

Article

Enhancing Short-Term Wind Power Forecasting through Multiresolution Analysis and Echo State Networks

Hugo Tavares Vieira Gouveia ^{1,*} , Ronaldo Ribeiro Barbosa de Aquino ¹ 
and Aida Araújo Ferreira ² 

¹ Department of Electrical Engineering, Federal University of Pernambuco (UFPE), Recife 50740-533, PE, Brazil; rrba@ufpe.br

² Federal Institute of Education, Science and Technology of Pernambuco (IFPE), Recife 50740-545, PE, Brazil; aidaaf@gmail.com

* Correspondence: hugotvg@gmail.com; Tel.: +55-81-2126-7103

Received: 5 March 2018; Accepted: 30 March 2018; Published: 3 April 2018



Abstract: This article suggests the application of multiresolution analysis by Wavelet Transform—WT and Echo State Networks—ESN for the development of tools capable of providing wind speed and power generation forecasting. The models were developed to forecast the hourly mean wind speeds, which are applied to the wind turbine's power curve to obtain wind power forecasts with horizons ranging from 1 to 24 h ahead, for three different locations of the Brazilian Northeast. The average improvement of Normalized Mean Absolute Error—NMAE for the first six, twelve, eighteen and twenty-four hourly power generation forecasts obtained by using the models proposed in this article were 70.87%, 71.99%, 67.77% and 58.52%, respectively. These results of improvements in relation to the Persistence Model—PM are among the best published results to date for wind power forecasting. The adopted methodology was adequate, assuring statistically reliable forecasts. When comparing the performance of fully-connected feedforward Artificial Neural Networks—ANN and ESN, it was observed that both are powerful time series forecasting tools, but the ESN proved to be more suited for wind power forecasting.

Keywords: Echo State Network; neural networks; reservoir computing; wavelet transform; wind forecasting

1. Introduction

Nowadays, wind energy is one of the most successful sources of renewable energy used around the world. The incentive policies adopted by several countries are among the drivers for using wind power, since they guarantee the purchase of energy produced by wind farms even if the price of energy is not competitive. The first countries that applied incentive policies to stimulate wind power development were Germany and Denmark. Then, other countries also adopted incentive policies, as in the case of Brazil, with the creation of the Incentive Program for Alternative Sources of Electric Energy—PROINFA [1].

Since wind generation is an intermittent source of generation, the use of models or techniques capable of predicting the energy produced by these type of sources is essential for an adequate wind power integration into electrical power systems.

Accurate wind power forecasting methods can assist power systems operators in the accurate balancing between demand and supply. The higher the penetration level of wind generation, the greater the need for forecasting tools of this type of generation, in order to maximize the integration between the generation from conventional sources (which is planned and programmed) and the forecast

electricity demand (which is predictable with sufficient accuracy). According to [2], even when using state-of-the-art wind forecasting methods, the hour-ahead prediction errors are around 10–15% for a single wind farm.

In interconnected electrical power systems, the balance between demand and generation in the short-term depends on the automatic generation control, which does not regulate power flows in transmission lines. Regional voltage controls typically regulate the primary buses, which may not result in improved voltage levels on other buses of the power system. Thus, due to the limitations of generation controllers and voltage regulators, a high level of wind power penetration can cause temporary overloads on transmission lines and violation of acceptable system voltage levels.

Regarding the different forecasting horizons, the day-ahead period will always be important for wind forecasting. The unit commitment process of the next-day is a core task for power system operators. As described in [3], the day-ahead time frame is typically when the overall operating plan for the next day is put in place, including the selection of dispatchable sources (like large thermal plants that may take several hours to start up).

The current tools used for forecasting wind speed and power may or may not involve a Numerical Weather Prediction—NWP model. The models that use NWP provide better predictions of temporal series for larger horizons, starting from 3 to 6 h, which makes them more appropriate for utilities usage. In Brazil, for example, the National Electric System Operator—ONS uses wind hourly forecasts on the heights of 10 m and 100 m, from the numerical model ETA with spatial resolution of 15 km, to several wind farms. This model is processed at Center for Weather Forecasting and Climate Studies—CPTEC.

For short-term prediction, there are basically two different types of modeling: physical and statistical. A combination of the two types is also used to make more reliable predictions, trying to use physical variables to obtain an adequate estimate of the local wind speed before use a statistical model (Model Output Statistics—MOS) to decrease the remaining error. Statistical modeling searches for strong relations between the historical values of the electrical energy production (and other meteorological parameters) and information measured in real time, usually using recursive techniques. For time series forecasting, there are statistical models that can be expressed analytically, such as Autoregressive Moving Average—ARMA and its variants, or “black box” models, which are not described analytically, such as ANN.

Several short-term forecasting models using ANN, Fuzzy Logic and Wavelets were analyzed in [1]. The forecasts results were compared with reference models. The performance gains of the best models proposed in relation to reference models were approximately 80% to the forecasting horizon of one hour ahead. It was shown that there are improvements when the average hourly wind speeds are preprocessed with the application of multiresolution analysis by Wavelet Transform—WT before providing them as inputs to neural networks, especially for the forecasting horizon in the range from 1 to 6 h. The analyzed forecasting horizon range of varied from 1 to 24 h ahead.

Based on the theories of wavelet, wavelet packet, time series analysis and ANN, three hybrid models for wind speed prediction were proposed in [4]. The models were compared with classical wind speed forecasting methods like PM. The results showed that the Wavelet Packet-ANN model is the best among them.

A hybrid neuro-fuzzy system for wind power prediction was proposed in [5]. A wireless sensor network measures and transmits the required parameters (air temperature, density and pressure, and wind speed) for the prediction model at the operations center. According to the authors, consideration of all these parameters increased the prediction accuracy of the proposed model.

A procedure to make judgment about the nonlinearity of wind speed time series was proposed in [6]. The proposed method solves the problem with quantizing the nonlinear deterministic components. The performance of some prediction were compared, with Grey–Markov presenting the best performance.

A prediction method based on wavelet networks, using multi-resolution analysis and adaptive wavelet neural networks, was implemented and evaluated by numerical simulations using real

wind speed profiles in [7]. According to the authors, the method allows reducing computational efforts without reduction in the accuracy, which is very important for prediction systems that uses digital processors.

Nonparametric methods using ANN and Fuzzy Inference Systems for modeling wind turbine's power curve were explored in [8]. The models were adjusted to predict the wind power in the range between 1 and 24 h ahead, for actual case studies from two wind farms in the Brazilian Northeast. The gain of the developed models lay in the range of 29 to 60% when compared to the PM in a twenty-four hours forecasting horizon.

Models presented in [9], based on techniques of ESN, ANN and Adaptive Neuro-Fuzzy Inference Systems—ANFIS, investigated the contribution of using climate variables as inputs seeking for the performance improvement of wind speed forecasting. The gain of some of the best models developed were approximately 50% in relation to the PM in a four hours ahead time window.

A method to find the best reservoir applied to the time series prediction task was proposed in [10]. The developed method, called RCDESIGN, combines an evolutionary algorithm with Reservoir Computing—RC and searches simultaneously for the best values of the ESN hyperparameters without rescaling the reservoir weights matrix by its spectral radius—SR. The method also considers the RC in all its nonlinearity, because it allows the use of all possible connections, rather than using only the required connections. The proposed method was applied in two real data sets of average hourly wind speed looking to verify the feasibility of the proposed method applied to wind speed forecasting.

A novel ESN that improves performance of the forecasting engine by decreasing the number of the internal states was proposed in [11]. The proposed topology has multiple sub-reservoirs with nonlinear relations between the internal states, resulting in reduced orders of the weight matrices. Simulation results show faster learning speed, fewer computations, and improved prediction performance as compared to ESN with fixed sizes and topologies.

The authors in [12] propose a Growing Echo State Network—GESN to automatically design ESN. The GESN makes use of multiple sub-reservoirs to reduce the coupling effects between units in a single reservoir. It was shown that the Echo State Property could be guaranteed without post-scaling of reservoir weight matrix. Experimental results on four time-series benchmarks demonstrate that the GESN performs better than some ESN with fixed hyperparameters.

An ESN consists of a large recurrent neural network, the reservoir (Figure 1), whose weights of connections are often set randomly and sparse. As discussed in [13], the reservoir states are driven by the input signal and then projected to output units. The present article suggests the application of multiresolution analysis by WT and ESN for the development of tools capable of providing wind speed and power generation forecasting. The generated tools seek to support planning of a multisource electrical power system, composed essentially by hydro, thermal and wind generation. The main objective is to provide further insight into the process used to develop a new model that uses ESN for wind speed forecasting. The performance of this model is then compared with one of the different models proposed in [1]. The wind power forecasts for each hour of the day are done through the application of the forecasted speeds into the wind turbine's power curve.

