Electrical load, wind speed and power forecasting with feed-forward neural networks

Eduardo Praun Machado
eduardopraun@gmail.com
Instituto Superior Técnico, Universidade de Lisboa, Portugal
October 2021

***Abstract* – The higher share of renewable energy sources (RES) in the electrical grid and the electrification of important sectors, such as transport and heating, are imposing a tremendous challenge to the operation of power system. In recent years, the use of machine learning (ML) models has been gaining increased attention from the power sector as they can provide accurate forecasts of the system behavior from energy generation to consumption, helping all the stakeholders to optimize their activities. The aim of this work is to develop a methodology to enhance the load demand and the wind speed forecasts generated by a ML model, namely a feed-forward neural network (FFNN), by incorporating a forecasting error correction step which involves the prediction of the initial errors by another FFNN. The results showed that the proposed methodology was able to significantly improve the load demand forecasts while any notable performance difference was verified for the wind speed forecasts. The forecasted wind speed was applied to a wind turbine power curve, which was modeled with the adoption of a clustering method, to estimate the wind power output. The results achieved using the proposed model forecasts were more accurate than the ones that used the benchmark model forecasts.**

Keywords – Feed-forward neural networks, Forecasting, Load demand, Machine Learning, Wind power, Wind speed.

1 Introduction

The raise of greenhouse gases (GHG) concentration in the atmosphere is most likely the main driver of the changes on the Earth's climate. The energy sector accounts for a quarter of the world GHG emissions [1], being the one with the largest share. Therefore, it is not possible to discuss climate change without considering the energy sector.

The ageing of conventional power plants, technological advances and cost reductions are allowing cleaner sources, mainly solar and wind-based systems, to boost its share in the electricity mix at the expense of fossil fuels [2]. Simultaneously, the electrification of significant sectors, such as transport and heating, are increasing the load demand while the system decentralization is altering the load patterns and the energy flow, as consumers are changing their roles to become prosumers, i.e., someone who both produces and consumes energy [2]. These changes are imposing a great challenge to the sector and require several adaptation measures which involves technical, economic and political issues [3] that must be applied to ensure a reliable, affordable and safe electricity.

The uncertain availability of natural resources and, consequently, of energy supply can cause grid instability issues, such as overvoltage and frequency deviations. To overcome this situation, the system must be flexible and resilient enough in order to cope with rapid generation and load changes and balance them at every moment. One way to ensure more trustworthiness and better manage of the system is by anticipating its behavior. When accurate forecasts are made, the decisions regarding the power system operation, maintenance and planning become more efficient [4, 5].

The technological developments in computing and the ever-growing data availability from the system operators, supervisory, control and data acquisition (SCADA) systems and weather forecasts allow the creation of sophisticated and efficient computational methods that can be applied in the energy sector. More specifically, ML models have shown great capability to generate accurate predictions of the system behavior from energy generation to consumption and can benefit all stakeholders to optimize their activities.

The main goal of this work is to investigate and propose a ML method to generate forecasts of electrical load demand and wind speed in different time scales by incorporating an error correction step on the initial models results. Also, the wind power output was estimated by using the wind speed forecasts and a wind power curve clustering and adjustment method.

This paper is divided in 7 sections. After this section, in Section 2, an overview of ML models and their application for electrical load, wind speed and power forecasting are introduced. Sections 3 and 4 present the databases and methodology for the electrical load and wind speed and power forecasts, respectively. In Sections 5 and 6, the results obtained with the load demand and the wind speed and power models are discussed respectively. Finally, in Section 7, the major findings of this research are summarized and suggestions for future works are proposed.

2 Background

2.1 Machine Learning

ML refers to computer algorithms that are able to learn from data and previous experiences without being explicitly programmed to do so [6]. ML is a subset of artificial intelligence, which concerns the ability that a computer has to perform tasks that are associated with and normally require human intelligence [7]. With data being collected at such a fast rate and large amount, it became extremely hard for a human or simpler statistical models to handle all the data. On the other hand, these systems can process the data in a much quicker pace and help to improve the decision-making process.

Tasks that are usually associated with ML models are: image and speech recognition, medical diagnosis, products recommendations, spam filtering and traffic prediction. Depending on the task and on the available data, ML problems can be classified as: supervised learning, unsupervised learning and reinforcement learning [8].

