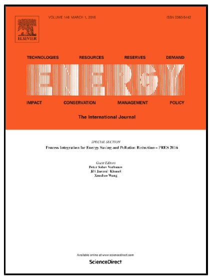
Accepted Manuscript



Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks

Henrique do Nascimento Camelo, Paulo Sérgio Lucio, João Bosco Verçosa Leal Junior, Paulo Cesar Marques de Carvalho, Daniel von Glehn dos Santos

PII:	S0360-5442(18)30486-9
DOI:	10.1016/j.energy.2018.03.077
Reference:	EGY 12535
To appear in:	*Energy*
Received Date:	23 December 2017
Revised Date:	08 February 2018
Accepted Date:	14 March 2018

Please cite this article as: Henrique do Nascimento Camelo, Paulo Sérgio Lucio, João Bosco Verçosa Leal Junior, Paulo Cesar Marques de Carvalho, Daniel von Glehn dos Santos, Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks, *Energy* (2018), doi: 10.1016/j.energy.2018.03.077

Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks

Henrique do Nascimento Camelo

Programa de Pós-Graduação em Ciências Climáticas (PPGCC), Universidade Federal do Rio Grande do Norte (UFRN), Campus Universitário Lagoa Nova, CEP 59078-970, Caixa Postal 1524, Natal, RN, Brazil. **e-mail: henriquecamelo13@yahoo.com.br (corresponding author)**

Paulo Sérgio Lucio

Programa de Pós-Graduação em Ciências Climáticas (PPGCC), Universidade Federal do Rio Grande do Norte (UFRN), Campus Universitário Lagoa Nova, CEP 59078-970, Caixa Postal 1524, Natal, RN, Brazil. e-mail: pslucio@ccet.ufrn.br

João Bosco Verçosa Leal Junior

Mestrado Acadêmico em Ciências Físicas Aplicadas (MACFA), Universidade Estadual do Ceará (UECE), Campus do Itaperi, Av. Dr. Silas Munguba, 1700, CEP 60714-903, Fortaleza, CE, Brazil.

e-mail: bosco.leal@uece.br

Paulo Cesar Marques de Carvalho

Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), Universidade Federal do Ceará (UFC), Campus do Pici, CEP 60440-554, Caixa Postal 6001, Fortaleza, CE, Brazil.

e-mail: carvalho@dee.ufc.br

Daniel von Glehn dos Santos

Mestrado Acadêmico em Ciências Físicas Aplicadas (MACFA), Universidade Estadual do Ceará (UECE), Campus do Itaperi, Av. Dr. Silas Munguba, 1700, CEP 60714-903, Fortaleza, CE, Brazil.

e-mail: danielvonglehn@gmail.com

Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks

ABSTRACT

This work shows two innovative hybrid methodologies capable of performing short and long term wind speed predictions from the mathematical junction of two classical time series models the Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) and the Holt-Winters (HW), both combined with Artificial Neural Networks (ANN). The first hybrid model (ARIMAX and ANN) is made from the physical relations between pressure, temperature and precipitation with the wind speed, that is, this model is considered as multivariate. The second hybrid model (HW and ANN) is considered as univariate, i.e. allowing only wind speed inputs. By means of statistical analysis of error it is verified that the proposed hybrid models offer perfect adjustments to the observed data at the regions of study, and thus, better comparisons with traditional ones from the literature. It is possible to find in this analysis percentage error of 5.0% and efficiency coefficient (Nash-Sutcliffe) of approximately 0.96. The confirmation of accuracy by the hybrid models reveals that they provide time series that are able to follow the observed time series profiles with similarities of maximum and minimum values between both series. Therefore, it became an important indicative in the representation of characteristics of seasonality by the models.

Keywords: Predictability; Time Series; Artificial Neural Networks; Exogenous Variables; Wind Generation.

1. Introduction

In recent years, there has been an increase in the usage of decentralized renewable energy sources as a means for electric power generation, such as wind energy power in Brazil. According to a study done by the Brazilian Wind Energy Association (ABEEólica[1]), the wind capacity installed at the end of 2016 in the country was approximately of 11 GW, which represents a significant increase when compared to 2005, when it was approximately 27 MW. Still, in accordance to ABEEólica, it is estimated that by 2020 Brazil will have around 18 GW of installed wind capacity, which will contribute to the country's energy security. It should be noted that Brazil has a large capacity to produce electricity from renewable energy sources [1].

Encouraging the usage of renewable energy resources, such as wind energy, is one of the strategies used to mitigate greenhouse gases from human activities in the atmosphere [2,3,4]. The Intergovernmental Panel on Climate Change (IPCC[2]) in its most recent report released in 2014 [5], pointed out that global warming is a reality and human contribution is significant to the occurrence of this phenomena. Thus, the increase of renewable energies in the different nations' matrices around the world can contribute to healthier environments [6].

The emergence of new techniques may further enhance the expansion of the wind sector, and the usage of forecasting methods may help further this growth. This will allow for the possibility of knowing the wind regime of a given region in the future, and thus, provide guarantees of electric power generation. In literature, there are several studies focused forecasting, however, only a few uses the analysis of wind speed with other meteorological variables to provide precise data. Kavasseri and Seetharaman [7] use the Auto-Regressive Integrated Moving Average (ARIMA) model, also known as Box-Jenkins modeling [8], for wind speed prediction of a 24-hour horizon in four locations in the North Dakota region of the USA.

Recently, with the development of artificial intelligence (AI), various new AI methods for wind speed prediction were also created. The newly developed methods include Artificial Neural Network (ANN), adaptive neuro-fuzzy inference system (ANFIS), fuzzy logic methods, support vector machine (SVM), neuro-fuzzy network, and evolutionary optimization algorithms. ANN could deal with non-linear and complex problems in terms of classification or forecasting. The ANN models can represent a complex nonlinear relationship and extract the dependence between variables through the training process. ANN based methods include back propagation neural networks, recurrent neural networks, radial basis function (RBF) neural networks, ridgelet neural

---

[1] http://www.abeeolica.org.br/en/dados-abeeolica/. (Accessed 20 June 2017).

[2] More details on http://www.ipcc.ch/. (Accessed 20 June 2017).

network, and adaptive linear element neural network. The ANN based method is a method that can be applied into the problem for forecasting wind power. Palomares-Salas et al. [9] used an ARIMA model for time-series forecast involving wind speed measurements. The paper presents the process of model validation, along with a regression analysis, based on real-life data. Results show that ARIMA model is better than back propagation neural network for short time-intervals forecasting.

In Liu et al. [10] the authors emphasize that wind speed prediction is important to protect the security of wind power integration. According to the authors, the performance of hybrid methods is always better than them separately when it comes to wind speed prediction. Thus, in this study, the two hybrid models based on Time Series, ANN and Kalman Filter (KF), performances were compared. In intelligent models, the ANN and the KF are popular due to their good nonlinear performance. For this particular analysis, they are chosen to do a multi-step prediction for two sections of non-stationary wind speed series from a wind farm in China. Wind speed data containing 500 measurements at the 40 m height level was used, the initial 300 measurements were used to fit the model, and the 200 final measurements were used for model validation. Based on literature reports the authors mentioned that a three-layer ANN model can handle any nonlinear data if the numbers of neurons are selected right, being that to short the training and forecasting time of an ANN model, in this study the three-layer network was chosen. In the hybrid ARIMA-ANN model, the ARIMA model is used to decide the structure of an ANN model, and in hybrid the ARIMA-Kalman model, the ARIMA model is employed to initialize the Kalman Measurement and the state equations for a Kalman model. Both cases portrayed good performances, which can thus be applied to the non-stationary wind speed prediction in wind power systems.

This article intends to contribute to the elaboration of wind predictability methodologies from the usage of innovative time series hybrid models (involving the exogenous variables of pressure, temperature and precipitation) with artificial intelligence. These hybrid models could assist in the power generation sector, for example, by providing information on wind potential for electricity generation.

2. Collected data and mathematical formulation

This section will focus on the studied regions' data, as well as on the forecast models used. The level of accuracy obtained is presented to illustrate the quality of adjustments of the models when compared to the observed time series. All the calculations produced in this study, as well as the graphical section, were executed by the software *R*[3].

[3] https://www.r-project.org/ (Accessed 20 June 2017).

2.1 Case study

The regions being studied are Fortaleza (3.77°S and 38.53°W) and Natal (5.75°S and 35.35°W), all in the Northeast of Brazil, as shown in Fig. 1. The meteorological data used are: wind speed (measured in m/s), pressure, temperature and precipitation (measured in hpa, ^{0}C and mm, respectively). All these data are made up of two forms: (1) the variables (wind speed, temperature, pressure and precipitation) in terms of monthly averages, 144 measurements in total for each variable were used starting in January 2003 and ending in December 2014; and (2) in terms of hourly averages, where the variables (wind speed, temperature and pressure) are used with all the hours on the year 2014, which consisted of 8760 measurements. Wind speed as well as pressure, temperature and precipitation data are originally measured at a level of 10 m tall and collected in anemometer towers at the airports of each location. These data donated by the Airmetar[4] project, first arose from the need for meteorological tools to consult and analyze information through the decoding of the METAR[5] data provided by the Brazilian aeronautics.

With the aim of forecasting monthly and hourly wind speed averages by applying the meteorological data into the proposed models, the following steps were taken: (I) adjustment phase – in terms of monthly averages, it consists of introducing the observed data (wind speed, pressure, temperature and precipitation) into the forecast models so that the respective adjustments can be made by using the January 2003 to December 2014 period, and in terms of hourly averages it consists of introducing the observed data (wind speed, pressure and temperature) into the forecasting models so that the respective adjustments can be made from 01/01/2014 to 31/12/2014; (II) adjustment quality phase – this investigation is performed through the usage of accuracy statistics, which will be described later on; (III) forecasting phase – serves to illustrate the ability of the models to do wind speed projections for the same period of the observed data through the usage of the most accurate models adjusted in two ways: (1) in terms of monthly averages for the months of 2015; and (2) in terms of hourly averages for the date of 01/01/2015.

