Received: 8 July 2021 | Revised: 19 January 2022 | Accepted: 22 January 2022

DOI: 10.1002/er.7728

RESEARCH ARTICLE

INTERNATIONAL JOURNAL OF ENERGY RESEARCH WILEY

Methodology for integration of wind resource forecasts based on artificial neural networks

Tatiane C. Carneiro[1] | **Marcello A. Ferreira Batista Lima**[2] | **Paulo C. Marques de Carvalho**[3] | **Josias Guimarães Batista**[2] | **Luis M. Fernández-Ramírez**[4]

[1]Environmental Engineering Course, Federal University of Maranhão, São Luís, Brazil

[2]Department of Industrial Mechatronics, Federal Institute of Education, Science and Technology, Limoeiro do Norte, Brazil

[3]Electrical Engineering Department, Federal University of Ceará, Fortaleza, Brazil

[4]Research Group in Sustainable and Renewable Electrical Technologies (PAIDI-TEP023), Department of Electrical Engineering, University of Cadiz, ETSI Algeciras, Algeciras, Spain

Correspondence
Tatiane C. Carneiro, Environmental Engineering Course, Federal University of Maranhão, São Luís, Brazil.
Email: tatianecarolyne@ufma.br

Funding information
National Council for Scientific and Technological Development (CNPq)

Summary

An adaptation of the portfolio theory (PT) is proposed in this article, denoted as PrevPT, "*Previsão*" (in Portuguese) by PT, to integrate the three artificial neural networks, namely multilayer perceptron (MLP) backpropagation, radial basis function (RBF), and self-organizing map (SOM), based forecasting techniques, aiming to analyze the impact of wind speed forecasting errors and achieve more accurate results. In its first use, the PT goal was to maximize a financial return, at any risk, through the diversification of securities or investments that are not positively correlated. Based on the development of PrevPT, which was used until this work only for solar forecasting, the proposed technique is applied in this paper to integrate and improve the results of individual wind forecasts. Four-year wind speed data (January 2007 to December 2010) from two different locations (Algeciras, Spain and Petrolina, Brazil) were used. Our methodology develops a topology that integrates the forecasts obtained by MLP, RBF, and SOM aiming to obtain smaller forecast errors. By diversifying the forecasted asset, when one of the assets has negative prediction errors, another compensates for them and, thus, the total or partial cancellation of the errors occurs. PrevPT obtains a mean absolute percentage error of 1.13% for Spain and 2.35% for Brazil. PrevPT surpassed the results obtained by the three techniques applied individually in the two locations. The main innovations of the methodology are the significant reduction of errors and optimization of resource planning, and the beneficial features compared to other predictor integration techniques.

KEYWORDS

portfolio theory, wind energy forecast, wind energy potential

1 | INTRODUCTION

Among the main renewable energy sources capable of meeting the global electricity demand, wind energy is an important component of society's path toward sustainable development. However, intermittency is inherent to wind resources and should be considered to operate electrical systems reliably and adequately.[1-3] In this context, wind energy forecasting can be relevant to energy imbalance markets, reducing backup resources.

Advanced wind speed forecasting techniques have the potential to improve the use of resources, reduce transmission problems, and contribute to the integration of wind farms in the energy market, contributing to the balance between supply and demand.[4] Accuracy in the predictions of electricity production from intermittent sources is extremely necessary: a forecast error of 2 GWh from intermittent sources in Germany resulted in an additional cost of 2.20 euros per MWh.[5] Forecasting models are commonly used to predict other types of series, such as electricity demand, solar irradiation, and rainfall.[6,7]

Accurate wind speed forecasting has been highlighted as a significant reducer of intermittency impacts.[8] Errors of forecasting techniques applied individually are often higher than errors found from the integration of techniques (hybrid models) or from the use of tools such as portfolio theory (PT).[9] Several methodologies that integrate predictions from individually implemented models have been found in the literature, including prediction errors to improve the results,[9] some set models mixed standard models to achieve a higher performance final product,[10] some set methods applied directly,[11,12] and others combined the results obtained separately by different models.[13]

Considering wind energy, "*Previsão*" (in Portuguese) by PT (PrevPT) allows to integrate different prediction methodologies and compensates for forecasting errors. When one methodology obtains high forecasting errors, the other methodology compensates them through predefined weights.[9] PrevPT is divided into the following stages: data collection, wind speed forecasting using individual methodologies, forecasting error study, data processing of forecasting errors by PrevPT, the final definition of assets weighting by PrevPT, and forecasting error verification using the methodologies integrated by the proposed methodology in comparison with individual assets. Hence, our objective is to apply PrevPT to integrate multilayer perceptron (MLP) backpropagation, radial basis function (RBF), and self-organizing map (SOM) to perform wind speed forecasting with high precision, using data from two sites: Petrolina, Brazil and Algeciras, Spain.

We propose a methodology that integrates, to obtain smaller forecast errors, the forecasts obtained by artificial neural networks (ANNs) defined as MLP, RBF, and SOM. Through an adapted topology and by proposing interconnected points, these minor prediction errors are achieved. PrevPT is fed by the outputs of each wind forecaster, making from then on, an optimal weighting of the forecasters, seeking a reduction in forecast errors.

The cancellation whether total or partial of the errors is obtained with the variability of assets (when one of them becomes negative, the other compensates it). Hence, the innovative aspects of our application can be summarized as follows.

1. First-time PrevPT was applied to wind data.
2. The use of PrevPT integrating three different ANNs for wind speed forecasting, resulted in significant error reduction and optimization of resource planning.
3. It is tested at two locations with different latitudes and consequently different climatic profiles and presented good performance, showing that it is a viable methodology worldwide.
4. The application of a technique commonly applied in the economic sector to optimize intermittent resource forecasting methods.
5. The ease of replicability and simplicity is also a distinguishing feature of the solution proposed in this work.

The PrevPT application proposed here is relevant for its low cost and easy applicability when compared, for example, to the Bayesian model averaging (BMA) and decision trees; in addition to obtaining better performance in relation to the forecasting methodologies applied individually.[9]

Wind predictability needs to be improved, with positive consequences for the impacts generated on the electricity grid by the intermittence of the aforementioned source. The correct forecast of this resource facilitates the management of the electricity network, with an improvement in the approximation between consumption and generation. Considering that the installation of PrevPT has a low computational cost and, consequently, low hardware cost, the technique presented in this study can contribute significantly to the electricity sector.

The paper is structured as follows: Section 2 introduces the state-of-the-art corrective approaches using time series forecasting. Section 3 describes the proposed methodology for wind-speed forecasting. Section 4 illustrates the results and discussion of the results. Finally, Section 5 presents the conclusions of the study.

2 | STATE-OF-THE-ART OF CORRECTIVE APPROACHES USING TIME SERIES FORECASTING

Different approaches can be found in the literature to develop forecasting strategies with great precision using the errors of individual forecasting. A corrective approach using solar forecasting errors was proposed in Reference 14. The proposal was tested against 24-hour solar

persistence forecasts with 1 month of PV substation generation and load data. The goal was to correct errors, reduce the impacts of solar variability, and improve the leveling of the net load in buildings in Southern California. Intelligent electric vehicle programming was proposed using quadratic programming to optimize and support the approach. A receding horizon optimization algorithm was used to modify the control action in relation to the expected PV generation and satisfy the restrictions in a fixed-length time window (24 hours). In conclusion, the ideal load strategy and accurate hand forecasts reduced the net load variability.

A forecasting hybrid system combining PT and different computational intelligence techniques was illustrated in Reference 9. PT was used to integrate and improve the results of individualized forecasts. For applications in Spain and Brazil, the mean absolute percentage error (MAPE) obtained by PrevPT was 1.56%, in both cases, below the MAPE obtained by the deep learning (DL) model. An adjustment approach to improving combined forecasting and predict ramp events more accurately was found in Reference 15. Ensemble learning was applied as a post-processing approach to minimize the residuals of the forecast ramp rates by combining different forecasts and their rates of increase.

A heuristic approach to improve and correct the errors of isolated load forecasting in power systems through trajectory tracking stability theory was proposed in Reference 16. The method was able to converge the forecast error into the theory and improve the forecasting performance. An autoregressive model adjusted by data to capture the stochastic behavior of forecast errors for the next day and produce autocorrelated time series forecasts was used in Reference 17. In addition, the authors sought to understand how these errors can influence storage performance. A simulation comparing the next day's solar forecast errors in regions with different geographic sizes can be found in Reference 18. A transformed multivariate autoregressive moving average (ARMA) model was used in Reference 19 to model hourly average forecasting errors for the next day.

The probabilistic ensemble model output statistics (EMOS) technique was developed in Reference 20 for local adaptation and correction of variations in the sample set, which can be super or sub-dispersed in grid irradiance estimates derived from satellite and reanalysis. The technique was superior in both probabilistic and deterministic estimates. BMA was implemented in Reference 21 to predict PV power in different time horizons. BMA post-processing was performed for the numerical weather prediction (NWP) assemblies. It was applied to 11 utility-scale PV plants in Texas. The methodology reduced the sub-dispersion of the raw set and improved the forecast calibration with FSS by 2% to 36% when compared to a persistence set, the raw NWP set, and one state-of-the-art post-processed EMOS parametric forecasting.

Liu et al[22] developed a spatial-temporal set DL model called the Bayesian convolutional recurrent unit (VB-ConvGRU). In the application, the DL model was used, and the advantages of the recurrent unit with block (GRU) and convolutional neural network (CNN) were combined for the prediction of solar irradiation.

