

Research Article

Wind Speed Prediction Based on Univariate ARIMA and OLS on the Colombian Caribbean Coast**Kevin Palomino¹, Fabiola Reyes¹, José Núñez², Guillermo Valencia^{1,*} and Roberto Herrera Acosta¹**¹Departamento de Ingeniería Industrial, Universidad del Atlántico, Carrera 30 Número 8-49, Puerto Colombia 080007, Colombia²Universidad de la Costa CUC, Energy Optimization Research Group GIOPEN, Cl. 58 #66, Barranquilla, Atlántico, Colombia

Received 10 January 2020; Accepted 10 June 2020

Abstract

Greater incorporation of wind energy into power systems has necessitated the development of accurate and reliable techniques for wind speed forecasting. However, although there are multiple studies, none are set up for the Colombia Caribbean coast. This is a disadvantage because the potential of wind resources in this region is greater than the hydroelectric potential of the whole country, but all this potential has yet to be developed. In this paper, based on time series, Autoregressive Integrated Moving Average (ARIMA), and Multiple Regression with Ordinary Least Squares (OLS) in the study, two models are proposed and their performance for wind speed prediction is compared. The data were collected in the meteorological station located in the experimental farm of the Atlantic University, in Barranquilla, Colombia, and variables analyzed included wind speed, wind direction, temperature, relative humidity, solar radiation, and pressure. The results of the two approaches indicated that among all the involved models, the ARIMA model has the best predicting performance. Also, it is essential to highlight that through this work, decision-makers would explore the local wind potential, allowing for the possibility of predicting future wind speed, and thus giving them the ability to plan the production and the interaction of other sources of energy.

Keywords: wind speed prediction, ARIMA, OLS, Sustainable energy.

1. Introduction

Renewable energy sources such as solar, biomass, geothermal, and wind are having an increasing part in the production of energy due to the environmental impact that currently exists in the world [1]. For example, wind energy is clean, affordable, efficient, abundant, and is experiencing rapid growth worldwide [2]. Furthermore, the results show that wind energy has contributed to the diversification of the energy matrix [3], increasing the share of renewable generation sources, and giving priority for the rational use of energy, and preservation of the environment [4].

In Colombia, the generation of energy through wind power has shown great potential in the Caribbean region, located in the North of Colombia and South America and it is composed by the states of Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, Sucre and the state composed by the archipelago of San Andrés, Providencia, and Santa Catalina island [5]. In this region, there are locations that register average values of multi-year wind speeds higher than 4.5 m/s at a height of 10 m, and a wind power density higher than 16 W/m² [6]. However, due to the few initiatives related to assessing the potential of wind power and predicting both wind speed and direction in this region, not enough feasibility studies on wind farms have been carried out, as a comprehensive database and accurate forecasts with a good level of confidence are required [7].

Consequently, many researchers have been proposing several methods to predict the wind potential of a specific wind field and several of them are based on advanced statistical approaches and techniques, such as Artificial Neural Networks (ANN) [8,9], Kriging Methods [9], pure autoregressive (AR) or moving average (MA) models (ARMA) [10], autoregressive integrated moving average (ARIMA) [11], Ordinary Least Squares (OLS) linear regression models [12], and hybrid models [13]. It is important to keep in mind that these methods have shown good performance in studies with many variables and complex associations between variables, but the major downside of them is their computational intensity, requiring high-performance computational resources and long training times. Also, to obtain proper and efficient wind power utilization, the wind speed prediction plays a significant part in forecasting [14], and within each type of methods a variety of specific methods and algorithms exist [15]; Besides, results indicate that statistical methods perform well in short term prediction [7].

Wind speed prediction using ARIMA and OLS has comparative advantages over other approaches, not only due to its relatively easy application and interpretation but also to minimize processing errors. Thus, wind speed forecasting studies have been conducted in several locations using these methods [16-17]. Nevertheless, although there are multiple studies using ARIMA and OLS, none are set up for the Colombia Caribbean coast. This is a disadvantage because the potential of wind resources in this region is greater than the hydroelectric potential of the whole country, but all this potential has yet to be developed.