The main difference between the compared models is that rather than using ESN, the authors in [1] used Multilayer Feedforward fully connected Artificial Neural Networks. The model proposed in [1], that will be used here, was named by their authors of WTANNLM. The "WT" prefix means Wavelet Transform and "ANN" is due to the use of Multilayer Feedforward fully connected Artificial Neural Networks. The suffix "LM" refers to the use of the Levenberg-Marquardt—LM training algorithm. These networks architecture consists of an input, an intermediate (hidden) and an output layer, as shown in Figure 2. The goal of the ANN training algorithm is to adjust all weights of the connections at all existing layers. In the case of ESN, the training algorithm is used to adjust only the connections from the reservoir to the output units.

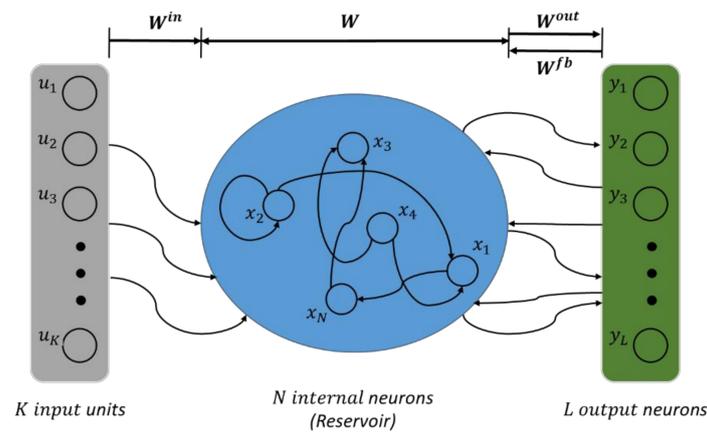


Figure 1. Architecture of an Echo State Network.

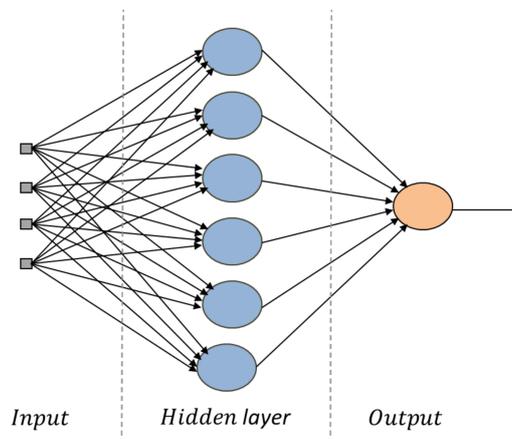


Figure 2. Architecture of a Multilayer Feedforward ANN.

In this article, the networks were replaced by ESN, and the proposed model is called WTESN. The results show significant improvements compared to the forecasts presented in previous articles for the same databases.

Due to the existence of extensive literature addressing the ANN, ESN and WT, their detailed description is not the scope of this work. For more details, we suggest reading the references listed at the end of this article.

2. Materials and Methods

The materials and methods used for performing this article will be detailed in this section. First, will be presented an overview of the wind speed time series used for the development of the forecasting models and the error criteria used to evaluate their performance. After that, there is a brief description of the data preprocessing through multiresolution analysis with WT. Next, will be shown the structures of the forecasting models and the description of the BASE MATRIX—BM, which serves to facilitate the pattern separation for adjusting the models. In sequence, using a flow chart as reference, will be detailed each one of the principal steps carried out for a better understanding of the developed models. Finally, will be presented the methodology used to convert the forecasted wind speeds to power generation through the power curve of the wind turbine.

The data handling and the development of the proposed models was done with MATLAB® in its version 7.10.0.499 (R2010a). The 64-bit processor used was the Intel (R) Core (TM) 2 Duo CPU T6600 @ 2.20 GHz, with 4.00 GB of RAM, and the operating system, Windows 7—64 bit.

2.1. Arrangement for the Wind Speed Time Series

The models developed in this article are univariate, i.e., wind speed is the only variable used as input for the forecasting models.

The wind speed series used to develop the models are the hourly average values measured in three different cities (Macau, Mossoro and Natal) in the Brazilian state of Rio Grande do Norte. The data were measured at a 10 m height, between January 2008 and December 2008, and each of the three series is composed by 8784 data samples. The statistics of the time series are shown in Table 1.

Table 1. Statistics of the wind speed time series from Macau, Mossoro and Natal.

Site	Minimum (m/s)	Maximum (m/s)	Average (m/s)	Standard Deviation (m/s)
Macau	0.20	11.60	4.48	1.99
Mossoro	0.10	9.40	3.32	1.95
Natal	0.09	10.50	4.70	1.81

Figures 3–5, complete the characterization of the three temporal series by the speed data and histograms, the typical day and the seasonality.

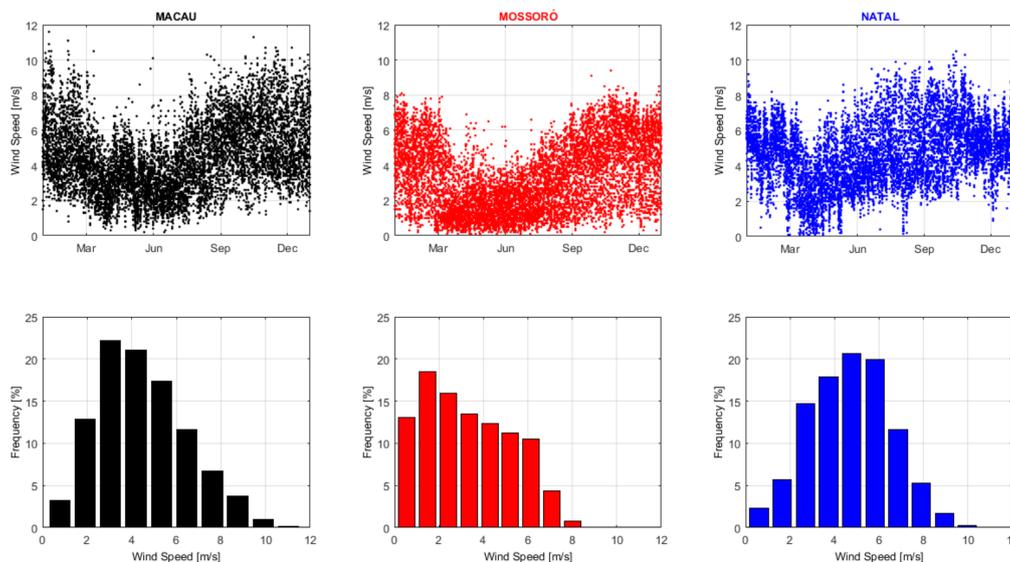


Figure 3. Wind speed time series and histograms. From left to right: Macau, Mossoro and Natal.

The composition of a typical day was obtained by calculating the arithmetic average of the data corresponding to each time of the day throughout all the studied year. It is clearly observed in Figure 4 that the wind speed has a daily cyclical variation. The values decrease throughout the day, reaching its minimum at approximately 09:00 (Coordinated Universal Time—UTC). Subsequently, the values increase until reaching its peak between 15:00 and 19:00 (UTC).

Seasonality can be seen in the graph of monthly average wind speeds (Figure 5). The seasonal behavior of wind in the Brazilian northeast region is extremely important for wind power generation, as there is the possibility of using this type of generation as a way to supplement hydroelectric generation, since in periods of low rainfall, winds are more favorable, and during wet periods winds are weaker.

As the autocorrelation directly affects the performance of univariate models, it was used for the definition of the forecasting model inputs. Figure 6 depicts the autocorrelation function of the wind speed. Because of the winds' daily cyclic behavior, as expected, it can be observed that the series autocorrelation has another maximum 24 h after the initial peak.

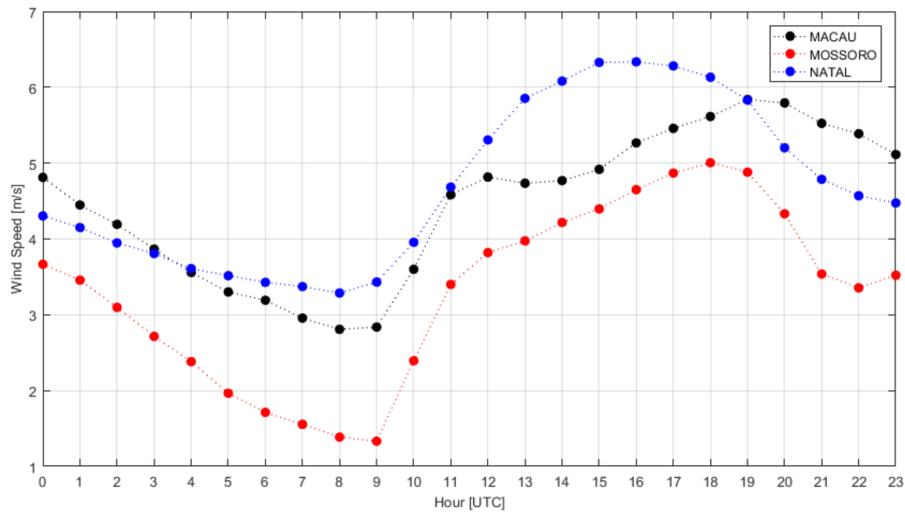


Figure 4. Wind speed for a typical day. Local time = UTC−3 h.

