IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 24, 2018, accepted September 19, 2018, date of publication October 1, 2018, date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872720

Hybrid Approach Combining SARIMA and Neural Networks for Multi-Step Ahead Wind Speed Forecasting in Brazil

DAVID B. ALENCAR[1], CAROLINA M. AFFONSO[1], (Member, IEEE), ROBERTO C. L. OLIVEIRA[1], AND JOSE C. R. FILHO[2]

[1]Faculty of Electrical Engineering, Federal University of Pará, Belém 66075-110, Brazil
[2]Faculty of Technology, Federal University of Amazonas, Manaus 66075-110, Brazil

Corresponding author: Carolina M. Affonso (carolina@ufpa.br)

This work was supported in part by FAPEAM, in part ITEGAM, and in part UFPA.

ABSTRACT This paper proposes a hybrid approach based on seasonal autoregressive integrated moving average (SARIMA) and neural networks for multi-step ahead wind speed forecasting using explanatory variables. In the proposed model, explanatory variables are first predicted, and wind speed forecasting is performed taking into account these forecasted values and wind speed historical series. The multi-step ahead forecasting is achieved recursively, by using the first forecasted value as input to obtain the next forecasting value. The proposed approach is tested using historical records of meteorological data collected from two real-world locations in Brazil. In order to demonstrate accuracy and effectiveness of the proposed approach, the results are compared with other techniques, such as neural networks, SARIMA, and SARIMA + wavelet. Simulation results reveal that the proposed hybrid forecasting method outperforms these popular algorithms for different forecasting horizons with higher accuracy.

INDEX TERMS SARIMA model, explanatory variable selection, multi-step ahead, neural networks, wind speed forecasting.

I. INTRODUCTION

The use of renewable energies has been growing considerably in the last years due to the increase in fossil fuel prices and the necessity to reduce greenhouse gas emissions. Among renewable energy alternatives, wind generation is one of the fastest growing sources in the world. In 2017, the global cumulative wind installed capacity reached 539 GW [1]. Brazil added over 2 GW of new wind energy capacity in 2017, and reached the seventh position in the ranking of the countries with largest wind generation in the world (12.7 GW of total installed capacity).

Since wind power generation is highly dependent on weather conditions and has stochastic and intermittent characteristics, accurate wind power forecasting plays an important role to its proper integration in the system. Wind power forecasting can have different time horizon according to its application. The literature presents different time-scale classification, but it can be mainly categorized into four time periods: ultra-short-term (few seconds to minutes), short-term (30 minutes to 24 hours), medium-term (24 hours to 1 week) and long-term (1 week to years) [2]–[4].

Wind power forecasting is mainly based on wind speed forecast, and several methods have been reported in the literature during the last years. They can be categorized into physical methods, traditional statistical methods, artificial intelligent methods, and hybrid approaches. Physical models are based on numerical weather prediction (NWP) data. They require physical description of wind farms and need considerable computational resources [5]. The traditional statistical methods typically use historical series to find the relationship of the measured data, such as autoregressive integrated moving average (ARIMA) models [6], [7]. The intelligent methods cover the use of neural networks (NN), evolutionary algorithms, fuzzy logic and others [8], [9].

The hybrid methods combine different models to obtain advantages of each one, presenting superior performance. For instance, [10] proposes a model combining neural networks with wavelet decomposition for wind speed prediction with a time horizon of 2 hours ahead. The obtained root mean square error (RMSE) varies from 1.054 to 1.945 considering a wind farm plant of East China. Reference [11] proposes a method based on wavelet decomposition and NN. The wind speed data is collected from a weather station in Canada, and the results present RMSE of 13.76 to a forecasting time horizon of 5 hours. Reference [12] proposes a hybrid model for multistep ahead wind speed forecasting based on variational mode decomposition, phase space reconstruction and wavelet neural network optimized by genetic algorithm. The proposed model is tested using wind speed series collected during spring and autumn from a wind farm located in Xinjiang, China. The RMSE obtained varies from 0.340 to 0.368 for a prediction horizon of 6 hours. In [13], a hybrid approach is proposed for multi-step-ahead wind speed forecasting using optimal feature selection and a NN optimized by a modified bat algorithm with cognition strategy. Based on a wind speed series from China, the average RMSE obtained is 2.0424 for a prediction horizon of 6 hours with 60-min sampling interval. Reference [14] proposes a method for wind power forecasting based on particle swarm optimization algorithm and type-2 fuzzy neural network using data from a wind farm located in Canada. The proposed method results in RMSE of 12.04 for a forecasting time horizon of 24 hours. In [15], several machine learning algorithms are applied to forecast wind speed using data from a wind farm located in the South of Brazil. The results show RMSE varying from 1.12 to 3.3 considering a forecasting time horizon of 3 hours.

As can be seen in the aforementioned literature, although wind speed forecasting models have improved considerably in the last years, relatively high errors are still obtained, requiring more accurate and efficient wind speed forecasting methods. According to ERCOT (Electric Reliability Council of Texas) report data, their root mean squared wind forecast error is 14% of installed wind capacity [16].

This paper proposes a hybrid approach combining seasonal autoregressive integrated moving average (SARIMA) and Neural Networks for wind speed forecasting in Brazil using explanatory variables, with prediction horizons of one-step (6 hours) and multi-step ahead (up to 24 hours). In the proposed model, the wind speed prediction is performed taking in account wind speed historical series and the forecasted values of the explanatory variables. The multi-step ahead forecasting is performed recursively, by using the first forecasted value as input to obtain the next forecasted value. The accuracy of the proposed method is analyzed in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results are compared with other commonly used methods such as SARIMA, Neural Networks and SARIMA+wavelet.