EMOS was implemented in Reference 23 to provide locally calibrated and spatially coherent probabilistic forecasts of the hot-dry-windy index. Forecast scenarios were generated via equidistant quantile ECC-Q variations. The authors concluded that post-processing is beneficial during all seasons of the year for forecast horizons of up to 2 weeks.

In Reference 24, an ensemble learning method that integrates the low-rank matrix completion model and the ridge regression model to predict anticancer drug response in cancer cells was implemented (Table 1). The authors used two datasets (Cancer Cell Line Encyclopedia and Genomics of Drug Sensitivity in Cancer).

3 | METHODOLOGY

3.1 | Study area and data collection

The Photovoltaic Geographical Information System was used to collect data from the city of Algeciras.[25] Anemometric data from Petrolina (Brazil) were obtained from the *Sistema de Organização Nacional de Dados Ambientais* database. This system is part of a project of the *Instituto Nacional de Pesquisas Espaciais* to implement physical infrastructure and human resources to raise and improve the database of solar and wind energy resources in Brazil.[26]

The data representation presented in this section follows the methodology proposed in Reference 27. In a total of 35 064 hourly values (4 years of data), covering the period January 2007-December 2010, the wind resource data were obtained. Using the year 2007 as a reference, Figures 1 and 2 show the wind speed data in the sites throughout the 12 months of the year and 24 hours a day.

Algeciras had the highest wind speed values between March and May. When the daily patterns were analyzed, the site had the highest average wind speed between 10 and 18 hours in all the observed months. In months

TABLE 1 Summary of the state-of-the-art corrective approaches using time series forecasting

References	Forecast goal	Approaches involved	Corrective approach or post-processing technique
14	PV power	Persistence, receding horizon optimization algorithm	Quadratic programming
9	Solar resource	MLP; RBF; support vector regression (SVR); e deep learning (DL)	Portfolio theory
15	Ramp events	Multiple linear regression (MLR), ANN, or support vector regression (SVR)	Ensemble learning
16	Load	ARMA and neural networks using backpropagation algorithm (BPNN)	Trajectory tracking stability theory
17	Energy storage; wind power	Monte Carlo	Autoregressive processes
18	PV power	Multivariate vector autoregressive moving average (VARMA); vector autoregressive (VAR)	Transformation-based methodology
19	Regional PV power production	—	ARMA
20	Grid irradiance forecasting derived from satellite and reanalysis	—	EMOS
21	PV power	NWP	BMA
22	Solar irradiation	DL, GRU, and CNN	VB-ConvGRU
23	Probabilistic forecasts of the hot-dry-windy index	Ensemble copula coupling (ECC)	EMOS
24	Anticancer drug-response	Low-rank matrix completion model and the ridge regression model	Ensemble learning method

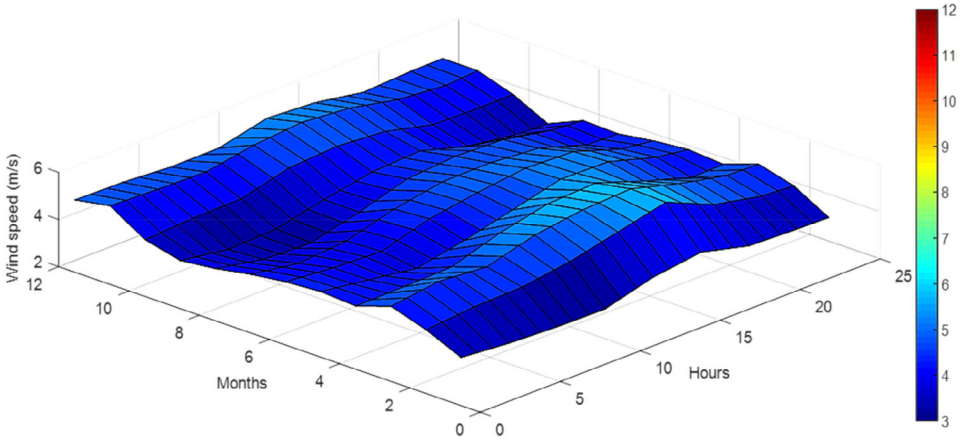

FIGURE 1 Behavior of the wind speed in Algeciras during the 12 months of the year and 24 hours a day (2007)

with the highest wind speed averages, values between 5 and 6 m/s were found.

Petrolina had the highest wind speed values between July and November. The site was characterized by low variations during the day in all months. During most of the second semester, a small reduction in values was observed throughout the morning, while during the rest of the day, values greater than 7 m/s were found, indicating low variability.

Figure 3 shows the characteristic days for all months in 2007 in Algeciras. In March, the site had the highest average wind speed.

Figure 4 shows the characteristic days for all months of 2007 in Petrolina. September had the highest average wind speed. In the first half of the year, wind speeds tended to have lower average values.

Understanding the behavior of wind resources and forecasting methodologies is relevant to help reduce the

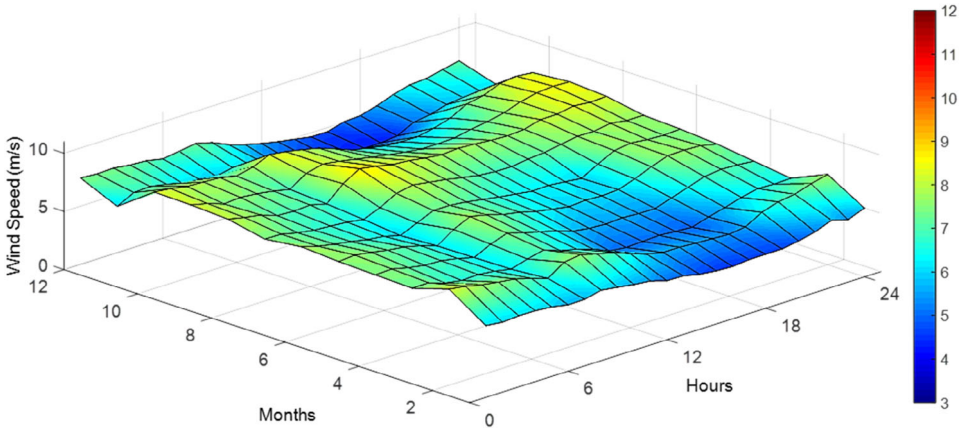

FIGURE 2 Behavior of the wind speed in Petrolina during the 12 months of the year and the 24 hours a day (2007)

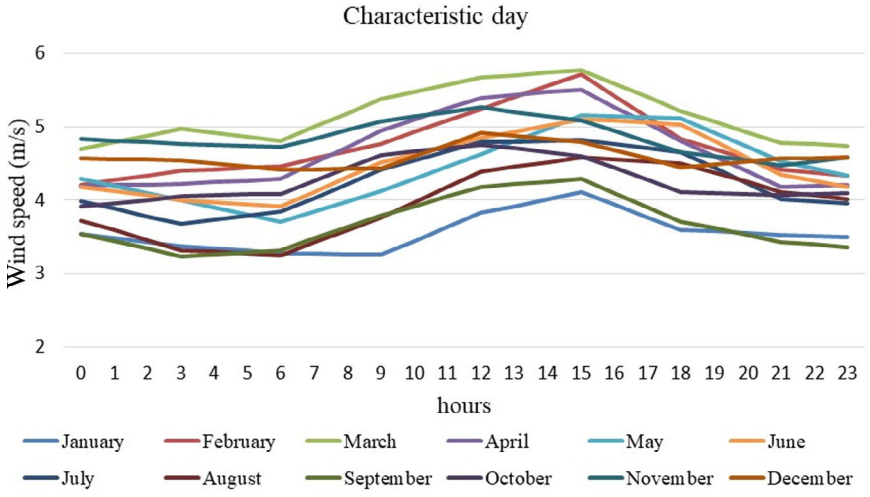

FIGURE 3 Characteristic days for all months of the year 2007 in Algeciras

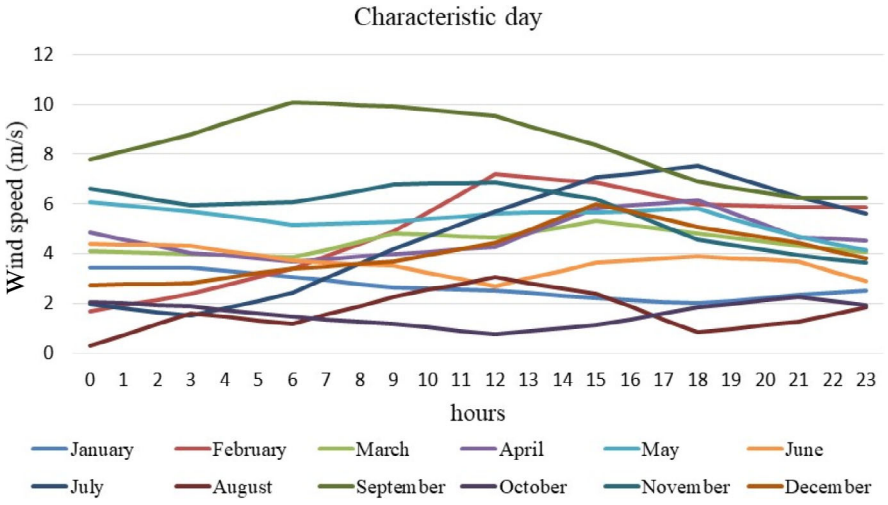

FIGURE 4 Characteristic days for all months of the year 2007 in Petrolina

negative effects of intermittent sources, such as wind-generated on the electricity grid. Among the main effects we can list: no controllability of the generation due to being dependent on wind speed, variability in electrical generation, need for the operator to be able to define, and prepare energy reserves to circumvent potential variations between the expected load and the actual load, among others.