2.2 Modeling Box - Jenkins and Box - Tiao

In order to carry out predictions of the wind speed in the studied regions, the Box–Jenkins [8] model was used, which refers to the systematic method of identifying, adjusting, checking and using the Auto-Regressive Integrated Moving Average model, whose abbreviation is commonly given by ARIMA. The usage of the ARIMA models is a powerful approach to solving many

[4] Further details on the website http://www.airmetar.com.br/. (Accessed 20 June 2017).
[5] Abbreviation for METeorological Aerodrome Report.

forecasting problems as it can provide extremely accurate predictions of time series [11]. In order to estimate the parameters of the ARIMA model, the values of p and q can be estimated, where p parameters φ (that is, the coefficient on the moving averages filter), q parameters θ (that is, the coefficient on the filter autoregressive stationary), and the errors of the model ε_t, also called residues, according to Eq. (1), where W_t represents the time series for projection, which in this work is the speed of the wind:

$$\begin{aligned} W_t &= \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \ldots + \varphi_p W_{t-p} + \ldots \\ &\ldots + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \end{aligned} \quad . \tag{1}$$

Another method used is the Box-Tiao model, commonly known as ARIMAX [12], and represents an expansion of the ARIMA models with the addition of a linear component, which includes the observations of the covariance (also called exogenous variables), thus making the ARIMAX model a multivariate. The main difference between the two is that the Auto-Regressive Integrated Moving Average with eXogenous inputs includes also an exogenous entrance besides for the autoregressive and moving average parameters [13]. The ARIMAX model can be understood as the combination of the Autoregressive model AR (p), Integrated (d), Moving Average -MA(q) and Exogenous X(r), which can then be symbolized by ARIMAX (p, d, q, r). A simplified form of the mathematical representation of this model can be found in Eq. (2):

$$y_t = \rho + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{r} \omega_j w_j + \sum_{j=1}^{q} (\theta_j \varepsilon_{t-j}) + \varepsilon_t , \tag{2}$$

where y_t is a dependent variable at time t, and the project intention (that is, wind speed); ρ is a constant; y_{t-i} is a dependent variable (which is also the wind speed) lagged by the time steps, i; β_i is a coefficient of y_{t-i}; p is the maximum number of time slots; w_j represents the exogenous variables (in this case it was included in the model through two forms: (I) - pressure, temperature and precipitation to get adjustments and forecasts for monthly averages; (II) – pressure and temperature to get adjustments and forecasts for hourly averages); ω_j represents the coefficients of the exogenous variables; r is the maximum number of exogenous variables; θ_j is the coefficient of the term ε_{t-j}, which represents the error in time t lagged from j. Lastly, ε_t is the error component of the model, with $\varepsilon_t \sim N(o,\sigma^2)$. The coefficients of the models are estimated by regression, more details on Equation (2) can be obtained at [13]. The orders of the ARIMA and ARIMAX models, that is, the numbers that will give rise to the operation of the functions were found through the functions *auto.arima*(TimeSeries.ts) and *arimax*(TimeSeries.ts), this function is part of the software package (*forecast*) *R*, created by Hyndman and Khandakar [14]. The choice of the best ARIMA and ARIMAX models take into account a number of criteria to provide the best adjustments, such as, the Akaike Information Criterion (AIC), which is based on identifying the lowest value for AIC, which, theoretically results in the best model for the adjustments of the observed data [14].

In order to evaluate if the models of the Box-Jenkins and Box-Tiao methodologies have feasibility for wind speed prediction, a residue analysis was performed. According to Brockwell and Davis [11], for the ARIMA and ARIMAX models to be viable regarding the adjustments of observed data, the error term ε_t of this model should behave as a white noise, that is, have a zero mean, constant variance (homoscedastic) and be uncorrelated, thus the errors are independent. Also, the term ε_t must follow a normal distribution. To identify these assumptions there are some possible tests that can be applied to the waste. Amongst them, the residue normality verification test from the Shapiro-Wilk Test, the residue independence test from the Durbin-Watson Test, and the test that verifies the equality of the residue variance, that is, the homoscedasticity hypothesis from the Breusch-Pagan Test, more details about the tests in Brockwell and Davis [11]. Table 1 shows the summarized representation of the used tests, which were applied at the significance level of α = 0.05.

Regarding the choice of exogenous variables applied to the ARIMAX model, it is important to note that the wind speed relates to the displacement of air masses, which, in turn, occurs due to the differences in pressure in the atmosphere [15]. The movement always occurs from high to low pressure zones, and thus determines the general dynamics of the atmospheric circulation and the formations of the different climatic types. It is important to note, that air masses can be defined as portions or volumes of the atmosphere that have generally the same characteristics of pressure, temperature and humidity due to their location, and are also quite thick and homogeneous [16]. In addition, air is composed mainly of nitrogen, oxygen and argon, which together make up most of the gases in the atmosphere. Other gases, such as water vapor, carbon dioxide, methane, nitrous oxide, and ozone form the greenhouse gases. Variables such as pressure, volume and temperature are considered essential to any gas, thus, due to the importance of pressure and temperature in relation to the displacement of air masses, it was decided to focus on both these variable in the application of the proposed model [16].

In addition, one of the motivations for choosing precipitation is due to the fact that in literature there are several studies on wind potential that show differences in the intensity of the winds over the northeastern region of Brazil due to the different precipitation period, that is, rainy or dry periods, as in [17,18,19], thus, it is necessary to include this variable in the ARIMAX model, so that it can provide adjustments and predictions of the wind speed taking into account the rainy and dry periods. It is important to note that variation in wind speed intensities are directly responsible for variation in rainfall intensities, for example, one of the large-scale dynamic mechanisms that produce rainfall on the northeast region of Brazil is its location in the Intertropical Convergence Zone (ITCZ) [20]. The ITCZ is a band of clouds that surrounds the equatorial line of the terrestrial globe, and it is formed mainly by the confluence of the trade winds from the northern hemisphere with the trade winds from the southern hemisphere. To simplify, it can be said that the convergence of the winds causes the hot, humid air to rise, carrying moisture from the ocean to the high levels of the atmosphere leading to the formation of clouds. The ITCZ plays a fundamental role in the occurrence of precipitation over the Brazilian Northeast (NNE), especially between the months of February and May in the states of Ceará and Rio Grande do Norte [21]. During this period, on the Atlantic Ocean, the ITCZ migrates to a position in the south, at about 4ºS favoring the existence of the rainy season on the NNE. It should be noted that the variations of the Northeast and Southeast trade winds seem to be one of the reasons for the alteration in intensity and positioning of convergence in the ITCZ [22].

2.3 Holt-Winters Model

Holt-Winters (HW) in the year of 1957, Holt expanded the simple exponential smoothing model to deal with the data that showed a linear tendency, thus making predictions that were more accurate than those performed with only the simple exponential smoothing. In 1960, Winters extended the Holt model, which included a new equation that predicts the behavior of the data´s seasonal component, thus generating the Holt-Winters (HW) [23]. The equation referring to the exponential smoothing method with seasonality and linear tendency, that is, with the seasonal component being treated in an additive way, is represented as follows:

$$Y_{t+n} = a_t + b_t . h + s_{t+p+n}, \tag{3}$$

where: a_t (level of the series, whose unit in this particular work is of m/s, shows how the expected time series evolves over time, being that it can vary slowly over time or undergo sudden variation. For example, the predicted series may have slow-growing motion in a linear form); b_t (tendency, whose unit here is m/s^2, this relates to the fact that the predicted time series can have increasing or decreasing motions in different time intervals. For example, a hypothetical forecasted time series can experience a linear growth in its first year, while on its second year it can overcome a linear decrease); s_t (seasonal component, represented here by m/s, which is related to the fact that the expected time series has cyclical patterns of variation that repeat at relatively constant time intervals. For example, a hypothetical predicted time series during one year shows undulating behavior every three months); Y_{t+n} (forecast for period n, and has a unit of m/s); p (seasonal period); $n = 1, 2, 3, \ldots, h$ (horizontal forecast).

2.4 Artificial Intelligence

The Artificial Neural Networks (ANN) are part of the so-called artificial intelligence, a branch of research in computer science. It was first presented in 1943 by the neurophysiologist McCulloch and the mathematician Walter Pitts. This model consisted of only one output and one input function, calculated by the weighted sum of a variety of values [24]. The model resembles the human neuron, where signals are received with different intensities, provoking a certain reaction on the nerve cell. Although biologically inspired in the human neuron, the Artificial Neural Networks have been applied to different scientific areas.

One of the computational intelligence techniques commonly used in attempting to predict time series is the ANN training, which is based on the architecture and learning of the human brain. ANNs work conceptually similar to the human brain, that is, it tries to recognize regularities and patterns of data while being able to learn from experience and making generalizations based on their previously accumulated knowledge. The ANNs in their structure may have both nonlinear and linear models and thus obtain better results when compared to other forecasting models [25].

An ANN can be associated with a network of "neurons" organized in layers. The predictors (or inputs $\boldsymbol{x_{t-i}}$) form the lower layer, and the predictions (or outputs $\boldsymbol{y_{t+h}}$) form the upper layer. There may be intermediate layers that contain "hidden neurons". There is an ANN structure known as the multilayered feed-forward network, where each layer of nodes receives inputs from the previous layers. Fig. 2 shows an example of an ANN structure with 5 inputs and 1 hidden layer. The coefficients representing the predictors are called "weights", and commonly represented by w_i. When dealing with time series data, lagged values of the time series can be used as inputs to a neural network. In this article, it was considered feed-forward networks with one hidden layer by using the notation NNAR(p,k) to indicate there are p lagged inputs and k nodes in the hidden layer. For example, a NNAR(9,5) model is a neural network with the last nine observations (x_{t-1}, x_{t-2}, …, x_{t-9}) used as inputs to forecast the output y_t, and with five neurons in the hidden layer. With seasonal data, it is useful to also add the last observed values from the same season as inputs. For example, NNAR$(3,1,2)_{12}$ model has inputs x_{t-1}, x_{t-2}, x_{t-3}, and x_{t-12}, and two neurons in the hidden layer. More generally, an NNAR$(p,P,k)_m$ model has inputs (x_{t-1}, x_{t-2}, …, x_{t-p}, x_{t-m}, x_{t-2m}, x_{t-Pm}) and k neurons in the hidden layer. It should be noted that in software *R* for the use of ANN, it is necessary to install the package (*forecast*) developed by Hyndman and Khandakar [14], and the function used is *nnetar*(TimeSeries.ts). This function provides the best choice of NNAR$(p,P,k)_m$ model based on observed wind speed data, if the values of p and P are not specified, they are automatically selected.

