 doi:10.20944/preprints201807.0501.v1

Article

Short-Term Forecast of Wind Speed through Mathematical Models

Moniki Ferreira [1], Alexandre Santos [2] and Paulo Lucio [3,*]

[1] Postgraduate Program in Climate Sciences, Department of Atmospheric Sciences and Climate, Federal University of Rio Grande do Norte, Brazil; monikidara@gmail.com
[2] Applied Research Unit, Centre for Gas Technology and Renewable Energy (UNPA/CTGAS-ER), Natal, Brazil; alexandretorres@ctgas-er.com.br
[3] Department of Atmospheric Sciences and Climate, Federal University of Rio Grande do Norte, Brazil
* Correspondence: pslucio@ccet.ufrn.br; Tel.: +55-84-999536920

Abstract: The predictability of wind information in a given location is essential for the evaluation of a wind power project. Predicting wind speed accurately improves the planning of wind power generation, reducing costs and improving the use of resources. This paper seeks to predict the mean hourly wind speed in anemometric towers (at a height of 50 meters) at two locations: a coastal region and one with complex terrain characteristics. To this end, the Holt-Winters (HW), Artificial Neural Networks (ANN) and Hybrid time-series models were used. Observational data evaluated by the Modern-Era Retrospective analysis for Research and Applications-Version 2 (MERRA-2) reanalysis at the same height of the towers. The results show that the hybrid model had a better performance in relation to the others, including when compared to the evaluation with MERRA-2. For example, in terms of statistical residuals, RMSE and MAE were 0.91 and 0.62 m/s, respectively. As such, the hybrid models are a good method to forecast wind speed data for wind generation.

Keywords: wind speed; ANN model; hybrid model

---

1. Introduction

In recent years researches from several nations have warmed of the possible consequences of global warming across the planet, encouraging the use of font's renewable energy resources, for examples wind energy, is one of the strategies used to mitigate greenhouse gases from human activities in the atmosphere [1,2]. The wind energy capacity installed at the end of 2016 in Brazil was approximately of 11 GW, with estimate for 2020 will have around 18 GW of installed wind energy capacity, which will contribute to the country's energy security [3].

Most research in the field of wind is being applied to wind power (conversion of kinetic energy into electrical energy) for the planning and development of wind farms (set of wind turbines) [4]. Wind power is a renewable and clean source of energy, available with variable behavior in each part of the world because of the different climatic and geographical characteristics of each region [4]. Wind is caused by solar energy hitting the earth's surface and atmosphere, which together with the planetary rotation results in an uneven heating of the atmosphere and, consequently, differentials in atmospheric pressures [5].

The predictability of wind information at a particular location is essential for the assessment of a wind energy exploitation project. In this context, accurately predicting the wind speed means improving the planning of wind power generating facilities, thus reducing mistakes and economic costs [4].

The development of Artificial Intelligence (AI) with various methods for forecast wind speed was also created. The methods were: Artificial Neural Network (ANN); adaptive neuro-fuzzy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

inference system (ANFIS); fuzz logic methods; support vector machine (SVM), υ–support vector machine (υ-SVM) and neuro-fuzzy network. Accurate forecasts of short-term wind speed and generation of energy are essential for the effective operation of a wind farm. The short-term forecast of wind speed is also critical to the operation of wind turbines so that dynamic controls can be accomplished to increase the energy conversion efficiency [6]. Reference [6] proposed a method using improved radial basis function neural network-based model with an error feedback scheme (IRBFNN-EF) for forecasting short-term wind speed and power of a wind farm. Results showed that the proposed model IRBFNN-EF leads to better accuracy for forecasting wind speed and wind power, compared with those obtained by four other artificial neural network-based forecasting methods. Reference [7] compared the performances of autoregressive integrated moving average (ARIMA), Auto-Regressive Integrated Moving Average with Exogenous (ARIMAX), ANN and hybrids (ARIMA-ANN and ARIMAX-ANN) models in wind speed predictions in the Brazilian Northeast region (Fortaleza, Natal and Paraiba). The hybrid model proposed in this study was efficient in reducing statistical errors, especially when compared to traditional models (ARIMA, ARIMAX and ANN), with lowest percentage error between the observed and the adjusted series, of only about 8%.

Some recent studies have been developed about the short-term predictability of wind speeds with the use of dynamic, mathematical and statistical tools using Numerical Weather Prediction (NWP) [8,9], stochastic [10] and hybrid [3,7,11] models. Reference [12] investigated the prediction of hourly wind speeds at 30 m above ground with the atmospheric model WRF (Weather Research Forecasting) for the State of Alagoas-Brazil. The difficulties encountered in this study included: (i) forecasting extreme values; (ii) minimums and maximums and (iii) forecasts in rainy periods. In the study, the same difficulties were also found for the forecast of wind speed for wind power generation through the outputs of mesoscale weather models in the Northeast Region of Brazil (NEB). Other feasible wind speed studies can be found in [13], which proposed methods to fill gaps in wind speed data - located in Rio Grande do Norte (Northeast Region of Brazil) - reducing the propagation of residuals in the results. Reference [10] studied and developed forecasting models, of short-term wind speed. The methods of exponential smoothing, in particular the method Holt-Winters (stochastic model) and its variations, are suitable in this context because of its high adaptability and robustness for apply the methodology in the wind speed data anemometric tower in city of São João do Cariri (State of Paraíba-Brazil), which will be compared with models: persistence and neuro-fuzz (ANFIS). Results showed that the proposed Holt-Winters additive model were satisfactory, compared with those obtained by two tested models (persistence and neuro-fuzzy) in the same wind speed time-series for short-term forecasting. Reference [14] proposed a hybrid model that consists of the EWT (Empirical Wavelet Transform), CSA (Coupled Simulated Annealing) and LSSVM (Least Square Support Vector Machine) for enhancing the accuracy of short-term wind speed forecasting. Results suggest that the developed forecasting method better compared with those other models, which indicates that the hybrid model exhibits stronger forecasting ability.

The spatial and temporal variability of the wind is difficult to simulate with precision. This is a result of the heterogeneity of regions regarding such factors as: surface roughness, vegetation variability and soil use and occupation [15]. In addition, several meteorological and climatic phenomena may influence the atmospheric dynamics of the NEB region [16]. Systematic errors occur with a certain frequency in the simulations of dynamic models, which means models with greater accuracy need to be developed for the short-term wind speed forecasting of a particular site [17]. Reference [18] presented a system consists of a numerical weather prediction model and ANN of a novel day-ahead wind power forecasting in China, in addition Kalman filter integrated to reduce the systematic errors in wind speed from WRF and enhance the forecasting accuracy. Results showed the Normalized Root Mean Square Error (NRMSE) has a month average value of 16.47%, which is an acceptable error margin. To reduce these systematic errors in the outputs of an NWP, lots of approaches have been derived from model output statistics (MOS) based on statistical methods [18].

This type of statistical relationship could be modeled by various methods for wind speed variable, including the ANNs [6,19,20], Kalman filter [18], ARIMA [7] and hybrid method [11,14,21,22].

Wind energy is considered to be technically feasible when its power density is greater than or equal to 500 W/m², for a height equal to or exceeding 50 m above the ground, which requires a minimum wind speed between 7-8 m/s. Only 13% of the earth's surface has average wind speeds greater than or equal to 7 m/s at the height of 50 m above ground [23,24].

In this context, this study seeks to present different methodologies that may assist in this regard, while at the same time proposing a model that has less error in relation to the dynamic models (more systematic errors) in the short term wind speed forecasts at a height of 50 meters above ground in two localities with different geographic and climatic characteristics in the NEB region. One of these locations can be found in the heartland of the continent with a complex terrain (Anemometric Tower 1 - TA01) and the other is located in the coastal region with a flat terrain (Anemometric Tower 2 - TA02). The objective of this study is to present mathematical and statistical models as suitable tools for the forecasting of wind speed variability during the diurnal cycle (24 hours), which could be used as an early warning system for wind power generation in the same time scale.

The content of this paper is organized as follows: Section 1 reviews the current status of wind speed variable for short-term forecasting with applicability in wind power; Section 2 describes the regions of study, dataset and theory of the Holt-Winters, ANN, hybrid models and its working principles; Section 3 analyzes the errors of the short-term predicted results; Section 4 description of discussion to explore the significance of the study; Section 5 the conclusions of the study.

2. Materials and Methods

In this section we comment on the data of the regions of study, as well as on the forecasting models. The measurement data is used to identify the accuracy of the forecast calculated by the mathematical models, by comparing the output of each model to the time series observed. All the calculations produced in this paper as well as the graphical part were executed with the software R (https://www.r-project.org/).

2.1. Region of study and dataset

The measured wind data were obtained from the TA-01 and TA-02 towers, the first located in Belo Jardim - state of Pernambuco, and the second in Camocim - state of Ceará, all located in the Northeast of Brazil, as shown Figure 1. The Belo Jardim data were obtained through the Sistema de Organização Nacional de dados Ambientais (System for the National Organization of Environmental Data, SONDA, http://sonda.ccst.inpe.br/) and the Camocim data through the Infrastructure department of the state of Ceará (SEINFRA/CE, www.seinfra.ce.gov.br/).

The anemometric sensors were installed at heights of 25 and 50 meters from the ground, taking measurements every 10 minutes, within the period from October 01 to December 31, 2004, selected for this study in both locations. The wind speed and direction sensor of the TA-01 tower is the Wind Monitor-MA model 05106 (R. M. Young Company) - it takes measurements of horizontal speeds between 0 to 60 m/s and resists gusts of up to 100 m/s. For the TA-02 tower, the data was recorded with a computerized anemometer; model NRG 9200Plus, manufactured by NRG Systems Inc. The temporal series were integrated in an interval of 1 hour at a height of 50 m above ground level - agl (TA-01 and TA-02) to validate the models evaluated in this study.