*E-mail address: guillermovalencia@mail.uniatlantico.edu.co
ISSN: 1791-2377 © 2020 School of Science, IHU. All rights reserved.
doi:10.25103/jestr.133.22

Therefore, the purpose of this research was to make a comparison of Autoregressive Integrated Moving Average (ARIMA) and multiple regression with Ordinary Least Squares (OLS) models for wind speed prediction in the metropolitan area of Barranquilla, the largest city and port on the north Caribbean coast and the fourth largest city in the entire country. In this way, this work seeks to minimize the prediction error by including different regressors, as well as to contribute to studies related to wind resource assessment on the Caribbean coast. This manuscript is organized as follows: The experimental data is presented in Section 2; Section 3 explains each model; Section 4 makes a comparison between both models; Section 5 concludes this paper.

2. Experimental data

The data were collected in the meteorological station located in the experimental farm of the Atlantic University, in Barranquilla, Colombia, at 10 m above ground level from 2017 and 2018. Variables analyzed included wind speed (m/s), wind direction (degrees), temperature (°C), relative humidity (%), solar radiation (W/m²), and pressure (mbar). First of all, the wind speed time series is presented in Figure 1. It can be seen that there are no trends for periodic or

seasonal variations of wind speed in time series. In addition, the average wind speed was 4.21 m/s with a standard deviation of 1.97.

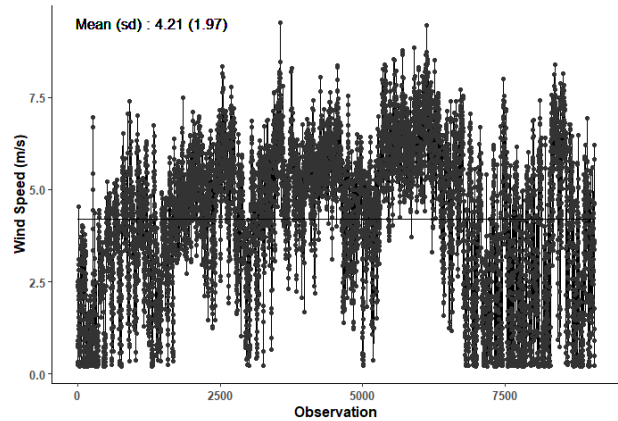


Fig. 1. Wind speed time series from Barranquilla, Colombia (ten-minute averages).

In the same way, Figure 2 shows that 10-minutes average (mean) air temperature is 26.75 °C; barometric pressure: 1008 mbar; solar radiation: 0.46 W/m²; and relative humidity: 83.10 %.