The outputs of nodes in one layer are inputs to the next layer. The inputs of each node are combined using a weighted linear combination. The result is then modified by a nonlinear function before becoming output. For example, the inputs into hidden neuron j in Fig. 2 are linearly combined to give

$$Z_j = b_j + \sum_{i=1}^{3} w_{i,j} y_i \,, \quad (4)$$

where: b_j, and $w_{i,j}$ are parameters discovered during the "learning" stage from the observed data of the time series being studied. The weights take random values to begin with, which are then updated using the observed data. Consequently, there is an element of randomness in the predictions produced by a neural network. Therefore, the network is usually trained several times using different random starting points, and the results are averaged. In the hidden layer, Eq. (4) is modified using a nonlinear activation function, such as a sigmoid, given by the equation:

$$s(z) = \frac{1}{1+e^{-z}} \,. \quad (5)$$

Eq. (5), however, serves as input to the next layer, and this strategy tends to reduce the effect of extreme input values, providing a better performance by ANN. More details on the usage of ANN to predict time series can be obtained from [25]. In literature, it is possible to highlight works that use ANN to predict time series of wind speed, as for example in Li and Shi [26], where the authors performed the prediction of wind speed by using three different forms of Artificial Neural Networks.

2.5 Hybrid Modeling

In this work three forms of hybrid modeling were tested for wind speed prediction, as described below:

(I) Hybrid model (ARIMA + ANN)

From the combination of the ARIMA model with ANN, that is, using a time series model, in this case, the ARIMA, combined with an artificial intelligence model, the ANN. In Zhang [27], the author proposes the combination of the ARIMA and ANN models by stating that this model can capture different patterns in the data. Statistical ARIMA models are capable of capturing linear patterns. However, because of its nonlinear modeling capacity, the use of ANN for time series forecasts has been strongly used [28]. The hybrid model of seasonal time series forecast, by combining the ARIMA and ANN models, is represented as follows:

$$\text{hybrid}(1) = \text{ARIMA}_{\text{forecast}} + \text{ANN}_{\text{forecast/ARIMA}}, \tag{6}$$

where hybrid(1) represents the prediction of the hybrid model, which is the result of adding the linear component predicted by the $\text{ARIMA}_{\text{forecast}}$ model, to the nonlinear component predicted by the $\text{ANN}_{\text{forecast/ARIMA}}$, being that the latter component is modeled from the residues of the ARIMA model. It should be emphasized that this method has been used in several studies of wind speed time series, for example, in Cadenas and Rivera [29], the authors proposed a hybrid model based on ARIMA and ANN for wind speed prediction in three different locations in Mexico.

(II) Hybrid models (ARIMAX+ANN) and (HW+ANN)

The hybrid models proposed in this article possess similarities to that described previously in Zhang [27], that is, the first is composed of a linear component (the ARIMAX model uses the exogenous variables of pressure, temperature and precipitation), and a nonlinear components (that utilizes the residues of the ARIMAX model in the ANN), being represented by the following expression:

$$\text{hybrid}(2) = \text{ARIMAX}_{\text{forecast}} + \text{ANN}_{\text{forecast/ARIMAX}}, \tag{7}$$

where hybrid(2), represents the prediction of the hybrid model, which is the result of adding the linear component predicted by the $\text{ARIMAX}_{\text{forecast}}$, with the nonlinear component predicted by the $\text{ANN}_{\text{forecast/ARIMAX}}$, being that the latter component is modeled from the residues of the ARIMAX model. The next hybrid model is proposed from the combination of the HW models (this being the linear component) and ANN (the nonlinear component), the mathematical representation of this model is given by the following expression:

$$\text{hybrid}(3) = \text{HW}_{\text{forecast}} + \text{ANN}_{\text{forecast/HW}}, \tag{8}$$

where: hybrid(3), represents the prediction of the hybrid model, which is the result of the sum of the linear component predicted by the $\text{HW}_{\text{forecast}}$ model, with the nonlinear component predicted by the $\text{ANN}_{\text{forecast/HW}}$ model, being that the latter component is modeled from the residues of the HW model. It is worth highlighting the innovative feature of this hybrid model in the attempt to make time series forecasts, especially for the wind speed variable.

2.6 Measures of Accuracy

One way to verify the accuracy of the proposed forecasting models is through the statistical analysis of errors. The Mean Absolute Error represents the average error value between the observed and the adjusted series. In this analysis, the variable will be represented by MAE, whose mathematical representation is given by,

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|v_{adj} - v_{obs}\right|, \tag{9}$$

where v_{adj} represents the individual value of the predicted time series, and v_{obs} represents the individual value of the observed time series, and n is the order of the series. In this study, the MAE has a unit of measurement of meters per second, or (m/s). The Root Mean Squared Error (RMSE) represents the individual quadratic differences between the observed and adjusted time series of wind speed.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(v_{adj} - v_{obs}\right)^{2}}\ . \tag{10}$$

Likewise, the RMSE in Eq. (10) has a unit of m/s. It is important to note that the RMSE can be interpreted as follows: when the value of this variable is high, the error margin of the adjusted variables is also high, while values that are close to zero indicate a near perfect fit [30]. Another way to measure error is through the Mean Absolute Percentage Error (MAPE). The advantage of using this expression is that it uses percentages (%) to illustrate the data, which allows for an easy and quick understanding. However, a disadvantage is that, if by any chance, the observed value is too small, any discrepancy will cause the MAPE to "explode". The expression used with this variable is represented by,

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{v_{adj} - v_{obs}}{v_{obs}}\right| \times 100\ . \tag{11}$$

Another measure used to identify the quality of the adjustments is the Nash-Sutcliffe efficiency coefficient (NS), which can vary from - ∞ to 1, being that values close to 1 indicate a perfect fit, while a NS value larger than 0.75 indicates that the performance of the model was considered good. For values between 0.36 and 0.75, the performance is considered acceptable, while NS values smaller than 0.36 makes the model seem unacceptable [31]. The expression used to calculate NS is given by,

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n} \left(v_{obs} - v_{adj}\right)^2}{\sum_{i=1}^{n} \left(v_{obs} - \bar{v}_{obs}\right)^2}, \qquad (12)$$

where $\bar{v}_{obs}$ represents the mean of the time series observed.

2.7. Models schematics

Fig. 3 shows the steps necessary to obtain the forecasts for the hourly and monthly averages by the models (a) hybrid(2) and (b) hybrid(3). The strategy to perform wind speed predictions is similar for both hybrid models, and the difference is related to the fact that the hybrid(2) model (composed of the ARIMAX and ANN models) is considered multivariate. That is, to obtain the wind speed values (values - Fitted and Residuals) by the ARIMAX model, it is necessary to first input the exogenous variables (temperature, pressure and precipitation). However, the hybrid(3) model (composed of the HW and ANN models) is considered univariate, thus, in order to obtain the wind speed values (values - Fitted and Residuals) by the HW model, it is only required the input of the wind speed variable.

All calculations and graphical illustrations were produced by the statistical software R, which for its full functionality, in some situations requires the installation of other supporting packages. The *forecast* package elaborated by Hyndman and Khandakar [14] was the main one used for the construction of the HW and ANN models, and consequently the Hybrid model proposed in this study. In the case of the HW model the *HoltWinters*(TimeSeries.ts) function was used, which provides all the coefficients necessary for the calculation of the forecast. To calculate the forecast, the function *arimax*(TimeSeries.ts) was responsible for obtaining the best ARIMAX model. Finally, the ANNs use the function *nnetar*(TimeSeries.ts), which provides the best possible network in accordance with the input data.

3. Results and Discussion

3.1 ANNs' characteristics obtained from the wind speed data

Table 2 numerically represents the characteristics of all the ANNs obtained in each study location. The results generated by the R software, through the input of the observed wind speed data, provided the best choices for the ANNs. It is important to highlight that the time series inputs on the artificial network are only affected by the wind speed variable, as allowed by the function *nnetar*(TimeSeries.ts). In order to carry out its later projection, this function works with the auto-regressive mechanism of the time series in question. It is also worth noting that in the case of the hybrid models, the ANNs have the ability to model only the residues that are nonlinear and provided by the ARIMA, ARIMAX and HW models. In order to illustrate the significance of the ANN's structure, the ANN relative to the residues of the hybrid(2) model for the hourly period in Fortaleza presents the following characteristics: p = 39 represents the quantity of wind speed inputs necessary to provide an output, P = 1 shows that the time series in question has seasonality with frequency m = 24, being that the ANN has k = 20 neurons in the hidden layer. Finally, these parameters were calculated with an average of 20 neural networks each containing 821 weights. Recalling that the residues obtained by the ANN from the hybrid(2) model were modeled from the residues generated by the ARIMAX model.

3.2. Validation of models

The analysis of the residuals, which are shown by the differences between the time series observed and adjusted by the ARIMA and ARIMAX models, is important in the identification of the quality of the models. Table 3 shows the results of the tests applied to the residuals to identify if the assumptions of the models were met, considering the level of significance of 0.05. The numbers shown are relative to the (p-value), where in the case of the monthly data, all the assumptions were met, because they were higher than the significance level. However, with respect to the hourly data the same cannot be said, especially regarding the KS normality test that was not applied in the three study regions. This may be due to the fact that the amount of waste was very high, around 8760 values, making it difficult to apply to the tests. However, the results of the error statistics analysis will assist in the reliability of the ARIMA and ARIMAX models for the hourly scenario.