The next sections will present the techniques applied individually, and then the proposed integration methodology to improve the accuracy of the wind forecast developed here. The PrevPT applied to Spain

managed to reduce the forecast developed in this article were implemented using MATLAB R2015b Software and also Microsoft Excel.

3.2 | Individual techniques

We used three techniques to forecast wind speed in Algeciras (Spain) and Petrolina (Brazil): MLP backpropagation, RBF, and SOM. The main differences between the three types of ANNs are presented in Table 2.

Data pre-processing was performed using the three techniques mentioned above, involving the form of acquisition, pre-processing, normalization, training, and test set definition. Data were normalized using linear normalization, according to Equation (1), leaving the data between 0 and 1, so that they could be applied to the activation functions properly.

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}}, \tag{1}$$

where X_n is the normalized value, X_i is the observed value, X_{min} is the smallest value, and X_{max} is the largest value of the dataset.

To define which inputs are beneficial for wind predictability, several simulations and different combinations of time-delayed inputs were performed, in addition to the application of different numbers of neurons in the intermediate layers. From the results, the inputs and quantities of neurons with the smallest errors in each application were defined.

3.2.1 | Self organizing map

The main objective of SOM is to assemble vectors with similar characteristics in the same class (winning neuron) or in similar classes (neighboring neurons).[28] SOM architectures contain a multidimensional input layer and an output layer that is usually one-dimensional or two-dimensional. In the output layer (competitive layer), neurons compete and only one of them is considered the winner, that is, the most suitable class for a given input vector x. Additionally, each input vector element is connected to all output layer elements, and the connection strength is measured by means of weights w_{ij} between the input neurons j and the output layer neurons i.

The Euclidean distances DI_i between the input vector and the weights connected to each of the output neurons are calculated in the SOM model training phase, according to Equation (2).

$$DI_i = \sqrt{\sum_{j=1}^{J} \left(x_j - w_{ij}\right)^2}; \text{ para } i = 1, 2, ..., M, \tag{2}$$

where x_j is the jth component of the input vector $\mathbf{x}$, J is the dimension of the input vector x, and M is the total number of neurons in the output layer.

The winning neuron is the output neuron i, which shows the shortest DI_i to the input vector. The weights connected to this neuron i^* and neurons within a Vi^* neighborhood radius are then updated by Kohonen's rule,[29] according to Equation (3).

$$W_{ij}(n) = W_{ij}(n-1) + \propto \cdot \left[x_j(n) - W_{ij}(n-1)\right]; \\ \text{para } i \in Vi * \text{e} j = 1, 2, ..., J, \tag{3}$$

TABLE 2 Summary of the main differences between the three artificial neural networks (ANNs) applied individually

	RBF	MLP	SOM
Multiple layers	Yes	Yes	No
Type of learning	Supervised	Supervised	Competitive learning
Parameters	Number of radial basis functions, initial center and scattering, rates, and weights	Weights, rates, and thresholds	Weights, rates, and dimensions of the competitive layer
Process of building the model	One stage	Two stages	One stage (divided into initialization, competition, cooperation, and adaptation)
Number of hidden layers	Only one	One or more (can be defined)	Does not contain hidden layers (only input layer and output layer or competitive layer)
Learning time	Fast	Slow	Slow

where $\propto$ is the learning rate (rates of 0.25% and 0.5% were used for Spain and Brazil, respectively), and n is the index representing the sequence of sample presentation to the network.

Kohonen's rule forces the weights attached to the winning neuron and its neighbors to move in the direction of the network input vector, reducing DI_i values and allowing these neurons to learn to classify similar vectors. SOM learning occurs without an external supervisor, who presents the desired outputs to the network. To determine the neighborhood, the distances between the output layer neurons can be defined in several ways[29]; we used the square-based bubble function.[30]

In the first training phase, called ordering, we used 60% of the iterations. The neighborhood radius was started at a given distance that decreased to the unit value. Such measures allow the neuron weights to be organized in the entry space according to their positions. In the second training phase, called an adjustment, we used the remainder of the number of iterations. At this stage, the neighborhood radius is equal to one. Hence, only the winning neuron weights are updated. During the second phase, the weights are expected to change uniformly in the entry space, maintaining the topology defined in the first phase.[29]

After the training, SOM can be used as a forecasting tool.[31] The input vector should be considered in the absence of the variable to be predicted, following these steps:

1. Calculate *the* DI_i between the input vectors and weights connected to the output neuron disregarding element j to be predicted. This step can be performed by including a Boolean variable m_j, according to Equation (4). This variable was used to include ($m\_j = 1$) or exclude ($m\_j = 0$) the contribution of a given element j from the input vector.
2. Determine the winning neuron based on the shortest DI_i.
3. The weight of the winning neuron attached to the missing element j of the input vector was used as the forecasting.

$$DI_i = \sqrt{\sum_{j=1}^{J} m\_j \left(x_j - w_{ij}\right)^2};\ \text{para}\ i = 1, 2, ..., M. \quad (4)$$

Our applications for Brazil and Spain present their input layer vectors with five neurons (four delays), representing the hourly wind speed values in $V(t-4)$, $V(t-3)$, $V(t-2)$, $V(t-1)$, $V(t)$, and output $V(t+1)$. A two-dimensional output layer composed of 25 neurons was used, as shown in Figure 5. Training occurred in sequential mode.

3.2.2 | Multilayer perceptron

MLP is used as an alternative strategy to solve nonlinear problems, which neural networks such as single-layer perceptron and adaline could not solve.[30] These nonlinearities are incorporated into the neural models through the activation functions (non-linear) of each neuron and the composition of its structure in successive layers. MLP is a perceptron neural network with at least one intermediate layer that has high connectivity and uses nonlinear and differentiable activation functions. The MLP architecture used for wind forecasting in Spain is shown in Figure 6.

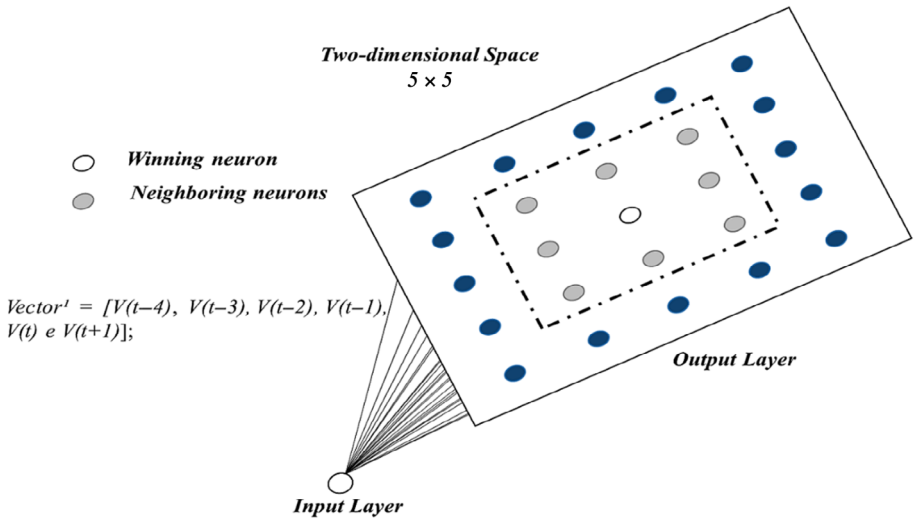

FIGURE 5 Self-organizing map (SOM) architecture for wind speed forecasting in Spain and Brazil

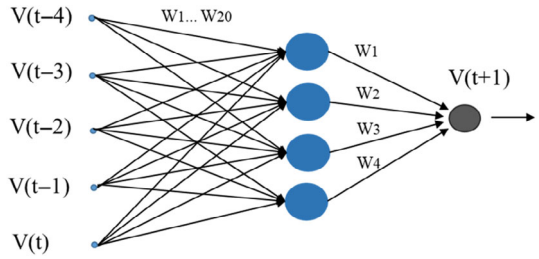

FIGURE 6 Multilayer perceptron (MLP) architecture for wind forecasting in Algeciras (Spain)

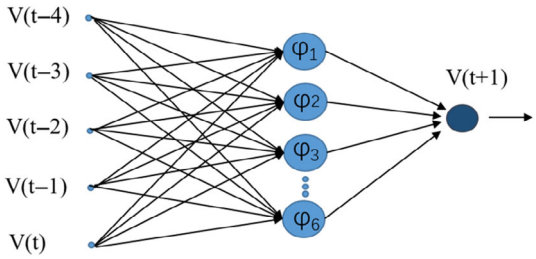

FIGURE 7 Radial basis function (RBF) architecture for wind speed forecasting in Petrolina (Brazil)

The input vector of the wind speed forecast in Petrolina (Brazil) had the following composition and 4 neurons in the hidden layer:

$$[V(t-6), V(t-5), V(t-4), V(t-3), V(t-2), V(t-1), V(t)].$$

The activation function used is the logistic sigmoid represented by Equation (5)

$$y_j = \frac{1}{1+\exp(-u_j)}, \tag{5}$$

where y_j is the neuron output and u_j is the weighted sum of all inputs.

The training algorithm used in MLP was "backpropagation." The back-propagation principle uses a downward gradient and estimates the intermediate layer error, as well as the effect of these errors on the output layer error. The output layer error is calculated and is backpropagated to the previous layers, allowing the weight adjustment proportional to the connection values between layers, according to Equation (6).

$$w_{ij}^m(t+1) = w_{ij}^m(t) - \propto \cdot \delta_i^m \cdot y_j^{m-1}, \tag{6}$$

where $w_{ij}^m(t+1)$ is the updated weight, $w_{ij}^m(t)$ is the weight in the iteration, $\propto$ is the ANN learning rate, δ_i^m is the sensitivity, y_j^{m-1} is the anterior layer output, and i and j identify the neurons of the connection posterior and anterior layer, respectively. Learning rates of 0.15 and 0.65 were used in the MLP networks applied for wind speed forecasting in Spain and Brazil, respectively.