The proposed approach is a modified version of the prediction model developed in a previous study for electricity price forecasting [17], and a continuation of the work developed in [18], focusing on the short-term wind speed forecasting and providing improvements such as the use of SARIMA model instead of ARIMA, a more detailed analysis regarding the use of explanatory variables, and the use of a database from another location in Brazil. The main contributions of this paper are: (i) present a hybrid forecasting approach based on SARIMA and NN for accurate wind speed forecasting using explanatory variables; (ii) apply a multi-step ahead procedure in order to make predictions up to 24 hours ahead; (iii) use Brazilian database for wind speed forecasting, which was explored only by few studies to the best of authors knowledge; (iv) develop an effective forecasting approach for multiple seasons of the year with higher accuracy.

This paper is organized as follows. Section II presents the database used in this paper and its statistical summary. In Section III, details of the proposed wind speed forecasting approach is presented. Section IV provides different criterions used to evaluate forecasting accuracy, and Section V presents the procedure to select the explanatory variables. Section VI shows the results obtained for one-step and multi-step ahead forecasting. Section VII outlines the conclusions.

II. STUDY AREA AND DATA COLLECTION

Two different sites in the northeast region of Brazil were considered in this paper as shown in Fig. 1: Macau (latitude 5° 9' 3.726'' on south, longitude 36° 34' 23.3112'' on west) and Petrolina (latitude 09° 04' 08'' on south, longitude 40° 19' 11'' on west). The northeast region of Brazil is a potential area for wind generation with constant wind speed without extreme gusts. It has a tropical climate with high ambient temperatures throughout the year and only two distinct seasons: wet and dry. According to the historical records, the temperature in Petrolina varies from 22.3°C to 29.86°C, and the temperature in Macau varies from 25.6°C to 32.3°C throughout the year of 2016 and 2017, as shown in Fig. 2.

FIGURE 1. Location of the weather stations analyzed in Brazil.

Meteorological records from Petrolina station were obtained from SONDA database (National Organization

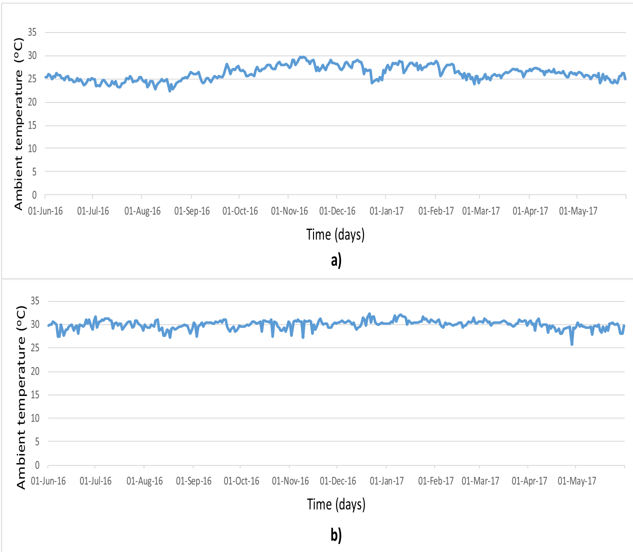

FIGURE 2. Historical ambient temperature series from Petrolina and Macau, Brazil.

System of Environment Data), which accounts for over 7 anemometric stations and 18 solarimetric stations across the Brazilian territory [19]. The historical data from Macau station were collected from INMET, the Brazilian National Institute of Meteorology [20]. Both database are publicly available for download at their respectively website.

The following variables were collected to each location: air temperature, air humidity, atmospheric pressure, wind direction and wind speed. The time series data are from 01 June, 2016, to May 31, 2017. Fig. 3 shows wind speed historical series in Petrolina and Macau. Table 1 and 2 show the summary statistics of the potential explanatory variables in Petrolina and Macau, respectively, with minimum value (Min), mean, maximum (Max), standard deviation (SD) and variance (VAR).

TABLE 1. Variables summary statistics - Petrolina.

Variable	Min	Mean	Max	SD	Var
Temperature (ºC)	22.3	26.19	29.86	1.54	2.4
Humidity (%)	37.17	58.13	83.54	8.86	78.64
Pressure (mb)	966.05	970.16	975.33	2.27	5.16
Wind direction (deg)	104.84	138.01	210.92	14.51	210.86
Wind speed (m/s)	5.30	8.45	12.73	1.73	2.99

TABLE 2. Variables summary statistics - Macau.

Variable	Min	Mean	Max	SD	Var
Temperature (ºC)	25.67	29.98	32.33	0.92	0.85
Humidity (%)	42.0	62.29	93.67	8.10	65.63
Pressure (mb)	1,006.2	1,011.3	1,016.2	1.85	3.45
Wind direction (deg)	5.00	9.83	19.67	3.17	10.04
Wind speed (m/s)	2.10	5.332	9.510	1.46	2.14

Since the original series was sampled by 10-minutes interval, the average hourly value was evaluated to perform the short-term prediction with a time sampling interval of 1 hour. The database used comprehends 365 days sampled by each 1 hour, totalizing 8,760 points to each variable and 43,800 points in total.