In order to analyze the accuracy of the tested models, and thus, identify which model offered the best adjustments to the observed data, Table 4 presents a statistical analysis of errors for the monthly data. In Fortaleza and Natal the model that provided the smallest error, and therefore, the best adjustments was the hybrid(2) model, that is, for this model the MAPE values for the two locations were 8.03% and 7.21%, respectively. These values were lower than those found by the two models that make up hybrid(2) separately, that is, the ARIMAX and ANN. For example, in Fortaleza the values of MAPE calculated by ARIMAX and ANN models separately were 8.48% and 10.29%, respectively. For Natal the MAPE value found by the ARIMAX and ANN were of 9.44% and 8.14%, respectively. The fact that the hybrid(2) model showed the lowest error values when compared to the other models used in this article, may favor greater similarities between the adjusted and observed time series, mainly in terms of maximum and minimum wind speed, thus producing more accurate forecasts for this particular variable.

Regarding the hourly wind speed data, Table 5 shows the statistics of the wind speed where it is possible to note that the hybrid(2) model provides the lowest error measurements as well as the best adjustments to the observed data. For example, the MAPE values for Fortaleza and Natal were approximately 7.20% and 7.22%, respectively. Similarly to the monthly data, these low values can provide better predictions of expected wind speeds, which will be shown below. The hybrid(2) model presented larger errors for the hourly measurements in comparison to the monthly measurements, which must be due to the different amount of data used, that is, for the hourly analysis around 8760 measurements were used for the year 2014, which is much higher than the monthly data, with only 144 measurements. However, it is important to highlight that the error values in both cases are low and very close to each other, thus emphasizing the quality of the model in providing good adjustments to the observed data.

3.3. Comparison with literature results

The results shown in the statistical analysis of errors (Tables 4 and 5) for the hybrid models proposed are very encouraging as they illustrate the possibility of providing precision in the predictions of wind speed for the studied region. It is important to note that these results are in agreement with the values found in similar literature studies, for example, Cadenas and Rivera [32] compared the performance of ARIMA and ANN models in an attempt to forecast monthly wind speed averages in the coastal region of Mexico. According to the authors, for this particular case the ARIMA model presented greater sensitivity to wind speed adjustment and prediction, the MAPE value found was approximately of 13.40%. However, it is likely that with the increase in the number of training vectors for the ANN model there will be an improvement in the adjustments, allowing for smaller statistical errors than that presented for the MAPE, which was 20.70%.

In [7], propose the usage of a time series forecasting models such as the ARIMA model to estimate the wind speed for a 24 hours basis in the regions of Minnesota-North Dakota, USA. The error statistic in this work presented an error percentage of approximately 12.00%. In [29], besides for the usage of the ARIMA model to forecast wind speed, the authors use ANN and a hybrid model composed of two models to estimate the hourly averages for three different locations in Mexico. The hybrid model showed the lowest values of error statistics, for example, for the MAE it was possible to find values of 0.06 m/s, versus that of 0.60 m/s and 0.67 m/s, for the models of ARIMA and ANN respectively. In terms of wind speed prediction for the monthly averages using the ARIMA model it is possible to cite [33], where the authors evaluate the prediction for three wind farms in northwest China. The results obtained show a MAPE value of approximately 13.60% for the period between January 2001 and December 2006.

Fadare [34] used wind speed monthly average data of 28 weather stations for a period of 20 years. During which, data of 18 stations were used for the training of the model and data from 10 stations were used to test it. The proposed ANN consists of 4 inputs, 2 hidden layers, and one output. The author used information about latitude, longitude, altitude and month of the year as inputs for the model, and the wind speed was the output of ANN. The results indicate that the proposed topology shows high accuracy in the predicting of wind speed monthly average, reaching a total MAPE of 8.90%. Most importantly, the correlation coefficient between the predicted and the observed was of 0.98, which shows the effectiveness of this model. Liu et al. [10] presented a combined hybrid model of the ARIMA and ANN models to predict the wind speed in terms of hourly means, whose adjusted time series was obtained from 500 measurements. The proposed model presented good accuracy, with a MAPE value of approximately 3.32%.

3.4. Time series graphical illustrations

Fig. 4 shows the comparisons between the monthly averages of the observed and adjusted time series (relative to the best models based on the statistical analysis of errors for each location), being in Fortaleza, the hybrid(2) model shown in (Fig. 4a) and Natal by hybrid(2) model in (Fig. 4b). The Nash-Sutcliffe (NS) efficiency coefficients presented values of 0.86 and 0.79, which demonstrate the efficiency of the hybrid(2) models in providing good adjustments to the observed data of the studied regions. In both regions it is possible to identify the similarities between the two time series (observed and adjusted), especially in terms of maximum and minimum wind speed values for the majority of the months. For example, in Fortaleza for the two time series (adjusted and observed) the minimum wind speed for the year of 2011 both occurred in the month of April, at approximately 2.5 m/s. In terms of maximum wind speed values for 2011, the month of September represented the period of maximum values for both time series (adjusted and observed) with numbers close to 6.0 m/s.

According to a study by the National Institute of Meteorology (INMET[6]), in Fortaleza the months of greatest precipitation, commonly known as the rainy season, are between February and May, with peak in the month of April. During this period, according to Camelo et al. [35], the monthly average of the wind speed intensity in Fortaleza is lower. However, during the period of

[6] Further details on: <http://www.inmet.gov.br/portal/index.php?r=clima/normaisclimatologicas>. (Accessed 20 June 2017).

low precipitation, also known as dry season, that occurs between the months of September to December the monthly wind speed average is high. In regards to the analysis of the hybrid models, it is possible to identify in (Fig. 4a), that the model hybrid(2) represents seasonal characteristics of the wind speed in Fortaleza, that is, lower intensities during the rainy season, and higher intensities during the dry season. Still according to INMET, the historical series of precipitation in the city of Natal shows a rainy period during the first half of the year, between March and July (with a peak in April). Like in Fortaleza, the hybrid(2) model adjustments for Natal also portrayed wind seasonality, where lower intensities occurred in the rainy season and higher intensities in the dry season.

Once the quality of the adjustments produced by the hybrid models is proven, it is possible to expect accuracy in the wind speed predictions, as shown in Fig. 5. The forecasts for the following year that is for 2015 were compared to their respective observations for that same year in Fortaleza (Fig. 5a) and Natal (Fig. 5b). The forecasts follow the profile of the series observed for both regions, presenting values very close to each other, for example, in Fortaleza the minimum values of the forecast and observation occurred in the same month of April and at approximately 4.0 m/s, likewise, the maximum values for the forecast and observed are also similar at approximately 7.0 m/s and both occurred in October. Another detail about the forecasts refers to the fact that they also illustrate the seasonality of the observed time series, that is, lower and higher wind speed intensities according to the different rainy and dry seasons.

Regarding the hourly wind speed time series in terms of observations, adjustments and predictions, it was verified by the statistical analysis of errors that the hybrid(2) model also provided the lowest values of error statistics. Fig. 6 shows the comparisons in Fortaleza, being that Figs. 6a and 6b represents the observations and adjustments respectively, for all the hours relative to 2015, and Fig. 6c illustrates the comparison (for the day following the adjustments, that is, 01/01/2015) between the predicted time series and the observed time series. In relation to the comparison between adjustments and observations in Figs. 6a and 6b, even though there are 8760 measurements of both series making it difficult to visualize, it is possible to identify similarities between the two series throughout the entire analyzed period. For example, between June 1st at 0:00h to August 1st at 0:00h the adjusted series reproduced the existing pattern of the observed series with variation of the wind speed, for the majority of the values, between 5.0 m/s to 10.0 m/s.

Another relevant detail that highlights the quality of the hybrid(2) model in Fortaleza, refers to the fact that it makes it possible to identify the adjusted time series with the same wind speed seasonality characteristics as for the observed data, that is, rainy months (February to April) have lower hourly wind speeds averages, on the other hand, the driest months (September to December) show higher values of hourly wind speed averages. This result illustrated the ability of the model to represent the differences in wind speed intensity when taking into account the seasonal characteristics, shown by the NS coefficient of efficiency of 0.96, thus a perfect adjustment to the observed data.

Fig. 6c shows the comparison between observation and forecast in terms of hourly averages, both relative to the day following the adjusted time series, that is, 01/01/2015. It is possible to identify that the forecast follows the profile of the observation with values very close to each other, for example, the forecast and observed peak coincide at 17:00h at values of approximately 7.0 m/s and 7.5 m/s, respectively. The most intense values for the predicted hourly wind speed averages occurred between the hours of 6:00h and 5:00h, which also agrees to the most intense values found for the observed series. It should be pointed out that the highest intensity for the hourly wind speed average occurs during the daytime in the northeastern region of Brazil because of the presence of sea breezes, as explained by Molion and Bernardo [36].

Fig. 7 shows a comparative among the time series for Natal in terms of wind speed hourly averages: observed, shown in Fig. 7a through a black line; adjusted, shown in Fig. 7b through a grey line; as well as the adjustments, forecasts and observations for the hours of 01/01/2015 represented in Fig. 7c. The value of the NS efficiency coefficient is 0.97, which shows a perfect fit to the observed data. Due to this result, it is possible to note that the adjusted time series can reproduce the maximum and minimum wind speed patterns of the observed time series during practically the entire period. Regarding the forecast, it is possible to identify that it follows the profile of the observed data with similar values, especially during daylight hours between 8:00h and 4:00h, for example, at 8:00h both series have a minimum value of approximately 4.0 m/s. The most intense hourly averages of the wind speed forecast are mostly during daylight hours between 6:00h and 10:00h and between 12:00h and 18:00h, which is also the most intense values for the time series observed.

It is important to note that the results for the wind speed predictions in this article were analyzed at a height of 10 m tall, when in reality the turbines in Brazil are over 100 m tall, according to ABEEólica[7] the prediction models do not possess dependency on specific heights; therefore it can be used for any height, however, from the availability of data with higher levels of height, the viability of the proposed hybrid models can be verified. The result of this work can serve as a tool in several areas, such as in the wind energy sector, by providing further knowledge of wind speed forecasting in the region for decision makers.