3.2.3 | Radial basis function

RBF techniques use the distance between input vectors to activate neurons in the middle layer. In general, RBF networks have only one intermediate layer. The RBF output is calculated by the linear combination of the activations of the intermediate layer and the weights that interconnect the two layers. Hence, each RBF intermediate neuron defines a hyperellipsoid in the input pattern space, which are local approximators.[28]

In the first RBF training stage, the number of radial base functions was estimated, and the parameters (center and spread) were defined by unsupervised learning. The second RBF training stage adjusted the output neuron weights. In our study, two RBF networks were trained and applied to hourly wind speed forecasting using data from Brazil and Spain. The network structure for wind speed forecasting in Brazil is shown in Figure 7.

The initial layer has five inputs: (a) $V(t-4)$, $V(t-3)$, (b) $V(t-2)$, (c) $V(t-1)$, and (d) $V(t)$. Simulations with different time delays were performed, and the best results were obtained with 4-time delays, which is why four entries were chosen. The number of neurons in the hidden layer (in our case, equal to six) is obtained through a trial-and-error process, each contains a RBF. A single neuron (the velocity value $V(t+1)$) composes the RBF output.

The input vector of the wind resource forecasting in Algeciras (Spain) has four hidden neurons. The composition of inputs is

$$[V(t-3), V(t-2), V(t-1), V(t)].$$

Equations (7) and (8) are considered using the Gaussian function.

$$\varphi = \exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{7}$$

$$r = ||x - t||, \tag{8}$$

where φ is the output of the hidden neurons, r is the subtraction of the input x by the center t, and σ is the scattering curve.

The result of the output (y) of the RBF is calculated through Equation (9) through the weighted sum of the hidden layer outputs by their weights (w).

$$y = \sum w_j \varphi_j. \tag{9}$$

3.3 | PrevPT

During the definition of the integration weights, that is, during the PrevPT training, each of the input techniques (SOM, MLP, and RBF) are compared with the measured wind speeds. With this, it is possible to understand the accuracy and variability of the forecasting technique in relation to wind speed.

The percentage of errors is calculated by comparing the ANNs applied individually and using the impact factor defined in Equations (10) and (11). These errors related to wind forecasts can occur at different times of the day when the resource is available on site.[5]

$$E_{pp} = \frac{(D_p - D_m)100}{D_m} F_i \tag{10}$$

$$F_i = \frac{D_m}{M_r}, \tag{11}$$

where D_p is the predicted data, D_m is the observed data, M_r largest resource, and F_i is the factor impact.

In its first use, the PT goal was to maximize a financial return, at any risk, through the diversification of securities or investments that are not positively correlated.[32-34] According to,[35] wind forecast techniques can improve the planning, operation, and management of electrical grids. The PrevPT methodology can integrate and improve the results of the individual wind forecasts.

Forecast risk can be defined as the difficulty of a specific technique or a combination of techniques in obtaining the forecast result. The lower the risk, the greater is the possibility that the forecast will get its result right. An application of PT was found in Reference 36. The capital asset pricing model is used to understand whether it is possible to own a single large wind farm or a portfolio of smaller wind farms. The conclusion reached was that production risk can be reduced by up to 30% by maintaining a portfolio of wind farms instead of a single large wind farm. However, the author highlighted that this depends on the correlation between the parks and that it depends on the location.

Equation (12) defines the calculation of asset variability through correlation.

$$p_{xy} = \frac{\mathrm{Cov}(y,x)}{\sigma_x \cdot \sigma_y}, \tag{12}$$

where p_{xy} is the correlation between x and y (assets), $\mathrm{Cov}(y,x)$ is the covariance between x and y, σ_x is the SD of x, and σ_y is the SD of y.

The covariance between x and y is calculated using Equation (13).

$$\mathrm{Cov}(y,x) = \frac{\sum_{i=1}^{n} (x\_i - \bar{x})(y\_i - \bar{y})}{n}, \tag{13}$$

where x_i are the asset values of x, $\bar{x}$ is the average value of asset x, y_i are the asset values of y, $\bar{y}$ is the average value of asset y, and n represents the amount of asset values.

Portfolio risk can be determined by calculating a statistical tool called SD, which reveals the probability of a certain result. To define the percentage of each asset and therewith reduce forecasting errors, it is necessary to set up an efficient frontier chart. Those boundaries are lines formed by different assets proportions in a portfolio, representing graphically how a particular return can promote the lowest possible risk. Figure 8 shows the hypothetical situation of a diversified application using assets A and B.

An absolute value for acceptable risk cannot be defined. In this case, risk, measured by the SD, is inherent in the forecasting process. The SD (risk) will always exist and PrevPT always seeks to reduce diversifiable risk by combining the forecasters.

Figure 9 shows the used PrevPT structure: SOM, MLP, and RBF techniques and their integrations through the connection points *psm* (SOM and MLP integration), *psr* (SOM and RBF integration), and *pmr* (integration

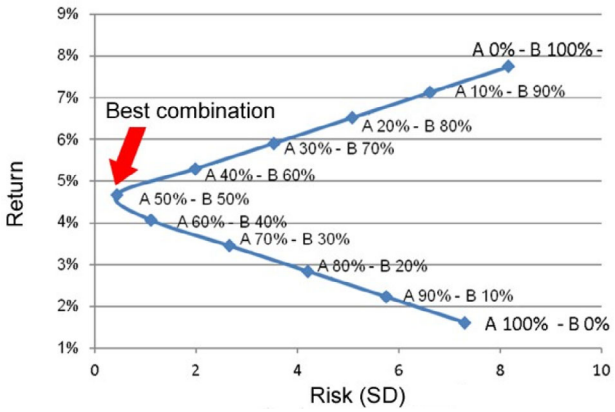

FIGURE 8 Efficient frontier

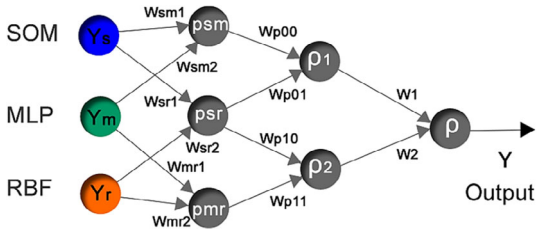

FIGURE 9 PrevPT structure

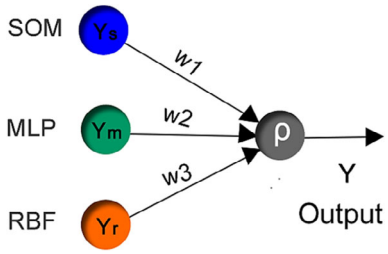

FIGURE 10 PrevPT simplified structure

between MLP and RBF). Sequentially, *p1* and *p2* represent the two best connections (minor forecasting errors) between *psm*, *psr*, and *pmr*, and *p* displays the PrevPT forecasting value.

Figure 10 is a simplified representation of PrevPT that is used after training and where the final weights interconnect the prediction methodologies applied individually. The contribution of each individual model (Ys, Ym, and Yr) in the simplified structure of the PrevPT is obtained from the sum of the likelihood of the paths to the output Y, through the connections *psm*, *psr*, *pmr*, $\rho 1$, and $\rho 2$.

According to our proposal, PrevPT performs the parallel integration of forecasting methodologies. After the topology formed from the PrevPT training, the data processing structure can be summarized as follows: input with information window on the wind speed in each technique (SOM, MLP, and RBF), processing of forecasts by said ANNs, and application of the integration with weightings made from the efficient frontiers found by the PrevPT.

The data processing steps are shown in Figure 11: (a) wind speed data collection, training, and application of SOM, MLP, and RBF forecasters and the calculation and analysis of their respective forecast errors; (b) production of efficient frontiers that are used for the integration of techniques; and (c) application of weights; and income statement.

Several statistical indicators were used to analyze the results. The MAPE was used, which is a highly accepted indicator in the academic world. It can detect, in percentage values, the amount of error in the proposed forecast system, without canceling between positive and negative errors. The mean percentage error was also used. It

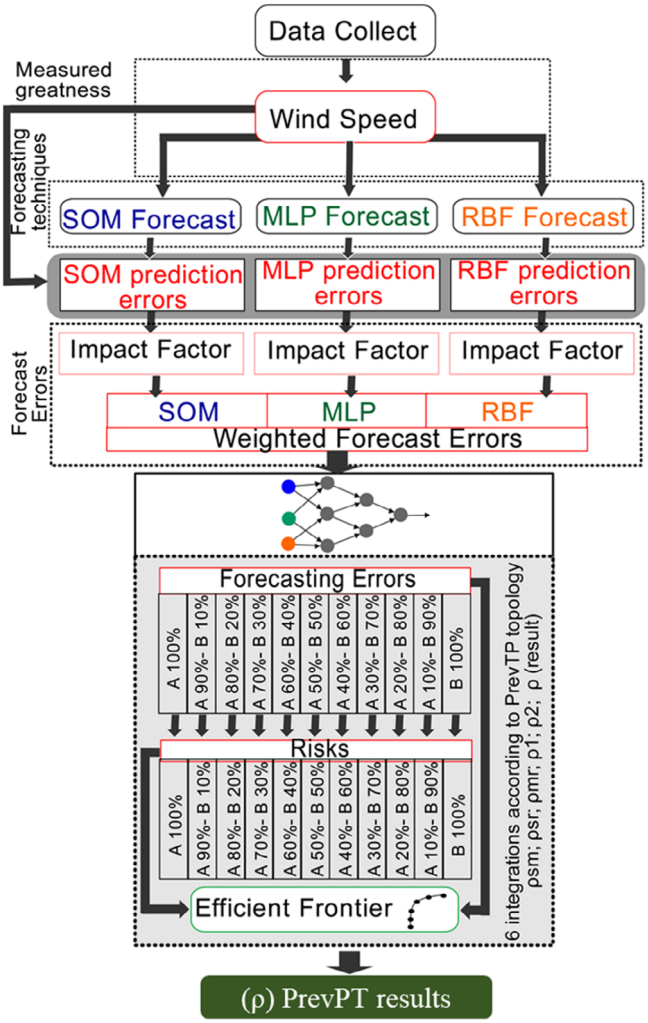

FIGURE 11 PrevPT data processing steps. Adapted from Reference 9

allows to observe the trend of forecasts, that is, if they are close to 0, if they are on the positive or negative axis, and thus, determine if there are trends of overestimation or underestimation of the wind resource.