III. PROPOSED HYBRID APPROACH

The main structure of proposed hybrid approach is presented in Fig. 4, and is composed by SARIMA and NN models arranged in series, combined with PCA and balancing procedures. The major steps for the proposed wind speed approach are as follows. First, a large database is created with historical records of meteorological data. Then, the explanatory variable selection is performed. SARIMA model is first used as an auxiliary linear predictor to forecast future values of the explanatory variables. The dimension of the output data is reduced through principal component analysis (PCA) and balancing procedures, referred in Fig. 4 as PCA+BAL. Then, a first NN (NN1) is used for each predicted variable in order to catch nonlinear relations between data. In order to guarantee generalization capacity, it is necessary to reduce the dimension of input data again. Finally, a second NN (NN2) is used to predict wind speed series based on past values of wind speed and data from the previous step.

The multi-step ahead forecasting is obtained recursively by feeding input variables with the forecasted outputs. Basically, the model combines 4 identical one-step ahead linear predictor structures to produce a series of predictions. After the explanatory variables are obtained for 1-step ahead (6-hours ahead), it is used as input for 2-step ahead prediction (12-hours ahead), and this procedure is repeated until explanatory variables of the next 24 hours is forecasted. The same recursive procedure is adopted for wind speed prediction, using previous wind speed prediction as additional input information for further predictions until 24-hours is forecasted. The details of the proposed approach are explained next.

A. SARIMA MODEL OVERVIEW

SARIMA model is a classical statistical method used to forecast future values as a linear function of past observations. It is an extension of the ARIMA model proposed by Box and Jenkins, used to handle time series data with seasonal behavior [21]. As wind speed series has periodic variations, SARIMA model is adopted in this paper.

SARIMA model can be represented as SARIMA(p,d,q) $(P,D,Q)_S$. The lowercase notations represent the non-seasonal parts of the model, as p represents auto regression parameter, d represents the differencing parameter, and q represents moving average parameter. The uppercase notations represent the seasonal parts of the model, and s refers to the number of periods per season. For example, an annual cycle is expressed as $s = 12$.

The SARIMA model is described mathematically as follows:

$$\phi_p(B)\,\Phi_P(B^S)\nabla^d\nabla_S^D y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \qquad (1)$$

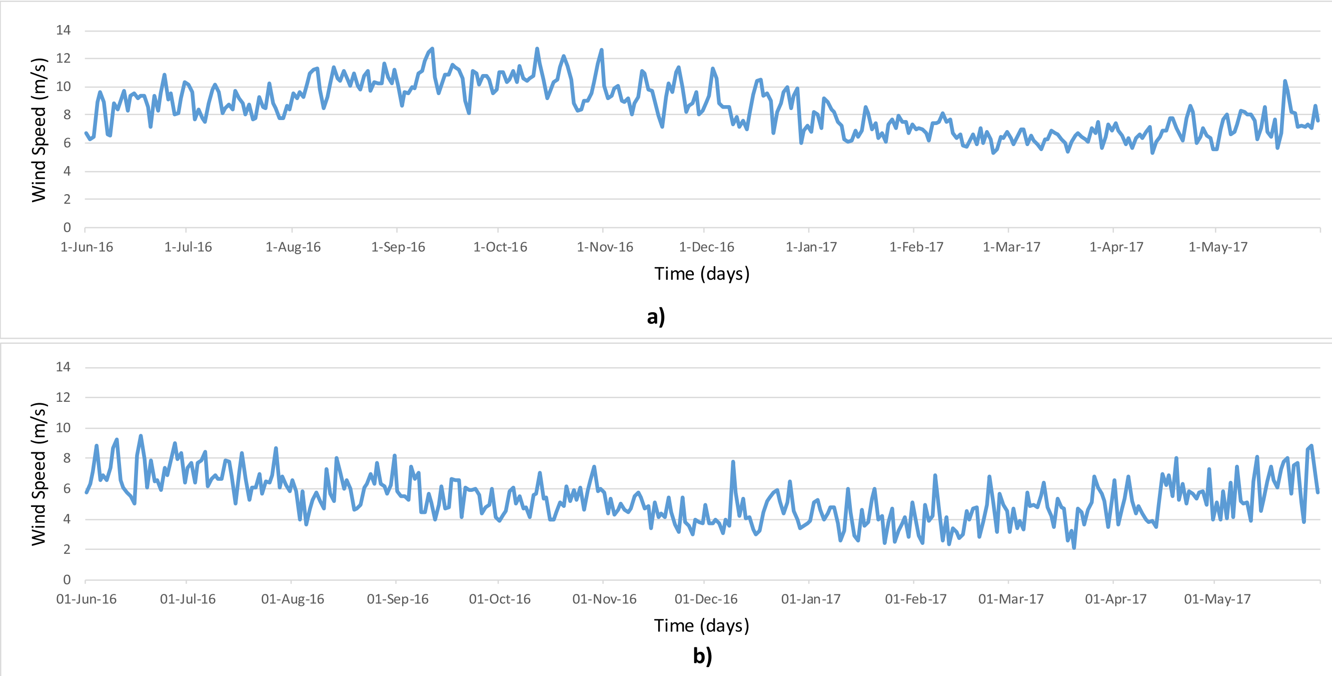

FIGURE 3. Historical wind speed series from Petrolina and Macau, Brazil.

where ϕ is the regular AR polynomial of order p, Φ is the seasonal AR polynomial of order P, θ is the regular MA polynomial of order q, Θ is the seasonal MA polynomial of order Q, ∇^d is the differentiating operator, ∇_S^D is the seasonal differentiating operator, y_t is the wind speed at time t, ε_t is the residual error at time t, and B is the backshift operator as $B^k(y_t) = y_{t-1}$.