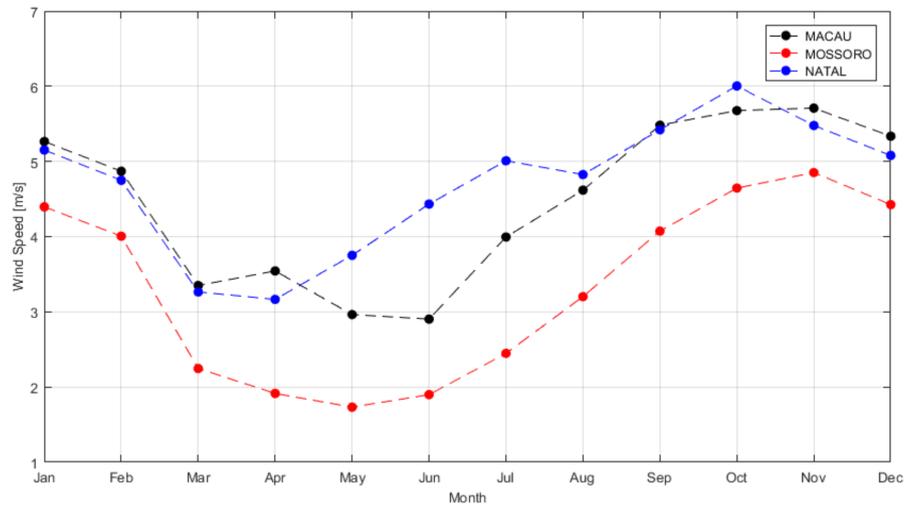


Figure 5. Wind speed seasonality.

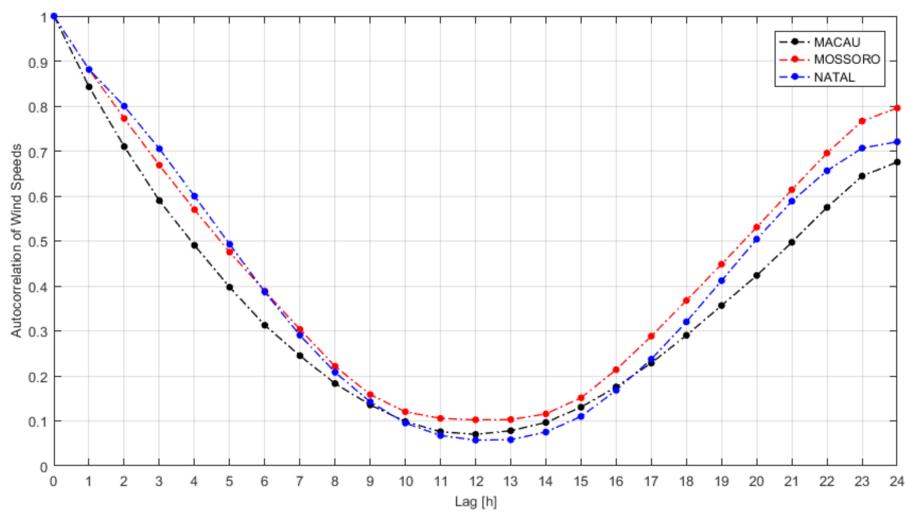


Figure 6. Autocorrelation of the wind speed.

2.2. Performance Evaluation

The error metrics used in this article to evaluate the performance of a forecasting model were Mean Squared Error—MSE, and the Mean Absolute Error—MAE, defined as follows:

$$\text{MSE}_s(k) = \frac{1}{N} \sum_{t=1}^N e_s(t+k|t)^2 \quad (1)$$

$$\text{MAE}_s(k) = \frac{1}{N} \sum_{t=1}^N |e_s(t+k|t)| \quad (2)$$

where k —forecasting horizon (number of time-steps); N —number of data used for the model evaluation; $e_s(t+k|t)$ —error corresponding to time $t+k$ for the wind speed forecast made at origin time t .

The errors defined in (1) and (2) refer to the wind speed forecasts. The Normalized Mean Absolute Error—NMAE was used to illustrate the performance of the power generation. The equation used to calculate it is defined as follows:

$$\text{NMAE}_p(k) = \left(\frac{100}{N \times P_{inst}} \right) \times \sum_{t=1}^N |e_p(t+k|t)| \quad (3)$$

where k —forecasting horizon (number of time-steps); N —number of data used for the model evaluation; P_{inst} —installed wind power capacity; $e_p(t+k|t)$ —error corresponding to time $t+k$ for the power generation forecast made at origin time t .

Probably the most common reference model used in the time series prediction is the Persistence Model [14]. This simple predictor tool states that the variable's future value will be the same as the last one measured, i.e.;

$$y_{PERS}(t+k|t) = y(t) \quad (4)$$

where $y_{PERS}(t+k|t)$ —wind speed (or power generation) forecast for time $t+k$ made at origin time t ; $y(t)$ —measured wind speed (or power generation) at time t .

In order to calculate the improvement of the forecasts obtained with the proposed models in relation to PM, the following equation was used:

$$I_{ref,EC}(k) = 100 \times \left(\frac{EC_{ref}(k) - EC(k)}{EC_{ref}(k)} \right) \quad (5)$$

where k —forecasting horizon (number of time-steps); EC —considered Evaluation Criteria, which can be MSE, MAE or NMAE.

2.3. Multiresolution Analysis

The wind speed have a highly oscillatory behavior, so developing models that provide satisfactory forecasts is a very complex task. In order to reduce the noise caused by the wind variability, a multiresolution analysis by WT is proposed in this article.

The multiresolution analysis decomposes time series into different time scales. As result, the decomposed signals are both an “approximation” and “details” of an original signal.

After preliminary tests, it was found that the Daubechies family wavelets were more robust for the models developed in this article. The level of decomposition was set to 3. It should be emphasized that the results are strongly influenced by both the level of decomposition and the kind of wavelet mother.

2.4. Structure of the Forecasting Models

The WTANNLM and WTESN models have the same input and output patterns, i.e., sixteen inputs and one output. The input data are the wavelet coefficients from the multiresolution analysis of the last four hourly average wind speeds, using the Daubechies wavelet family (db10) decomposed at level 3. The output is the forecast of the hourly average wind speed for a given forecasting horizon k . Figure 7 schematically illustrates the models.

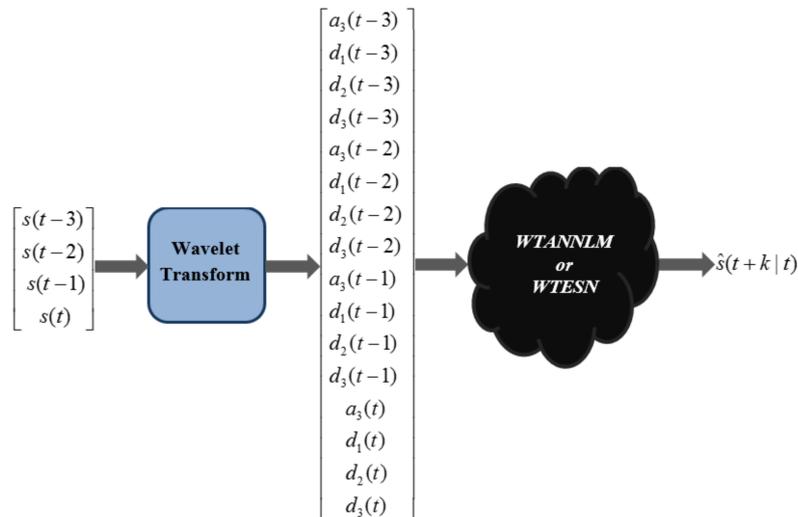


Figure 7. Illustration of models (WTANNLM or WTESN). Variables a and d are described in Section 2.5.

2.5. Mounting of the BASE MATRIX

Before performing the models’ training, a matrix of input and output patterns was set up for each series of wind speeds, and for each forecasting horizon (1 to 24 h ahead). The main purpose of these matrices is to further the manipulation of the training patterns. In total, 72 matrices were mounted (3 locations and 24 different hours ahead).