In supervised learning problems, the available data comprises features that are associated with labeled data. The name supervised learning actually comes from the fact that every input has a target value to learn from [7]. Normally, they are used for classification tasks, where the model labels the inputs (discrete output); and regression tasks, where the program returns a numerical value given some input (continuous output). Examples of classification and regression problems are, respectively, spam filtering and energy price prediction.

In unsupervised learning problems, the available data is not labeled, and the system tries to identify useful properties and patterns on the data by itself. Commonly, they are used to learn a probability distribution function on the dataset. The learned pattern can then be used to denoise or divide the data into groups of similar examples (clustering). One typical application of these methods is to identify credit card frauds [7].

Finally, the reinforcement learning models does not use solely the provided data and also interact with the environment by means of a feedback loop that evaluate its output [9]. Based on the evaluation of previous experiences, it must learn by itself the best strategy to avoid bad feedbacks and maximize good ones. Examples of reinforcement learning are inventory management and traffic signal control.

Since the object of this research is a supervised learning problem and, more specifically, a regression task to predict electrical loads, wind speed and power, the next sections present recent studies within these fields.

2.2 Electrical Load Forecasting

The increased need for electricity, the change of power generation mix and load pattern are some of the already observed transformations. With all this, the electricity grid behavior becomes more unpredictable and its operation and control more complex which can lead to greater supply instability. Therefore, techniques and methods capable of increasing system reliability are extremely important.

One way to ensure more trustworthiness and better manage the system is by anticipating the load demand. When accurate forecasts are made, the decisions regarding the power system operation, maintenance and planning become more efficient [4, 5]. Furthermore, improvements on energy policies and tariffs can be achieved. In recent years, much of the research have been focused on the development of models to forecast the electrical load in different time horizons. These periods are often classified as short-term, which goes up to 1 week ahead; medium-term, from weeks to 1 year; and long-term, for future years [10]. Also, each of these timeframes have different applications, with the first being more important for daily operation and cost minimization [11] and the others for fuel reserves estimation, maintenance and capacity expansion planning [12].

Several approaches have been employed recently to make these forecasts. These approaches can be separated into three categories: statistical based, computational intelligence based and hybrid approaches [13]. Statistical models usually embrace uni or multivariate time-series models and regression techniques, such as Autoregressive Integrated Moving

Average (ARIMA) and linear regression, while computational intelligence models are mainly related to ML approaches. Commonly, statistical based methods are less memory intensive due to its simplicity and, thus, faster to execute. On the other hand, ML models are capable of identifying nonlinear relationships between inputs and outputs and can be extremely time consuming. However, this level of complexity can be necessary to achieve better results [10]. Finally, hybrid models combine features from statistical and computational intelligence models. They generally use the former to preprocess and/or select the input data that will be fed to the latter.

These forecasting models can be implemented using a large range of inputs that can be divided into four major categories: socio-economic, such as the region average income and GDP; environmental, such as mean temperature; building and occupancy, which is related to building sizes and dwelling types; and time index, that is related to the date stamps used as inputs [14]. Also, electricity demand historical data is generally taken into consideration. However, the choice of these inputs will depend on the time scale and type of the region of the study. Usually, historical, environmental and time index data are more common for short-term forecasts in a region scale [14].

2.3 Wind Speed and Power Forecasting

Over the last decades, power production from RES have experienced a sharp expansion. With growing concerns about the global climate crisis, the expectation is that this trend will continue to escalate. Amid those sources, wind energy arises as one of the most attractive due to its high generation capacity, efficiency and cost-benefit ratio [15]. However, as others RES, wind power generation also suffers from resource stochasticity and intermittency which impose a challenge to its large-scale penetration as it can undermine the whole electrical system operation [17].

To surpass these issues, accurate wind energy forecasts can play an essential role. They can, for example, help to optimize the market prices, producer profits [17] and electricity supply reliability [18]. Several factors, such as the environmental conditions, weather and time of the day can affect the predictions [20]. Nonetheless, the volatile characteristics of the wind hamper forecasts models to be precise [20]. Hence, minimizing the uncertainty associated with the wind is a keystone to improve wind energy forecasts [19].