Table 1 presents the local characteristics and the geographical coordinates of the used wind towers.

 doi:10.20944/preprints201807.0501.v1

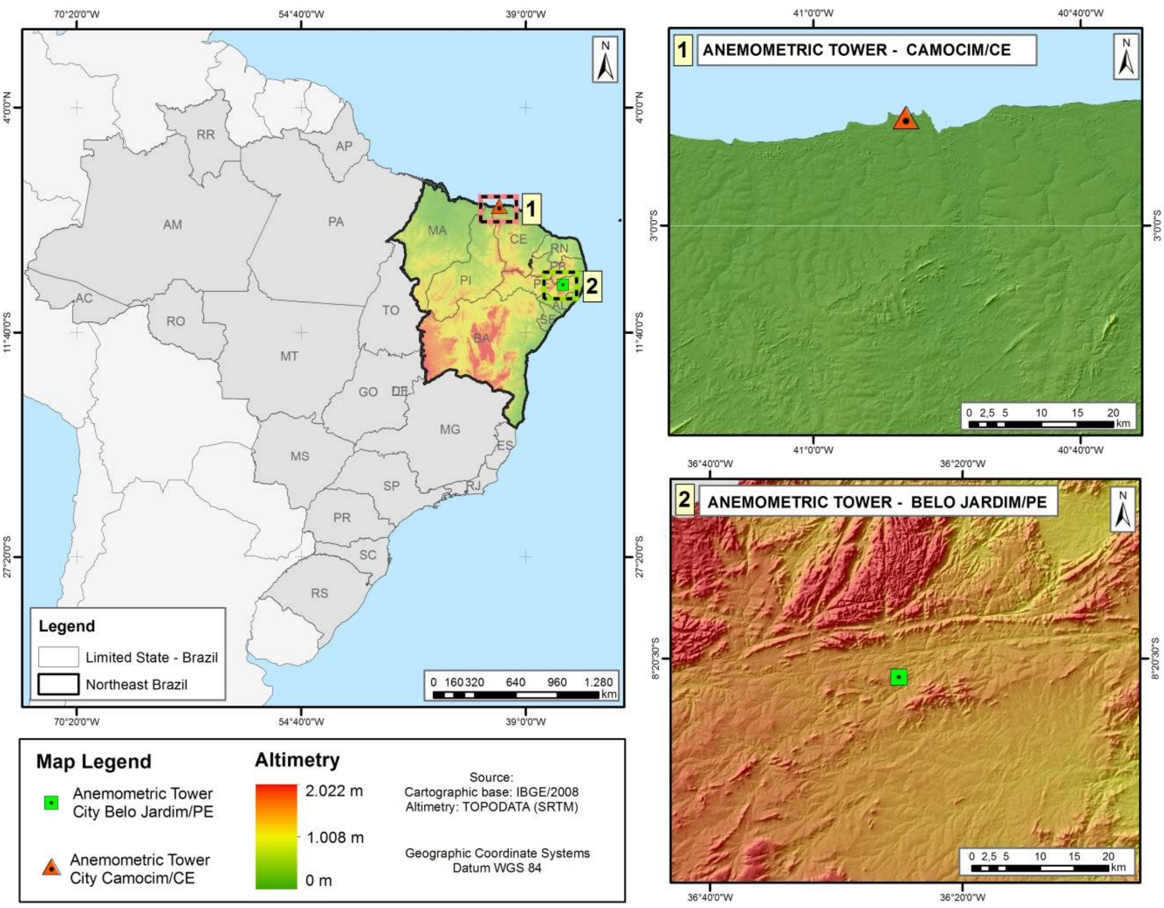

Figure 1. Location of each anemometric tower (TA) used in the presented study in Northeast Brazil.

Table 1. Local characteristics and geographical location of the wind towers.

City/State	Altitude (m)	Terrain	Vegetation	Geographic coordinates
Belo Jardim/PE	718	Plateau	Caatinga	8°22′S/36°25′W
Camocim/CE	8	Lowland	Caatinga	2°51′56,7°S/40°53′09,2°W

2.2. Methods

The forecasts for 24 hours were made with the Holt-Winters, Artificial Neural Networks (ANN) and hybrid models, applying them to the temporal series of the TA-01 and TA-02 towers with the use of the R software. The forecasting methodology consisted in making a division in the series in three steps: *(i)* the first step used 30 days of data observed in the period from October 01 to 30, 2004, to predict the daily wind speed cycle for the day of October 31, 2004; *(ii)* The second step used 60 days of data observed in the period from October 01 to November 29, 2004, to predict the day of November 30, 2004; and *(iii)* the third and last step used 90 days of data observed in the period from October 01 to December 29, 2004, to forecast the day of December 30, 2004.

There are numerous techniques to forecast wind speeds, such as the NWP methods, statistical methods, ANN and hybrid approaches. The NWP methods may be the most accurate technique for forecasting in the short term. In general, however, statistical methods, ANNs or hybrid approaches have smaller errors in short term forecasts [22].

2.2.1. Holt-Winters Model

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 [25]. In year 1960, Winters extended the Holt model, which included a new equation that predicts the behavior of the data's seasonal component, transforming at Holt-Winters model [3,26].

 doi:10.20944/preprints201807.0501.v1

Holt-Winters is one of the most used methods for the prediction of meteorological variables due to its simplicity, low operating costs, good accuracy and ability to make automatic adjustments and quick changes to the temporal series. This model has the following smoothing coefficients: level, linear trend, seasonal factor and an unpredictable residual element called random error. The exponential adjustment method, also called "exponential smoothing", is used for the estimation of these factors. The name "smoothing" comes from the fact that after its reduction to structural components, a series will have a smaller number of abrupt variations, revealing a smoother behavior. The term "exponential" is used because the smoothing processes involve the weighted arithmetic mean, where the weights decrease exponentially as you progress into the past [10].

The prediction equations are allocated in two ways: additively or multiplicatively, according to the nature of the series. To calculate the forecasts of future values of a series, the level and the trend of the series in the current period have to be estimated, just as the seasonal factor values corresponding to the last period of seasonality. These estimates are performed through the following equations:

Additive Seasonality – The equation referring to the exponential smoothing method with seasonality and linear tendency, with seasonal component being treated in an additive way, is represented as follows:

$$y_{t+h} = (a_t + h.b_t) + s_{(t\text{-}p+1+1(h-1)mod\ p)} \tag{1}$$

where: a_t is level of the series, whose unit in this particular work is of m/s, shows how the expected time series evolves over time; b_t is tendency, unit here is m/s, this relates to the fact that the predicted time series can have increasing or decreasing motions in different time intervals; s_t is 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, h is forecast for period and p is seasonal period with h = 1,2,3,…,n (horizontal forecast).

where a_t , b_t and s_t are given by:

$$a_t = \alpha(y_t\text{-}s_{t\text{-}p}) + (1-\alpha)(a_{t\text{-}1} + b_{t-1}) \tag{2}$$

$$b_t = \beta(a_t\text{-}a_{t\text{-}1}) + (1-\beta)b_{t-1} \tag{3}$$

$$s_t = \gamma(y_t\text{-}a_t) + (1-\gamma)s_{t\text{-}p} \tag{4}$$

Multiplicative Seasonality: – Is similar to the Holt-Winters additive. Holt-Winters Multiplicative method also calculates exponentially smoothed values for level, trend, and seasonal adjustment to the forecast. This method is best for data with trend and with seasonality that increases over time. It results in a curved forecast that reproduces the seasonal changes in the data. The multiplicative Holt-Winters prediction function is:

$$y_{t+h} = (a_t + h.b_t)s_{(t\text{-}p+1+1(h-1)mod\ p)} \tag{5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

where a_t , b_t and s_t are given by:

$$a_t = \alpha\left(\frac{y_t}{s_{t\text{-}p}}\right) + (1-\alpha)(a_{t\text{-}1} + b_{t-1}) \tag{6}$$

$$b_t = \beta(a_t\text{-}a_{t\text{-}1}) + (1-\beta)b_{t-1} \tag{7}$$

$$s_t = \gamma\left(\frac{y_t}{a_t}\right) + (1-\gamma)s_{t\text{-}p} \tag{8}$$

where α, β and γ are damping constants. At the end of each period t, the estimate of the "step" (trend) and the seasonal component are given by b_t and s_t, respectively. The level component, on the other hand, is denoted by a_t.

The additive Holt-Winters model was used in the Camocim/CE data for the temporal series of 30, 60 and 90 days and the Belo Jardim/PE data for the series of 30 and 60 days. For the Belo Jardim data set with 90 days, therefore, the multiplicative Holt-Winters model was used, since the additive prediction produced negative values.

2.2.2. Artificial Neural Network model

Artificial Neural Networks (ANNs) are computational models based on the neural structure of intelligent organisms [7]. Their behavior emerges from the interactions between processing units, which compute certain mathematical (usually non-linear) functions. These processing neurons can be distributed in one or more layers and are linked by a large number of connections, which store the knowledge in the model and weigh each input received over the network.