The polynomials and all operators are defined as follows:

$$\Phi_P\left(B^S\right) = 1 - \Phi_1\left(B^S\right) - \Phi_2\left(B^{2S}\right) - \ldots - \Phi_P\left(B^{PS}\right) \tag{2}$$

$$\Theta_Q\left(B^S\right) = 1 - \Theta_1\left(B^S\right) - \Theta_2\left(B^{2S}\right) - \ldots - \Theta_Q\left(B^{QS}\right) \tag{3}$$

$$\phi_p\left(B^S\right) = 1 - \phi_p(B) - \phi_2\left(B^2\right), \ldots, \phi_p\left(B^p\right) \tag{4}$$

$$\theta_q\left(B^S\right) = 1 - \theta_q(B) - \theta_2\left(B^2\right), \ldots, \theta_q\left(B^q\right) \tag{5}$$

$$\nabla^d = (1 - B)^d \tag{6}$$

$$\nabla_S^D = \left(1 - B^S\right)^D \tag{7}$$

The methodology to obtain SARIMA model can be summarized in four steps: identification, estimation, diagnostic checking and forecasting [21].

- Identification of seasonal and non-seasonal orders of SARIMA model using the autocorrelation (ACF) and partial autocorrelation (PACF) functions;
- Estimate the coefficients of the model;
- Diagnostic check to verify the adequacy of the model examining the residuals;
- Forecasting the future outcomes based on the historical data.

B. NEURAL NETWORK MODEL OVERVIEW

Neural Network (NN) model is inspired by the way human brain processes information [22]. It has been largely used in different fields due to its high capability of generalization, such as image identification, recognizing diseases, energy price and demand forecasting, and market area.

The network architecture used in this paper is the multilayer perceptron, which is composed by an input layer, one or more hidden layers and one output layer as shown in Fig. 5. Each layer is composed by units called neurons, and each neuron is connected to all units from the previous layer. The inputs are multiplied by weights and summed to the bias, and the results pass through a linear function as shown in (8) and (9) [22].

$$v_k = \sum\nolimits_{j=1}^{m} w_{kj} a_j + b_k \tag{8}$$

$$z_k = \varphi(v_k) \tag{9}$$

where w_{kj} is the weight that goes from the input k to the hidden neuron j, b_k is the bias, a_j is the input of the neuron, φ is the activation function and z_k is the output of the neuron.

During the training stage, the learning process is iterative. The inputs and targets are presented to the network, and the weights and biases are adjusted continuously until the error between the produced output and target value is minimized within a specified tolerance criteria. After training, the model is validated by checking whether it produces accurate output or not.

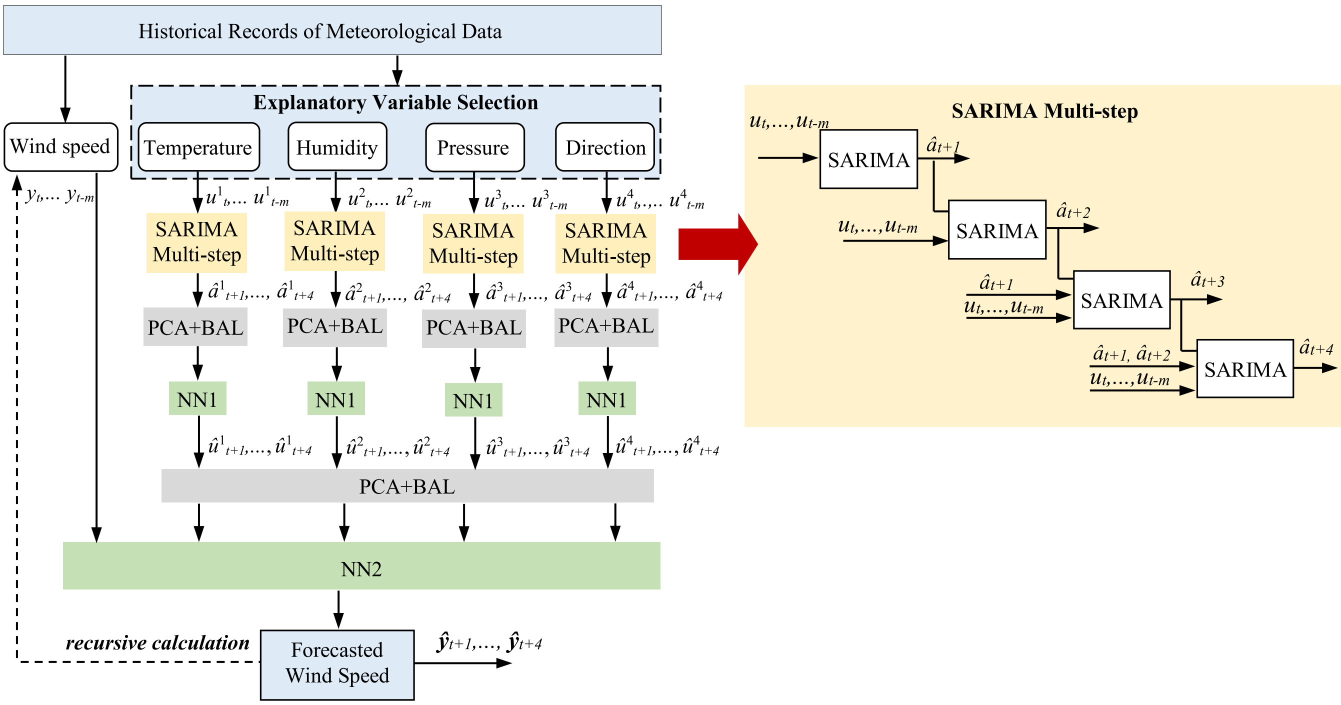

FIGURE 4. Flowchart of the proposed hybrid approach.