As for electrical load forecasts, the wind speed and power forecasting models can be classified according to the time horizons: very short-term, for forecasts up to 30 minutes ahead; short-term, which goes from 30 minutes to 6 hours ahead; medium-term, from 6 hours to 1 day ahead; long-term, from 1 day to 3 days ahead; and very long-term, from 3 days to longer forecasts [18,21,22]. The application of these forecasts goes from real-time market and grid operation for shorter periods and unit commitment decisions, maintenance planning and reserve requirements for longer horizons.

With regard to the methodology, the models can be divided into physical, statistical, ML and hybrid models. Physical models work with complex mathematical equations that relate the terrain, atmosphere and external conditions to generate weather forecasts [18]. They are also known as NWP models and can make forecasts for different time horizons, height levels and parameters. However, they tend to be extremely time consuming due to the high level of complexity [21]. On the other hand, statistical models are much simpler and can run faster. They usually have good results for the short-term but their performance quickly worsens for longer terms [18]. As mentioned in Section 2.2, ML models can identify nonlinear relationships between inputs and targets by working with algorithms that enable them to learn from data. They usually are not as time consuming as physical models and have better results than statistical methods. Finally, hybrid models can combine some techniques to make forecasts.

3 Electrical Load Forecasting Models

3.1 Database

The database for the electrical load forecasts consisted of historical data collected from Enedis[1], which is the major distribution system operator in France. This data is freely available, and it is published in a consolidated form, so that the information from producers and consumers is kept anonymous. The collected data refers to the measured load demand in an industrial area connected to the medium voltage grid (for subscribed power up to 36kVA) and was sampled at every 10-minutes. Furthermore, it came in a .csv format, which could be directly used by the Python scripts that

[1] Webpage: https://data.enedis.fr/

were developed later. The retrieved period ranges from 01-10-2016 to 31-03-2017 in a total of 26208 samples.

3.2 Methodology

3.2.1 Feed-Forward Neural Networks

The considered FFNN is composed of nodes (or neurons) that are distributed across different layers, namely input, hidden and output layers. Each node in a layer is linked to the ones in the next by means of a weight parameter that measures the strength of that connection, forming a fully connected network structure that reminds the nervous system. The operating principle of neural networks can be described as a sequence of functional transformations [8]. For a given layer $l \in \{1, \dots, L\}$, where L is the number of layers, a quantity called the activation value of the next layer $A^{[l]} = \left(a_1^{[l]}, \dots, a_j^{[l]}\right)$ can be calculated as a linear combination of inputs $X^{[l-1]} = \left(x_1^{[l-1]}, \dots, x_i^{[l-1]}\right)^T$ and weights $W^{[l]}$ in the form

$$A^{[l]} = W^{[l]} X^{[l-1]} + b^{[l]} \tag{3.1}$$

where

$$W^{[l]} = \begin{pmatrix} w_{1,1}^{[l]} & w_{1,2}^{[l]} & \cdots & w_{1,j}^{[l]} \\ w_{2,1}^{[l]} & w_{2,2}^{[l]} & \cdots & w_{2,j}^{[l]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,1}^{[l]} & w_{i,2}^{[l]} & \dots & w_{i,j}^{[l]} \end{pmatrix} \tag{3.2}$$

and $b^{[l]} = \left(b_1^{[l]}, \dots, b_D^{[l]}\right)^T$ is a parameter called bias, which is used to adjust the output. The subscripts i and j represents the number of nodes or dimension of layers $l-1$ and l, respectively.

The activation value $A^{[l]}$ is transformed by a nonlinear, differentiable function $h^{[l]}$ that is named activation function as in equation (3.3), resulting in the next layer input vector $X^{[l]}$. For hidden layers, the activation function is a logistic sigmoid or hyperbolic tangent function, while for the output layer it is the identity function.

$$X^{[l]} = h^{[l]}\left(A^{[l]}\right) \tag{3.3}$$

Equations (3.1) and (3.3) present recursive calculations that constitutes a process known as forward propagation [8]. This name comes from the fact that the information is flowing forward through the network, and this is the reason why this type of model is called FFNN.