An ANN can be thought of as a network of "neurons" organized in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may be intermediate layers that contain "hidden neurons". The predictors (or inputs y_{t+i}) form the lower layer, and the predictions (or outputs y_{t+h}) form the upper layer. There may be intermediate layers containing hidden neurons [7]. Figure 2 shows an example of an ANN structure with 4 inputs and 1 hidden layer. The coefficients related to the predictors are called "weights" and commonly represented by w_i. The weights are selected through a "learning algorithm." This study used the backpropagation algorithm, which is based on the backpropagation of errors to adjust the weights of the intermediate layers, which minimizes the error between the predicted and observed temporal series.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

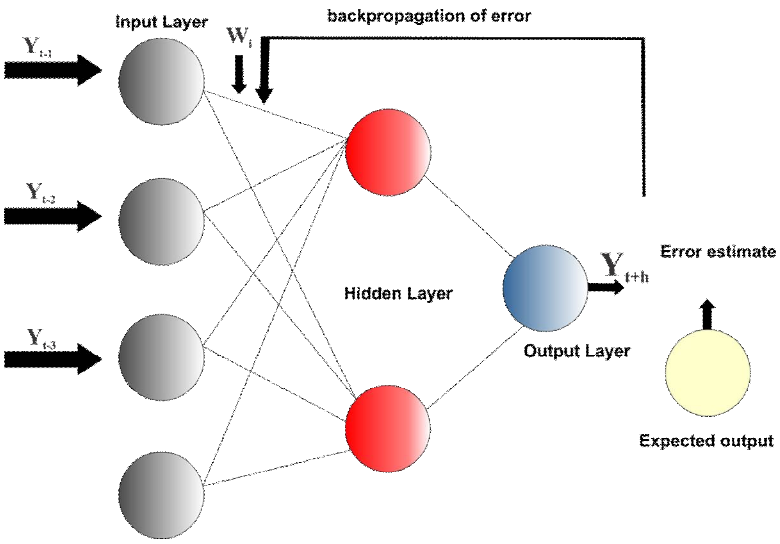

Figure 2. Example of an ANN structure with 4 inputs layers, 1 hidden layer and 1 output layer.

In terms of use of the ANN by software R, the *forecast* package also allows for this possibility through the use of the function *nnetar()* [27]. This study used R's *forecast* package based on the *nnetar()* function, which in turn is symbolized by notation $NNAR(p, P, k)_m$, where p represents lagged inputs (the quantity of inputs), for example, y_{t-1}, y_{t-2}, y_{t-3},..., y_{t-p}, y_{t-m} , y_{t-2m}, y_{t-Pm} , refers to seasonal data, k represents the number of neurons in the hidden layer and P refers to seasonal data. This function $NNAR(p, P, k)_m$ model based on observed wind speed data, if the values of p and P are not specified, they are automatically selected [3]. For example, $NNAR(9,1,4)_{24}$ model has inputs y_{t-1}, y_{t-2}, y_{t-3}..., and y_{t-24}, and four neurons in the hidden layer.

The number of network to fit with different random starting weights was equivalent to 20 and is based on the backpropagation learning algorithm for dynamic processing. These are then averaged when producing forecasts.

2.2.3 Hybrid model

The hybrid model used the *hybridModel*(y_t, $models$) function of R's *forecastHybrid* package, with multiple adjustments of the individual models to generate ensemble forecasts. Where y_t is a numeric vector or time series (this work was wind speed time series) and models (combination of model types) by default, the models forecasts generated are from the *auto.arima(), ets(), thetam(), nnetar(), stlm(),* and *tbats()* functions, can be combined with equal weights, weights based on in-sample errors [28]. Cross validation for time series data and user-supplied models and forecasting functions is also supported to evaluate model accuracy.

The default setting of this package, which works well in most cases, allows for the combination of two to five models through the following arguments:

- *n.args*: adjusts a univariate neural network model using the *nnetar()* function by forecast package [27];

- *a.args*: fit best ARIMA model to univariate time series using the *auto.arima()* function by *forecast* package. The function conducts a search over possible model within the order constraints provided [27].

- *s.args*: is based on the *stlm* model, which combines the forecast with the adjusted seasonal decomposition in the series. The *stlm()* function takes a time series y, applies an STL decomposition, and models the seasonally adjusted data using the model passed or specified using method. It returns an object that includes the original STL decomposition and a time series model fitted to the seasonally adjusted data [27].;

- *t.args*: the *tbats()* function couples the exponential smoothing model with the Box-Cox transformation, ARMA and the seasonal and trend components [28].
- *e.args*: Exponential smoothing state space model. Based on the classification of methods as described in Hyndman et al [29]. The only required argument for *ets()* is the time series [27]. The model is chosen automatically if not specified.

In this work, three "*nst*" *models* were adjusted for wind speed forecasting short-term, which were coupled with the arguments: *n.args, s.args* and *t.args*. After adjustment of the models by *hybridModel()* function used the *forecast()* function with h (horizontal forecast) equal to 24 (24 hours) for short-tem forecast of wind speed at each time series proposed in this study.

2.2.4. Modern-Era Retrospective analysis for Research and Applications-Version 2 reanalysis dataset

The MERRA-2 reanalysis is generated by combining the data assimilation techniques of the models that use dynamic numerical prediction models with the data observed by the global meteorological observation network. The MERRA-2 reanalysis was introduced in the study to evaluate its results in relation to the data observed by the wind towers and predicted by the statistical and mathematical models proposed in this study. The MERRA-2 reanalysis is a freely-available product through the MDISC (Modeling and Assimilation Data and Information Services Center) portal and it is an update to the MERRA project [30].

The wind speed data at 50 m agl of MERRA-2 were obtained for the period from October 01 to December 31, 2004 to compare with predicted data from the mathematical and statistical models and observed data (TA-01 and TA-02).

2.3. Evaluation of the Models

The evaluation of the performance of the forecasting techniques was done through the Pearson correlation coefficient (r), the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) [31].

2.3.1. Pearson Correlation Coefficient

The Pearson Correlation Coefficient is calculated mathematically thus:

$$r = \frac{\sum_{i=1}^{n}(y_i\text{-}\bar{y})(x_i\text{-}\bar{x})}{\left(\sqrt{\sum_{i=1}^{n} y_i\text{-}\bar{y}}\right)\left(\sqrt{\sum_{i=1}^{n} x_i\text{-}\bar{x}}\right)} \tag{9}$$

where:

- r represents the linear correlation coefficient for a sample;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

- n is number of paired values considered;
- y_i is prediction value (wind speed - m/s);
- $\bar{y}$ is mean of prediction value;
- x_i is observed value (wind speed – m/s);
- $\bar{x}$ is mean of observed value.

The magnitude of the Pearson correlation coefficient ranges between the values of -1 and 1. This magnitude measures the "intensity" of the relationship between two variables. As such, a coefficient equal to 0.6 has a greater degree of linear dependence than one equal to 0.3. A coefficient with the value of zero indicates the total absence of a linear relationship between the variables and coefficients with the values 1 and -1 suggest a perfect linear dependence. It is therefore used to measure the correlation between the observed and modeled data in order to obtain the degree of linear relationship between both.

2.3.2. Mean Absolute Error

To evaluate the degree of dispersion between modeled (y_i) and observed (x_i) values, they are used as indexes that provide the performance of the model in relation to the observation. The closer the value is to zero, the greater the accuracy of the model and, consequently, the lower its error. The mean absolute error is a measure of the forecasting skill between the observed and predicted series. It is given as a module and is represented mathematically by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{y_i}\text{-}\mathrm{x_i}| \tag{10}$$

2.3.3. Root Mean Square Error

The RMSE was another error index used. It represents the difference between the prediction (y_i) and the observed value (x_i), presenting error values in the same dimensions as the analyzed variable (which in this case is m/s). It can be mathematically defined by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{y_i}\text{-}\mathrm{x_i})^2}{n}} \tag{11}$$

3. Results

 doi:10.20944/preprints201807.0501.v1

3.1. Available Observational data

These placed of study have anemometric towers in Camocim-Ceará and Belo Jardim-Pernambuco, located in Brazil. Figure 1 shows the topography of Camocim (coastal region) and Belo Jardim (complex topography) and the location of the study site. The data set used in paper was collected form an anemometer tower, which measuring height is 50 m agl. The available data are from 00:00, 01/10/2004 to 23:50, 31/12/2004, with a time interval of 10 minutes (Figures 3 and 5). Tables 2 and 3 presented the basic statistical description of the available data. Figures 4 and 6 displays the frequencies distributions of the available data and the probability distribution function by the two-parameter Weibull distribution k and c, where are of shape and scale parameters, respectively. Applying the maximum likelihood method (ML), the shape and scale parameters are 3.702/6.406 m/s and 3.096/11.337 m/s for TA-01 and TA-02, respectively.

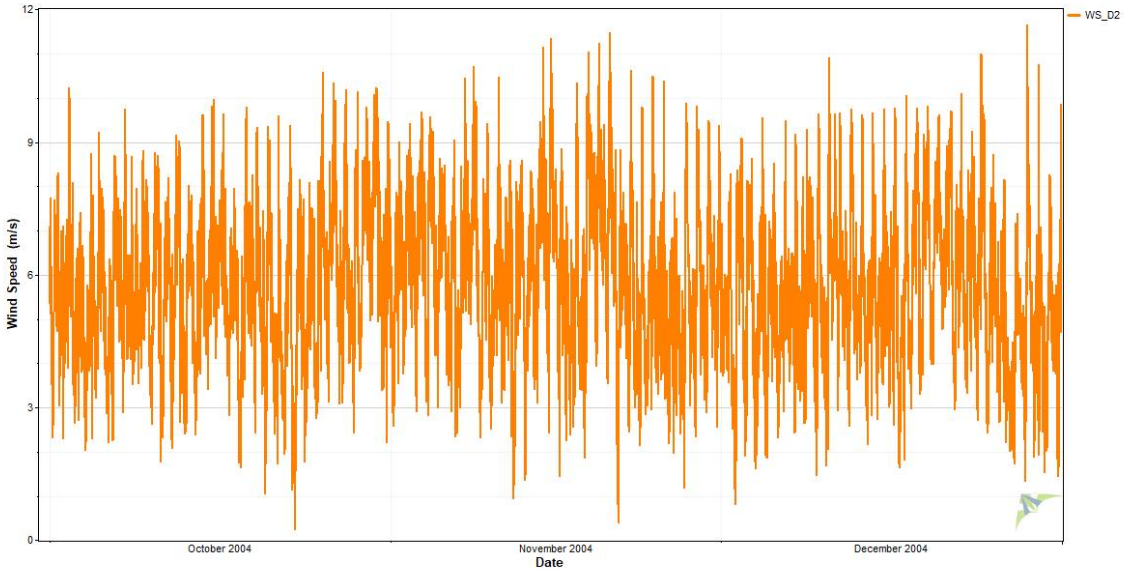

Figure 3. Time series of wind speed (interval of 10 minutes) measurements by the anemometer at a height of 50 m agl of the TA-01 for the period from 01/10/2004 to 31/12/2004.