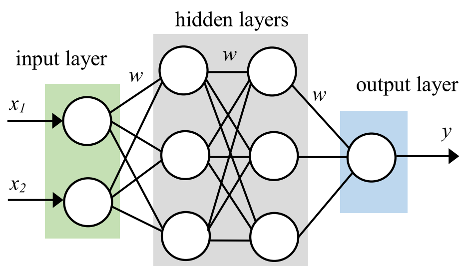

FIGURE 5. Basic structure of NN model.

The NN architecture adopted in this study is composed of one input layer with 11 neurons, 2 hidden layers with 8 neurons each, and an output layer with one neuron, with sigmoid tangent $(1/1 + e^{-x})$ activation function in all layers. The number of hidden layers and neurons in each layer are determined by research and experiment. The Levenberg–Marquardt learning method was applied, and the database was divided into two sets, 80% for training and the remaining 20% for testing procedure.

C. PRINCIPAL COMPONENT ANALYSIS AND BALANCING

It is well known that NN may not be effective when using high dimensional data. It can cause difficulties in the learning process, increase the training time, and reduce or even deteriorate forecasting accuracy. Also, time series data is highly imbalanced. This occurs when certain ranges of values are over-represented in comparison to others that are severely underrepresented. When subject to unbalanced data, standard learning algorithms bias the models toward the more frequent situation, and it may cause performance degradation.

In this sense, two pre-processing techniques are applied in the forecasting approach: Principal Component Analysis (PCA) and balancing. PCA reduces the dimension of a large data set with correlated variables into a smaller set with uncorrelated variables, retaining most of the sample's information [23]. The unbalanced problem is avoided applying resampling strategies, a common solution used with unbalanced data. It changes data distribution and balance the number of rare and normal cases [24].

IV. ACCURACY EVALUATION METRICS

Different popular metrics were adopted to evaluate the accuracy of the proposed forecasting approach: mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). These performance indexes are computed as follows:

$$MAE = \frac{1}{N}\sum\nolimits_{t=1}^{N}\left|\hat{y}(t) - y(t)\right| \tag{10}$$

$$MAPE = 100 \times \frac{1}{N}\sum\nolimits_{t=1}^{N}\left|\frac{\hat{y}(t) - y(t)}{y(t)}\right| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{N}\sum\nolimits_{t=1}^{N}\left(\hat{y}(t) - y(t)\right)^2} \tag{12}$$

where N is the number of samples, $\hat{y}(t)$ and $y(t)$ are wind speed forecasted and observed values at time t respectively.

V. EXPLANATORY VARIABLE SELECTION

The forecasting models can be developed using only past values of wind speed series [12], [25], or taking also in account historical data of other variables such as wind direction, temperature, and humidity [14], [26]. These variables are called explanatory, and the selection of the most relevant set is crucial to develop an accurate forecasting model. Redundant variables can potentially degrade the learning process and require extra computational memory.

Based on a priori knowledge, five explanatory variables are considered as potential inputs to wind speed forecasting: wind speed (past values), wind direction, air temperature, humidity, and atmospheric pressure. At a second stage, a correlation analysis is performed. The correlation matrix to Petrolina database is presented in Table 3, and the results show a stronger relationship between humidity and wind direction with wind speed (0.427 and 0.389), and a weaker relationship between temperature and atmospheric pressure with wind speed (0.097 and 0.125).

TABLE 3. Correlation matrix of explanatory variables to Petrolina.

	Temp.	Humidity	Pressure	Wind direction	Wind speed
Temperature	1				
Humidity	-0.663	1			
Pressure	-0.182	0.042	1		
Direction	-0.313	0.536	-0.005	1	
Wind speed	-0.097	-0.427	0.125	-0.389	1

To avoid redundant information, some authors discard explanatory variables that have weaker relationship with the predicted variable, and consider only those who seem more influential. Ma *et al.* [25] used only historical series of wind speed and wind direction to predict wind power generation. In order to prove the importance of using atmospheric pressure and temperature on wind speed forecasting, two models were developed:

- Model A: this model is developed using 5 explanatory variables (wind speed, wind direction, air temperature, humidity, and atmospheric pressure);
- Model B: this model is developed using only 3 explanatory variables (wind speed, wind direction and humidity). In this case, air temperature and atmospheric pressure were not considered.

The prediction errors obtained with the proposed hybrid approach using models A and B are shown in Table 4 considering a forecast horizon of 1-step ahead (6 hours). The results show that the proposed forecast model performs considerably better when using all explanatory variables, having lowest MAE, RMSE and MAPE values. The same behavior was verified using Macau database. For this reason, the forecasting approach proposed in this paper will consider for both locations Petrolina and Macau all set of potential explanatory variables: temperature, humidity, pressure, direction, and wind speed.

TABLE 4. One-step ahead forecasting errors to Petrolina (Models A, B).