3.5. Important research highlights

---

[7]http://www.portalabeeolica.org.br/noticias/5283-torres-mais-altas-elevam-em-6-vezes-potencial-elico.html. (Accessed 20 June 2017).

In order to emphasize the results obtained in this article, below is a list of relevant statements:

- The present article deals with the elaboration of research that can help in the development of wind energy generation, that is, an attempt to encourage the use of renewable resources for energy generation;

- Innovative hybrid models capable of performing short or long-term wind time series forecasts are presented in this research, which provides high precision;

- From the combination of time-series models with artificial intelligence models it is possible to create innovative hybrid models, there have been improvements in the wind speed predictions of the study regions when compared to predictions provided by traditional literature models ;

- One of the innovative hybrid models, that is, the ARIMAX + ANN model, takes into account local meteorological characteristics (relation of the wind speed with the variables of temperature, pressure, and precipitation). The accuracy of this model can be related to the inclusion of these variables, considering that, when using the traditional hybrid model, ARIMA + ANN (only the input of wind speed is taken into account);

- The other innovative hybrid model, the HW + ANN presents better values of accuracy when compared to the traditional literary HW. This hybrid model can be very useful in the situation of only having wind speed as the input variable;

- The hybrid ARIMAX + ANN model can represent characteristics of local seasonality of wind speed, being this factor important for possible energy planning.

- Innovative hybrid models can be tested in the wind power generation sector in order to be identified as useful tools for regional energy planning.

4. Conclusions

The hybrid models proposed in this article, the ARIMAX + ANN, and the HW + ANN, created from combining a time series model with artificial intelligence, in order to perform monthly and hourly wind speed averages, were efficient in producing adjustments to the observed data. This assumption is based on the calculation of the NS efficiency coefficient, whose values were 0.96 and 0.97 respectively for Fortaleza and Natal. These values illustrate the similarities in the maximum and minimum wind speed of both adjusted and observed time series for each of the studied location. Another conclusive factor that shows the quality of the proposed hybrid models is the low statistical analysis of errors values, especially when compared to the other models, for example, in Fortaleza it was approximately 8.03% and 7.20% for the monthly and hourly averages, respectively. This analysis is certainly responsible for providing greater precision for the predicted time series in both monthly and hourly averages. It was also clear that the proposed hybrid models can represent the wind speed's seasonal characteristics in the studied region, that is, lower intensities during the rainy season, and higher intensities during the dry season. The most intense hours of wind speed averages for the predicted time series in the studied region occurred during the daytime between 6:00h to 18:00h, being that these hours are also in accordance with the most intense observed time series values. This research can influence the wind energy sector, by providing more guarantees to decision makers of the possibility of forecasting local wind speed intensity, and thus optimize strategies to meet the demand for electricity from wind farms.

5. Acknowledgements

The authors would like to thank the Postgraduate Program in Climate Sciences (PPGCC) of the Federal University of Rio Grande do Norte (UFRN). This work was supported by the National Petroleum Agency (ANP), with grant number PRH-51; and the National Council for Scientific and Technological Development (CNPq), with grant numbers PQ2 Proc. 301553/2016-5, and Proc. 400053/2014-4.

6. References

[1] Fontoura CF, Brandão LE, Gomes LL. Elephant grass biorefineries: towards a cleaner Brazilian energy matrix?. J Clean Prod 2015;96:85-93. https://doi.org/10.1016/j.jclepro.2014.02.062.

[2] Dong K, Sun R., Hochman G. Do natural gas and renewable energy consumption lead to less CO 2 emission? Empirical evidence from a panel of BRICS countries. Energy 2017;141: 1466-78. https://doi.org/10.1016/j.energy.2017.11.092.

[3] Squalli J. Renewable energy, coal as a baseload power source, and greenhouse gas emissions: Evidence from US state-level data. Energy 2017;127:479-88. https://doi.org/10.1016/j.energy.2017.03.156.

[4] González MOA, Gonçalves JS, Vasconcelos RM. Sustainable development: Case study in the implementation of renewable energy in Brazil. J Clean Prod 2017;142:461-75. https://doi.org/10.1016/j.jclepro.2016.10.052.

[5] Pachauri RK, Meyer L, Plattner GK, Stocker T. IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. http://boris.unibe.ch/71642/. (Accessed 11 June 2017), 2015.

[6] Lin B, Ankrah I, Manu SA. Brazilian energy efficiency and energy substitution: a road to cleaner national energy system. J Clean Prod 2017;162:1275–84. https://doi.org/10.1016/j.jclepro.2017.06.011.

[7] Kavasseri R, Seetharaman K. Day-ahead wind speed forecasting using *f*-ARIMA models. Renew. Energy 2009;34:1388-93. https://doi.org/10.1016/j.renene.2008.09.006.

[8] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. New York: John Wiley & Sons; 2015.

[9] Palomares-Salas JC, de la Rosa JJG, Ramiro JG, Melgar J, Aguera A, Moreno A. ARIMA vs. Neural Networks for Wind Speed Forecasting. In: Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Hong Kong, 11-13 May 2009:129-133.

[10] Liu H, Tian H, Li Y. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl. Energy 2012:98;415–24. https://doi.org/10.1016/j.apenergy.2012.04.001.

[11] Brockwell PJ, Davis RA. Introduction to time series and forecasting. 4th ed. New York: Springer; 2016.

[12] Dannecker L. Energy Time Series Forecasting: Efficient and Accurate Forecasting of Evolving Time Series from the Energy Domain. New York: Springer; 2015.

[13] Li Y, Su Y, Shu L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. Renew Energy 2014;66:78-89. https://doi.org/10.1016/j.renene.2013.11.067.

[14] Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. J Stat Softw 2008;23:1-22. https://doi.org/10.18637/jss.v027.i03.

[15] Ahrens CD. Meteorology today: an introduction to weather, climate, and the environment. 9th ed. Boston: Cengage Learning; 2012.

[16] Iribarne JV, Godson WL. Atmospheric thermodynamics (Vol. 6). 2th ed. Boston: Springer Science & Business Media; 2012.

[17] Martins FR, Pereira EB. Enhancing information for solar and wind energy technology deployment in Brazil. Energ Policy 2011;39:4378-90. https://doi.org/10.1016/j.enpol.2011.04.058.

[18] Rocha PAC, de Sousa RC, de Andrade CF, da Silva MEV. Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. Appl Energy 2012;89:395-400. https://doi.org/10.1016/j.apenergy.2011.08.003.

[19] de Andrade CF, Neto HFM, Rocha PAC, da Silva MEV. An efficiency comparison of numerical methods for determining Weibull parameters for wind energy applications: A new approach applied to the northeast region of Brazil. Energy Convers Manag 2014;86:801-8. https://doi.org/10.1016/j.enconman.2014.06.046.

[20] Misra V. A sensitivity study of the coupled simulation of the Northeast Brazil rainfall variability. J Geophys Res 2007;112:1-16. https://doi.org/10.1029/2006JD008093.

[21] Hounsou-Gbo GA, Araujo M, Bourles B, Veleda D, Servain J. Tropical Atlantic contributions to strong rainfall variability along the Northeast Brazilian coast. Adv Meteorol 2015;2015:1-13. http://dx.doi.org/10.1155/2015/902084.

[22] Souza P, Cavalcanti IFA. Atmospheric centres of action associated with the Atlantic ITCZ position. Int J Climatol 2009;29:2091-105. http://dx.doi.org/10.1002/joc.1823.

[23] Makridakis SG, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. 4th ed. New York: John Wiley & Sons; 2008.

[24] Rojas R. Neural networks: a systematic introduction. Boston: Springer Science & Business Media; 2013.

[25] Khashei M, Bijari M. An artificial neural network (p, d, q) model for time series forecasting. Expert Syst Appl 2010;37:479-89. https://doi.org/10.1016/j.eswa.2009.05.044.

[26] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl. Energy 2010;87:2313-20. https://doi.org/10.1016/j.apenergy.2009.12.013.

[27] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159-75. https://doi.org/10.1016/S0925-2312(01)00702-0.

[28] Aladag CH, Egrioglu E, Kadilar C. Forecasting nonlinear time series with a hybrid methodology, Appl Math Lett 2009;22:1467-70. https://doi.org/10.1016/j.aml.2009.02.006.

[29] Cadenas E, Rivera W. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. Renew Energy 2010;35:2732-38. https://doi.org/10.1016/j.renene.2010.04.022.

[30] Wilcox RR. Fundamentals of modern statistical methods: Substantially improving power and accuracy. New York: Springer Science & Business Media; 2010.

[31] Jain SK, Sudheer KP. Fitting of hydrologic models: a close look at the Nash–Sutcliffe index. J Hydrol Eng 2008;13:981-86. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:10(981).

[32] Cadenas E, Rivera W. Wind speed forecasting in the south coast of Oaxaca, Mexico. Renew. Energy 2007;32:2116-28. https://doi.org/10.1016/j.renene.2006.10.005.

[33] Hu J, Wang J, Zeng G. A hybrid forecasting approach applied to wind speed time series. Renew. Energy 2013;60:185-94. https://doi.org/10.1016/j.renene.2013.05.012.

[34] Fadare DA. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. Appl Energy 2010;87:934–42. https://doi.org/10.1016/j.apenergy.2009.09.005.

[35] Camelo HN, Carvalho PCM, Leal Junior JBV, Accioly Filho BP. Statistical analysis of the wind speed of the state of Ceará. Magazine Technology 2008;29:211-23.

[36] Molion LCB, Bernardo SDO. A review of rainfall dynamics in the Brazilian Northeast. Brazilian Journal of Meteorology 2002;17:1-10.