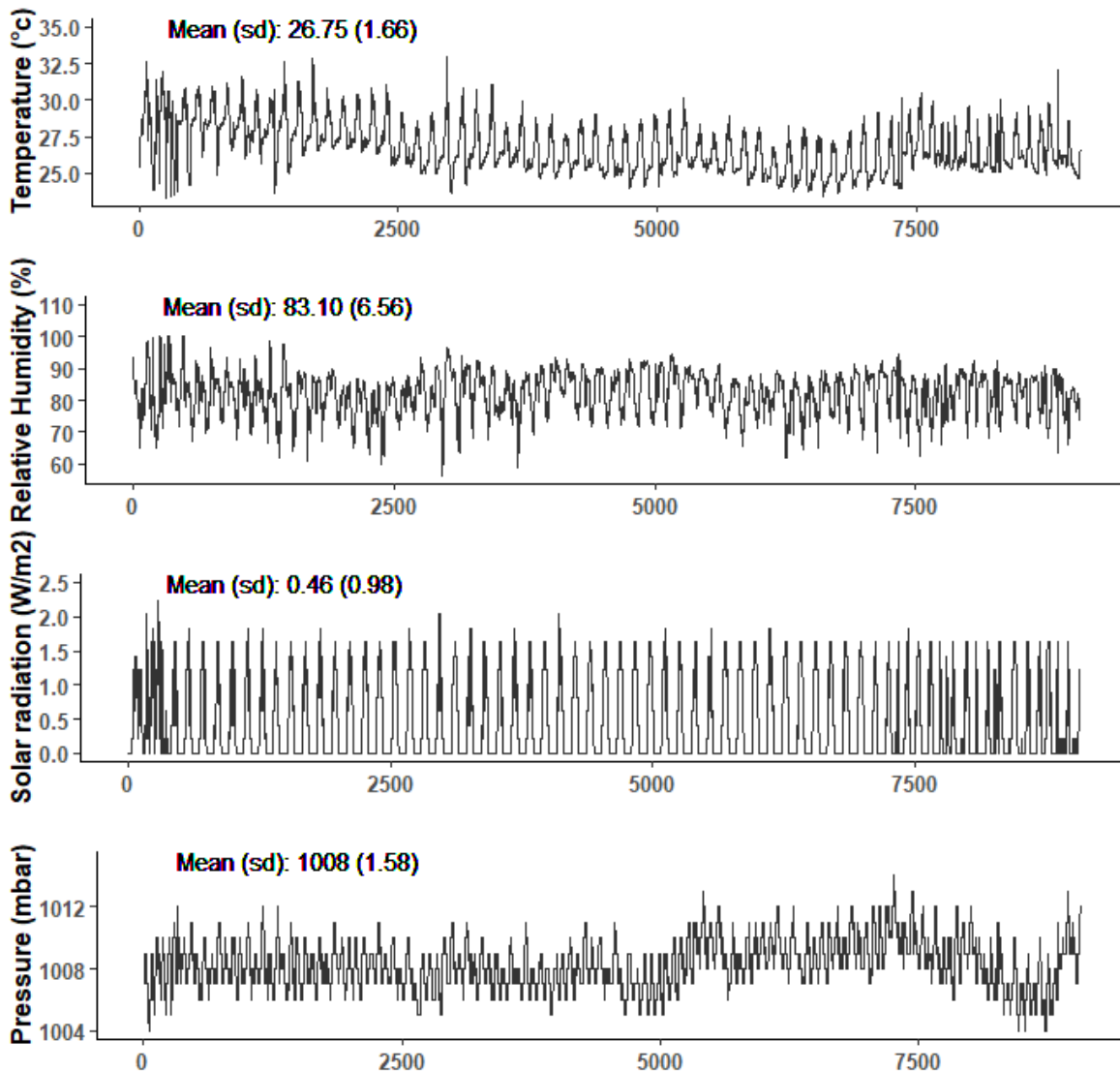


Fig. 2. Temperature, relative humidity, solar radiation, and pressure time series from Barranquilla, Colombia (ten-minute averages).

Finally, the annual wind-roses for 2017 and 2018 are shown in Figures 3, where 0°, 90°, 180° and 270° represent North, East, South and West directions, respectively. The prevailing wind direction is similar for both months, showing

higher intensities on the east side, and a lower intensity on the northeast side, which is associated with an arid coastal fringe, windy plains, and coastline on the Caribbean Sea.

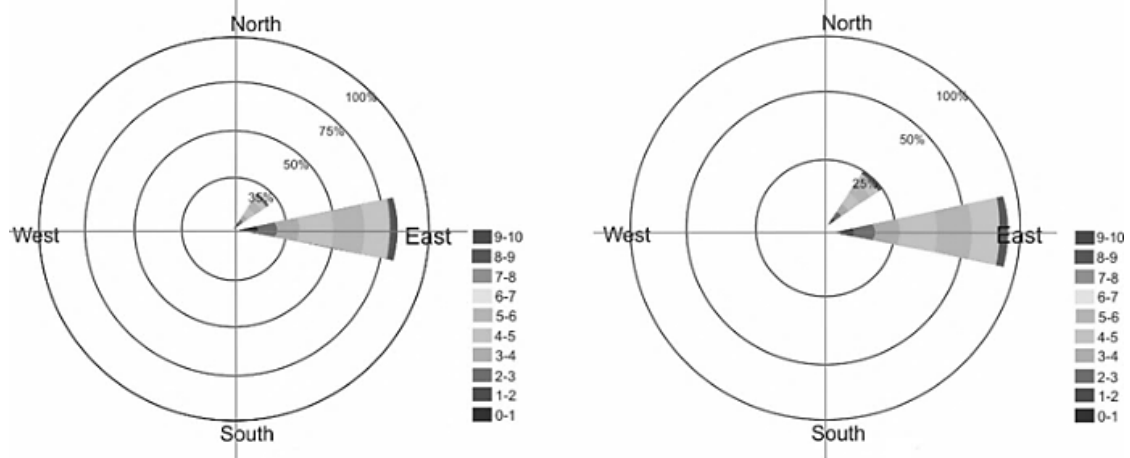


Fig. 2. Wind roses of the reference station, a) 2017, b) 2018.

3. Time series model

A time series (y_t) model uses the information in a time series to forecast future values of that series. In this way it is possible to build a simplified model of the time series that represents its randomness, making it useful for prediction [18]. Correspondingly, the first model studied was based on a univariate method and employed an Autoregressive Integrated Moving Average (ARIMA) with the wind speed as the variable of interest. The second approach is defined by a multivariate method based on an Ordinary Least Square (OLS) model using wind direction and velocity, air temperature, barometric pressure, solar radiation, and relative humidity.