Table 2. Statistical description of wind speed for the 3 months sampling period of the available data of Belo Jardim (TA-01).

Possible Data Points	Valid Data Points	Recovery Rate (%)	Mean (m/s)	Median (m/s)	Minimum (m/s)	Maximum (m/s)	Standard Deviation
13.249	13.249	100	5.780	5.730	0.250	11.650	1.725

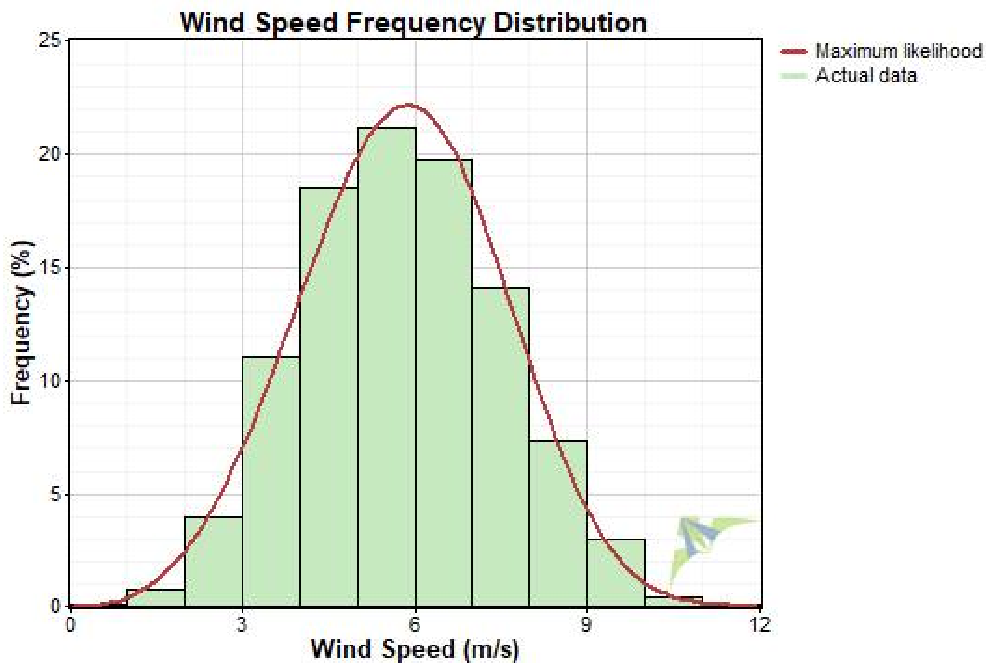

Figure 4. Histogram derived from the estimate Weibull probability density function (PDF) by maximum likelihood method compared with the histogram of the wind speed data (bar diagram) for the period of 3 months at wind site Belo Jardim (TA-01). Applying the ML method, the shape and scale parameters are $k = 3.702$ and $c = 6.406\ m/s$.

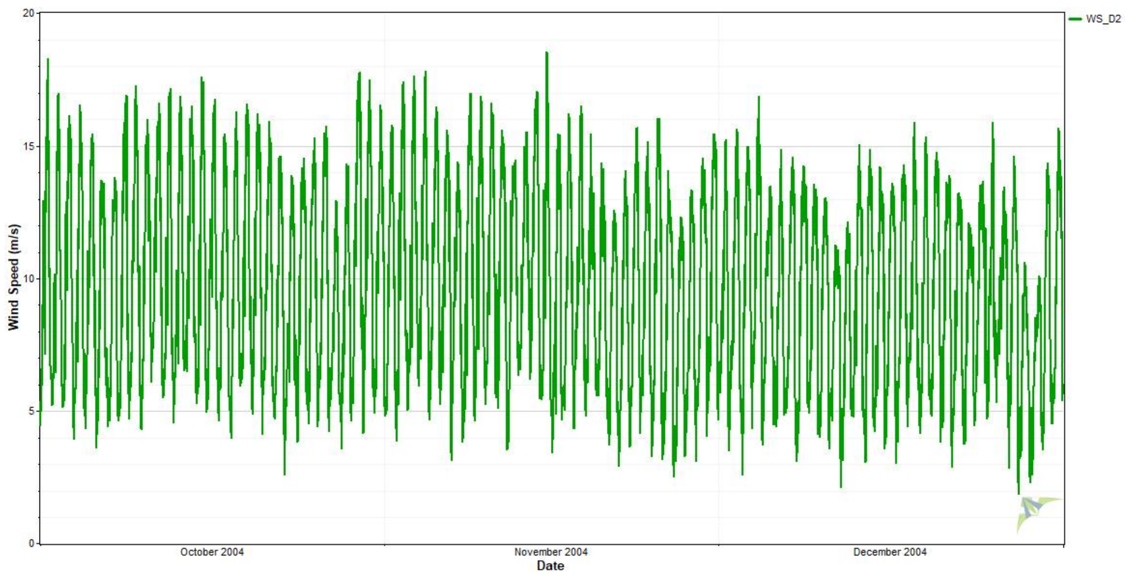

Figure 5. Time series of wind speed (interval of 10 minutes) measurements by the anemometer at a height of 50 m agl of the TA-02 (Camocim) for the period from 01/10/2004 to 31/12/2004.

Table 3. Statistical description of wind speed for the 3 months sampling period of the available data of Camocim (TA-02).

Possible Data Points	Valid Data Points	Recovery Rate (%)	Mean (m/s)	Median (m/s)	Minimum (m/s)	Maximum (m/s)	Standard Deviation
13.249	13.249	100	7.771	7.747	1.789	14.579	2.773

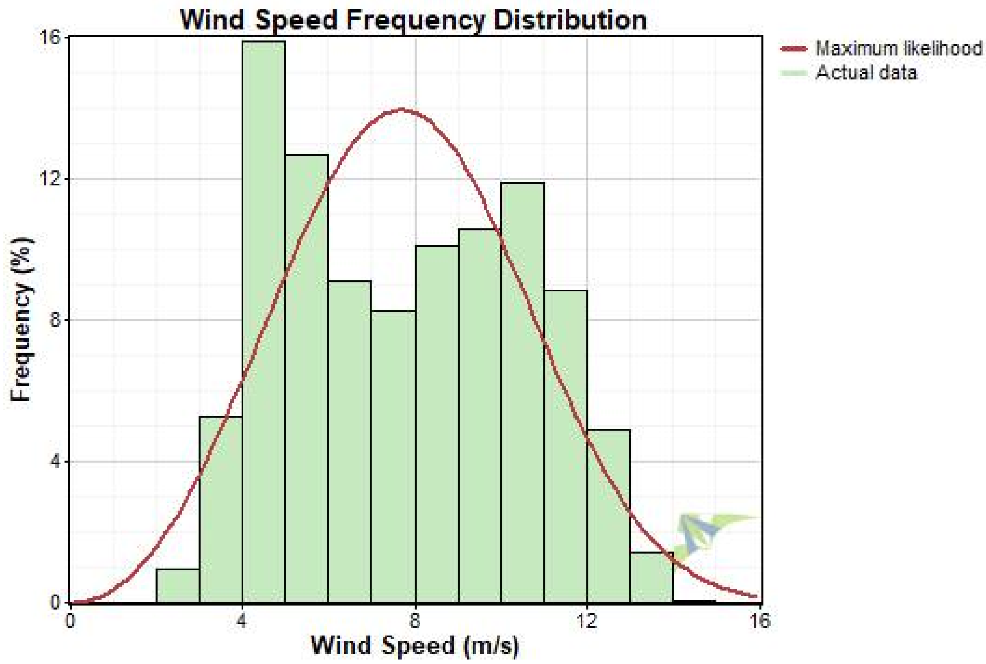

Figure 6. Histogram derived from the estimate Weibull probability density function (PDF) by maximum likelihood method compared with the histogram of the wind speed data (bar diagram) for the period of 3 months at wind site Camocim (TA-02). Applying the ML method, the shape and scale parameters are $k = 3.096$ and $c = 11.337\ m/s$.

3.2. Quantitative and Qualitative Analysis of the 24-Hour Forecast

3.2.1. Based on 30 days of data

Figures 7 and 8 present the forecast for the day 31/10/2004 using 30 days of observations (01/10/2004 to 30/10/2004) for the locations of the TA-01 and TA-02 towers, respectively. The Holt-Winters, Artificial Neural Network and Hybrid models and the MERRA-2 reanalysis are shown in Figures 3 and 4 along with the observations.

