xiao, X., Liu, F. (2016). An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. Applied Energy, 162: 808-826. https://doi.org/10.1016/j.apenergy.2015.10.145

[4] Chen, J., Zeng, G.Q., Zhou, W., Du, W., Lu, K.D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Conversion and Management, 165: 681-695. https://doi.org/10.1016/j.enconman.2018.03.098

[5] Demuyakor, J. (2020). Coronavirus (COVID-19) and online learning in higher institutions of education: A survey of the perceptions of Ghanaian international students in China. Online Journal of Communication and Media Technologies, 10(3): e202018. https://doi.org/10.29333/ojcmt/8286

[6] Gardner Jr, E.S. (2006). Exponential smoothing: The state of the art—Part II. International Journal of Forecasting, 22(4): 637-666. https://doi.org/10.1016/j.ijforecast.2006.03.005

[7] Akcan, S. (2017). Wind speed forecasting using time series analysis methods. Çukurova Univ. J. Fac. Eng. Archit., 32(2): 161-172.

[8] Kusiak, A., Zhang, Z. (2010). Short-horizon prediction of wind power: A data-driven approach. IEEE Transactions on Energy Conversion, 25(4): 1112-1122. https://doi.org/10.1109/TEC.2010.2043436

[9] Grigonytė, E., Butkevičiūtė, E. (2016). Short-term wind speed forecasting using ARIMA model. Energetika, 62(1-2): 45-55. https://doi.org/10.6001/energetika.v62i1-2.3313

[10] Palomares-Salas, J.C., De La Rosa, J.J.G., Ramiro, J.G.,

Melgar, J., Aguera, A., Moreno, A. (2009). ARIMA vs. Neural networks for wind speed forecasting. In 2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, pp. 129-133. https://doi.org/10.1109/CIMSA.2009.5069932

[11] Cadenas, E., Rivera, W. (2010). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. Renewable Energy, 35(12): 2732-2738. https://doi.org/10.1016/j.renene.2010.04.022

[12] Ye, R., Suganthan, P.N., Srikanth, N., Sarkar, S. (2013). A hybrid ARIMA-DENFIS method for wind speed forecasting. In 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-6. https://doi.org/10.1109/FUZZ-IEEE.2013.6622503

[13] Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

[14] Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50: 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

NOMENCLATURE

p-value	Probability value
H_0	Null hypothesis
H_1	Alternative hypothesis
L_t	First component
N_t	Second component
e_t	Residuals
$\hat{E}_t$	Estimated residuals
$\hat{Y}_t$	Forecast value from hybrid model

Greek symbols

$\phi_p(B)$	Nonseasonal autoregressive operator of order p
$\theta_q(B)$	Nonseasonal moving average operator of order q
B	Backward shift operator
$\nabla^d y_t$	Degree of differencing
a_t	White noise process with a mean equal to 0 and a variance equal to σ_a^2
DF_τ	ADF test value
$\hat{\pi}$	Estimation of π
$SE(\hat{\pi})$	Standard error of coefficient
F_{t+1}	Forecast value for the next period t
y_t	Observed value in period t
$\hat{y}_t$	Predicted value in period t
F_t	Forecast value in period t
β	Smoothing constant
π	Mean
α	Critical value

Subscripts

p	Autoregressive lag
d	Degree of differencing
q	Moving average lag