3.1. Integrated self-regression moving average models

ARIMA is a widely known and frequently applied approach that considers time series data as a random, time-dependent variable. Thus, an ARIMA model is a mathematical model that allows us to identify and construct prediction models of a time series based on historical data at a moment t due to randomness [19]. Also, ARIMA (p, d, q) is defined as an integrated autoregressive moving average process that allows describing a series of observations, where $p, d,$ and q correspond to the meaning of the order of autoregression (AR), degree of difference (I), and order of moving average (MA), respectively [20].

Therefore, Equation (1) describes the ARIMA model (p, d, q), where ϕ_i is the variance stabilizing or auto-regression parameter, θ_j is the j -th moving average parameter, ϵ_t is known as the error term in time t , e_{t-i} is the error value of the dependent variable at the instant $t - i$, and y_{t-i} is the value of the dependent variable at the time $t - i$.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + \epsilon_t \quad (1)$$

The Box-Jenkins (1979) method was used to model the time series [18]. The ARIMA framework to forecasting

originally developed by Box-Jenkins involves an iterative three-stage process:

- i) Model-identification. The first stage consists of determining the type of stochastic process that generated the data. In other words, finding the most appropriate values for $p, d,$ and q in the ARIMA (p, d, q) model. However, it is necessary to verify whether the series is stationary and, in the case of not meeting this criterion (stationarity), a log transformation of the data can be performed to stabilize the variance.
- ii) Parameter estimation. Following the selection of a potential model (s), estimates of the parameters are calculated. In this stage, the values of autoregression and moving average included in the model are estimated, and the number of lags p and q . The approaches used are maximum likelihood and characteristic function-based methods.
- iii) Model diagnostic checking. With a model selected and parameters estimated, the adequacy of it must be checked to verify whether the model assumptions are met and how well the model fits the data. All models are wrong at some level of abstraction since they are simplified and idealized representations of reality. Also, it should be noted that there may be another ARIMA model that fits better, and for that reason, it is advisable to check the model to determine the expected performance of the forecast.

Proposed ARIMA Model: Following the Box-Jenkins methodology, stationarity, seasonality, and invertibility were identified. Regarding seasonality for the wind speed series, there were no peaks at seasonal frequencies. Hence, there were no trends for seasonal wind speed, as shown in Figure 4. Furthermore, the Dickey-Fuller test was employed to test that that wind speed data may be considered a stationary time series. The series had no unit roots and it was characterized as stationary. Therefore, for this model, the null hypothesis was rejected at a 99% confidence level (Dickey-Fuller test: -44.174, $P < 0.01\%$), as shown in Figure 5.

Also, assumptions of linearity and homoscedasticity were assessed with the standardized residuals plotted against predicted values, while the normality of the model was

assessed with the Durbin-Watson test. The Durbin-Watson test statistic, which is used to test the assumption of independent errors (residuals), yielded a value of 2.001 and showed that the assumption of independent errors was confirmed.

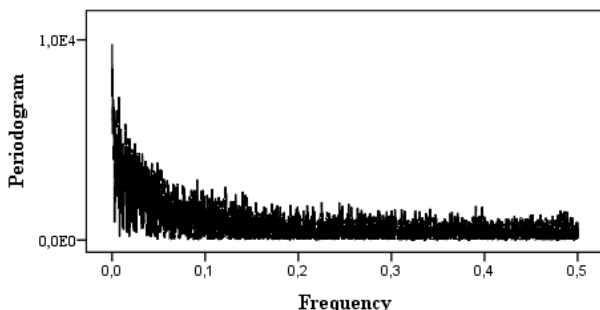


Fig. 4. Periodogram of wind speed.

According to the identification of the model parameters for wind speed, the partial auto-correlation function calculated in Figure 5 was used, where the parameters $p = 1, d = 1,$ and $q = 2$ were obtained. Therefore, it was concluded that the best fit for the wind speed series was ARIMA (1,1,2), which means an ARIMA model with 1 lag in the AR term and 2 lags in the MA term, as shown in Table 1.