Error	Model A (5 explanatory var.)	Model B (3 explanatory var.)
MAE	0.141	0.216
RMSE	0.316	0.483
MAPE (%)	1.862	2.891

VI. SIMULATION RESULTS

This section presents the results obtained with the proposed hybrid approach based on SARIMA-NN models for wind speed forecasting. Two different sites in the northeast region of Brazil were considered, and different prediction time horizons were analyzed: one step-ahead (6 hours) and multi-step ahead forecasting (up to 24 hours). The forecasting model is developed using SPSS Modeler software, which includes different data mining tools [27]. The results are presented as follows.

A. ONE-STEP AHEAD FORECASTING

This section presents the forecasting results obtained with the proposed hybrid approach for the time horizon of one-step ahead. First, parameters of SARIMA model were automatically obtained by SPSS modeler software, identifying the best-fitting model to each historic series using autocorrelation function, partial autocorrelation function and statistic measures such as MAPE, MAE and normalized Bayesian information criterion (BIC). Table 5 shows the best SARIMA model obtained to each explanatory variable. Then, the proposed hybrid approach presented on Section III was applied and wind speed forecasting was obtained.

TABLE 5. SARIMA model to the explanatory variables.

Variable	SARIMA Model $(p, d, q)\times(P, D, Q)_S$ Petrolina	SARIMA Model $(p, d, q)\times(P, D, Q)_S$ Macau
Temperature	$(2,1,10)(1,0,1)_{12}$	$(0,1,3)(0,0,1)_{12}$
Humidity	$(1,1,6)(1,1,1)_{12}$	$(1,1,3)(1,0,0)_{12}$
Pressure	$(2,0,1)(2,0,1)_{12}$	$(3,1,5)(1,0,0)_{12}$
Direction	$(1,0,11)(1,0,1)_{12}$	$(2,0,2)(0,0,0)_{12}$

To evaluate the performance of the proposed hybrid model, other commonly used techniques such as SARIMA, SARIMA+Wavelet, and NN were employed for comparison purposes. The forecasting errors obtained are presented on Table 6. The results show that the proposed model reduces the forecasting errors considerable in both sites Petrolina and Macau, indicating a great improving on forecasting accuracy. Fig. 6 illustrates wind speed observed and forecasted values. The proposed approach outperforms the other methods, and the predicted values correctly follow the trend of wind speed variation.

B. MULTI-STEP AHEAD FORECASTING

This section presents multi step-ahead forecasting results obtained with the proposed hybrid method. The following

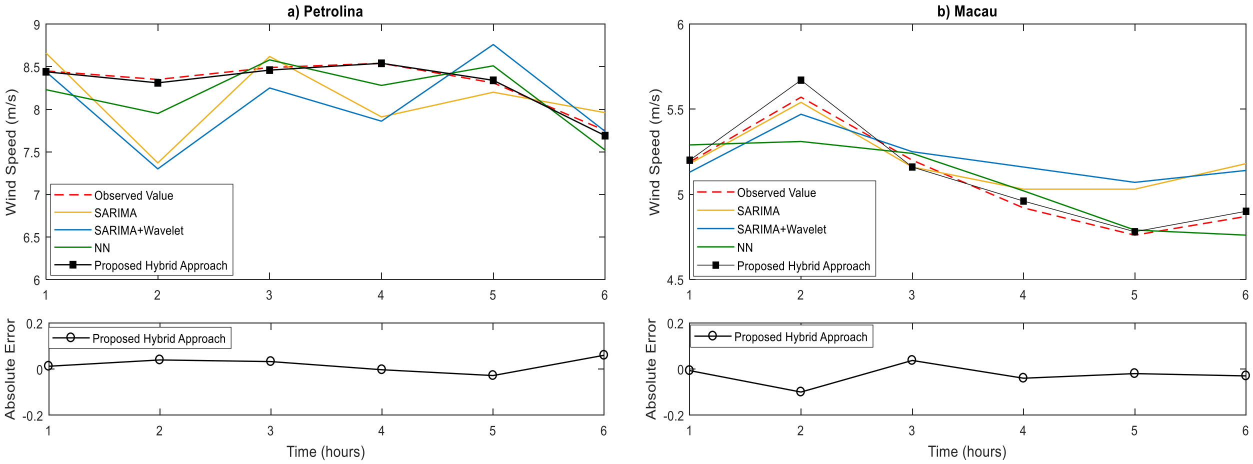

FIGURE 6. Observed wind speed series and one-step ahead forecasting: a) Petrolina, b) Macau.

TABLE 6. One-step ahead forecasting errors to Petrolina and Macau.

Error	SARIMA	SARIMA+Wavelet	NN	Hybrid
		Petrolina		
MAE	0.733	0.674	0.387	**0.141**
RMSE	1.638	1.474	0.829	**0.316**
MAPE (%)	9.686	8.814	4.844	**1.862**
		Macau		
MAE	1.122	1.032	0.576	**0.202**
RMSE	5.836	5.428	2.811	**1.740**
MAPE (%)	14.391	13.096	7.791	**2.815**