In addition, we can set the range from −10 (%) to 10 (%), for which we can set the amount of hits within a specific range of low forecast errors. To the above, this study is compatible with,[9] where the proposed theory was applied for solar forecasting, and thus, in this paper, the similarities and differences between these two studies are analyzed. Furthermore, to facilitate the understanding of the results, linear and boxplot regressions are presented for both study sites.

4 | RESULTS

PrevPT was tested in different locations, which resulted in specific training with data from each location, as well as an optimal topology distinguished for each region

studied. The following subsections show the results obtained.

4.1 | Forecasting errors (Spain)

Figure 12 shows the behavior of velocity data in Algeciras (Spain), the month of October 2009, showing the fluctuations that can occur due to several factors, including cloudiness.

Forecast errors for a window of 200 samples taken from the forecasts of the individual forecast methodologies (SOM, MLP, and RBF) are shown in Figure 13.

The highest wind speed values foreseen by SOM, MLP, and RBF were 12.38, 15.17, and 14.12 m/s, respectively. The highest measured wind speed was 17.71 m/s.

The SOM technique underestimated the wind resource forecast by 55.13% and overestimated it by 44.87%. The highest negative and positive errors were −44.16% and 31.32%, respectively. A share of 44.26% of the forecasting showed forecasting errors in the range of −10% to 10%. The mean of the forecast error was −0.68% and the MAPE was 2.95%.

The MLP technique underestimated the wind resource by 41.54% and overestimated it by 58.65%. The highest negative and positive errors were −27.4% and 19.12%, respectively. A share of 58.41% of the forecasting showed forecasting errors in the range of −10% to 10%. The mean of the forecast error was 0.05%, and the MAPE was 1.55%.

The RBF technique underestimates the forecasting by 43.82% and overestimates it by 56.18%. The highest negative and positive errors were −20.47% and 8.36%, respectively. A share of 56.18% of the forecasting shows forecasting errors in the range of −10% to 10%. The mean of the forecast error was 0.01%, and the MAPE was 1.54%.

Considering individualized forecasts in Spain, MLP and RBF overestimated most of the forecasts, SOM underestimated more than 55% of them, and RBF achieved the lowest MAPE (1.54%).

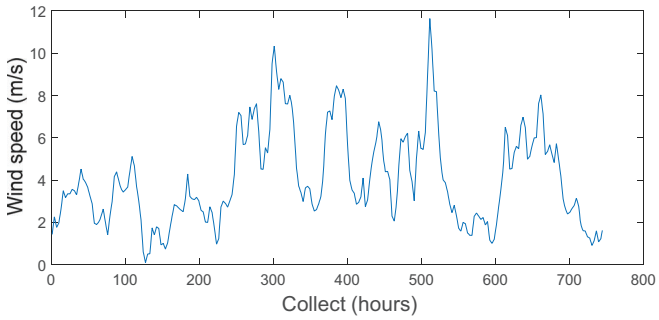

FIGURE 12 Wind speed data in Algeciras (Spain). Data: October 2009

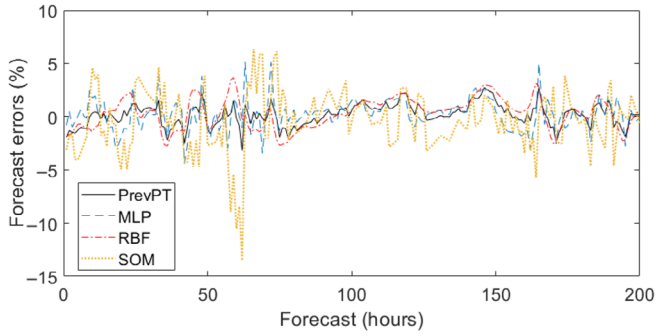

FIGURE 13 Forecasting errors in Algeciras (Spain). Data: October 2009

4.2 | Integration of forecasting methodologies in Spain

The first, second, and last efficient frontiers are shown in Figures 14-16 (Spain). An efficient frontier is defined for the reduction of forecast errors, this was done using 21 combinations (connections and weights).

The PrevPT first layer was defined with connection *pmr* 40% MLP—60% RBF, connection *psm* 80% MLP—20% SOM, and connection *psr* 75% RBF—25% SOM. The second layer was defined with connection *p1* 35% *pmr*—65% *psr* and connection *p2* 25% *psm*—75% *psr*. PrevPT output was defined as 45% *p1* and 55% *p2*.

The optimized PrevPT obtained values of 20.375% SOM, 17.3% MLP, and 62.325% RBF as the best results aiming at the predictive capacity. The best fit is shown in Figure 17 (Spain).

The details of the errors/risk calculated with the individual models at the input to the PrevPT, at the existing connections and at the output are described in Table 3.

Analyzing the results shown in Table 3, it can be seen that, when connections and integrations occur, errors and risk are reduced. The lower the risk, the better the predictability of the method or combination. It is also observed that PrevPT uses the combination of errors and risk aiming at the best configuration. Figure 17A illustrates the high contribution of the *psr* connection (the one with the lowest risk and MAPE). *psr* presents 75% of the RBF model, which contains lower risk and lower MAPE between the individual techniques.

A comparison of the forecasting error distributions is shown in Figure 18. PrevPT underestimated wind resource availability by 43.34% and overestimated it by 56.66%. The highest negative and positive errors were −22.18% and 5.74%, respectively. Of the forecasts, 56.67%

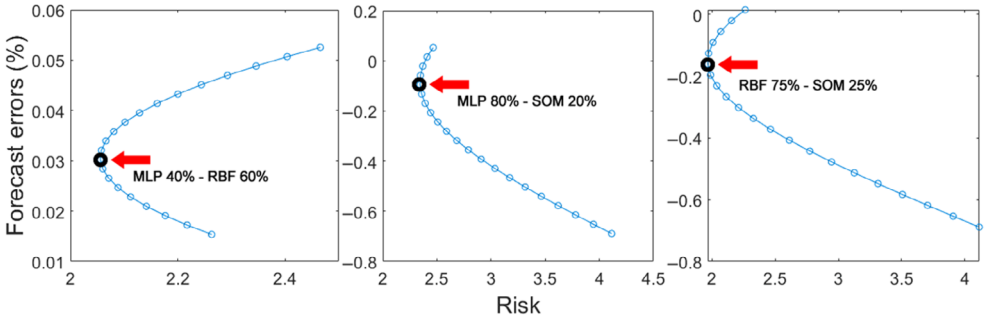

FIGURE 14 Efficient frontiers—first layer

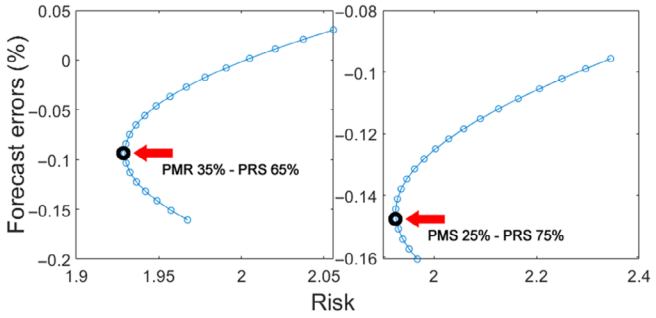

FIGURE 15 Efficient frontiers—second layer

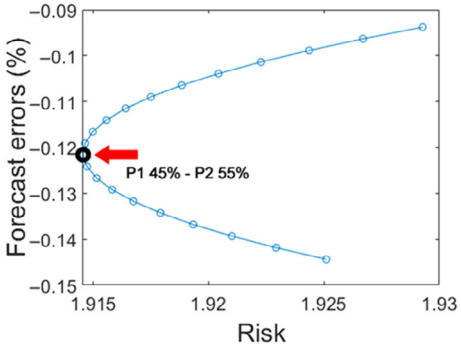

FIGURE 16 Efficient frontiers—output

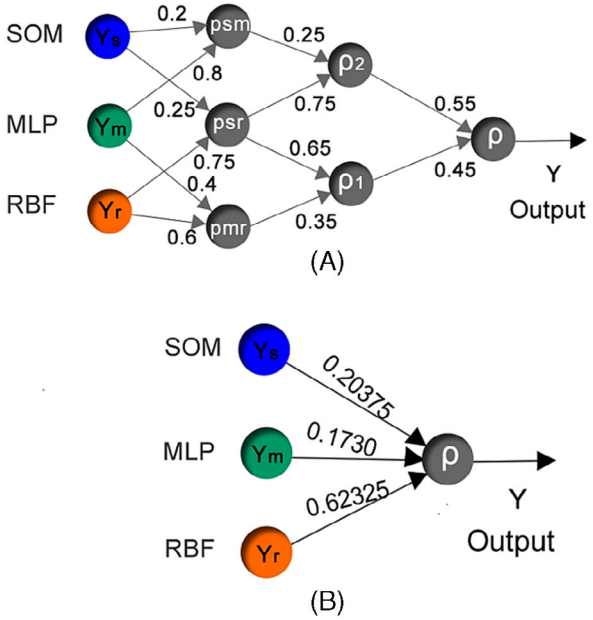

FIGURE 17 PrevPT Architecture—Spain: (A) full version and (B) simplified version

showed forecasting errors in the range of −10% to 10%. The mean of the forecast error was −0.12% and the MAPE was 1.13%, demonstrating that the integration of the methodologies applied individually results in better precision predictions, the MAPE obtained by the MLP, RBF, and SOM ANNs were 1.55%, 1.54%, and 2.95%, respectively.