The BASE MATRIX—BM is a matrix with m rows and n columns. The number of rows depends on the size of the wind speed and of the forecasting horizon. The number of columns is determined by summing the number of inputs and outputs (targets) of the forecasting models.

Given a forecasting horizon k , and a wind speed time series with w elements, the number of rows will be determined by the following expression: $m = w - 3 - k$, with $k = 1, 2, \dots, 24$, and $w \geq 28$.

As there are sixteen inputs and one output for the networks, then, the number of columns is equal to seventeen ($n = 17$). Let us call the elements of the time series as s_1, s_2, \dots, s_w . Then, for each element s_i , there are four corresponding elements $a_{i,3}, d_{i,1}, d_{i,2}, d_{i,3}$, where $a_{i,j}$ and $d_{i,j}$ corresponds, respectively, to the approximation and detail coefficients of the element s_i for the level j . The BM is filled as presented in Table 2.

Table 2. BM structure for models WTANNLM and WTESN.

Inputs													Targets
$a_{1,3}$	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$a_{2,3}$	$d_{2,1}$	$d_{2,2}$...	$d_{3,3}$	$a_{4,3}$	$d_{4,1}$	$d_{4,2}$	$d_{4,3}$	s_{k+4}
$a_{2,3}$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$a_{3,3}$	$d_{3,1}$	$d_{3,2}$...	$d_{4,3}$	$a_{5,3}$	$d_{5,1}$	$d_{5,2}$	$d_{5,3}$	s_{k+5}
$a_{3,3}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	$a_{4,3}$	$d_{4,1}$	$d_{4,2}$...	$d_{5,3}$	$a_{6,3}$	$d_{6,1}$	$d_{6,2}$	$d_{6,3}$	s_{k+6}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	...	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$a_{m,3}$	$d_{m,1}$	$d_{m,2}$	$d_{m,3}$	$a_{m+1,3}$	$d_{m+1,1}$	$d_{m+1,2}$...	$d_{m+2,3}$	$a_{m+3,3}$	$d_{m+3,1}$	$d_{m+3,2}$	$d_{m+3,3}$	s_w

2.6. Development of the WTANNLM Model

A flow chart that illustrates the development process for the WTANNLM model is presented in Figure 8. For a better understanding of the process, we will explain in the following subsections the steps outlined on the diagram. All the procedures implemented in the following subsections aim to improve the generalization power of the neural networks, as addressed in [15,16], avoiding the memorization of the wind speeds patterns during the training of the models.



Figure 8. Development process for the model WTANNLM.

2.6.1. Normalization

Models based on ANN demand a pre-treatment of the data to ensure good performance during network training. Data normalization is the most commonly used method in preparation for ANN training and testing, because it tries to ensure that the same importance will be given to all inputs and targets of the neural network in weights' tuning. And this is also why artificial neuron activation functions are limited. In this article, the BM elements were normalized to values in the range from 0 to 1, according the following equation:

$$\bar{S}(t) = \frac{S(t) - S_{\min}}{S_{\max} - S_{\min}}, \quad (6)$$

where S and \bar{S} are the wind speed non-normalized and normalized, respectively; S_{\max} and S_{\min} are the maximum and the minimum absolute value of the wind speed, respectively.

Due to the use of preprocessing with wavelets, normalization was performed separately on three different groups: one group formed by columns of the BM referring to the coefficients of approximation (columns 1, 5, 9 and 13). The second group was formed by columns referring to the coefficients of details (columns 2, 3, 4, 6, 7, 8, 10, 11, 12, 14, 15 and 16). The third group refers to outputs (column 17).

2.6.2. Shuffle

As the training algorithm is based on a gradient descent method, if the data is given in some meaningful order, this can bias the gradient and lead to poor convergence. Thus, after normalization, the NORMALIZED BASE MATRIX—NBM rows were shuffled.

The shuffling of the NBM rows was done as follows: using the MATLAB function “rand”, we created a vector whose number of elements is equal to the number of rows of the NBM. Then, applying the MATLAB function “sort” to the generated vector, we stored the indexes of the original elements positions; and finally, we used these indexes to reorder the NBM rows.

2.6.3. Database Partition

The NBM shuffled rows were used to create the datasets for adjusting the weights and bias of the neural networks. The database was divided into 60%, 30%, and 10% for training, validation and testing sets, respectively. Calling m the number of rows of the NBM, the separation sequence was defined as follows: training set—rows 1 to $0.60 m$; validation set—rows $(0.60 m + 1)$ to $0.90 m$; test set—rows $(0.90 m + 1)$ to m .

2.6.4. Selection of the Best Architecture

The RNA learning process is strongly influenced both by the number of patterns in the training set and by the free parameters (weights and biases), and also by the complexity of the problem addressed,

as shown in [16]. Although the relationship between these variables can be defined deterministically, part of the scientific community prefers to use some practical rules for this definition. In this work, the variables are restricted to the number of free parameters, i.e., to the ANN architecture, because the database's size already limits the number of training patterns and the complexity is intrinsically defined by the kind of problem handled.

In this regard, the ANN architecture selection was defined by changing the amount of neurons in the hidden layer from 3 to 15, and selecting the one with the best performance in the training validation set, as described in [1]. For each hidden layer size, the weights and bias of the neural network were initialized ten times. The maximum number of training epochs was equal to 500. After the end of the process, the number of neurons in the hidden layer was defined by the lowest average *MSE* for the initializations.

2.6.5. Selection of the Best ANN

The 10-fold cross-validation method was applied to select the best RNA for each neural model. The training, validation and test sets were defined from ten mutually exclusive partitions composed by the random division of the rows of NBM. Each experiment, in total 10 for each forecasting horizon of a specific wind speed series, was performed by setting the partition $d = 1, 2, 3, \dots, 10$ to test the ANN; 6 of the remaining partitions for training; and the last 3 for validation.

When applying this technique, it is expected that the MAE of the ANN predictions will tend to the MAE for the experiment's test set, as discussed in [15,16]. In this article, the selection of the best ANN was done based in the fold with lower MAE, averaged in ten initializations. Again, the maximum number of training epochs was equal to 500.

2.7. Development of the WTESN Model

A flow chart that illustrates the development process for the model WTESN is presented in Figure 9. For a better understanding of the process, we will explain in the following subsections the steps outlined on the diagram. All the procedures implemented in the following subsections aim to improve the generalization power of the Echo State Network.

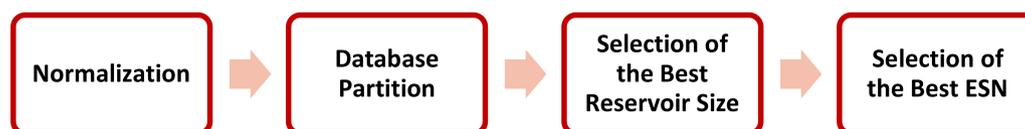


Figure 9. Development process for the WTESN model.

2.7.1. Normalization

Normalization is performed exactly as presented for the WTANNLM model.

2.7.2. Database Partition

The NBM shuffled rows were used to create the datasets for adjusting the weights of the readout. The database was divided into 90% and 10%, for training and testing sets, respectively. Calling m the number of rows of the NBM, the separation sequence was defined as follows: training set—rows 1 to $0.90 m$; test set—rows $(0.90 m + 1)$ to m .

2.7.3. Selection of the Best Reservoir Size

The best reservoir size (i.e., the number of neurons in reservoir) was defined by changing the number of neurons in the reservoir, and selecting the one with the best performance in the test set during the training process. The procedure for each location and for each forecasting horizon was performed as follows:

- I. It was created an ESN containing 100 neurons in the reservoir, with hyperbolic tangent as activation functions;
- II. Random weights were assigned to the reservoir weight matrix, and the matrix was rescaled to yield a spectral radius equal to 0.8;
- III. Supervised training was performed through the MATLAB function “train_esn”, specific function from the toolbox in [17], in offline mode and with parameter “nForgetPoints” equal to 100;
- IV. The MSE value for the test set was calculated;
- V. The number of neurons in the reservoir was increased by 100 units, it was returned to the step II, and this “loop” continued until the step IV of the reservoir with 1000 neurons was performed;
- VI. The best reservoir size was set equal to that one which presented the lowest average MSE for the test set.

After completing the steps above, the number of reservoir neurons for each of considered sites (Macau, Mossoro and Natal) was determined. For all of them, the least average MSE value was obtained with 1000 neurons in the reservoir. It was observed in all cases that increasing the number of neurons resulted in reduction of average MSE, however runtimes to training increased significantly. Some more training with a higher quantity of neurons were performed, but the reduction in average MSE was not significant.

2.7.4. Selection of the Best ESN

The next step after determining the reservoir size, which provides the final architecture of ESN, is the definition of the ESN that presents better generalization ability to forecasting wind speeds for each location and for each of the different forecasting horizons.