For the TA-01 tower (Figure 7), it is possible to identify that the models follow the diurnal behavior of the observations, but they do not adequately capture the extreme values for minimum variation at certain times, showing a smoother pattern. The MERRA-2 reanalysis overestimates the observations between the 2h to 19h range and at other times it fits with the observations. The ANN model underestimates the first hours and overestimates results from 5h to 19h. The Holt-Winters model approximates the observed data after 7h. The hybrid model was the one with the best fit with the observations in the whole series, satisfactorily capturing the wind speed variability during the diurnal cycle.

doi:10.20944/preprints201807.0501.v1

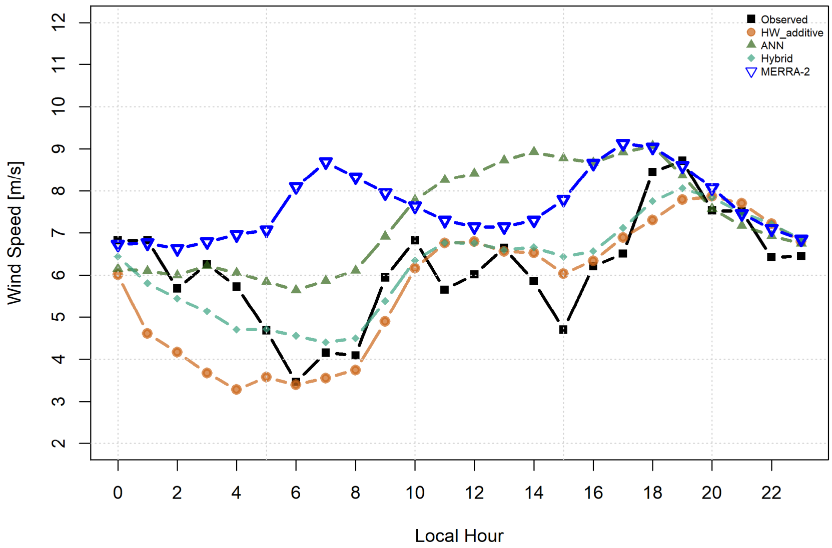

Figure 7. Comparison between the observed series, those predicted by the models and the MERRA-2 reanalysis for the TA-01 tower on 31/10/2004.

Table 4 shows the forecast skills between the models and reanalysis with observations for TA-01. The best results, according to the errors indexes, are presented by the hybrid model.

Table 4. Residues of the models and the reanalysis with the observations (TA-01) for 30 days of forecasting, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	1.14	1.73	**0.74**	2.07
MAE	0.90	1.35	**0.62**	1.53
r	0.73	0.46	**0.81**	0.07

For the TA-02 tower (Figure 8), one can see that the models can represent the behavior of the observations, unlike the MERRA-2 reanalysis. The models still have difficulty capturing the extreme values (minimum and maximum), except for the ANN and hybrid models, but, in general, the diurnal variability is well represented. The Holt-Winters model underestimates most observations, approaching them in the first hours and also near 6:00 and 18:00 h. The ANN model is able to identify extreme variability in the range of 6:00 and 9:00am. In the remaining time, the ANN model can track the variability of the wind for the period of one day. The hybrid model fits the observations properly, capturing the extreme value signals in the predicted range of 6h to 9h.

Table 5 shows the forecast skills between the models and the MERRA-2 reanalysis data with the observations. The Pearson correlation coefficient reveals that the values had excellent correlations (0.96). The MERRA-2 data had a low correlation coefficient (-0.10). The residuals were reasonable, with the result of the hybrid model having the lowest errors. The MERRA-2 reanalysis, with low correlation, obtained larger errors (RMSE: 3.9 m/s and MAE: 3.25 m/s).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

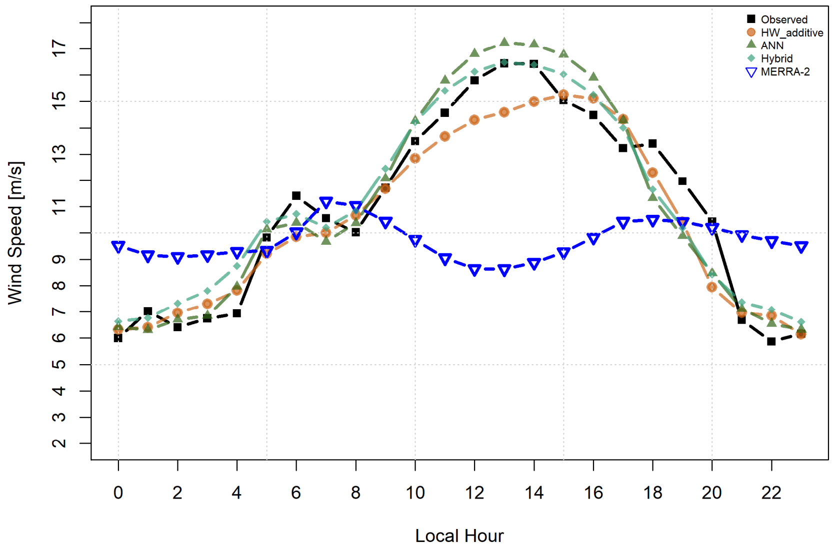

Figure 8. Comparison between the observed series, those predicted by the models and the MERRA-2 reanalysis for the TA-02 tower on 31/10/2004.

Table 5. Residuals of the models and the reanalysis with the observations (TA-02) for 30 days of forecasting, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	1.06	1.07	**1.00**	3.90
MAE	0.88	0.90	**0.84**	3.25
r	0.96	0.96	**0.96**	-0.10

3.2.2. Based on 60 days of data

Figures 9 and 10 present the forecast for the day 31/11/2004 using 60 days of the temporal series (01/10/2004 to 29/11/2004) for the measurements of the TA-01 and TA-02 towers, respectively.

For the TA-02 tower, the ANN model can be seen to follow the variability of the observed wind speed data for up to 8 predicted hours. The model is unable to capture the small wind speed variations that occur during the day, smoothing its results. The stochastic Holt-Winters model overestimates after 11 predicted hours, but it follows the trends of the observations until 18 predicted hours, capturing the variability of wind. The hybrid model, therefore, showed a smoothed pattern of the predicted wind speed with respect to the observations, and even so it underestimated results most of the time. MERRA-2 overestimates results in relation to the observed data in the temporal series, representing the behavior only in the first hours (until 6am) and last hours (after 21pm) predicted for the day 30/11/2004.

Table 6 shows the forecast skills of the models and the reanalysis with the observations. The lowest values can be found in the hybrid model, whose RMSE and MAE were 0.91 m/s and 0.80 m/s. The largest results were found in the Holt-Winters model (RMSE: 2.09 m/s and MAE: 1.68 m/s). The correlation coefficient was also high in the hybrid model and low in the MERRA-2 reanalysis.

doi:10.20944/preprints201807.0501.v1

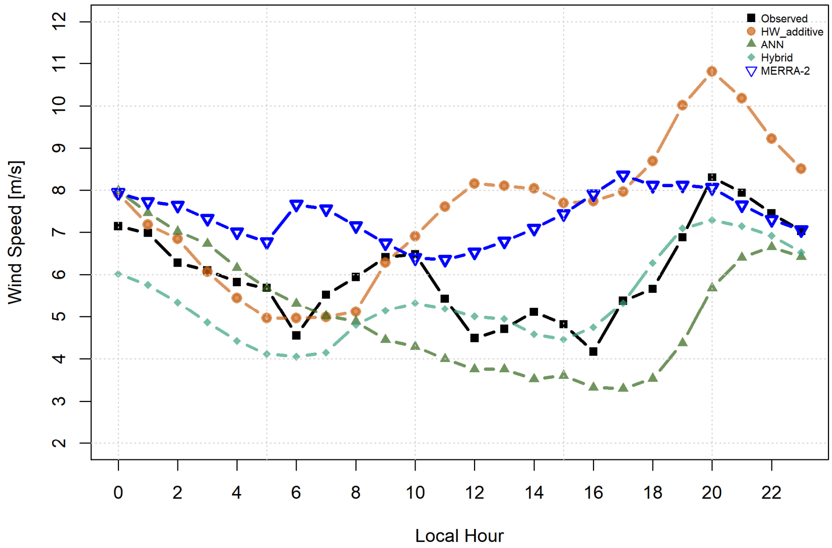

Figure 9. Comparison between the observed series, those predicted by the models and the MERRA-2 reanalysis for the TA-01 tower on 31/11/2004.

Table 6. Residuals of the models and the reanalysis with the observations (TA-01) for 60 days of forecasting, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	2.09	1.39	**0.91**	1.75
MAE	1.68	1.19	**0.80**	1.41
r	0.47	0.66	**0.80**	0.23

For the TA-02 tower (Figure 10), one can see that the models follow the observed diurnal wind speed data, but they underestimate and overestimate them at certain specific times. The ANN model underestimates the first hours and overestimates the predicted wind speeds from 11:00am to 18:00pm. The Holt-Winters model underestimates most of the period and follows a behavior consistent with the observed data between 8:00am and 20:00pm. The hybrid model approximates the observed data most of the time, mainly between 5:00am and 8:00am.

Table 7 shows the forecast skills of the models and the MERRA-2 reanalysis with the wind speed observations. The lowest values can be found in the hybrid model, whose RMSE and MAE were 1.12 m/s and 0.81 m/s. The highest values were found in the MERRA-2 reanalysis, with RMSE and MAE values of 3.60 m/s and 2.92 m/s. The correlation coefficients were higher for all models in this period.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

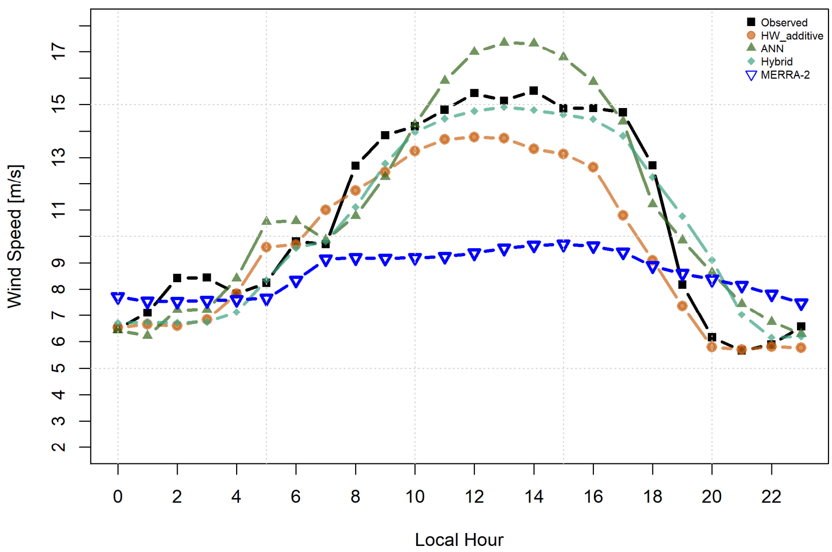

Figure 10. Comparison between the observed series, those predicted by the models and the MERRA-2 reanalysis for the TA-02 tower on 31/11/2004.