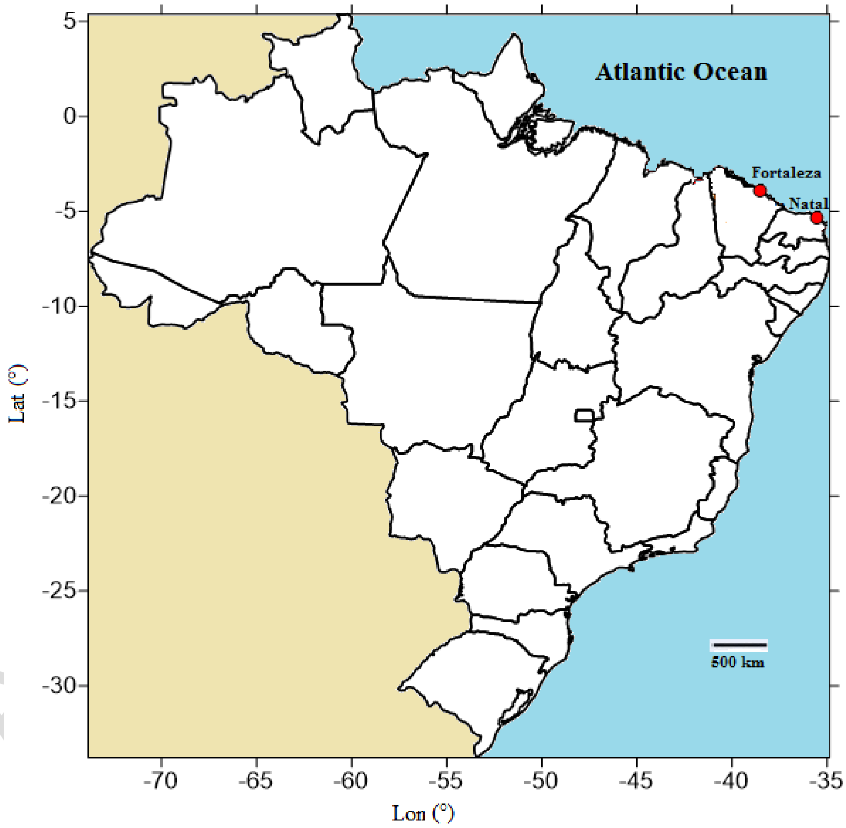


Fig. 1. Map of Brazil with representation of the study regions: Fortaleza and Natal

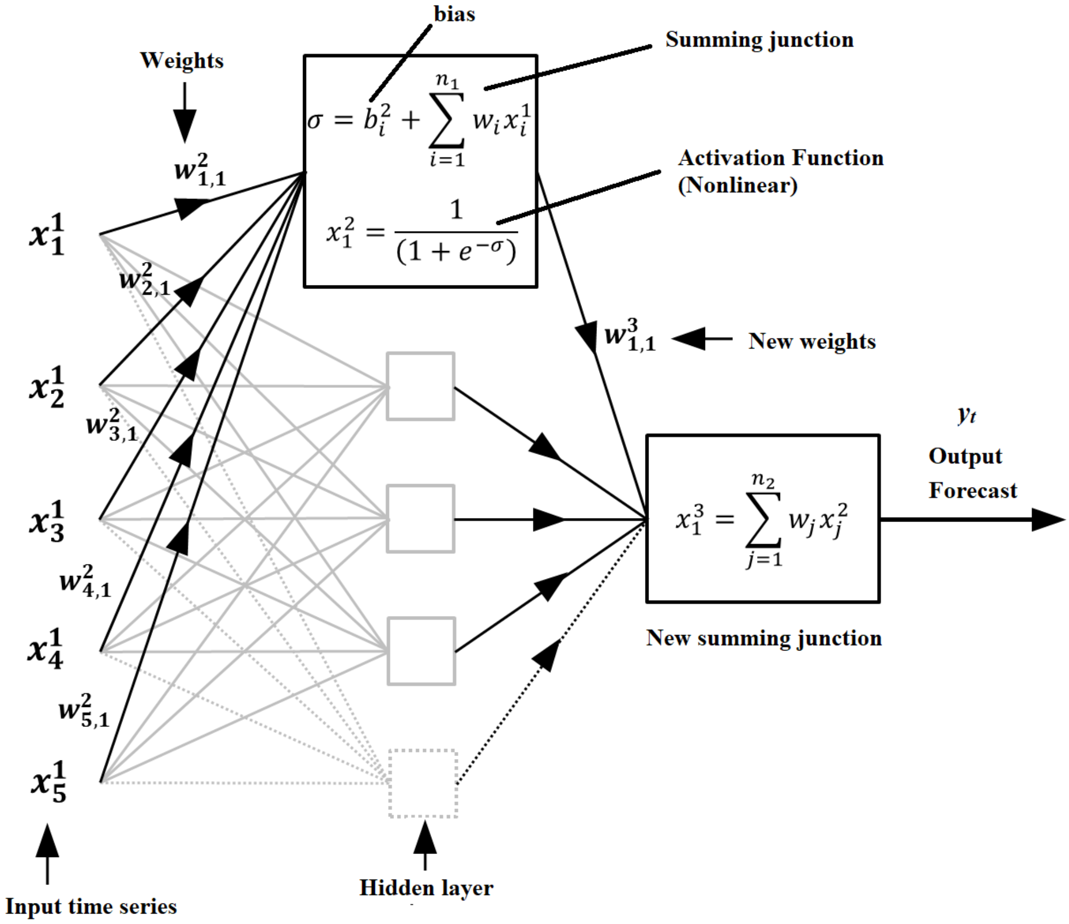


Fig. 2. Artificial Neural Network Structure.

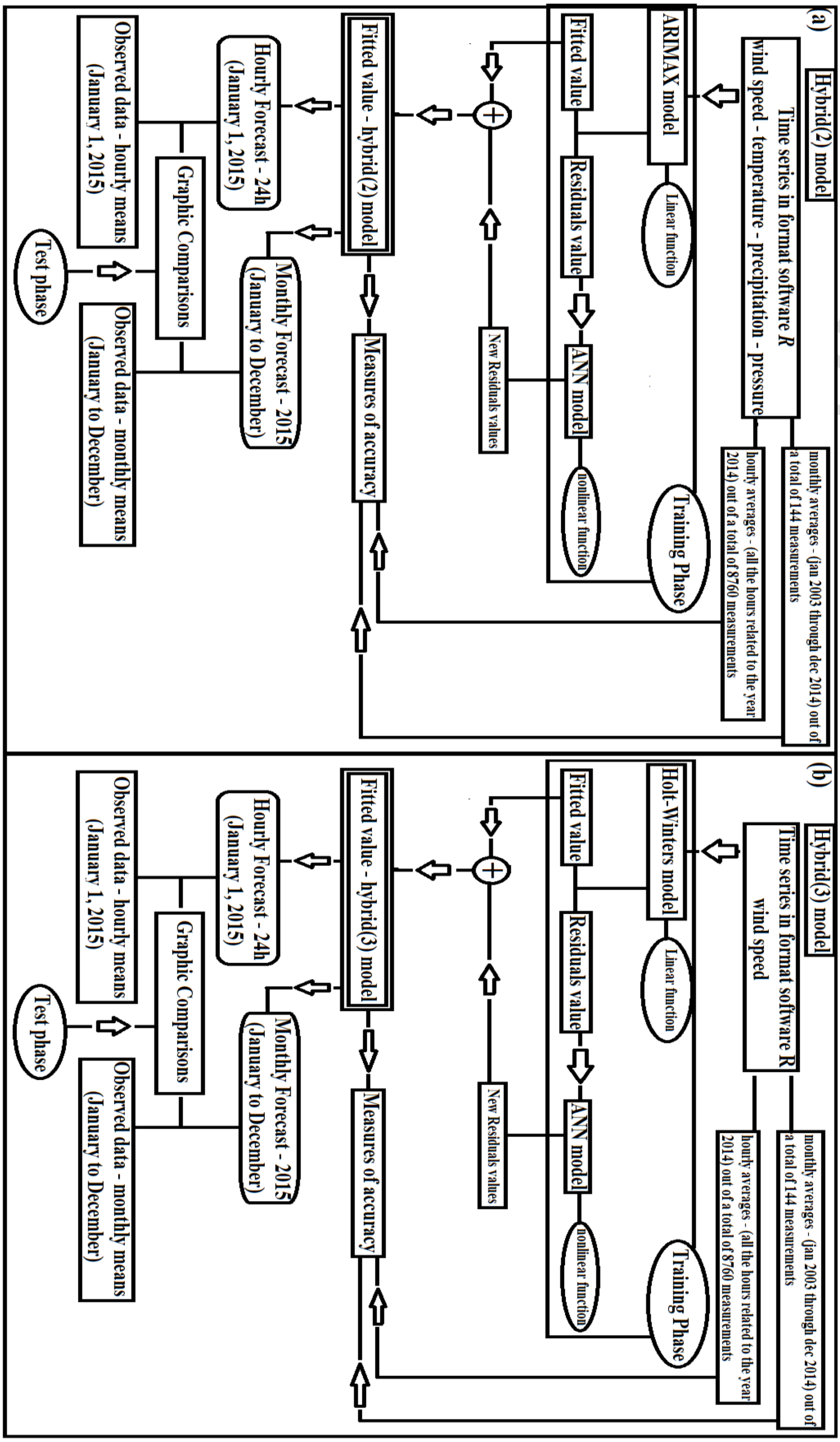


Fig. 3. Description of the models (a) hybrid(2) and (b) hybrid(3).