The best ESN was defined by changing the spectral radius of the reservoir weight matrix and the activation functions of its neurons, selecting the one with the best performance in the test set during the training process. The procedure for each location and for each forecasting horizon was performed as follows:

- I. It was created an ESN containing 1000 neurons in the reservoir, with hyperbolic tangent as activation functions;
- II. Random weights were assigned to the reservoir weight matrix, and the matrix was rescaled to yield a spectral radius equal to 0.1;
- III. Supervised training was performed through the MATLAB function “train_esn”, specific function from the toolbox in [17], in offline mode and with parameter “nForgetPoints” equal to 100;
- IV. The MSE value for the test set was calculated;
- V. The spectral radius was increased by 0.1, it was returned to the step II, and this “loop” continued until the step IV with spectral radius equal to 1.4 was performed;
- VI. It was returned to the step I changing the activation functions to logistic sigmoid (in the first time that step VI was reached) or identity (in the second time that step VI was reached);
- VII. The best ESN was set equal to that one which presented the lowest average MSE for the test set.

After completing the steps above, the best ESN for each of considered sites was determined. The average MSE are summarized in Table 3. It is presented that when utilizing hyperbolic tangent and logistic sigmoid, lower values for the MSE was obtained for spectral radii between 1.1 and 1.4. With the use of identity function, there were convergence problems for spectral radiuses greater or equal to 1 (one), and this is the reason why the last four rows of Table 3 are marked in grey color and with the text Did Not Converge (DNC). Despite this, for each one of the three sites, the best ESN (those that presented the lowest average MSE for the 24 h forecasting horizons of test sets, marked in green in Table 3), has spectral radius equal to 0.9 and uses the identity as activation function of the reservoir neurons.

Table 3. Average MSE for the test sets.

SR	MSE (m ² /s ²)								
	Macau			Mossoro			Natal		
	Tanh	Sigmoid	Identity	Tanh	Sigmoid	Identity	Tanh	Sigmoid	Identity
0.1	1.5966	1.5998	1.2776	0.9003	0.9054	0.8122	0.6989	0.7027	0.5820
0.2	1.5620	1.5697	1.2459	0.8879	0.8939	0.7854	0.6993	0.6998	0.5702
0.3	1.5420	1.5703	1.1685	0.8775	0.8956	0.7287	0.6958	0.6987	0.5499
0.4	1.5770	1.5749	1.0680	0.8661	0.8923	0.6738	0.6935	0.6987	0.5292
0.5	1.5418	1.5669	1.0190	0.8521	0.8812	0.6432	0.6936	0.7006	0.5114
0.6	1.5083	1.5404	0.9472	0.8274	0.8816	0.6116	0.6916	0.6952	0.4886
0.7	1.4666	1.5404	0.8932	0.8342	0.8780	0.5961	0.6841	0.6902	0.4666
0.8	1.4556	1.5574	0.8414	0.8309	0.8638	0.5806	0.6848	0.6892	0.4461
0.9	1.4211	1.5026	0.8136 ^a	0.8131	0.8657	0.5605	0.6849	0.6922	0.4260
1.0	1.4137	1.5137	1.1262	0.8148	0.8557	0.8164	0.6801	0.6879	0.5266
1.1	1.3824	1.4983	DNC ^b	0.8065	0.8563	DNC	0.6776	0.6907	DNC
1.2	1.3937	1.5119	DNC	0.7934	0.8567	DNC	0.6771	0.6898	DNC
1.3	1.3860	1.5070	DNC	0.8031	0.8384	DNC	0.6728	0.6857	DNC
1.4	1.3662	1.4745	DNC	0.8102	0.8397	DNC	0.6795	0.6868	DNC

^a The lowest MSE are marked in green; ^b DNC, marked in gray, means “Did Not Converge”.

The behavior of MSE for all tested ESN is presented in Figures 10–12. The rectangles colors are a Red-Green-Blue (RGB) mapping based on interpolated MSE values of their vertices, whose coordinates refer to the spectral radius (in the axis of the abscissa) and to the forecasting horizon (in the axis of the ordinates). The gray color was used to indicate that there was no convergence in at least one of the rectangle’s vertices.

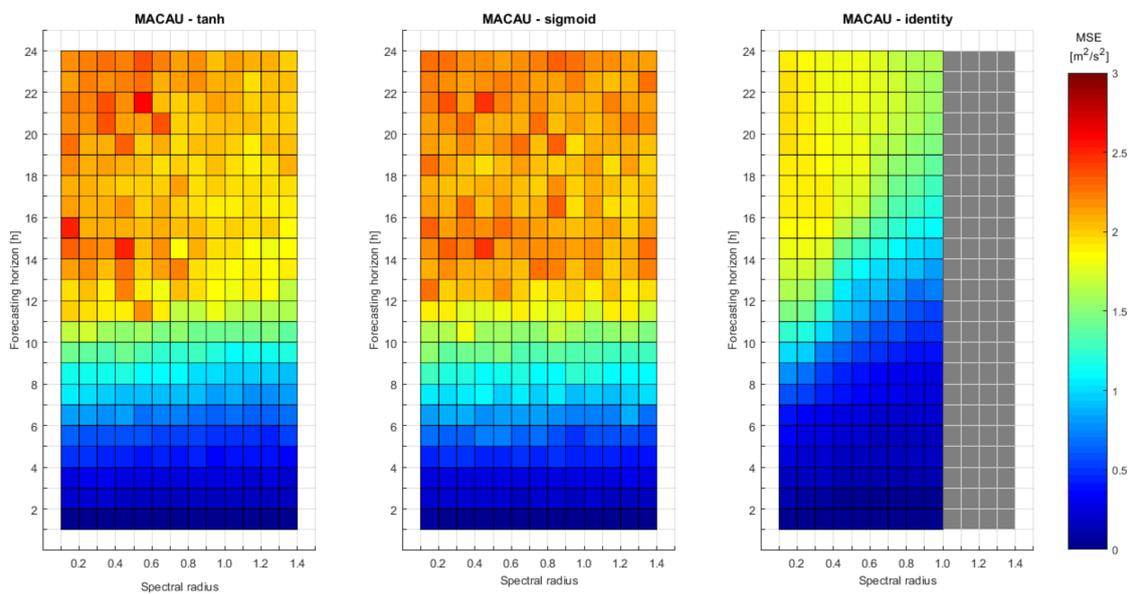


Figure 10. MSE for all tested ESN—Macau. From left to right: Hyperbolic tangent, Logistic sigmoid and Identity.

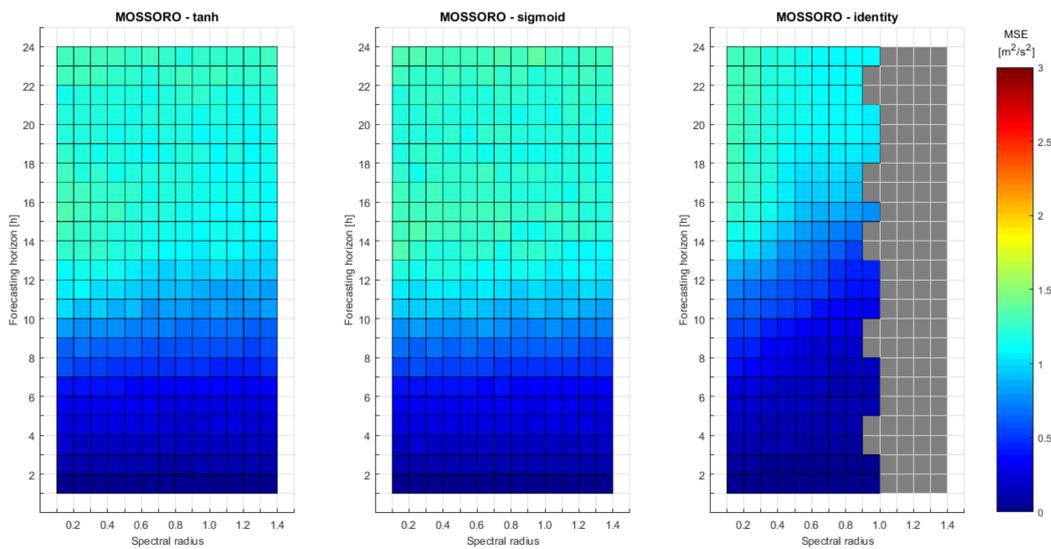


Figure 11. MSE for all tested ESN—Mossoro. From left to right: Hyperbolic tangent, Logistic sigmoid and Identity.