Table 7. Residuals of the models and the reanalysis with the observations (TA-02) for 60 days of forecasting, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	1.61	1.41	**1.12**	3.60
MAE	1.25	1.22	**0.81**	2.92
r	0.95	0.94	**0.95**	0.89

3.2.3. Based on 90 days of data

Figures 11 and 12 present the forecast for the day 31/12/2004 using 90 days of observations (01/10/2004 to 29/12/2004) for the locations of the TA-01 and TA-02 towers, respectively.

For the TA-01 tower (Figure 11), the hybrid model underestimates the data observed in the early hours and comes close to the observations between 11:00am and 14:00pm. The ANN model overestimates the observations between the predicted hours of 7:00am and 20:00pm. The Holt-Winters model underestimates the observed data until 15:00pm, coinciding with the original series between 16:00pm and 19:00pm, after which it underestimates results again. The MERRA-2 reanalysis overestimates the original data in most of the time series.

Table 8 shows the residuals of the models and the reanalysis with the observed data. The lowest errors can be found in the hybrid model and the largest in the MERRA-2 reanalysis. The Holt-Winters model showed a large error in this assessment.

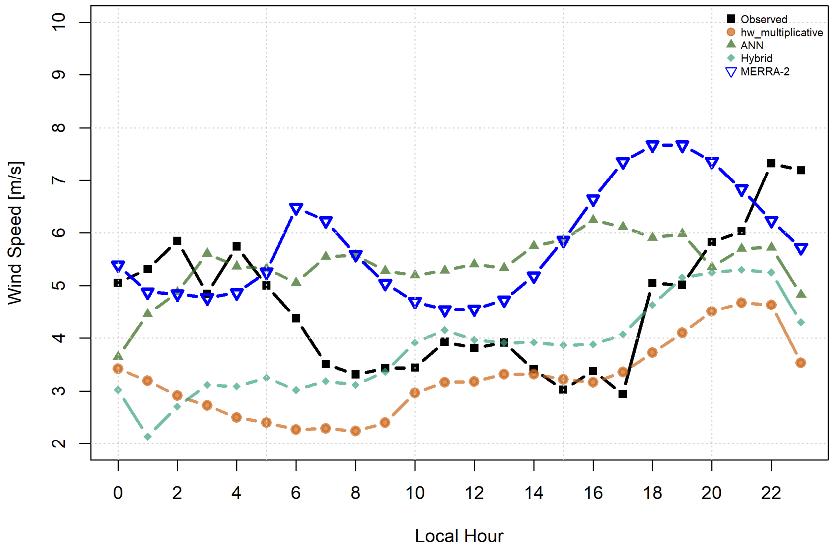

Figure 11. Comparison between the series observed by TA-01, those predicted by the models and the MERRA-2 reanalysis for the day 30/12/2004.

Table 7. Residuals of the models and the reanalysis with the observations (TA-01) for 90 days of data, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	1.76	1.70	**1.52**	1.90
MAE	1.45	1.48	**1.13**	1.57
r	**0.51**	-0.31	0.27	0.13

For the TA-02 tower, the Holt-Winters model overestimates the observed data at all times, but it follows the behavior of the observations. Between the predicted hours of 23:00pm and 24:00pm, the model coincides with the observed data. The hybrid and RNA models underestimate the observed data during most of the time series and don't follow the behavior as well as the Holt-Winters model (Figure 12).

Table 9 shows the residuals of the models and the reanalysis with the observations. The smallest forecast skills are of the Holt-Winters model, with a RMSE and MAE of 2.14 and 1.62 m/s, followed by the hybrid model with 2.27 and 1.84 m/s, respectively. The highest correlations are found in the Holt-Winters and hybrid models.

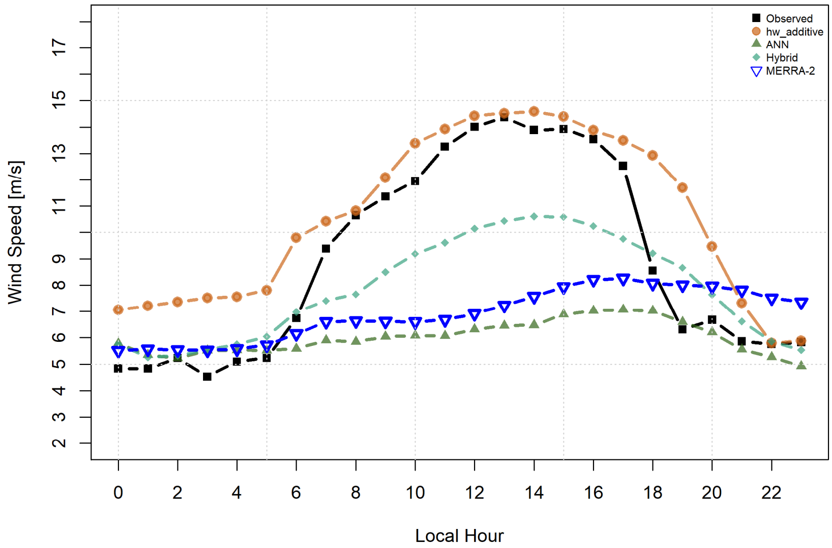

Figure 12. Comparison between the observed series, those predicted by the models and the MERRA-2 reanalysis for the TA-02 tower on 30/12/2004.

Table 9. Residuals of the models and the reanalysis with the observations (TA-02) for 90 days of data, in m/s.

	Holt-Winters	ANN	Hybrid	MERRA-2
RMSE	**2.14**	4.36	2.27	3.87
MAE	**1.62**	3.21	1.84	3.03
r	**0.93**	0.71	0.93	0.48

4. Discussion

4.1 Comparison with literature results

It is important to verify that these results of the errors in wind speed prediction this paper are in agreement with the values found in similar literature studies. Reference [3] presented a study involving the modeling of the monthly and hourly wind speed averages at a height of 10 meters in the coastal Northeast region of Brazil for the prediction of wind speed for power generation with the additive Holt-Winters model. According to the authors, when compared to the observed data, this model showed a good fit with errors of RMSE and MAE of 0.50/0.57 m/s and 0.40/0.45 m/s for monthly average wind speed forecasting, 1.38/1.55 m/s and 1.03/1.19 m/s for hourly average wind speed forecasting, respectively. However, in studies like these applied to wind power generation, the positioning of the measuring equipment must be installed at the top of the station (upper anemometer) at a height from the ground equal to the axes of the wind park turbines and at least 50 meters from the ground, in accordance with the technical standards [32]. Reference [33] indicate the measurement level of the primary wind speed measurement, which is mainly relevant for determining the hub height wind conditions, shall be at least 2/3 of the planned hub height. Currently the wind turbines in Brazil are being implanted at over 100 m height. Reference [10] utilized the method of exponential smoothing, in particular the methods Holt-Winters additive and multiplicative for 6 hours wind speed prediction at a height 50 m agl in the city São João Cariri (Paraíba-Brazil) in order to apply the methodology. According author, the additive Holt-Winters model presented better result than the multiplicative Holt-Winters model for 6 hours prediction in the same series and data period of wind speed with errors of RMSE 2.0365 and 2.6197 m/s.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

Reference [7] compare the performance of the ARIMA, ARIMAX and ANN models in an attempt to forecast monthly wind speed averages at 3 locality's in the coastal NEB (Fortaleza, Natal and Paraíba). According to the authors, the ARIMAX model presented greater sensitivity to the wind speed adjustment and prediction, the RMSE and MAE values found were 0.48(Fortaleza)/0.45(Natal)/0.71(Paraíba) m/s and 0.37(Fortaleza)/0.37(Natal)/0.54(Paraíba) m/s, respectively. The authors proposed that, likely, with the increase in the number of training vectors for the ANN model, its performance will improve and its statistical errors. Reference [6] proposed an improved radial basis function neural network with an error feedback scheme to forecast short-term wind speed and wind power, with parameter initialization method and the inclusion of the shape parameter in the Gaussian basis function of each hidden neuron, used to search better initial center and standard deviation values. Results show that the forecast accuracy by proposed model is better compared to the other neural network-based models. Reference [34] used wind speed monthly average data at 10 m agl of 28 meteorological stations operated by the Nigeria Meteorological Services (NIMET) where were used as training (18 stations) and testing (10 stations) in the ANN model. The ANN model consisted of 3-layered, feed-forward, back-propagation network with different configurations. The proposed is used ANN model in the predicting of wind speed monthly average. The results indicate high accuracy in the predicting of wind speed, with the correlation coefficient between the predicted and the observed of 0.938, which shows the effectiveness of this model.

In [17] developed a methodology for the short-term forecasting of wind power generation with the modified ARIMAX model, which is based on the Box-Jenkins methodology, through which an adjustment of models is obtained for the time series of observations so that the residuals are around zero. The predicted and observed wind data as well as the actual wind power data were used as inputs of the dynamic models. No satisfactory result was found among the models used, and this has made the develop more research on wind forecasting geared to wind power generation [17].