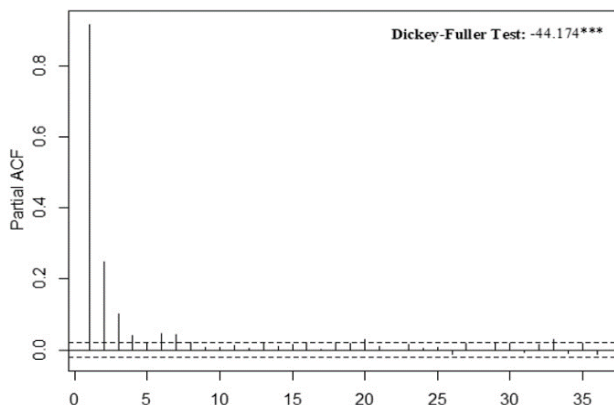


Fig. 5. Correlogram of the first difference for wind speed.

Table 1. ARIMA Model: Parameters and coefficients.

$$\text{Wind speed} = 0.922y_{t-1} + 1.278e_{t-1} - 0.284e_{t-2}$$

		Estimate	SE	t	Sig.	
Speed	AR	0.922	0.006	146.912	0.000	
	Lag 1	1				
I	MA	Lag 1	1.278	0.013	99.177	0.000
		Lag 1	-0.284	0.012	-23.406	0.000
		Lag 2				

3.2 Multiple regression model with Ordinary Least Squares (OLS)

Wind speed forecasting based on Ordinary Least Squares (OLS) considers not only observed values of wind speed over a period, but also the factors that influence speed variation, such as temperature, relative humidity, barometric pressure, and solar radiation. Thus, this approach is useful to include environmental variables or factors that appear to have a significant influence [21]. Multiple regression models present the wind speed as a continuous dependent variable, taking into account the trend of past information as a function of independent variables to predict the future. Therefore,

Equation (2) describes the OLS model, as follows:

$$Y_t = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon_t \quad (2)$$

where, Y_t = wind speed observation over time t , x_i = observations or values of the different independent variables influencing wind speed, β_0 and β_i = Constant regression coefficient and independent variable coefficients, and ε_t = prediction errors, that is, all factors that do not consider independent variables.

Ordinary Least Squares (OLS) is also referred to as Multiple Linear Regression (MLR) regression because it is also based on the minimization of the sum of the squares of the residuals. - As with any statistical test or method, the OLS model must meet certain main assumptions that justify the use of regression models for purposes of inference or prediction: linearity, statistical independence of the errors, homoscedasticity, and normality of the error distribution. Therefore, if the data are consistent with these hypotheses, the Gauss-Markov theorem states that the least-squares estimation model will produce consistent estimators, which will give the minimal variance for all parameters [22].

Proposed OLS model: OLS regression assumptions were examined by using the following methods. Firstly, the results indicated that the mean of the disturbances is equal to zero and that, also, the variance is constant, as shown in Figure 6. This is consistent with the assumptions of linearity and homoscedasticity. The assumptions of the residuals were visually checked by plotting predicted values against standardized residuals because known statistical tests for homoscedasticity are very sensitive to the violation of the normality assumption and cannot be used if the normality assumption fails. Next, the normality assumption for regression residuals was checked by plotting the regression residuals against the standard normal distribution. In brief, the residuals fitted the expected pattern well enough to indicate a relatively normal distribution. The Durbin-Watson test was performed to verify the error independence assumption. The Durbin-Watson statistic value was 0.183 which indicated that the assumption that the error terms are independent was not been completely fulfilled and, consequently, logarithmic transformations were employed.

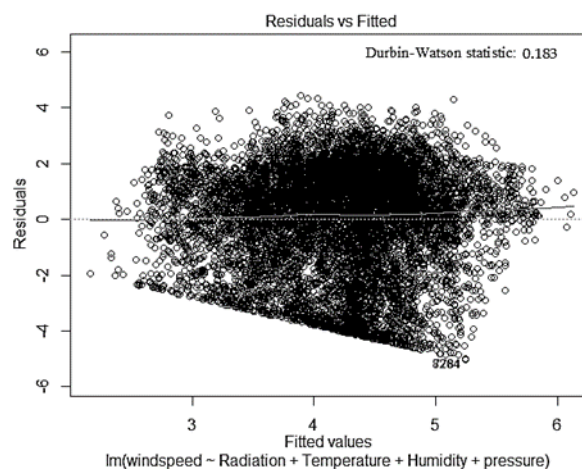


Fig. 6. Scatter plot of wind speed.