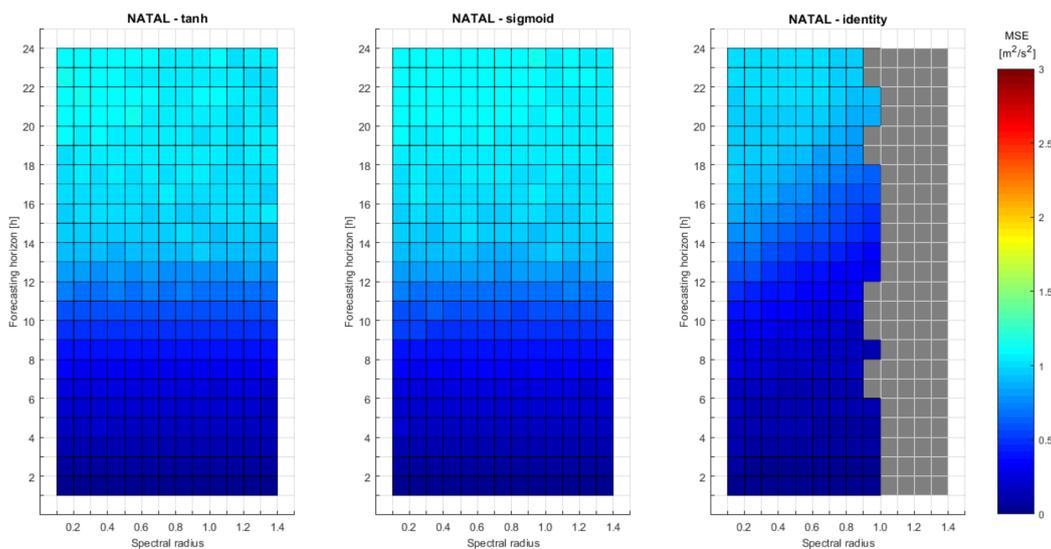


Figure 12. MSE for all tested ESN—Natal. From left to right: Hyperbolic tangent, Logistic sigmoid and Identity.

2.8. Conversion of Wind Speed to Wind Power Generation Forecasting

As the selected turbine has a hub height designed to 57 m and the wind speed series were measured at a 10 m height, and consequently the forecasts will be realized at this latter height too, it is necessary to convert the wind speed from 10 to 57 m before transforming it into wind power using the wind turbine manufacturer’s power curve. The change in speed or direction over some distance is called wind shear. According to [18], the wind shear at a ground-level surface increases the wind speed with height, as follows:

$$s_1 = s_0 \left(\frac{h_1}{h_0} \right)^\alpha \tag{7}$$

where s_0 is the wind speed measured at the reference height h_0 ; s_1 is the wind speed estimated at height h_1 ; and α is the friction coefficient (also known as Hellmann exponent), which represents the ground surface friction. In the present article, as in [1], the friction coefficient was equal to 0.1.

3. Results

As previously mentioned, the models were developed to forecast the hourly mean wind speeds with horizons ranging from 1 to 24 h ahead, for three different locations of Brazilian Northeast. These cities are: Macau, Mossoro and Natal.

The following subsections are divided in the following way: first, development of the models is presented; in sequence, comparisons between forecasts of the proposed models and PM are done. To avoid the use of several similar figures to illustrate the results of the simulations, it was chosen to present only some results for MACAU. Similar patterns and results were observed for all forecasting horizons and analyzed wind speed series.

3.1. Best Neural Networks for the WTANNLM Model

The MAE values for the test set with WTANNLM for one, six and twenty-four hours ahead are presented in Figure 13. For some initializations, due to a not adequate choice of the starting point, MAE values were higher. It occurs when the training algorithm did not converge, ending the process once in the first few training epochs.

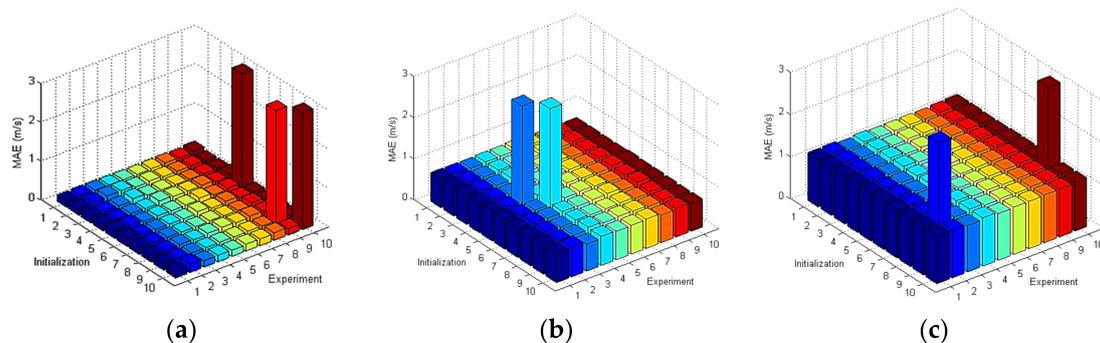


Figure 13. MAE values for the test set of Macau—(a) 1 h; (b) 6 h; and (c) 24 h ahead. Experiments are described in Section 2.6.5.

3.2. Best Echo State Network for the WTESN Model

The MAE of test sets for the best ESN are presented in Figure 14. It can be observed that for all considered locations, as expected, the error increases when increasing the forecasting horizon.

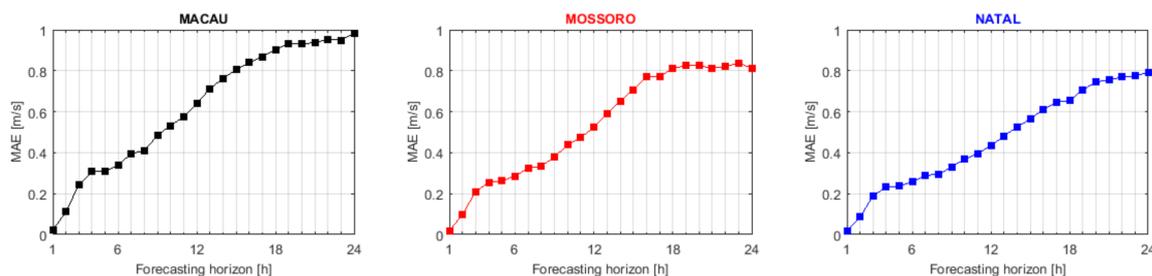


Figure 14. MAE of test sets for the best ESN. From left to right: Macau, Mossoro and Natal.

3.3. Forecasts and Comparisons between the Models

This subsection presents comparisons between the performances of the WTANNLM and WTESN models with those obtained using PM (4). The period chosen for the Macau forecasts evaluation of the models is formed by 8016 hourly average speeds, corresponding to zero hour (UTC) from 1 January 2009 to 23 (UTC) on 30 November 2009. For Mossoro, the chosen period is formed by

4416 hourly average speeds measured between 3 January 2009 and 31 August 2009. For Natal, were used 4368 hourly average speeds measured between 10 January 2009 and 31 March 2010.

The criteria for selecting the periods was based on the choice of months that had no gaps in the wind speed data. It is important to emphasize that the data for these periods have never been used for training the WTANNLM and WTESN models.

3.3.1. Wind Speed Forecasts

The MAE values of wind speed forecasts are presented in Figure 15. As can be observed, the WTESN model proposed in this article provided better forecasts than PM and WTANNLM for all look-ahead times. For lower forecasting horizons (mainly up to 12 h ahead), the average errors of forecasts with the WTESN model are significantly lower than those obtained with PM. From 12 to 24 h ahead, the average errors of the WTANNLM model hardly vary, while these errors reduce for the PM.

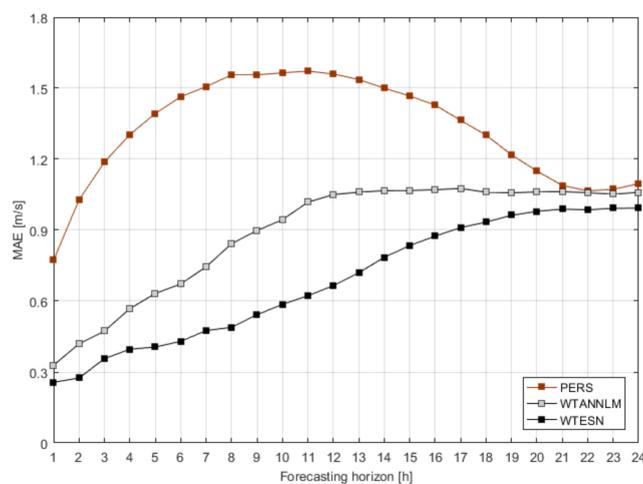


Figure 15. MAE of wind speed forecasts for Macau.