The parameters optimization is done with gradient descent-based calculations. With this approach, the required partial derivatives are related to the two parameters of a FFNN model: the weights and biases. Appling equation (3.4) to the last layer of this model results in (3.6) and (3.7).

$$p^{[i+1]} = p^{[i]} - \eta \bigtriangledown E(p) \tag{3.4}$$

where *p* represents the model parameters, η is the learning rate, and *E(p)* is the loss function (3.5), which is the Mean Squared Error (MSE). In this equation, x_i is the forecasted value, t_i is the target value and n is the number of points in the dataset. The search for the loss function minimum is commonly done by computing its gradient ($\bigtriangledown E(p)$), which is the vector containing the partial derivatives of $E(p)$ [7]. The partial derivative $\frac{\partial}{\partial p} E(p)$ indicates how the function changes with a small change in one of the parameters. Therefore, the gradient vector points to the direction of the steepest increase of the function. As the learning algorithm goal is to minimize the error, with this approach the parameters can be updated at each iteration i by going in the opposite direction.

$$E(p) = \mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(f(x_i, p) - t_i\right)^2 \tag{3.5}$$

$$W_{[i+1]}^{[L]} = W_{[i]}^{[L]} - \eta \frac{\partial E}{\partial W^{[L]}} \tag{3.6}$$

$$b_{[i+1]}^{[L]} = b_{[i]}^{[L]} - \eta \frac{\partial E}{\partial b^{[L]}} \tag{3.7}$$

By computing the partial derivatives in (3.6) and (3.7), one can notice that the algorithm goes through each layer in reverse, measuring the error contribution from each connection and updating the parameters accordingly [9], in a process known as backpropagation. This algorithm is usually more computationally efficient than other numerical methods such as finite differences [8] as it avoids unnecessary calculations by reusing some of the partial derivatives from previous layers.

3.2.2 Load Forecast

Preliminary tests using different ML models such as SVM, RF and FFNN demonstrated a better forecast performance of the latter. Thus, only FFNNs were further investigated in this work. The models were created with Python's TensorFlow library, together with scikit-learn, numpy and others for three different forecast horizons: 10-minutes ahead, 1-hour ahead and 12-hours ahead. For each horizon, six different inputs combinations were tested. These inputs included previous load measurements, which were chosen based on a preliminary analysis, and time indexes. At each timestep, the targets of the model were the measured electrical load.

In a first moment, 20% of the data was separated (≈5000 samples) to be tested by the initial electrical load forecasting model while the other 80% was used as the training set. In order to have more reliable results, the models were trained with 3-fold cross validation. Later, the training data was scaled to be between 0 and 1. Data scaling is a common pre-processing technique which can improve the model learning process. After the training, the models were ranked based on the average of the validation set Root Mean Squared Error (RMSE) over the 3 folds. Also, the Mean Average Error (MAE) was another metric used to evaluate the performances.

To find the best parameters set for the FFNN, a grid search was performed. The searched parameters were the number of nodes in the hidden layer, the activation function and the learning rate, in a total of 48 models for each forecast horizon. Also, the chosen optimizer was the Adam algorithm [23]. Finally, the configuration with the best results were compared with the persistence model in the 10-minutes ahead horizon and to a linear regression in the 1-hour and 12-hour horizons.

3.2.3 Error Forecast

A model to predict the errors of the initial load forecast model was developed for the three horizons. The data used in this step was provided by the errors observed on the test set, computed as the difference between the initial model forecasts and the actual values. After analyzing the errors and based on the error forecasting methodology proposed in [24], three different input configurations using previous errors were created. At each timestep, the targets of the model were the corresponding initial model error. The data preparation and the model training were conducted as described in the previous section, with a slight change on some parameters. A total of 24 models were evaluated for each horizon. Finally, the errors predicted by the best model were combined with the initial forecasts to achieve an adjusted electrical load estimation.

4 Wind Speed and Power Forecasting

4.1 Database

The database for the wind speed and power forecasts consisted of NWP data from the European Centre for Medium-Range Weather Forecasts (ECMWF[2]) and historical data from the SCADA system of three wind turbines in a wind farm in the south of South America.

The retrieved NWP data referred to 24 forecast hourly steps (steps 1 to 24) from 31-12-2018 at 12:00h UTC to 12-12-2020 at 12:00h UTC. The collected parameters were the U and V components of the wind, which are the components parallel to the x and y axis respectively, at the wind farm location and at pressure level 133 (equivalent to a geopotential and geometric altitude of 106.54m), as this height was the closest to the turbines' hub height. In addition to that, SCADA data sampled every 10-minutes from three turbines was collected. For two of them (Turbines A and B), the period ranged from 01-01-2019 to 10-12-2020 while for the other (Turbine C) the last date was 30-04-2019, totalizing 17040 and 2880 samples, respectively. To match the NWP sample rate, this data was resampled to a 1-hour basis. Furthermore, the available parameters were production status information, average wind speed and power.