In this way, to predict wind speed determined by factors like air temperature, relative humidity, solar radiation, and pressure, the model with the non-standardized coefficients

presented in Table 2 was used. It is worth mentioning that these coefficients were statistically significant ($P < 0.01$) at a 95% confidence level. Thus, using the constant and coefficients, the OLS model can be described by equation (3) as follows,

$$WS = 103.42 - 0.56T - 0.08H + 0.22R - 0.07P \quad (3)$$

where WS is the wind speed, T is the temperature, H is the relative humidity, R is the solar radiation, and P is the pressure.

Table 2. OLS model: Parameters and coefficients.

	Coefficient s	95.0% confidence interval for B		Collinearit y statistics (FIV)
		Low er limit	Upper limit	
(Constant)	103.421	77.59 6	129.24 6	
Temperatur e	-0.564***	-0.599	-0.528	2.337
Humidity	-0.084***	-0.094	-0.075	2.606
S. Radiation	0.227***	0.163	0.29	2.602
Pressure	-0.077***	-0.102	-0.051	1.092

Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

1. The contrast of the models MCO and ARIMA.

The most appropriate forecasting model was determined on the basis of accuracy and performance. In this research, several common accuracy methods were used in order to evaluate the forecast and choose the best model that can represent the collected data: Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and BIAS. Table 3 summarizes the forecasting performance measures for the two models. Overall, the best is clearly the ARIMA (1,1,2) model, in which all the performance measures are lower. In this way, the results indicate that the ARIMA model provides reliable and satisfactory predictions for predicting wind speed and future energy values and expected to be an alternative tool for practical assessment and justification. Figure 7 shows a comparison between observed and forecast wind speed, with a confidence level of 95%.

Table 3. Comparison of ARIMA and OLS models for wind speed.

	OLS Model	ARIMA Model
MSE	55.862	1.187
MAD	7.994	0.769
MAPE	443.310%	38.144%
BIAS	7.236	0.258

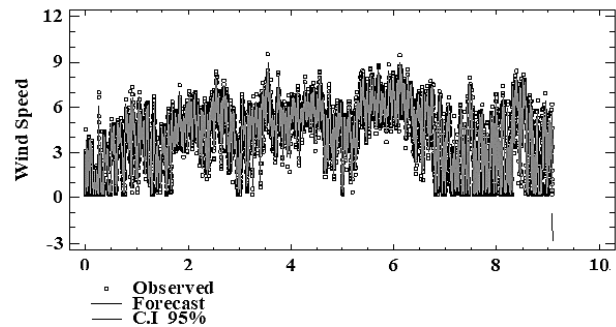


Fig. 7. Comparison between observed and forecast wind speed ARIMA (1,1,2).

5. Conclusions

A study to analyze and predict wind speed using important meteorological variables was developed. Firstly, using traditional statistical techniques, such as the ARIMA model, and, secondly, by using a multiple regression model with Ordinary Least Squares (OLS). Wind speed predictions given by both models were analyzed and compared. The results obtained show reasonable one step ahead wind speed prediction can be made with the univariate ARIMA model. On the other hand, by using a multivariate OLS model, less accurate results were obtained. Also, the ARIMA model evaluated in this paper was useful for enhancing prediction precision and did well in the study case, for it has several advantages. To begin with, the model can be a novel approach for wind speed prediction in the Colombian Caribbean coast. The concept of comparing the ARIMA approach with the OLS approach had not been attempted for wind speed prediction through the literature. Therefore, studying both the prediction approach to wind speed forecast was significant research. Besides, it is possible to use this research as a starting point to move towards much more elaborate models.

As far as future research is concerned, this research presents several aspects to consider. Firstly, it should be pointed out that with the growing applications of wind energy, power utilities in Colombia need to project the integration of wind power and other traditional powers [23]. Accurate prediction of wind speed plays an important role in socio-economic benefits derived from an appropriate power grid management in the country. As a second, this study can be a helpful tool in wind farm operations and programs, for the compared models can not only obtain a superior forecasting accuracy when compared with the traditional forecasting models but also be easily calculated. Last but not least, in future research work, it would be recommended to include additional meteorological variables, and focus on achieving better performance in different conditions and conducting incremental improvement according to different model parameters.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License



References

- [1] Taner, T. and Demirci, K.O., “Energy and economic analysis of the wind turbine plant’s draft for the Aksaray City”, Applied Ecology and Environmental Sciences, Vol. 2, No. 3, (2015), 82–85.