3.3.2. Power Generation Forecasts

The power curve of the wind turbine with rated capacity of 2300 kW used in this article is presented in Figure 16. As manufacturers only provide a few points of the curve, it was parametrized at four different intervals through the least squares method, enabling the power estimation at any speed between the cut-in and cut-out.

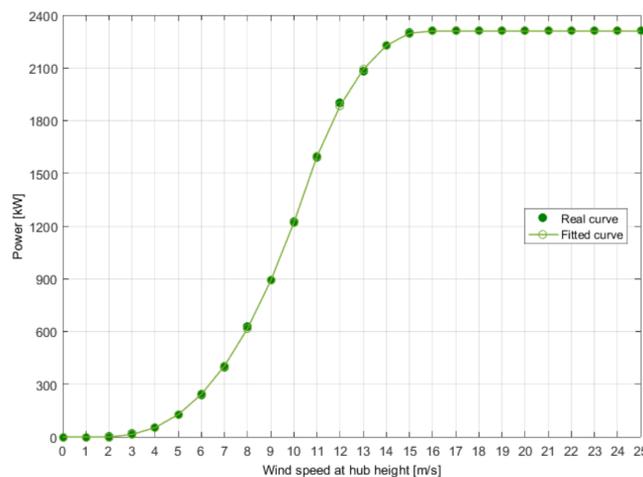


Figure 16. Wind turbine's power curve.

The NMAE for the power generation forecasts are presented in Figure 17 and its values are presented in Table 4, in which the minimum and maximum NMAE values marked in green and red, respectively. For the WTESN model, the maximum NMAE is 5.91% of the wind power installed capacity, while for the PM, the maximum NMAE is 13.13%. As expected, the minimum NMAE occurs for one hour ahead, with values equal to 1.88% for WTESN, 2.15% for WTANNLM and 4.73% for PM. The results for Mossoro and natal are also presented in Table 4.

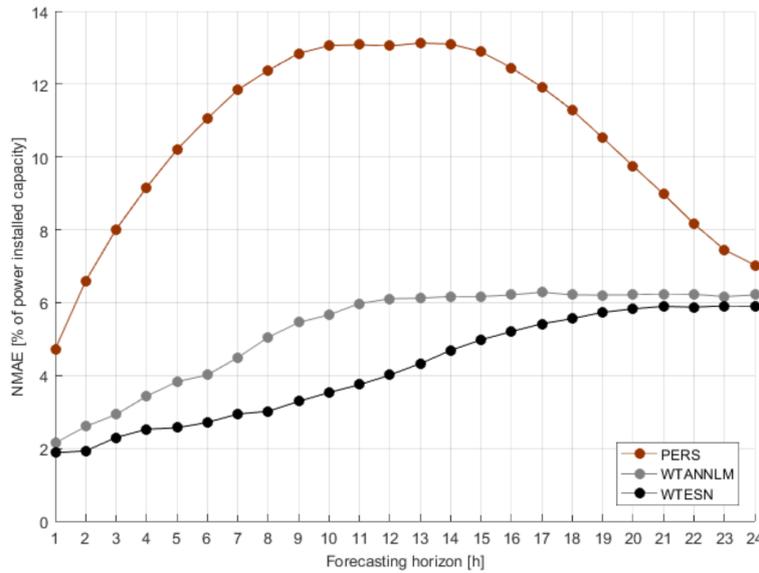


Figure 17. NMAE of wind power forecasts for Macau.

Table 4. NMAE of power generation forecasts.

Horizon (h)	NMAE (% of Power Installed Capacity)								
	Macau			Mossoro			Natal		
	(1 January 2009–30 November 2009)			(1 March 2009–31 August 2009)			(1 October 2009–31 March 2010)		
	PERS	WTANNLM	WTESN	PERS	WTANNLM	WTESN	PERS	WTANNLM	WTESN
1	4.73 ^a	2.15	1.88	0.99	0.17	0.03	3.50	0.62	0.11
2	6.60	2.60	1.93	1.32	0.36	0.14	4.34	1.36	0.61
3	8.01	2.93	2.30	1.55	0.47	0.33	5.21	1.77	1.32
4	9.16	3.43	2.53	1.75	0.57	0.40	6.14	2.22	1.62
5	10.20	3.83	2.56	1.90	0.64	0.42	6.95	2.38	1.65
6	11.06	4.01	2.71	2.04	0.68	0.45	7.67	2.65	1.80
7	11.84	4.49	2.94	2.17	0.74	0.51	8.28	3.02	1.98
8	12.38	5.05	3.02	2.27	0.82	0.53	8.81	3.35	2.02
9	12.84	5.46	3.29	2.33	0.87	0.59	9.22	3.78	2.28
10	13.06	5.67	3.53	2.36	0.93	0.63	9.41	4.06	2.47
11	13.08	5.98	3.75	2.37	0.99	0.65	9.55	4.47	2.64
12	13.05	6.11	4.01	2.38	1.01	0.70	9.63	4.82	2.88
13	13.13	6.12	4.33	2.38	1.03	0.77	9.65	4.85	3.19
14	13.10	6.17	4.69	2.39	1.10	0.81	9.56	4.95	3.51
15	12.89	6.17	4.98	2.36	1.11	0.87	9.38	4.87	3.86
16	12.45	6.21	5.21	2.30	1.10	0.91	9.09	4.93	4.08
17	11.92	6.29	5.42	2.20	1.14	0.95	8.64	4.91	4.30
18	11.28	6.21	5.56	2.09	1.16	0.98	8.14	4.89	4.50
19	10.53	6.21	5.73	1.97	1.14	1.01	7.57	4.86	4.68
20	9.75	6.22	5.83	1.87	1.11	1.02	7.00	4.86	4.77
21	8.98	6.24	5.90	1.76	1.12	1.05	6.48	4.90	4.80
22	8.17	6.23	5.88	1.60	1.14	1.04	6.00	4.88	4.88
23	7.45	6.17	5.91	1.45	1.13	1.03	5.72	4.87	4.89
24	7.03	6.21	5.91	1.38	1.15	1.05	5.58	4.85	4.92

^a The lowest NMAE are marked in green; ^b The highest NMAE are marked in red.

3.3.3. Improvements of Power Generation Forecasts

The improvements of NMAE (5) regarding PM (4) for the power generation forecasts are presented in Figure 18, in which can be observed that for forecasts up to 15 h ahead, improvements of the WTESN model are greater than 60%.

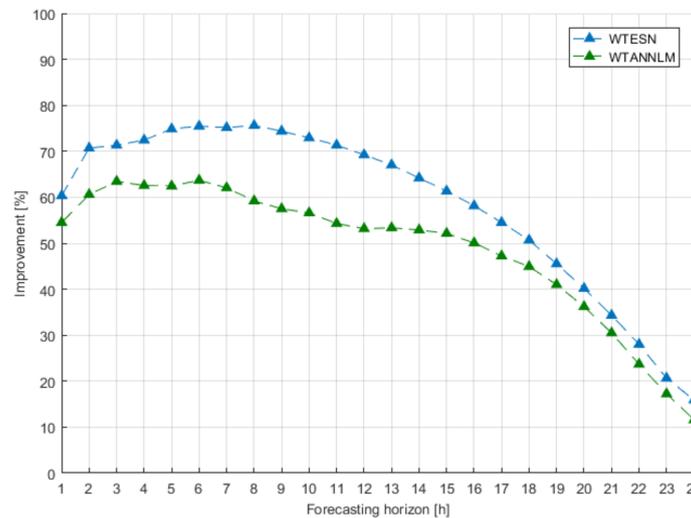


Figure 18. Improvements of NMAE with respect to PM of wind power forecasts for Macau.

4. Discussion

The models developed in this article are based on preprocessing the wind speed time series through multiresolution analysis by WT, with posterior wind speed forecasting by ESN and power forecasting obtained via wind turbine's power curve. The presented forecasting horizons fall within short-term forecasts up to twenty-four hours ahead, which is appropriate to support the operation planning of electrical power systems, since the startup of a thermal power plant must be planned in advance and the starting time varies from one plant to another.

According to [19], for wind farms in flat terrain, typical NMAE for wind power forecasting is around 6% of the installed capacity for the first 6 h, rising to 9–11% for 48 h ahead.

In [20], the NMAE is around 15% for six hours ahead, while in [21], the NMAE is around 9% at the same forecasting horizon. In [22], NMAE values for predictions from 1 to 4 h ahead are 5.98%, 9.47%, 12.00% and 15.21%, respectively.

The average NMAE value for 3 h ahead with the hybrid model proposed in [23] is equal to 1.65%, and the average NMAE value for the PM is 5.24%. In [24], the NMAE value varies from 2 to 16% for forecasting horizons between 1 to 24 h ahead, respectively.