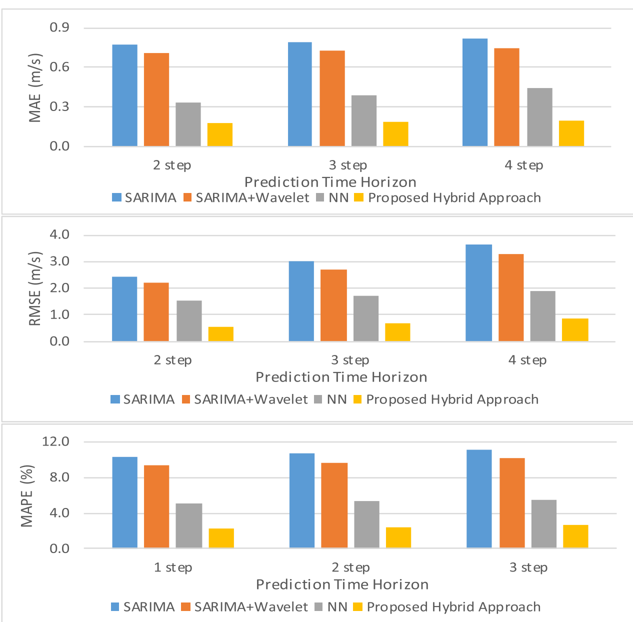

FIGURE 7. Performance measures for multi-step ahead forecasting in Petrolina: MAE, RMSE, and MAPE.

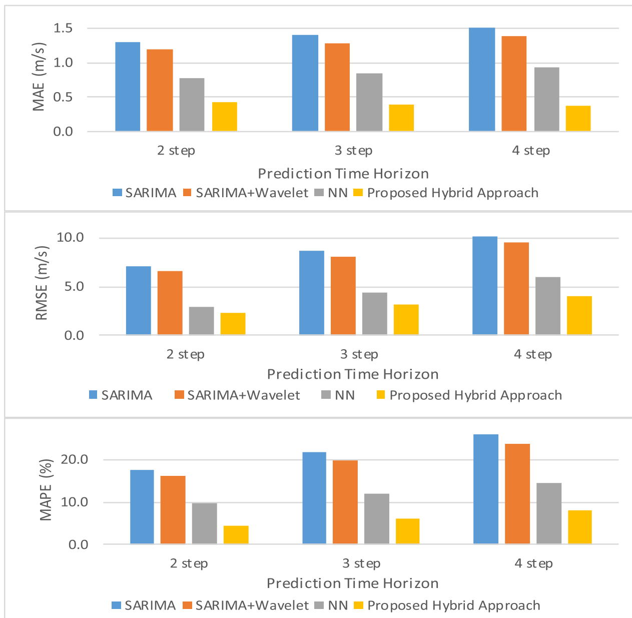

FIGURE 8. Performance measures for multi-step ahead forecasting in Macau: MAE, RMSE, and MAPE.

horizons were considered: two-step (12 hours), three-step (18 hours) and four-step (24 hours) ahead.

Fig. 7 and 8 shows forecasting performance metrics obtained with the proposed approach and other commonly used techniques over all time horizons considered. The results show that the proposed model outperforms all other methods for all horizons analyzed, presenting lower MAE, RMSE and MAPE errors. The proposed method demonstrates its effectiveness to both Petrolina and Macau sites, being a robust approach when applied for one-step and multi-step ahead forecasting.

Comparing with other methods presented in literature using other database, the proposed hybrid model provides reliable wind speed forecasting. In [15], the accuracy obtained in terms of RMSE varies from 1.12 to 3.3 for a prediction horizon of 3-hours ahead. In [12], the RMSE obtained for a prediction horizon of 6-hours ahead varies from 0.34 to 0.36, and in [13] the RMSE is 2.04. For a prediction horizon of 24-hours ahead, [14] presented RMSE of 12.04. The RMSE obtained with the proposed hybrid method varies from 1.740 to 0.316 for a prediction horizon of 6-hours ahead, and varies 0.861 to 4.049 for a prediction horizon of 24-hours ahead. Then, the maximum RMSE obtained in our study are comparatively lower than those obtained in [12]–[15].

VII. CONCLUSIONS

A hybrid approach is proposed for multi-step ahead wind speed forecasting based on SARIMA and NN models. This hybrid method combines the strengths of SARIMA and NN algorithms, which are capable of learning linear and nonlinear system behavior, respectively. Explanatory variables are first forecasted, and then used to predict wind speed series. The multi-step ahead prediction is performed recursively, using forecasted variables 1-step ahead (6 hours) as input for 2-step ahead (12 hours) prediction, and this procedure is repeated until a time horizon of 24 hours ahead. The methodology is applied using real meteorological data from two locations in Brazil. The validation of the proposed model is accomplished comparing the results with other commonly used techniques such as SARIMA, NN, SARIMA+Wavelet. The results show the proposed wind speed forecasting approach outperforms these methods with respect to all criteria adopted (MAE, MAPE, and RMSE) for one-step ahead and multi-step ahead horizons, indicating the robustness of the method.

REFERENCES

[1] Global Wind Energy Council (GWEC), *Global Statistics*. [Online]. Available: http://gwec.net/global-figures/graphs/

[2] A. Sfetsos, "A comprehensive overview of short term wind forecasting models based on time series analysis," in *Soft Computing in Green and Renewable Energy Systems* (Studies in Fuzziness and Soft Computing), vol. 269. Berlin, Germany: Springer, 2011, ch. 3, pp. 97–116.

[3] X. Zhao, S. Wang, and T. Li, "Review of evaluation criteria and main methods of wind power forecasting," *Energy Procedia*, vol. 12, pp. 761–769, Dec. 2011.

[4] M. Mao, J. Ling, L. Chang, N. D. Hatziargyriou, J. Zhang, and Y. Ding, "A novel short-term wind speed prediction based on MFEC," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 4, no. 4, pp. 1206–1216, Dec. 2016.

[5] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012.