- [2] Boutoubat, M., Mokrani, L., Machmoum, M., "Control of a wind energy conversion system equipped by a DFIG for active power generation and power quality improvement", *Renewable Energy*, Vol. 50, (2013), 378–386.
- [3] Brown, S.P.A. and Huntington, H.G., "Energy security and climate change protection: Complementarity or tradeoff?" *Energy Policy*, Vol. 36, No. 9, (2008), 3510–3513.
- [4] Congress of the Republic of Colombia: Law 1715 - 2014.
- [5] Valencia, G., Vanegas, M. and Polo, J., "Análisis estadístico de la velocidad y dirección del viento en la Costa Caribe colombiana con énfasis en La Guajira", Vol. 1, 1st ed., Universidad del Atlántico, Barranquilla, (2016), 150.
- [6] Valencia, G. and Vanegas, M., "Atlas Eólico de la Región Caribe Colombiana", Vol. 1, 1st ed., Universidad del Atlántico, Barranquilla, (2016), 1-45.
- [7] Ouyang, T., Zha, X., Qin, L., Xiong, Y. and Xia, T., "Wind power prediction method based on regime of switching kernel functions", *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 153, (2016), 26–33.
- [8] Cadenas, E. and Rivera, W., "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model", *Renewable Energy*, Vol. 35, No. 12, (2010), 2732–2738.
- [9] Lawrie, L.K. al., "ENERGYPLUS, a new-generation building energy simulation program". Proceedings of Building Simulation 1999, Kyoto, Japan, (Aug. 1999), 1999.
- [10] Liu, H., Erdem, E. and Shi, J., "Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed", *Applied Energy*, Vol. 88, No. 3, (2011), 724–732.
- [11] Kavasseri, R.G. and Seetharaman, K., "Day-ahead wind speed forecasting using f-ARIMA models", *Renewable Energy*, Vol. 34, No. 5, (2009), 1388–1393.
- [12] Ariza, A.M., "Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución", Vol. 1, 1st ed., Universidad Tecnológica de Pereira, Pereira, (2013), 15-145.
- [13] Wang, X., Guo, P. and Huang, X., "A Review of Wind Power Forecasting Models", *Energy Procedia*, Vol. 12, (2011), 770–778.
- [14] Barrozo, F., Valencia, G. and Escorcía, Y.C., "Hybrid PV and wind grid-connected renewable energy system to reduce the gas emission and operation cost", *Contemporary Engineering Sciences*, Vol. 10, No. 26, (2017), 1269-1278.
- [15] Haque, A.U., Mandal, P., Meng, J. and Negnevitsky, M., "Wind speed forecast model for wind farm based on a hybrid machine learning algorithm", *International Journal of Sustainable Energy*, Vol. 34, No. 1, (2015), 38–51.
- [16] Cassola, F. and Burlando, M., "Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output", *Applied Energy*, Vol. 99, (2012), 154–166.
- [17] Torres, J. L., García, A., De Blas, M. and De Francisco, A., "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)", *Solar Energy*, Vol. 79, No. 1, (2005), 65–77.
- [18] Box, G.E.P and Jenkins, G. M "Time Series Analysis Time series Analysis: Forecasting and control", Vol. 1, 3rd ed., Prentice-Hall, New Jersey, (1994), 614
- [19] Box, G.E.P, Jenkins, G. M. and Reinsel, G., "Time series Analysis: Forecasting and control", Vol. 1, 2nd ed., Holden-Day, New Jersey, (1976), 586
- [20] Makridakis, S. G., Wheelwright, S. C. and Hyndman, R. J., "Forecasting: Methods and Applications", Vol. 1, 3rd ed., John Wiley & Sons, New York, (1998), 656
- [21] Aguado, J., Quevedo, A., Castro, M., Arteaga, R., Vázquez, M.A, and Zamora, B.P., "Meteorological variables prediction through ARIMA models", *Agrociencia*, Vol. 50, No. 1, (2016), 1-13.
- [22] Rojo, J. M., "Regresión lineal múltiple", Instituto de Economía y Geografía Madrid, Madrid, (2007), 32.
- [23] Herrera, R., Palomino, K., Reyes, F. and Valencia, G., "Análisis Estadístico Descriptivo e Inferencial de la Velocidad y Dirección del viento en la Costa Caribe Colombiana", *Revista Espacios*, Vol. 39, No. 19, (2018), 3-15.