In [25], the authors present one hour ahead wind power forecasts based on random forests—RF for a one-year period. Results were compared with PM and ANN. The best NMAE values presented are equal to 3.49%, 3.67% and 4.25% for RF, ANN and PM, respectively.

The daily average NMAE for hourly wind power forecasts obtained with the proposed approach in [26] was equal to 0.48%, while for the PM, NMAE was equal to 1.01%. Care should be taken when analyzing such low values for NMAE. The authors in [26] did not emphasize, however, according to the graphs presented in that article, it can be observed that for a great part of the time wind power generation remains at very low or zero values. Thus, forecasting errors naturally tends to zero, as can be seen in the reduced NMAE for the PM.

The results obtained with WTESN are promising when compared with the state-of-the-art forecasters. As can be calculated by Table 4, the average NMAE of first six, twelve, eighteen and twenty-four hourly power generation forecasts are 2.32%, 2.87%, 3.59% and 4.16%, respectively. The average improvement of

NMAE for the first six, twelve, eighteen and twenty-four hourly power generation forecasts are 70.87%, 71.99%, 67.77% and 58.52%, respectively.

5. Conclusions

When compared to PM and a model proposed in literature that uses ANN, the WTESN model provided better forecasts for all the look-ahead times considered, especially for the shorter ones (up to 12 h ahead). The average improvement of NMAE for the horizon of twelve hours ahead was above 70%. These results of improvements in relation to the PM are among the best published results to date in the field of wind power generation forecasts. Here is observed that the forecasts quality is strongly influenced by the autocorrelation series for the PM because it does not use WT.

Decomposition of wind speed by WT allow models to extract relevant information about the series, which significantly improve the forecasts when the decomposed signals are used as inputs.

Through the analysis, it was found that the behavior of forecast errors for different horizons was similar for all the three sites (as presented in Table 4). The adopted methodology was adequate, assuring statistically reliable forecasts, that is, the models acquired generalization capacity.

When comparing the performance of ANN and ESN, the results show that both are powerful time series forecasting tools, but the ESN proved to be better suited for wind power forecasting.

Author Contributions: All authors conceived of the presented idea. Hugo Tavares Vieira Gouveia developed the theory and performed the computations. Aida Araújo Ferreira and Ronaldo Ribeiro Barbosa de Aquino verified the analytical methods as well as encouraged Hugo Tavares Vieira Gouveia to investigate the use of ESN for wind speed and power forecasting. All authors discussed the results and contributed to the final manuscript.

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

References

- De Aquino, R.R.B.; Gouveia, H.T.V.; Lira, M.M.S.; Ferreira, A.A.; Neto, O.N.; Carvalho, M.A. *Models Based on Neural Networks and Neuro-Fuzzy Systems for Wind Power Prediction Using Wavelet Transform as Data Preprocessing Method*; Communications in Computer and Information Science Series; Springer: Berlin/Heidelberg, Germany, 2012; pp. 272–281.
- Venayagamoorthy, G.K.; Rohrig, K.; Erlich, I. One Step Ahead: Short-Term Wind Power Forecasting and Intelligent Predictive Control Based on Data Analytics. *IEEE Power Energy Mag.* **2012**, *10*, 70–78. [[CrossRef](#)]
- Milligan, M.; Kirby, B.; Acker, T.; Ahlstrom, M.; Frew, B.; Goggin, M.; Lasher, W.; Marquis, M.; Osborn, D. *Review and Status of Wind Integration and Transmission in United States: Key Issues and Lessons Learned*; Technical Report; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2015.
- Liu, H.; Tian, H.; Pan, D.; Li, Y. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. *Appl. Energy* **2013**, *107*, 191–208. [[CrossRef](#)]
- Saleh, A.E.; Moustafa, M.S.; Abo-Al-Ez, K.M.; Abdullah, A.A. A hybrid neuro-fuzzy power prediction system for wind energy generation. *Int. J. Electr. Power Energy Syst.* **2016**, *74*, 384–395. [[CrossRef](#)]
- Samet, H.; Marzbani, F. Quantizing the deterministic nonlinearity in wind speed time series. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1143–1154. [[CrossRef](#)]
- Doucoure, B.; Agbossou, K.; Cardenas, A. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renew. Energy* **2016**, *92*, 202–211. [[CrossRef](#)]
- De Aquino, R.R.B.; Albuquerque, J.C.; Neto, O.N.; Ferreira, A.A.; Neto, A.C.; Lyra, M. Assessment of power curves in models of wind power forecasting. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, BC, Canada, 24–29 July 2016; pp. 3915–3922.
- De Aquino, R.R.B.; Souza, R.B.; Neto, O.N.; Ferreira, A.A.; Lyra, M. Echo state networks, artificial neural networks and fuzzy systems models for improve short-term wind speed forecasting. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2015), Killarney, Ireland, 12–17 July 2015; pp. 1–8.
- Ferreira, A.A.; Ludermir, T.B.; de Aquino, R.R.B. An approach to reservoir computing design and training. *Expert Syst. Appl.* **2013**, *40*, 4172–4182. [[CrossRef](#)]

11. Chitsazan, M.A.; Fadali, M.S.; Nelson, A.K.; Trzynadlowski, A.M. Wind speed forecasting using an echo state network with nonlinear output functions. In Proceedings of the 2017 American Control Conference (ACC), Seattle, WA, USA, 24–26 May 2017; pp. 5306–5311.
12. Qiao, J.; Li, F.; Han, H.; Li, W. Growing Echo-State Network with Multiple Subreservoirs. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 391–404. [[CrossRef](#)] [[PubMed](#)]
13. Yildiz, B.; Jaeger, H.; Kiebel, S.J. Re-visiting the echo state property. *Neural Netw.* **2012**, *35*, 1–9. [[CrossRef](#)] [[PubMed](#)]
14. Madsen, H.; Kariniotakis, G.; Nielsen, H.A.; Nielsen, T.S.; Pinson, P. A Protocol for Standardizing the Performance Evaluation of Short-Term Wind Power Prediction Models. 2004. Available online: <https://pdfs.semanticscholar.org/1a48/a5ba3159aa938e268cdf8955fff7b8642a72.pdf> (accessed on 2 April 2018).
15. Prechelt, L. Proben1: A Set of Neural Network Benchmark Problems and Benchmarking Rules. Available online: <https://publikationen.bibliothek.kit.edu/39794/2050> (accessed on 2 April 2018).
16. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
17. Jaeger, H. Simple Matlab Toolbox for Echo State Networks. Available online: <http://organic.elis.ugent.be/node/129> (accessed on 1 September 2015).
18. Patel, M.R. *Wind and Solar Power Systems: Design, Analysis, and Operation*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2006.
19. Giebel, G.; Brownsword, R.; Kariniotakis, G.; Denhard, M.; Draxl, C. The State of the Art in Short-Term Prediction of Wind Power a Literature Overview. Available online: http://orbit.dtu.dk/files/128933990/GGiebelEtAl_StateOfTheArtInShortTermPrediction_ANEMOSplus_2011.pdf (accessed on 2 April 2018).
20. Miranda, V.; Bessa, R.; Gama, J.; Conzelmann, G.; Botterud, A. New Concepts in Wind Power Forecasting Models. In Proceedings of the WindPower 2009 Conference and Exhibition Centre, Chicago, IL, USA, 4 May 2009.
21. Sideratos, G.; Hatziargyriou, N.D. An Advanced Statistical Method for Wind Power Forecasting. *IEEE Trans. Power Syst.* **2007**, *22*, 258–265. [[CrossRef](#)]
22. Kusiak, A.; Zheng, H.; Song, Z. Short-Term Prediction of Wind Farm Power: A Data Mining Approach. *IEEE Trans. Energy Convers.* **2009**, *24*, 125–136. [[CrossRef](#)]
23. Catalão, J.P.S.; Pousinho, H.M.I.; Mendes, V.M.F. Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. *IEEE Trans. Sustain. Energy* **2011**, *2*, 50–59.
24. Xydas, E.; Qadrdan, M.; Marmaras, C.; Cipcigan, L.; Jenkins, N.; Ameli, H. Probabilistic wind power forecasting and its application in the scheduling of gas-fired generators. *Appl. Energy* **2017**, *192*, 382–394. [[CrossRef](#)]
25. Lahouar, A.; Slama, J.B. Hour-ahead wind power forecast based on random forests. *Renew. Energy* **2017**, *109*, 529–541. [[CrossRef](#)]
26. Zheng, D.; Shi, M.; Wang, Y.; Eseye, A.T.; Zhang, J. Day-Ahead Wind Power Forecasting Using a Two-Stage Hybrid Modeling Approach Based on SCADA and Meteorological Information, and Evaluating the Impact of Input-Data Dependency on Forecasting Accuracy. *Energies* **2017**, *10*, 1988. [[CrossRef](#)]