4.2 Methodology

4.2.1 Wind Speed Forecast

The wind speed forecasts were made for two different horizons: 12 and 24-hour ahead. In every case, the inputs were the forecasted wind speeds by the NWP model and the output (target) was the SCADA wind speed average. For the 12-hour horizon, three different inputs combinations were tested, while for 24-hour horizon, only the configuration with the best result in the first case was appraised. The inputs were chosen based on a preliminary analysis of the data.

The data preparation and the model training were conducted similarly to the electrical load case. However, a 5-fold cross validation was used here instead of 3-fold and the activation function was the same in every case (rectified linear unit). Furthermore, the grid search was carried only for Turbine A, in a total of 48 models for 12-hour horizon and 16 for the 24-hour. For the other turbines, only the top-3 models in every forecast horizon for Turbine A were retrained and evaluated. The results were compared with the ones of the NWP model and of an adjusted version of the NWP model (ANWP), which was created by adding the average error of the NWP model over the training set on the test set. The error metrics used for the wind speed forecasts were the RMSE, the MAE and the R2 coefficient.

[2] Webpage: https://www.ecmwf.int/

4.2.2 Error Forecast

The error forecast for the wind speed model was performed only for Turbine A in the 12-hour horizon, as the initial models were proposed for this condition. The data for the error forecast model was obtained from the test set of the wind speed correction model by calculating the difference between the measured wind speed and the forecasts made by this model.

To predict the errors of the initial wind speed forecasting model, an approach similar to the one described in Section 3.2.3 was used. In this case, only one configuration was tested: for every 12 timesteps (12 hours), an input vector consisting of the previous 12 observed errors and the following 12 forecasts and an output with the next 12 errors (target vector), in a multioutput neural network approach. A total of 18 models were analyzed.

4.2.3 Wind Power Curve Modeling

Similarly to the error forecast, the wind power curve was obtained only for Turbine A but, here, only the SCADA data for 2019 was considered. A typical wind power curve can be divided into four different operating regions, as illustrated in Figure 1. The power curve modeling performed in this work refers to the region II, in which the wind speed is between the cut-in and rated speed of the turbine and its relationship with the output power is non-linear.

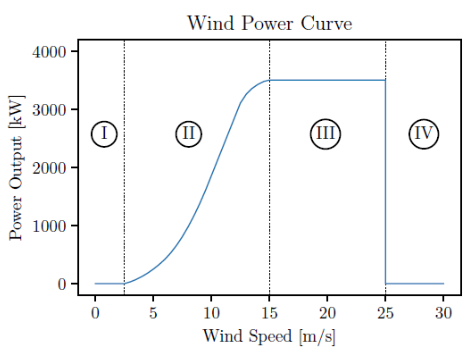

Figure 1: Typical Wind Power Curve

Using the SCADA measurements, samples outside region II and with any abnormal measurement, such as turbine brake active, misalignment between turbine yaw and wind speed or active event/alarm code, were removed from the dataset. However, after removing these points, it was verified that several non-representative samples were still in the dataset. Hence, another approach to filter those points was needed.

The chosen solution was the Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which is a clustering algorithm that aims to separate high density from low density areas of the dataset [25]. The working principle of this method is based on grouping together points that are sufficiently close to each other (neighbors' samples). To do that, two parameters must be controlled: a distance measurement, which can come from any distance function, and a minimum number of samples to constitute a cluster [26].

After applying the DBSCAN algorithm and removing the outliers from the dataset, it was possible to obtain the wind power curve. To do that, different polynomial curves were fitted by changing the degree from 3 to 7 and the best one chosen based on a visual analysis. Equation 4.1 presents the formulation for these curves, where a represents the coefficients and n indicates the degree of the polynomial. Finally, with the polynomial fit for region II and, considering the expected power output for regions I, III and IV, the wind power curve was defined.

$$y = \sum_{i=0}^{n} a_i x^i \qquad (4.1)$$

4.2.4 Wind Power Forecast

After computing the wind speed with the methodology described in Subsections 4.2.1 and 4.2.2 and obtaining the wind power curve as detailed in Subsection 4.2.3, the output power was calculated by using the estimated wind speed with the power equations. Finally, the results were compared to the ones observed using the original NWP forecasts and the SCADA wind speed with the proposed power curve.