Table 4 compares the individual AI results and PrevPT (Spain).

Figure 19 shows a boxplot that represents the variation between the methodologies applied individually and the PrevPT.

The interquartile range (50% of the data presented) analyzes the degree of data scattering around a measure of centrality, which allows to better understand the data behavior. In the case of application in Spain, the results

TABLE 3 Summary of errors and risks calculated at all stages of the application in Spain

Model/combination	MAE	MAPE	RISK
Ym—MLP	0.052545	1.5587	2.4656
Yr—RBF	0.015379	1.5448	2.2632
Ys—SOM	−0.6886	2.9589	4.1139
pmr	0.030245	1.3076	2.0557
psm	−0.095685	1.5095	2.3443
psr	−0.16062	1.1893	1.9674
ρ1	−0.093815	1.1659	1.9293
ρ2	−0.14438	1.1432	1.9251
ρ (PrevPT)	−0.12163	1.1385	1.9146

were −0.71% to 0.77% for PrevPT, −0.71% to 1.04% for MLP, −0.94% to 1.24% for RBF, and −2.66% to 1.76% for SOM. The best results were obtained by PrevPT. Closer to

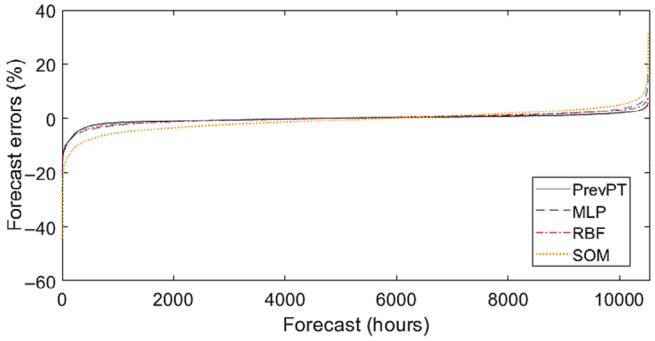

FIGURE 18 Forecasting errors (Spain)

TABLE 4 Metrics for comparing results (Spain)

	MAPE (%)	MAE (%)	Range −10% to 10%
MLP	1.55	0.05	58.41
RBF	1.54	0.01	56.18
SOM	2.95	−0.68	44.26
PrevPT	1.13	0.12	56.67

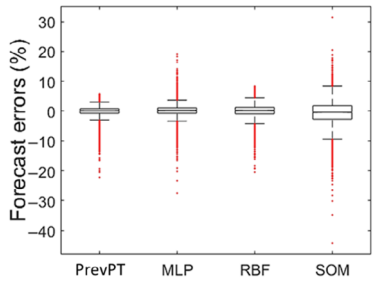

FIGURE 19 Forecast data variation boxplot for Spain

0, the PrevPT showed greater precision in estimating the values of the wind resource.

The interquartile range (50% of the data presented) is a boxplot tool that can be used to understand the placement of errors. Regarding the limits, the closer to zero, the greater the accuracy of the technique. Furthermore, it is possible to understand the tendencies of overestimation or underestimation of the wind resource by verifying the orientation of the interquartile in relation to the positive and negative axes.

Figure 20 shows the linear regression of the proposed AI technique. The integration of techniques carried out by PrevPT improved the fit between the predicted and observed data from Spain.

The efficiency of integrations is linked to the concept of correlation. This correlation coefficient varies between −1 and +1. At −1 the variables are negatively correlated and at +1 the variables are positively correlated (not favoring the diversification proposed by the theory), since

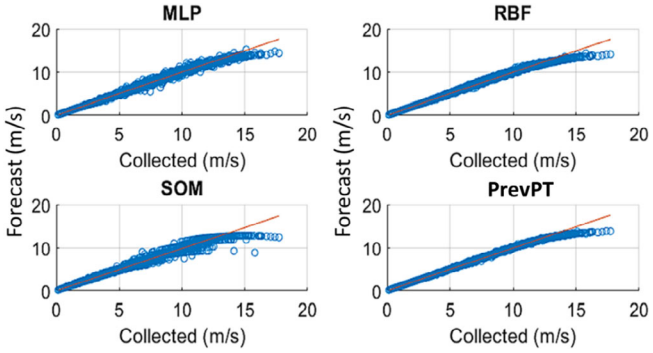

FIGURE 20 Linear regressions for the data from Spain

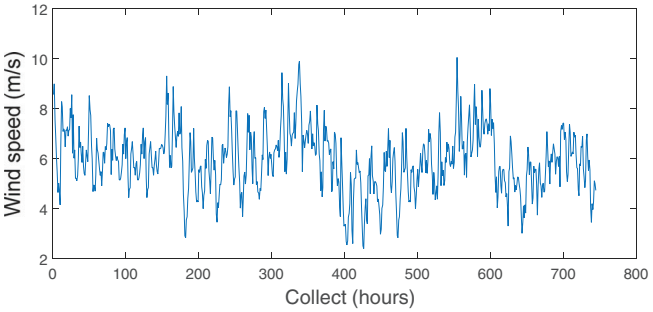

FIGURE 21 Wind speed data in Petrolina (Brazil). Data: October 2009

this characteristic makes that there is no compensation of one asset against a possible loss of the other. The same happens in the case of Brazil.

4.3 | Forecasting errors (Brazil)

Figure 21 shows the behavior of velocity data in Petrolina (Brazil), for the entire month of October 2009.

Forecast errors for a window of 200 samples taken from the forecasts of the individual forecast methodologies (SOM, MLP, and RBF) are shown in Figure 22.

The highest wind speed values foreseen by SOM, MLP, and RBF are 9.85, 9.52, and 9.75 m/s, respectively.

SOM underestimated the forecasts by 46.10% and overestimated them by 53.89%. The highest negative and positive errors were −17.29% and 12.73%, respectively. A percentage of 53.58% of the forecasting showed forecasting errors in the range of −10% to 10%. The mean of the forecast error was 0.26%, and the MAPE was 2.51%.

MLP underestimated the forecasting by 40.23% and overestimated it by 59.77%. The highest negative and positive errors were −13.18% and 14.93%, respectively. A percentage of 59.20% of the forecasting presented

forecasting errors in the range of −10% to 10%. The mean of the forecast error was 0.94%, and the MAPE was 2.39%.

RBF underestimated the forecasting by 13.41% and overestimated it by 86.59%. The highest negative and positive errors were −10.67% and 15.86%, respectively. A percentage of 85.46% of the forecasting showed forecasting errors in the range of −10% to 10%. The mean of the forecast error was 2.52%, and the MAPE was 2.82%.

MLP showed the smallest MAPE (2.39%), and SOM achieved the smallest average forecast error (0.26%). The three ANNs applied individually showed a tendency to overestimate their predictions.

4.4 | Integration of forecasting methodologies in Brazil

The first, second, and last efficient frontiers are shown in Figures 23-25 (Brazil). Defined in the same way as detailed data for Spain.

In this case, the PrevPT first layer is defined with connection *pmr* 0% MLP—100% RBF, connection *psm* 60% MLP—40% SOM, and connection *psr* 60% RBF—60% SOM. The second layer is defined with connection *p1* 5% *pmr*—95% *psr* and connection *p2* 0% *psm*—100% *psr*. The PrevPT output was defined as 100% *p1* and 0% *p2*. The complete structure is illustrated in Figure 26A.

The optimized PrevPT obtained values of 28.5% SOM, 0% MLP, and 71.5% RBF as the best results aiming at the predictive capacity. The best fit is shown in Figure 26B (Brazil).

As in the case of Spain, Table 5 illustrates the errors/risk calculated in the individual models at the input to the PrevPT, at the existing connections and at the output.

The errors and the risk are reduced after the integration in each connection (see Table 5). Figure 26A illustrates the significant contribution (70%) of the RBF model in the *prs* connection (the one with the lowest risk and MAPE), which was the most used in the optimized PrevPT structure. The RBF model was chosen with high contribution because despite obtaining the second-lowest MAPE, it has the lowest risk and highest predictability. The SOM model had low MAPE, but high risk, which could reduce the predictability of the PrevPT model.