In the study [22], the authors proposed a hybrid modeling method for the short-term wind speed forecast for wind power generation, using data collected from an anemometric tower at a height of 20 meters of (which also does not follow the standard) located in Beloit (Kansas/USA) in the period from 2003 to 2004. The results showed that the hybrid modeling method for this variable provides better predictions when compared to other methods, resulting in a MAE (m/s) ranging from 0.016 to 0.52 in this study. In our study, the hybrid methodology also produced satisfactory results. Reference [3] presented a combined hybrid model of the ARIMAX-ANN and Holt-Winters-ANN models to predict the wind speed in terms of hourly means, being efficient in the producing adjustments to the observed data of the studied regions. The authors showed the quality of the proposed hybrid model is the low statistical analysis of errors values, for example, the RMSE and MAE of 0.46/0.41 m/s and 0.35/0.32 m/s (locality: Fortaleza/Natal), especially when compared to the ARIMAX, ANN and Holt-Winters models with errors values high for the RMSE e MAE. In [21] proposed a hybrid short-term wind speed forecasting model at two cases studies. The results from two cases show that the proposed hybrid model offers greater accuracy which relationship to the other compared models (Persistent and ARIMA) in short-term wind speed forecasting. The hybrid model ($\mathrm{G} - [\upsilon - \mathrm{SVM}] - \mathrm{CS}$ (cuckoo search)) showed the lowest values of error statistics, for example, for the MAE it was possible to find values of 0.5132 m/s, versus that of 0.582 m/s and 0.5242 m/s, for the models of Persistent and ARIMA, respectively. Reference [14] proposed a hybrid forecasting approach involving the statistics techniques EWT-CSA-LSSVM for short-term wind speed prediction from a windmill farm located in northwestern China. The hybrid model showed results suggest that the developed for wind speed forecasting method yields better predictions compared with those of other models with the lower RMSE and MAE of 0.58 and 0.57 m/s errors.

 doi:10.20944/preprints201807.0501.v1

5. Conclusions

Most research with wind data in Brazil is done through databases with little coverage, especially with wind towers. Most of these studies are done by private companies and are not publicly available for research. On the other hand, the monitoring via meteorological stations features a series of errors. These facts contributed to make the development of this study difficult, since there are no long wind speed time series available observed at 50 meters in height.

Given the difficulty of obtaining good temporal series and likely good results in the studies, this article eventually became a mechanism for discovery and the exploration of alternative methods. As for the modeling of the diurnal wind variability, the MERRA-2 reanalysis, which is the most recent version, did not capture this variability well. The Holt-Winters model, which models the seasonality and trend components, showed a good fit, but time series of several years are needed to better capture these parameters and thus provide more satisfactory results. Finally, the best fits for the forecast were found using the hybrid methodology. This method can be applied for the planning of operations, maintenance and deployment of wind turbines, since accurate predictions minimize technical and financial risks.

6. Patents

Author Contributions: This article has three authors. Paulo Sérgio Lucio conceived of the research idea and gave suggestions. Moniki Dara de Melo Ferreira carried out the research, analyzed the data, developed the scripts and drafted the article. Alexandre Torres Silva dos Santos has contributed with the research of articles, the simulations, tips and revisions.

Acknowledgments: The Postgraduate Program of Climate and Atmospheric Sciences at Federal University of Rio Grande do Norte and the Laboratory of Maps and Energetic Resources Data at Center for Technologies of Gas and Renewable Energies supported this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Algorithm to Calculate wind speed forecasting short-term using Holt-Winters, ANN and Hybrid models.

R-Code:

```
# Algorithm to calculate using Holt-Winters, ANN and Hybrid models
# Authors: Moniki Melo (UFRN) and Alexandre Santos (CTGAS-ER)

# Load add-on packages

library(stats)
library(tseries)
library(forecast,warn.conflicts=TRUE)
library(forecastHybrid)

# Data wind speed of BELO JARDIM - TA01
# Open file wind speed observational data

data_TA01 <- read.csv("belojardim_data.csv",header=T,sep=";",dec=",")
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1


```
# Set the column of a matriz object

colnames(data_TA01)<-c("date","Wind_speed")

# The database is attached to the R search path (data wind speed of the TA01)
# This means that the database is searched by R when evaluating a variable, so objects in the datab
ase can be accessed by simply giving their names.
attach(data_TA01)
# Return the First or Last Part of an Object

head(data_TA01)
##                   date  Wind_speed
## 1 01/10/2004 00:00        6.182
## 2 01/10/2004 01:00        6.517
## 3 01/10/2004 02:00        5.845
## 4 01/10/2004 03:00        5.125
## 5 01/10/2004 04:00        5.137
## 6 01/10/2004 05:00        3.798
 # Holt-Winters-additive model

# 30 days - file based in this period of wind speed observed

data30days=data.frame(Wind_speed[1:720])

# Creating an time series with frequency of 24-hours

belojardim.ts=ts(data30days,freq=24)

# The Holt-Winters-additive filtering of wind speed time series.
# Unknown parameters are determined by minimizing the squared prediction error.
# Default configuration

belojardim.hw <- HoltWinters(belojardim.ts,seasonal="additive")

# Result of the Holt-Winters-additive

attributes(belojardim.hw)
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1


```
## $names
## [1] "fitted"        "x"             "alpha"         "beta"
## [5] "gamma"         "coefficients"  "seasonal"      "SSE"
## [9] "call"
##
## $class
## [1] "HoltWinters"

# stats package documentation – Holt-Winters Filtering
# Description

# An object of class "HoltWinters", a list with components:

# fitted : A multiple time series with one column for the filtered series
#           as well as for the level, trend and seasonal components, estimated
#           contemporaneously (that is at time t and not at the end of the series).
# x : The original series
# alpha  : alpha used for filtering
# beta : beta used for filtering
# gamma  :gamma used for filtering
# coefficients: A vector with named components a, b, s1, ..., sp containing the
#               estimated values for the level, trend and seasonal components.
# seasonal: The specified seasonal parameter
# SSE    : The final sum of squared errors achieved in optimizing
# call : The call used

# Result of the parameters: Holt-Winters-additive model

belojardim.hw$seasonal

## [1] "additive"

belojardim.hw$beta

## [1] 0

belojardim.hw$gamma

## [1] 0.9505823

belojardim.hw$SSE

## [1] 581.0709

# Forecast for the next 24-hours: forescat() function
# Default configuration
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1


```
forecast_24horas_hw <- forecast(belojardim.hw,h=24)

# Artificial Neural Networks model
# nnetar() function
# Default configuration
# Neural Network Time Series Forecasts

belojardim_ann <- nnetar(belojardim.ts)

# values of the arguments:  parameters of model

attributes(belojardim_ann)

## $names
##  [1] "x"         "m"         "p"         "P"         "scalex"
##  [6] "size"      "subset"    "model"     "nnetargs"  "fitted"
## [11] "residuals" "lags"      "series"    "method"    "call"
##
## $class
## [1] "nnetar"

# forecast package documentation – Neural Network Time Series Forecasts
# Description
# P - Number of seasonal lags used as inputs.

# p - Embedding dimension for non-seasonal time series. Number of non-seasonal lags used as input
s. For non-seasonal time series, the default is the optimal number of lags (according to the AIC) for
 a linear AR(p) model. For seasonal time series, the same method is used but applied to seasonally a
djusted data (from an stl decomposition).

# size - Number of nodes in the hidden layer. Default is half of the number of input nodes (includi
ng external regressors, if given) plus 1.

# model - Output from a previous call to nnetar. If model is passed, this same model is fitted to y
without re-estimating any parameters.

belojardim_ann$m

## [1] 24

belojardim_ann$P

## [1] 1

belojardim_ann$p

## [1] 27
```

 doi:10.20944/preprints201807.0501.v1


```
belojardim_ann$size

## [1] 14

belojardim_ann$model

## Average of 20 networks, each of which is
## a 27-14-1 network with 407 weights
## options were - linear output units

forecast_24horas_ann <- forecast(belojardim_ann,h=24)

# Hybrid Model
# Default configuration

belojardim_hybrid <- hybridModel(belojardim.ts,models="nst")

## Fitting the nnetar model
## Fitting the stlm model
## Fitting the tbats model

# Forecast for the next 24-hours: forescat() function
# Default configuration

forecast_24horas_ann <- forecast(belojardim_hybrid,h=24)

# End of code
```

Appendix B

Algorithm to calculate the errors of the predictions by indices RMSE and MAE and Person Correlation and plot of the graphics.

R-Code:

```
# Function : Calculate error

# Function to calculate  Root Mean Squared Error (RMSE)

RMSE <- function(result) {

    sqrt(mean(result^2))

}

# Function to calculate Mean Absolute Error (MAE)

MAE <-  function(result) {

      mean(abs(result))
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1


```
  }
# Calculate error
# result <- predicted - observed
# Error of the Holt-Winters model
result <- forecast_24horas_hw$mean - data_TA01$Wind_speed[721:744]
round(RMSE(result),2)
round(MAE(result),2)
# Calculate Pearson correlation coefficient

round(cor(forecast_24horas_hw$mean,data_TA01$Wind_speed[721:744]),2)
# Plot graphics and saving in format tiff (vector)
# Defining time for cycle 24 hours

time    <- c("00:00", "01:00", "02:00", "03:00", "04:00", "05:00", "06:00", "07:00",
             "08:00", "09:00", "10:00","11:00","12:00","13:00","14:00","15:00",
             "16:00","17:00","18:00","19:00","20:00","21:00","22:00","23:00")
hour.time <- strptime(time,"%H:%M")

# Open file wind speed observational data and predicted by models

dados <- read.csv("valores-belojardim24_30dias.csv",head=T,sep="",dec=",")
attach(dados)
head(dados)
##           data  hora   obs hw_aditivo hw_mult   rna hybrid merra2
## 1 2004-10-31 00:00 6.828      6.011  6.7630 6.151  6.440  6.729
## 2 2004-10-31 01:00 6.825      4.610  5.5218 6.100  5.809  6.761
## 3 2004-10-31 02:00 5.685      4.169  5.0910 5.998  5.447  6.626
## 4 2004-10-31 03:00 6.260      3.674  4.7000 6.225  5.138  6.789
## 5 2004-10-31 04:00 5.730      3.279  4.4010 6.057  4.701  6.961
## 6 2004-10-31 05:00 4.680      3.574  4.5070 5.845  4.707  7.071
# Saving file in vector (tiff)

tiff("belojardim_paper_30days.tiff", width = 8, height = 6, units = 'in', res = 300,compression =
 'lzw')

plot(0:23, obs, type='b', bty="l",col="black",lwd=3 , pch=15,ylim=c(2,10),axes=FALSE,
```