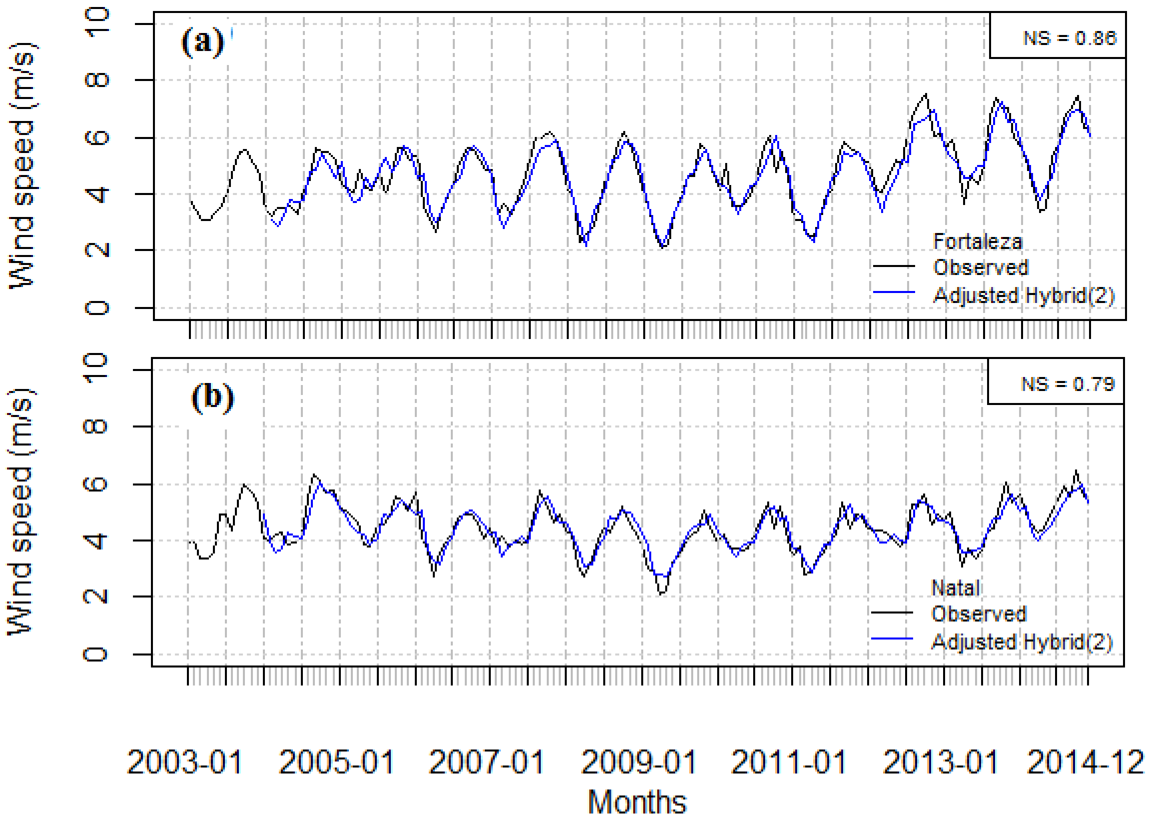


Fig. 4. Monthly wind speed averages (a) Fortaleza and (c) Natal, comparison between the observed time series (in continuous black line) and adjusted (in continuous blue line). Information on the NS value used to identify the quality of the adjustment for each region.

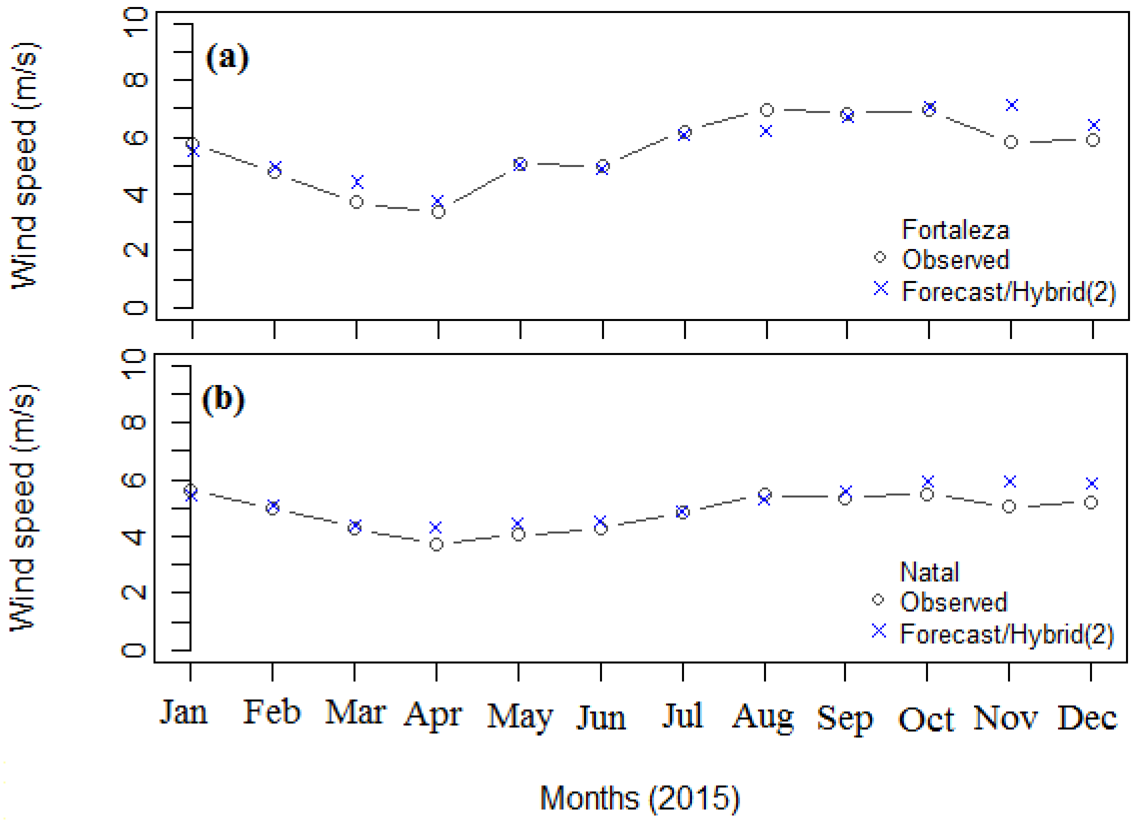


Fig. 5. Comparisons between the time series predicted by the hybrid(2) model (x-shaped in blue) with that of the observed time series (empty black circle) for the months of 2015. (a) Fortaleza and (b) Natal.

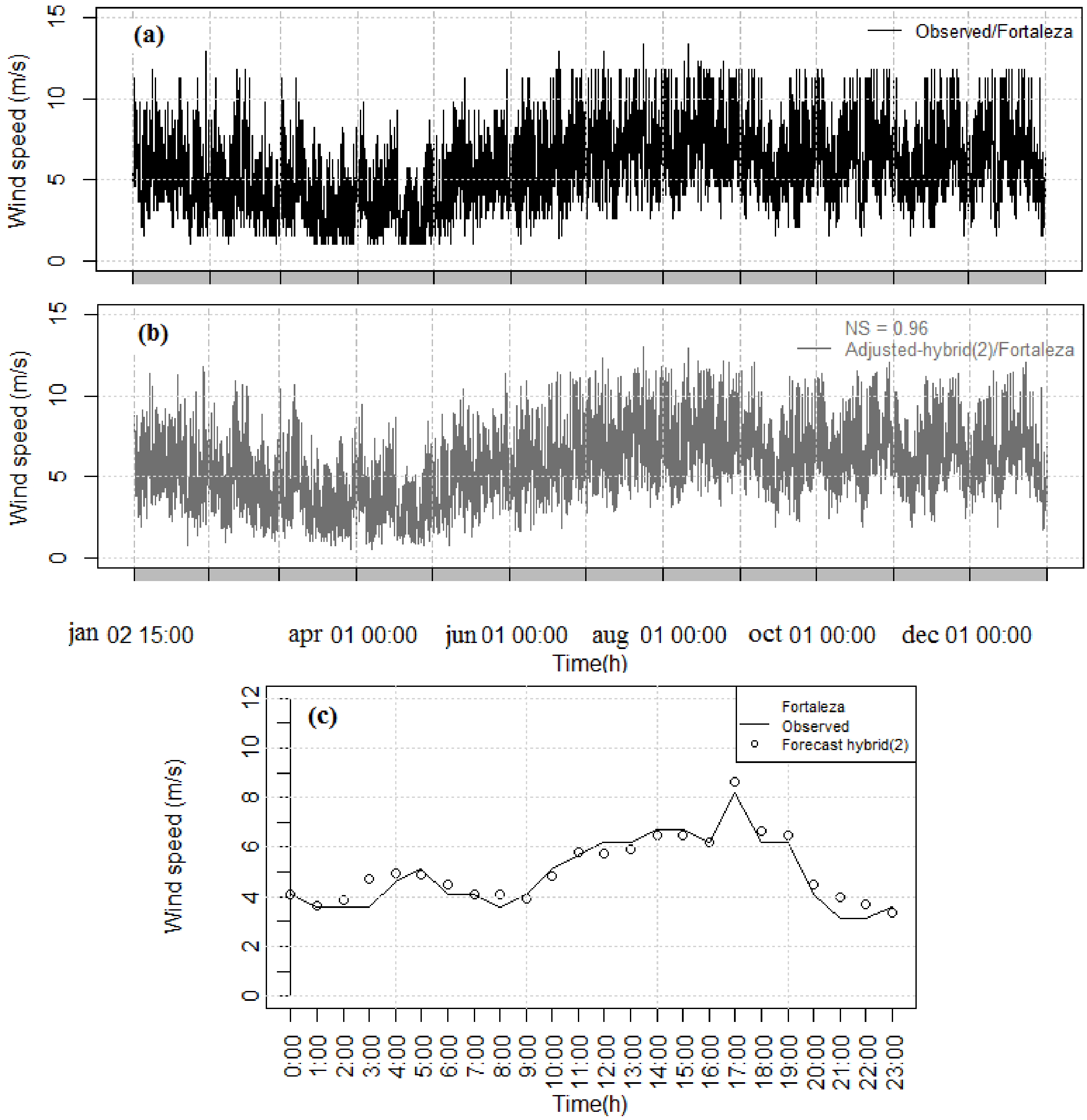


Fig. 6. Comparison between the time series in Fortaleza, in relation to the hours of the year 2015. (a) Observed (black line), (b) Adjustments by the hybrid(2) model (grey line) and (c) or the following day, Observed (black line) and forecast by the hybrid(2) model in (circle).