A comparison of the forecasting error distributions is shown in Figure 27. PrevPT underestimated wind resource availability by 18.99% and overestimated it by 81.01%. The highest negative and positive errors were −10.33% and 14.58%, respectively. A percentage of 80.3% of the forecasting presented forecasting errors between −10% and 10%. The MAPE of the techniques applied

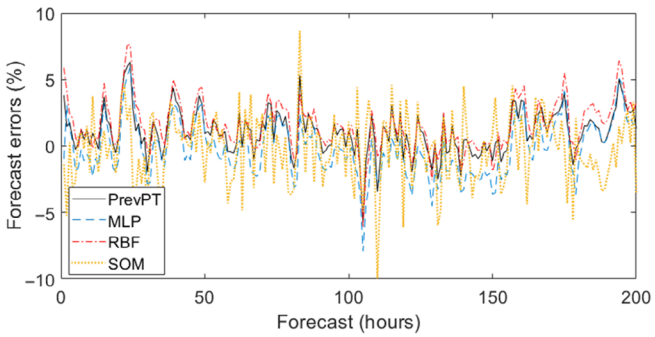

FIGURE 22 Forecast errors in Petrolina (Brazil). Data: October 2009

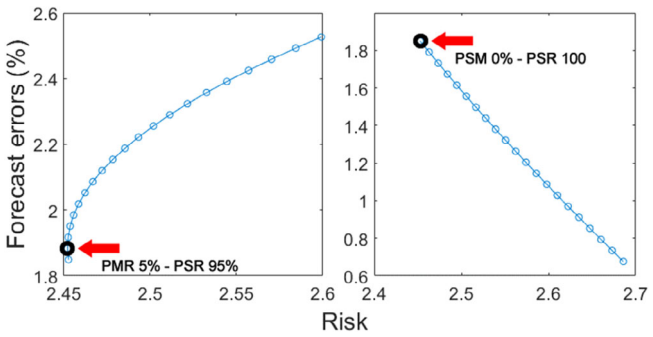

FIGURE 24 Efficient frontiers—second layer

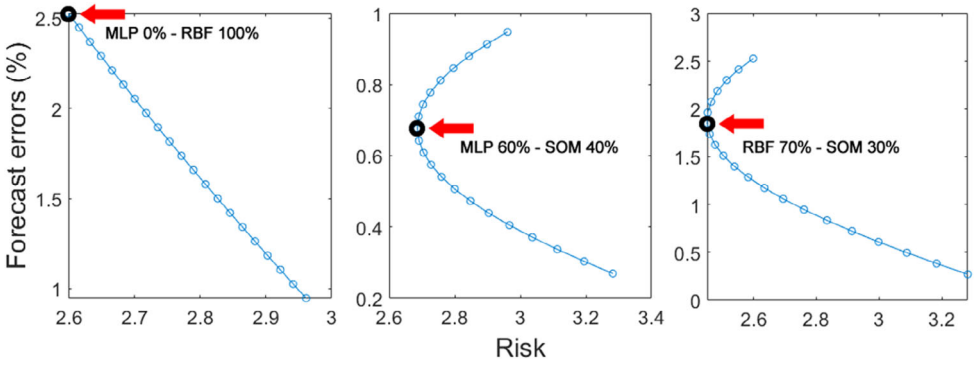

FIGURE 23 Efficient frontiers—first layer

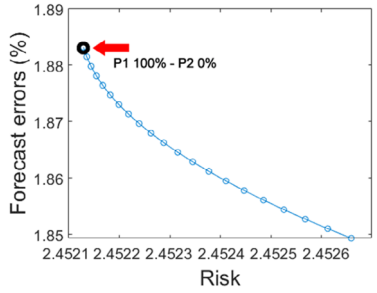

FIGURE 25 Efficient frontiers—output

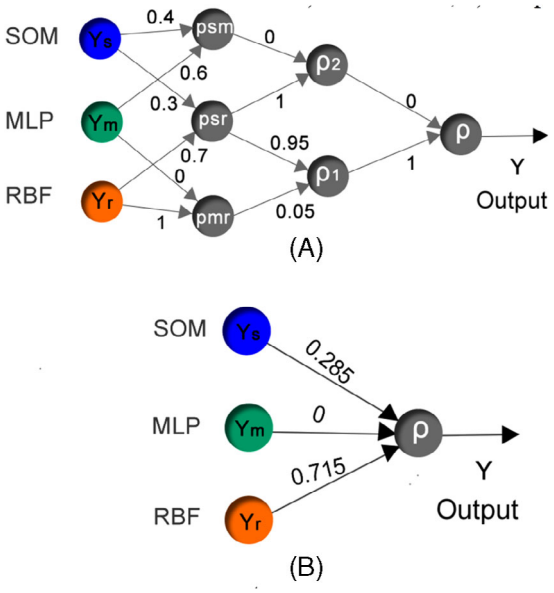

FIGURE 26 PrevPT Architecture—Brazil: (A) full version and (B) simplified version

TABLE 5 Summary of errors and risks calculated at all stages of the application in Brazil

Model/combination	MAE	MAPE	RISK
Ym—MLP	0.9475	2.3966	2.9611
Yr—RBF	2.5264	2.8267	2.5991
Ys—SOM	0.2695	2.5132	3.2821
Pmr	2.5264	2.8267	2.5991
Psm	0.6725	2.4432	3.0923
Psr	1.8593	2.3727	2.8040
$\rho 1$	1.9132	2.3714	2.5721
$\rho 2$	1.8994	2.3631	2.7527
ρ (PrevPT)	1.8832	2.3514	2.4521

individually was 1.55%, 1.54%, and 2.95% for MLP, RBF, and SOM, respectively. The integration methodology showed better results than the individual techniques, with an average forecast error of 1.88% and MAPE of 2.35%.

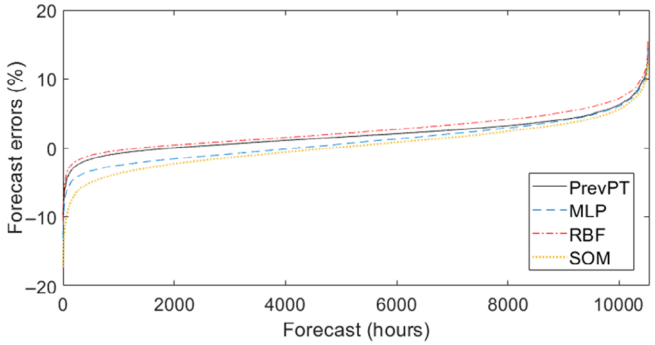

FIGURE 27 Forecasting errors (Brazil)

TABLE 6 Metrics for comparing results (Brazil)

	MAPE (%)	MAE (%)	Range −10% to 10%
MLP	2.39	0.94	59.20
RBF	2.82	2.52	85.46
SOM	2.51	0.26	53.58
PrevPT	2.35	1.88	80.3

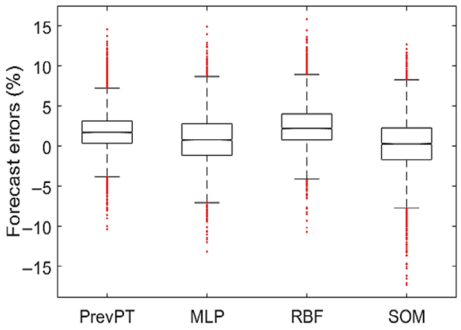

FIGURE 28 Forecast data variation boxplot for Brazil

Table 6 compares the individual AIs results and PrevPT (Brazil).

Figure 28 shows a boxplot representing the forecast data variation of the proposed AI techniques.

Analyzing the dispersion of data around a measure of centrality, the interquartile range (50% of the data presented) values were 0.35% to 3.12% for PrevPT, −1.14% to 2.79% for MLP, 0.74% to 4.01% for RBF, and −1.7% to 2.3% for SOM. The best results were obtained by PrevPT. Closer to 0, mainly on the negative side, the PrevPT showed greater precision in estimating the values of the wind resource.

Figure 29 shows the linear regressions of the proposed AI techniques. PrevPT integration obtained the best fit between the predicted and observed data from Brazil.

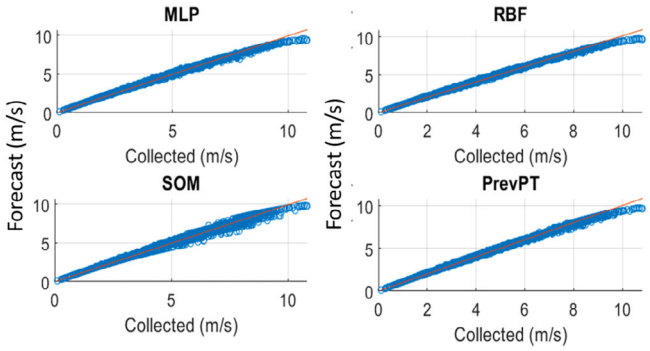

FIGURE 29 Linear regressions for Brazil

5 | CONCLUSIONS

The effects of intermittency, which is inherent to wind resources, represent a challenge for wind farm managers, who need to use appropriate tools to forecast the wind resource availability and to manage the wind farm. This work presented a methodology that uses an adaptation of PT, called PrevPT, to integrate forecasting techniques applied individually (MLP, RBF, and SOM, which are techniques based on RNA). The results showed that PrevPT has an improved ability to reduce forecast errors by increasing the modeling accuracy. The error resulting from the integration of the MLP, RBF, and SOM forecasts showed a better performance compared to the errors of the individual methodologies.

During the performance of the artificial intelligence (AI) tests, a large number of hours of exposure and processing of wind speed data were used by the techniques. This results in a long test time and, consequently, because of its use on a large time scale, it eliminates possible occasional errors of comparison of the techniques during the evaluation of the predictors. The larger the sample, the greater the possibility that the processed data will clearly represent the difference between the techniques studied.

For the wind forecasts in Brazil, PrevPT obtained a MAPE and MAE of 2.35% and 1.88%, respectively, while a MAPE and MAE of 2.39% and 0.94%, respectively, was achieved by MLP, 2.82% and 2.52% by RBF, and 2.51% and 0.26% by SOM, respectively. In the case of Spain, PrevPT obtained a MAPE and MAE of 1.13% and 0.12%, and the MAPE and MAE were 1.55% and 0.55% for MLP, 1.54% and 0.01% for RBF, and 2.95% and − 0.68% for SOM, respectively. For PrevPT, the limits of the interquartile range (50% of the data presented) were −0.71% to 0.77% for the case of Spain and 0.35% to 3.12% for Brazil.