5 Electrical Load Forecasting Results

As mentioned in Section 3, the electrical load forecasts were made for the three time horizons. However, due to space limits, only the results for the 1-hour ahead horizon are presented in detail.

5.1 Initial Results

The best model configuration for the initial load forecasts had the previous 3 periods, 3 days at the time of the forecast and the time index corresponding to the forecasted hour decomposed into sine and cosine components as inputs, in a total of 8 inputs. Also, it had 48 nodes in the hidden layer and the hyperbolic tangent as activation function.

In Table 1, the results obtained with this configuration and three different learning rates are presented with the metrics for the baseline model (linear regression). One can notice that the model B2.8a performed better than the other proposed models and, therefore, it was chosen to be evaluated using the test set. The results suggest that the higher learning rate enabled the model to be closer to a loss curve minimum at the end of the training. Furthermore, the proposed model achieved superior performance than the linear regression model in the order of 10.7% over the RMSE and of 11.3% over the MAE.

5.2 Error Forecast

The best model configuration for the error forecasts had the errors of the previous 4 periods and 4 days at the time of the forecasts as inputs. Also, it had 12 nodes in the hidden layer and the rectified linear unit as activation function. In Figure 2, the error forecasts made by the proposed model are shown. One can notice that the predictions were very accurate in general despite the spikes' underestimation.

Table 1: Results of the 1-hour ahead initial and baseline electrical load forecasting models

Model	LR	RMSE (kW)		MAE (kW)	
		Val.	Test	Val.	Test
B2.8a	0.01	339.43	**376.25**	261.33	**303.76**
B2.8b	0.005	357.58	-	276.05	-
B2.8c	0.001	1037.6	-	736.56	-
Bas.	-	-	421.48	-	342.3

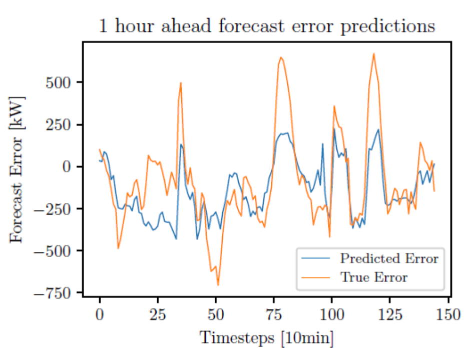

Figure 2: 1-hour ahead electrical load error forecasts

5.3 Adjusted Results

As described in Section 3, the adjusted results were obtained by combining the initial forecasts with the errors predictions. In Table 2, a comparison between the results obtained with the initial, adjusted and baseline forecast models for the 1-hour ahead scenario can be seen. Originally, the initial model had a performance only slightly superior to the one of the linear regression model (see Table 1). With the proposed methodology, the initial results were improved by 32.1% on the RMSE and 36.2% on the MAE. In Figure 3, it is possible to notice that the model correctly adjusted the initial forecasts between timesteps 0 and 50 and close to timestep 150, while it did not change the accurate predictions of some load spikes.

Table 2: Results of the 1-hour ahead electrical load forecasts

Model	RMSE (kW)	MAE (kW)
Initial	352.95	295.43
Adjusted	**239.34**	**188.49**
Baseline	371.58	307.11

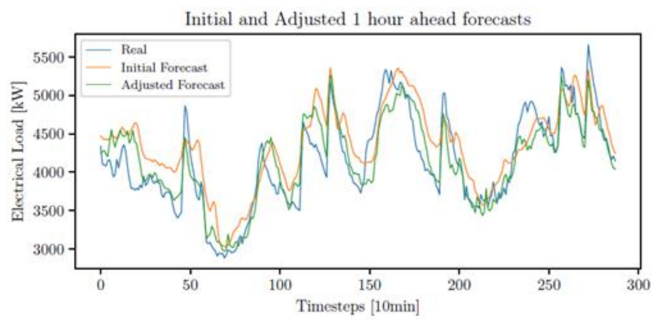

Figure 3: 1-hour ahead electrical load forecasts

6 Wind Speed and Power Forecasting Results

As mentioned in Section 4, the wind speed forecasts were made for three turbines in two time horizons. However, due to space limits, only the results for Turbine A in the 12-hour ahead horizon are shown here.