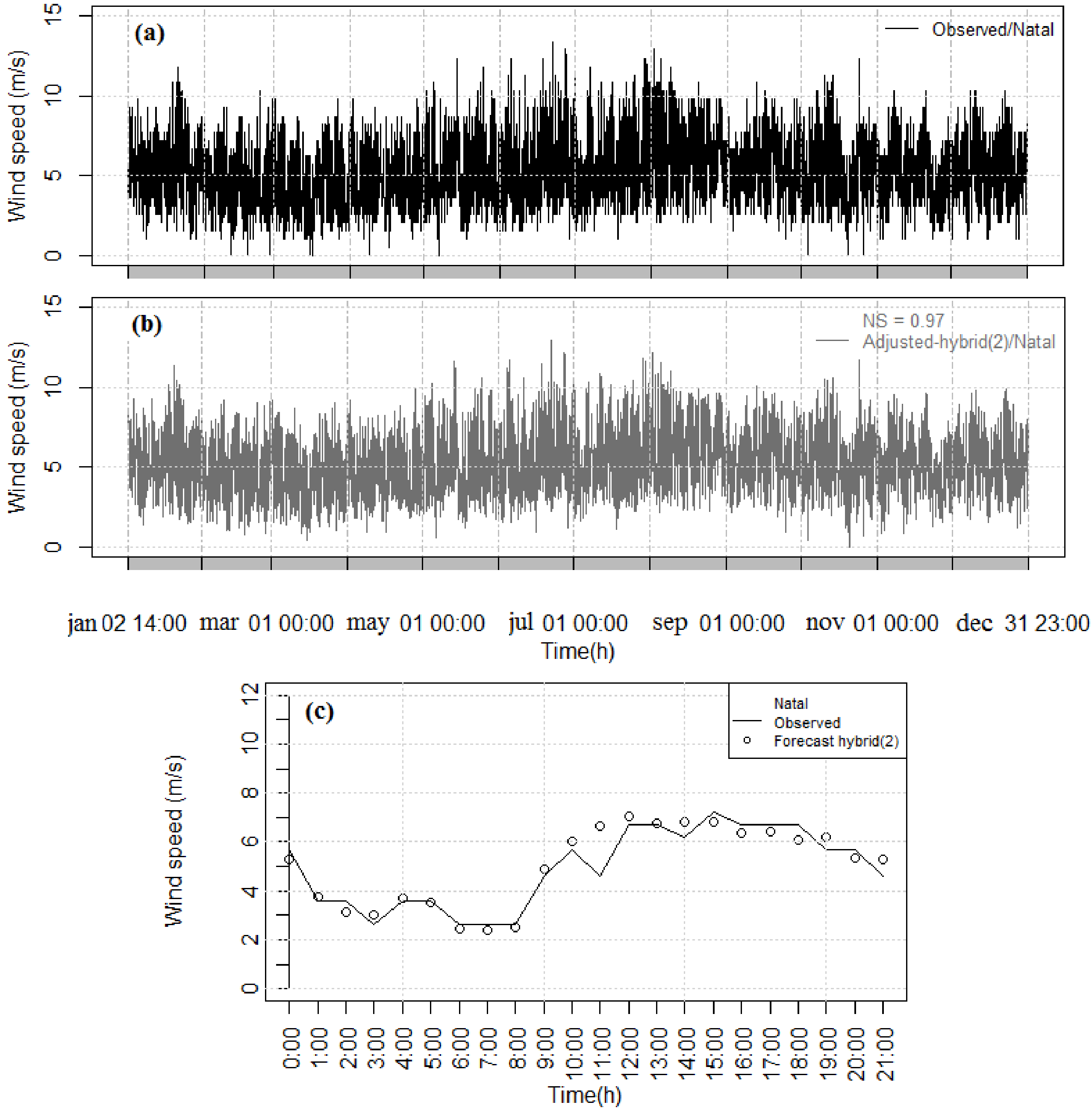


Fig. 7. Comparison between the time series in Natal, in relation to the hours of the year 2015. (a) Observed (black line), (b) Adjustments by the hybrid(2) model (grey line); and (c) For the following day, Observed (black line) and forecast by the hybrid(2) model in (circle).

Table 1

Presentation of tests that identify white noise in Box-Jenkins and Box-Tiao models.

Shapiro-Wilk Test
H_0: sample comes from a Normal population.
H_1: sample does not come from a Normal population.
Decision making: if the p-value is greater than α, i.e. $p > 0.05$ (do not reject H_0).
Durbin-Watson Test
H_0: the residues are independent.
H_1: the residues are not independent.
Decision making: if the p-value is greater than α, i.e. $p > 0.05$ (do not reject H_0).
Breusch-Pagan Test
H_0: the residues have homoscedasticity.
H_1: the residues have heteroscedasticity.
Decision making: if the p-value is greater than α, i.e. $p > 0.05$ (do not reject H_0).

Table 2

Main characteristics of ANNs produced in R software.

Fortaleza – Hour					
ANN model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
36	1	18	24	685	20
ARIMA+ANN – hybrid(1) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
32	1	16	24	545	20
ARIMAX+ANN – hybrid(2) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
39	1	20	24	821	20
HW+ANN – hybrid(3) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
38	1	20	24	801	20
Natal – Hour					
ANN model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
28	1	14	24	421	20
ARIMA+ANN – hybrid(1) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
36	1	18	24	685	20

ARIMAX+ANN – hybrid(2) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
38	1	20	24	801	20
HW+ANN – hybrid(3) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
36	1	18	24	685	20
Fortaleza – Month					
ANN model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
13	1	7	12	106	20
ARIMA+ANN – hybrid(1) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
12	1	6	12	85	20
ARIMAX+ANN – hybrid(2) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
13	1	7	12	106	20
HW+ANN – hybrid(3) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
13	1	7	12	106	20
Natal – Month					
ANN model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
13	1	7	12	106	20
ARIMA+ANN – hybrid(1) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
12	1	6	12	85	20
ARIMAX+ANN – hybrid(2) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
12	1	6	12	85	20
HW+ANN – hybrid(3) model NNAR(p,P,k)$_m$					
p	***P***	***K***	***m***	**Weights**	**Average of network**
12	1	6	12	85	20

Table 3

Tests applied to the residues to identify if the assumptions of the ARIMA and ARIMAX models were met.

	Month		
Local/ARIMA	**KS**	**DW**	**BP**
Fortaleza	0.499	0.771	0.054
Natal	0.514	0.466	0.575
Local/ARIMAX	**KS**	**DW**	**BP**
Fortaleza	0.970	0.615	0.551
Natal	0.987	0.799	0.509
	Hour		
Local/ARIMA	**KS**	**DW**	**BP**
Fortaleza	2.10^{-16}	0.671	0.051
Natal	5.10^{-7}	0.501	0.420
Local/ARIMAX	**KS**	**DW**	**BP**
Fortaleza	4.10^{-9}	0.489	0.103
Natal	2.10^{-10}	0.516	0.022

Table 4

Statistical analysis of errors used to identify the accuracy of the proposed models for the monthly data.

Erro – ARIMA	**Fortaleza**	**Natal**	**Erro – ARIMAX**	**Fortaleza**	**Natal**
MAE (m/s)	0.44	0.39	MAE (m/s)	0.37	0.37
RMSE (m/s)	0.57	0.50	RMSE (m/s)	0.48	0.45
MAPE (%)	9.89	9.10	MAPE (%)	8.48	8.47
Erro – HW	**Fortaleza**	**Natal**	**Erro – ANN**	**Fortaleza**	**Natal**
MAE (m/s)	0.45	0.40	MAE (m/s)	0.46	0.36
RMSE (m/s)	0.57	0.50	RMSE (m/s)	0.66	0.53
MAPE (%)	10.10	9.44	MAPE (%)	10.29	8.14
Erro – hybrid(1)	**Fortaleza**	**Natal**	**Erro – hybrid(2)**	**Fortaleza**	**Natal**
MAE (m/s)	0.39	0.32	MAE (m/s)	0.36	0.31
RMSE (m/s)	0.49	0.42	RMSE (m/s)	0.46	0,38
MAPE (%)	8.68	7.53	MAPE (%)	8.03	7.21
Erro – hybrid(3)	**Fortaleza**	**Natal**			
MAE (m/s)	0.44	0.37			
RMSE (m/s)	0.56	0.45			
MAPE (%)	10.0	8.70			

Table 5

Statistical analysis of errors to identify the accuracy of the proposed models for hourly data.

Erro – ARIMA	**Fortaleza**	**Natal**	**Erro – ARIMAX**	**Fortaleza**	**Natal**
MAE (m/s)	0.95	0.86	MAE (m/s)	0.89	0.81
RMSE (m/s)	1.27	1.17	RMSE (m/s)	1.19	1.08
MAPE (%)	18.75	19.42	MAPE (%)	18.84	17.34
Erro – HW	**Fortaleza**	**Natal**	**Erro – ANN**	**Fortaleza**	**Natal**
MAE (m/s)	1.19	1.03	MAE (m/s)	0.75	0.71
RMSE (m/s)	1.55	1.38	RMSE (m/s)	1.01	0.96
MAPE (%)	25.71	27.37	MAPE (%)	14.93	15.30
Erro – hybrid(1)	**Fortaleza**	**Natal**	**Erro – hybrid(2)**	**Fortaleza**	**Natal**
MAE (m/s)	0.36	0.36	MAE (m/s)	0.35	0.32
RMSE (m/s)	0.46	0.46	RMSE (m/s)	0.46	0.41
MAPE (%)	7.25	8.56	MAPE (%)	7.20	7.22
Erro – hybrid(3)	**Fortaleza**	**Natal**			
MAE (m/s)	0.73	0.66			
RMSE (m/s)	0.97	0.89			
MAPE (%)	15.10	14.50			

Highlights - Energy

1. The objective is to provide forecasts of wind speed in regions of the Brazilian.
2. Predictions from multivariate statistical models.
3. Neural Networks improve temporal series predictions combined with statistical models.
4. The models provide forecasts with better accuracy, when compared to others.
5. Wind speed forecast models can assist decision makers in the wind sector.