The PrevPT applied to Spain managed to reduce the forecast error by 61% when compared to RNA SOM and by 27% when compared to forecasts obtained by RNA MLP and RBF. In the case of Brazil, the reduction in forecast errors was 16%, 1.6%, and 6.3% in relation to the RBF, MLP, and SOM models, respectively. The results showed that when the models are combined, the errors are reduced (in terms of MAE and MAPE) and the risk is reduced. Thus, future work may aim at increasing the number of individual techniques to be integrated, with a chance of further increasing accuracy and reducing risks.

The large amount of energy generated from renewable and interminable energy sources, if not correctly predicted, can generate significant increases in the production costs of these sources. Thus, the proposed PrevPT was able to integrate three different ANN prediction methods to determine the optimal participation of each prediction technique, with the aim of reducing predictability errors. The training phase was carried out in parallel, and therefore, the ANN techniques used were processed simultaneously. The methodology is an alternative strategy for managing new configurations in the electricity sector, with a growing share of wind energy.

Future works are related to the application of mathematical models to determine the electricity generation capacity through the wind speed resources, foreseen and observed in this work, as well as the development of a study on the financial impacts of the errors.

ACKNOWLEDGEMENTS

We thank the National Council for Scientific and Technological Development (CNPq) for providing a research scholarship to the third author.

ORCID

Tatiane C. Carneiro https://orcid.org/0000-0001-7457-6836
Marcello A. Ferreira Batista Lima https://orcid.org/0000-0001-7609-2472
Paulo C. Marques de Carvalho https://orcid.org/0000-0002-0115-0807
Josias Guimarães Batista https://orcid.org/0000-0002-9879-5985
Luis M. Fernández-Ramírez https://orcid.org/0000-0002-4898-0680

REFERENCES

1. Liu J, Wang X, Lu Y. A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system. *Renew Energy*. 2017;103:620-629. doi: 10.1016/j.renene.2016.10.074
2. Ozkan MB, Karagoz P. Reducing the cost of wind resource assessment: using a regional wind power forecasting method for assessment. *Int J Energy Res*. 2021;45(9):13182-13197. doi: 10.1002/er.6645

3. Liu H, Chen C, Lv X, Wu X, Liu M. Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. *Energ Conver Manage*. 2019;195:328-345. doi:10.1016/j.enconman.2019.05.020
4. Kabouris J, Kanellos FD. Impacts of large-scale wind penetration on designing and operation of electric power systems. *IEEE Trans Sustain Energy*. 2010;1(2):107-114.
5. Speth V. Wind and solar portfolios and their impact on predictability: german case study 2010-2011. Paper presented at: 11th International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants; November 13-15, 2012; Lisbon, Portugal. Anais [...] Lisbon: Energynautics, 2012, pp. 1–6. Disponível em: https://windintegrationworkshop.org/lisbon2012/index.html. Acesso em set. 10 2018.
6. Mert YG, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. *Renew Sustain Energy Rev*. 2019;105:487-498. doi:10.1016/j.rser.2019.02.006
7. Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. *IEEE Trans Power Syst*. 2007;22(1):258-265.
8. Wu YK, Hong JS. A literature review of wind forecasting technology in the world. Paper presented at: 2007 IEEE Lausanne Power Tech; July 1-5, 2007; Lausanne, Switzerland:504-509.
9. Lima MAFB, Carvalho PCM, Fernandez-Ramírez LM, Braga APS. Improving solar forecasting using Deep Learning and Portfolio Theory integration. *Energy*. 2020;195:117016. doi:10.1016/j.energy.2020.117016
10. Zhang C, Ma Y, eds. *Ensemble Machine Learning*, 1st ed. Boston, MA: Springer US; 2012. doi:10.1007/978-1-4419-9326-7
11. Ahmed T, Sharma P, Karmaker CL, Nasir S. Warpage prediction of injection-molded PVC part using ensemble machine learning algorithm. *Mater Today: Proc*. 2020;50:565-569. doi:10.1016/j.matpr.2020.11.463
12. Saha S, Roy J, Pradhan B, Hembram TK. Hybrid ensemble machine learning approaches for landslide susceptibility mapping using different sampling ratios at East Sikkim Himalayan, India. *Adv Space Res*. 2021;68(7):2819-2840. doi:10.1016/j.asr.2021.05.018
13. Kar S, Purbey VK, Suradhaniwar S, et al. An ensemble machine learning approach for determination of the optimum sampling time for evapotranspiration assessment from high-throughput phenotyping data. *Comput Electron Agric*. 2021;182:105992. doi:10.1016/j.compag.2021.105992
14. Wang GC, Ratnam E, Haghi HV, Kleissl J. Corrective receding horizon EV charge scheduling using short-term solar forecasting. *Renew Energy*. 2019;130:1146-1158. doi:10.1016/j.renene.2018.08.056
15. Abuella M, Chowdhury B. Forecasting of solar power ramp events: a post-processing approach. *Renew Energy*. 2019;133:1380-1392. doi:10.1016/j.renene.2018.09.005
16. Song Y, Shen Z, Dai D, Qian Y, Wang Y. Short-term load forecasting in electrical power systems via trajectory tracking and error correcting approach. *J Renew Sustain Energy*. 2014;6:013112. doi:10.1063/1.4861784
17. Haessig P, Multon B, Ahmed HB, Lascaud S, Bondon P. Energy storage sizing for wind power: impact of the autocorrelation of day-ahead forecast errors. *Wind Energy*. 2015;18:43-57. doi:10.1002/we.1680
18. Nuño E, Koivisto M, Cutululis NA, Sørensen P. On the simulation of aggregated solar PV forecast errors. *IEEE Trans Sustain Energy*. 2018;9(4):1889-1898. doi:10.1109/TSTE.2018.2818727
19. Nuño E, Koivisto M, Cutululis NA, Sørensen P. Simulation of regional day-ahead PV power forecast scenarios. Paper presented at: 2017 IEEE Manchester PowerTech; June 18-22, 2017; Manchester, UK:1-6. doi:10.1109/PTC.2017.7981155
20. Yang D. Ensemble model output statistics as aprobabilistic site-adaptation tool for satellite derived and reanalysis solar irradiance. *J Renew Sustain Energy*. 2020;12:016102. doi:10.1063/1.5134731
21. Doubleday K, Jascourt S, Kleiber W, Hodge B. Probabilistic solar power forecasting using Bayesian model averaging. *IEEE Trans Sustain Energy*. 2021;12(1):325-337. doi:10.1109/TSTE.2020.2993524
22. Liu Y, Qin H, Zhang Z, et al. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. *Appl Energy*. 2019;253:113596.
23. Worsnop RP, Scheuerer M, Hamill TM. Extended-range probabilistic fire-weather forecasting based on ensemble model output statistics and ensemble copula coupling. *Mon Weather Rev*. 2020;148(2):499-521. doi:10.1175/MWR-D-19-0217.s1
24. Liu C, Wei D, Xiang J, et al. An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Mol Ther Nucleic Acids*. 2020;21:676-686. doi:10.1016/j.omtn.2020.07.003
25. Photovoltaic Geographical Information System. http://re.jrc.ec.europa.eu/pvgis/. Accessed May 10, 2020.
26. Sistema De Organização Nacional De Dados Ambientais. http://sonda.ccst.inpe.br. Accessed May 10, 2020.
27. Carneiro TC, Melo SP, Carvalho PCM, Braga APS. Particle Swarm Optimization method for estimation of Weibull parameters: a case study for the Brazilian northeast region. *Renew Energy*. 2016;86:751-759. doi:10.1016/j.renene.2015.08.060
28. Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 1999.
29. Beale M, Hagan M, Demuth H. *Neural Network Toolbox 7.0.3: User's Guide*. Natick, MA: The MathWorks Inc; 2012:404.
30. de Pádua Braga A, de Leon Ferreira Carvalho ACP, Ludermir TB. *Redes Neurais Artificiais: Teoria e Aplicações*. 2ª edição ed. Rio de Janeiro, Brazil: Editora LTC; 2007.
31. Farias CAS, Santos CAG, Lourenço AMG, Carneiro TC. Kohonen neural networks for rainfall-runoff modeling: case study of Piancó River Basin. *J Urban Environ Eng*. 2013;7(1):176-182. doi:10.4090/juee.2013.v7n1.176182
32. Markowitz HM. *Portfolio Selection – Efficient Diversification of Investments*. 2nd ed. Oxford, UK: John Wiley and Sons Ltd; 1991.
33. Markowitz HM. *Portfolio Selection: Efficient Diversification of Investments*. 1st ed. New York, NY: John Wiley & Sons; 1959.
34. Renner MC. Markowitz portfolio theory at times of crisis. *Specialization in Capital Markets*. Porto Alegre, Brazil: Federal University of Rio Grande do Sul; 2010.
35. Devaraj J, Elavarasan RM, Shafiullah GM, Jamal T, Khan I. A holistic review on energy forecasting using big data and deep learning models. *Int J Energy Res*. 2021;45(9):13489-13530. doi:10.1002/er.6679
36. Dunlop J. Modern portfolio theory meets wind farms. *J Private Equity*. 2004;7(2):83-95.

How to cite this article: Carneiro TC, Ferreira Batista Lima MA, Marques de Carvalho PC, Guimarães Batista J, Fernández-Ramírez LM. Methodology for integration of wind resource forecasts based on artificial neural networks. *Int J Energy Res*. 2022;46(6):8271-8287. doi:10.1002/er.7728