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1


```
      xlab="Local Hour",ylab="Wind Speed [m/s]")
lines(0:23,hw_aditivo,col=rgb(0.8,0.4,0.1,0.7) , lwd=3 , pch=19 , type="b" )
lines(0:23,rna,col=rgb(0.2,0.4,0.1,0.7), lwd=3 , pch=17 , type="b" )
lines(0:23,hybrid,col=rgb(0.225,0.64,0.5,0.7), lwd=3 , pch=18 , type="b" )
lines(0:23,merra2,col="blue", lwd=3 , pch=25 , type="b" )

axis(side=1, at=seq(0,23,by=2))
axis(side=2, at=seq(1, 10, by=1))
box()
grid()

# Legend of the graphics


legend("topright",
        legend = c("Observed", "hw_additive","ANN","Hybrid","MERRA-2"),
        col = c("black",rgb(0.8,0.4,0.1,0.7),
                 rgb(0.2,0.4,0.1,0.7),rgb(0.225,0.64,0.5,0.7),"blue"),
        pch = c(15,19,17,18,25),
        bty = "n",
        pt.cex = 1,
        cex = 0.6,
        text.col = "black",
        horiz = F,)

# End of the code

dev.off()
```

Appendix C

Tables: Main characteristics of ANNs and hybrid (*nnetar-stlm-tbats* functions) models produced in R software.

Belo Jardim - Hour					
Based on 30 days of data					
p	*P*	*Size(k)*	*m*	*Weights*	*Average of Network*
26	1	14	24	393	20
Based on 60 days of data					
p	*P*	*Size(K)*	*m*	*Weights*	*Average of Network*
25	1	13	24	352	20
Based on 90 days of data					
p	*P*	*Size(K)*	*m*	*Weights*	*Average of Network*
25	1	13	24	352	20

doi:10.20944/preprints201807.0501.v1

Camocim - Hour					
Based on 30 days of data					
p	*P*	*Size(k)*	*m*	*Weights*	*Average of Network*
11	1	6	24	85	20
Based on 60 days of data					
p	*P*	*Size(k)*	*m*	*Weights*	*Average of Network*
27	1	14	24	407	20
Based on 90 days of data					
p	*P*	*Size(k)*	*m*	*Weights*	*Average of Network*
27	1	14	24	407	20

References

1. Fontoura. C.F.; Brandão, L.E.; Gomes, L.L. 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. Crate, S.A.; Nuttall, M. In *Anthropology and climate change: from encounters to actions*. 2nd ed.; New York, New York, Routledge: London, England, 2016; p. 450, 1629580015.
3. Camelo, H.N.; Lucio, P.S.; Junior, J.B.V.L.; De Carvalho, P.C.; Santos, D.von G. 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.
4. Emeis, S. In *Wind Energy Meteorology, Green Energy and Technology*, 2013th ed.; Springer-Verlag, Berlin, Germany, 2013, p. 198, 978-3642305221.
5. Pinto, M.O. Fundamentos de Energia Eólica, 1st ed. Livros Tecnicos e Cientificos: São Paulo, Brazil, 2013, 392.
6. Chang, G.W.; Lu, H.J.; Chang, Y.R.; Lee, Y.D. An improved neural network-based approach for short-term wind speed and power forecast. *Ren Ene*, 2017, 107, 301-311, https://doi.org/10.1016/j.renene.2016.12.071.
7. Camelo, H.N.; Lucio, P.S.; Junior, J.B.V.L.; De Carvalho, P.C.M. A hybrid model based on time series models and neural network for forecasting wind speed in the Brazilian northeast region. *Sustainable Energy Technologies and Assessments*, 2018, 28, 65-72, https://doi.org/10.1016/j.seta.2018.06.009.
8. Cheng, W.Y.Y.; Liu, Y.; Bourgeois, A.J.; Wu, Y., Haupt, S.E. Short-Term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. *Renewable Energy*, 2017, 107, 340-351, https://doi.org/10.1016/j.renene.2017.02.014.
9. Martinez-Arellano, G.; Nolle, L. Short-term wind power forecasting with WRF-ARW model and genetic programing. 19th International Conference on Soft Computing, Brno, Czech Republic, 26-28 of June; Radomil Matousek; Publisher: Brno, Czech Republic, 2013.
10. Monteiro, C.M.N.; Souza, R. C. Modelo para previsão de curto prazo de velocidade usando Holt-Winters. Master degree. Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.
11. Krishnaveny, R.; Nair, V.; Vanitha, M. J. Forecasting of wind speed using ANN, ARIMA and Hybrid models. Intelligent Computing Instrumentation and Control Technologies International Conference, Kannur, India, 170-175, 2017.
12. Ramos, D.N.S.; Lyra, R.F.F.; Silva, J. R.S. Previsão do vento utilizando o modelo atmosférico WRF para o estado de Alagoas. Rev. bras. meteorol., 2013, 28, 163-172, http://dx.doi.org/10.1590/S0102-77862013000200005.
13. Ferreira, M.D.M.; Lucio, P.S.; Santos, A.T.S. Desempenho de métodos de preenchimento de falhas para velocidade do vento sobre o Rio Grande do Norte. *O Set. Elétrico*, 2017, 138, 21, 1-116,
14. Hu, J.; W. Jianzhou, Ma, Kailiang. A hybrid technique for short-term wind speed prediction. *Energy*, 2015, 81, 563-574, https://doi.org/10.1016/j.energy.2014.12.074.
15. Brower, M.C. Wind Resource Assessment: A Practical Guide to Developing a Wind Project. John Wiley & Sons, New York, (Ed.), 2012, 280, http://dx.doi.org/10.1002/9781118249864.
16. Santos, A. T. S.; Santos e Silva, C.M.; Lemos, D.F.A.; Oliveira, L.L.; Bezerra, L.A.C. Assessment of wind resources in two parts of Northeast Brazil with the use of numerical models. *Met. App.* 2016, 23, 563-573, https://doi.org/10.1002/met.1595.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0501.v1

17. Operador Nacional do Sistema Elétrico (ONS). Desenvolvimento Metodológico Para Previsão de Geração de Fonte Eólica. ONS (in press).
18. Zhao, P.; Wang, J.; Xia, J.; Dai, Y.; Sheng, Y.; Yue, J. Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China. *Ren. Ene.*, 2012, 43, 234-241, https://doi.org/10.1016/j.renene.2011.11.051 .
19. Gonçalves, A. R.; Pereira, E. B.; Martins, F. R. Previsão de Curto Prazo da Geração eólio-elétrica para o Nordeste Brasileiro por redes neurais artificiais. 3rd Solar Energy Conference, Belém, Brazil, 21-14 of September, 2010.
20. SOUZA, R. B. Avaliação de modelos de inteligência artificial para previsão da velocidade de vento em curto prazo. Master degree. Universidade Federal de Pernambuco, Recife, 2014.
21. Jiang, P.; Wang, Y.; Wang, J. Short-term wind speed forecasting using a hybrid model. *Energy*, 2017, 119, 561-577, https://doi.org/10.1016/j.energy.2016.10.040.
22. Tatinati, S; Veluvolu, K.C. A Hybrid Approach for Short-Term Forecasting of Wind Speed. *The Sci. Wor. Jour*, 2013, 2013, 1-8, http://dx.doi.org/10.1155/2013/548370.
23. Atlas de Energia Elétrica do Brasil. Available online: http://www2.aneel.gov.br/aplicacoes/atlas/pdf/06-energia_eolica (3).pdf (Accessed on 22 November 2017).
24. WU, Yuan-kang; HONG, Jing-shan. A literature review of wind forecasting technology in the world. Lausanne Power Tech Conference, Lausanne, Switzerland, 1-5 July, 2007.
25. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. Int. Jour. Fore, 2004, 20, 5-10, https://doi.org/10.1016/j.ijforecast.2003.09.015.
26. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Man. Sci.*, 1960, 324-242, http://dx.doi.org/10.1287/mnsc.6.3.324.
27. Khandakar, Y.; Hyndman, R.J. Automatic time series forecasting: the forecast Package for R. *Jour. of Sta. Soft*, 2008, 27, 1-22, http://dx.doi.org/10.18637/jss.v027.i03.
28. Shaub, D.; Ellis, P. Convenient functions for ensemble time series forecasts. R package version 2.2.12. Available online: https://cran.r-project.org/web/packages/forecastHybrid/forecastHybrid.pdf (accessed on 10 may 2018).
29. Hyndman, R.J.; Akram, M.D.; Archibald, B. The admissible parameter space for exponential smoothing models. *Ann. of Sta Mat*, 2008, 60(2), 407-426, https://doi.org/10.1007/s10463-006-0109-x.
30. Global Modeling and Assimilation Office (GMAO), 2016. Available online: http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl (accessed on 16 august 2017).
31. Wilks, D. S. Statistical Methods in the Atmospheric Sciences. 3rd ed. Academic Press: Amsterdam, Netherlands, 2011, p. 704, 9780123850232.
32. Instruções para as medições anemométricas e climatológicas em parques eólicos. Available online: http://www2.aneel.gov.br/aplicacoes/editais_geracao/documentos/Adendo_6_DEA%2016-13%20Medi%C3%A7%C3%B5es%20Anemom%C3%A9tricas%20LEN%20A-5_2013.pdf (Acessed on 23 november 2017).
33. MEASNET-Measuring Network of Wind Energy Institutes. MEASNET Procedure: Evaluation of Site-Specific Wind Conditions, Version 2, 2016, 56.
34. Fadare, D.A. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. *App. Ene*, 2010, 87(3), 934-942, https://doi.org/10.1016/j.apenergy.2009.09.005.