[6] O. Karakuş, E. E. Kuruoğlu, and M. A. Altinkaya, "One-day ahead wind speed/power prediction based on polynomial autoregressive model," *IET Renew. Power Gener.*, vol. 11, no. 11, pp. 1430–1439, Sep. 2017.

[7] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar, "Linear and non-linear autoregressive models for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 112, pp. 115–124, Mar. 2016.

[8] Z. Shi, H. Liang, and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1177–1187, Jul. 2018.

[9] G. W. Chang, H. J. Lu, Y. R. Chang, and Y. D. Lee, "An improved neural network-based approach for short-term wind speed and power forecast," *Renew. Energy*, vol. 105, pp. 301–311, May 2017.

[10] H. Shao, X. Deng, and F. Cui, "Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 11, pp. 2585–2592, Aug. 2016.

[11] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data," *Renew. Energy*, vol. 92, pp. 202–211, Jul. 2016.

[12] D. Wang, H. Luo, O. Grunder, and Y. Lin, "Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction," *Renew. Energy*, vol. 113, pp. 1345–1358, Dec. 2017.

[13] T. Niu, J. Wang, K. Zhang, and P. Du, "Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy," *Renew. Energy*, vol. 118, pp. 213–229, Apr. 2018.

[14] A. Sharifian, M. J. Ghadi, S. Ghavidel, L. Li, and J. F. Zhang, "A new method based on type-2 fuzzy neural network for accurate wind power forecasting under uncertain data," *Renew. Energy*, vol. 120, pp. 220–230, May 2018.

[15] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, Aug. 2018.

[16] B. Mauch, J. Apt, P. M. S. Carvalho, and P. Jaramillo, "What day-ahead reserves are needed in electric grids with high levels of wind power?" *Environ. Res. Lett.*, vol. 8, no. 3, pp. 1–11, Aug. 2013.

[17] J. C. R. Filho, C. de M. Affonso, and R. C. L. de Oliveira, "Energy price prediction multi-step ahead using hybrid model in the Brazilian market," *Electr. Power Syst. Res.*, vol. 117, pp. 15–122, Dec. 2014.

[18] D. B. de Alencar, C. de Mattos Affonso, R. C. L. de Oliveira, J. L. M. Rodríguez, J. C. Leite, and J. C. R. Filho, "Different models for forecasting wind power generation: Case study," *Energies*, vol. 10, no. 12, pp. 1–27, 2017.

[19] *SONDA Project Website*. Accessed: Aug. 24, 2018. [Online]. Available: http://sonda.ccst.inpe.br/

[20] *INMET*. Accessed: Aug. 24, 2018. [Online]. Available: http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep

[21] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2016.

[22] S. Haykin, *Neural Networks and Machine Learning*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

[23] I. T. Jolliffe, *Principal Component Analysis*, 2th ed. New York, NY, USA: Springer, 2002.

[24] N. Moniz, P. Branco, and L. Torgo, "Resampling strategies for imbalanced time series," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal.*, Oct. 2016, pp. 282–291.

[25] J. Ma, M. Yang, X. Han, and Z. Li, "Ultra-short-term wind generation forecast based on multivariate empirical dynamic modeling," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1029–1038, Mar. 2017.

[26] N. Amjady, F. Keynia, and H. Zareipour, "Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization," *IEEE Trans. Sustain. Energy*, vol. 2, no. 3, pp. 265–276, Jul. 2011.

[27] *IBM SPSS Modeler*. Accessed: Aug. 24, 2018. [Online]. Available: http://www.spss.com.hk/software/modeler/

DAVID B. ALENCAR received the B.S. degree in production engineering from the State University of Amazonas, Brazil, in 2007, and the master's degree in electrical engineering with an emphasis on industrial processes from the Federal University of Pará, Brazil, in 2011. He is currently pursuing the Ph.D. degree with the Federal University of Pará. He has over nine years of experience as an Engineer with Manaus Industrial Park, Brazil. He is currently an Assistant Professor with the Galileo Technological Institute of Amazon, Brazil. His current research interests include computational intelligence and data mining.

CAROLINA M. AFFONSO (S'03–M'07) received the B.S. degree in electrical engineering from the Federal University of Pará, Brazil, in 1998, the master's degree in electrical engineering from the Federal University of Santa Catarina in 2000, and the Ph.D. degree in electrical engineering from the State University of Campinas, Brazil, in 2004. Since 2005, she has been an Associate Professor with the Federal University of Pará, Brazil. She is currently a Visiting Professor with Texas A&M University, College Station, TX, USA. Her current research interests include smart grids, electric vehicles, distributed generation, and power system operation and planning.

ROBERTO C. L. OLIVEIRA was born in Belém, Pará, Brazil, in 1965. He received the master's degree in electrical engineering from the Aeronautical Institute of Technology, Brazil, in 1990, and the Ph.D. degree in electrical engineering from the Federal University of Santa Catarina, Brazil, in 1999. He is currently a Visiting Professor with Texas A&M University, College Station, TX, USA. He is also a Full Professor with the Federal University of Pará. He published over 200 refereed papers in journals and conferences and a book. His research interests include natural computing, computational intelligence, artificial neural networks, and evolutionary computing.

JOSE C. R. FILHO received the B.S. degree in electrical engineering from the Federal University of Amazonas, Brazil, in 1998, and the master's and Ph.D. degrees in electrical engineering from the Federal University of Pará, Brazil, in 2007 and 2014, respectively. He is currently a Professor of engineering with the Federal University of Amazonas. His research interest is on intelligent data analysis applied to electricity market.