6.1 Initial Results

The best model for Turbine A in the 12-hour ahead horizon had 15 nodes in the hidden layer and a learning rate of 0.001. However, it was noticed that all the models had similar performances regarding the error metrics. This is possibly explained by the use of the original NWP wind speed forecast at time t as an input in every model. This input had the highest correlation with the target (wind speed measured in SCADA). Also, using other timesteps close to t seemed to have added a little more information to the models while further timesteps did not.

In Table 3, the results obtained with the best proposed model and the baseline models is shown. One can notice that the proposed model had significant better results. The improvements over the initial NWP forecast were about 23.3% on the RMSE, 24.8% on the MAE and 41.3% on the R2, while for the ANWP they were 10.7%, 10.5% and 11.7%. This result suggests that the errors in the NWP model is not only caused by some bias and/or the difference between the model height and the turbine hub height. The use of an ANN was able to find and correct other error patterns and is, thus, a reasonable and efficient choice to enhance NWP wind speed forecasts.

6.2 Error Forecast

Differently than the results achieved with the electrical load error forecasting models, in this case, the proposed model was not able to find any pattern in the data (see Figure 4). One of the possible reasons is related to the chosen input parameters. For the load forecast case, the preliminary analysis showed a significant correlation between previous observations and the target values while, here, this relationship was not observed. Thus, it is reasonable to conclude that more features would be necessary to improve the error forecasts with this approach. Also, due to the stochastic nature of the wind, it is possible that the error pattern is random and, therefore, much harder to predict.

Table 3: Results of the 12-hour ahead initial and baseline wind speed forecasting models

Model	RMSE (m/s)	MAE (m/s)	R2
PB10.12.TA	**1.290**	**0.986**	**0.705**
ANWP	1.444	1.102	0.631
NWP	1.685	1.312	0.499

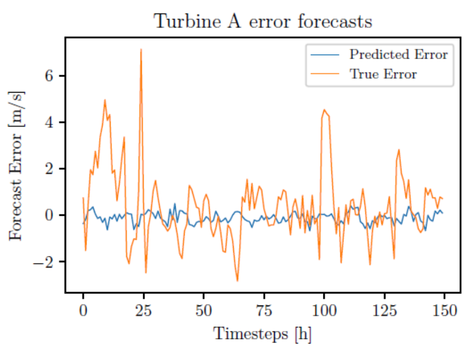

Figure 4: Turbine A 12-hour ahead wind speed error forecasts

6.3 Adjusted Results

In Table 4, the metrics for both initial and adjusted models are presented. One can notice that the proposed wind speed error forecast model was not able to improve the results as it was observed for the electrical load scenario. On the contrary, a slight performance decline was observed. This outcome was somehow expected after the findings in the previous section, which showed the poor performance of the proposed model to forecast the errors. However, it is worth mentioning that the two proposed models had better results than the NWP model. Finally, in Figure 5, the forecasts of the initial and NWP models are shown.

Table 4: Results of the 12-hour ahead wind speed forecasts

Model	RMSE (m/s)	MAE (m/s)
Initial	**1.386**	**1.050**
Adjusted	1.393	1.058
NWP	1.655	1.268

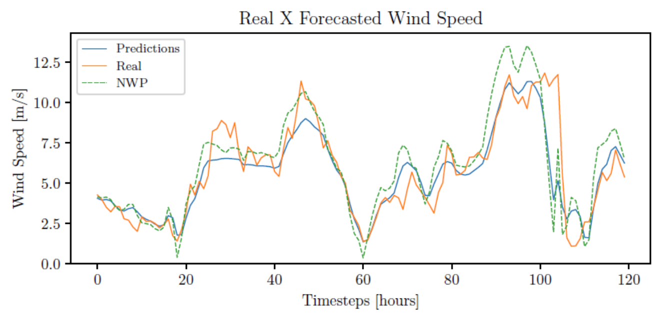

Figure 5: 12-hour ahead wind speed forecasts

6.4 Power Curve Modeling

As explained in Section 4, after removing measurements that did not follow regular operating conditions, several samples in the operating region II still behave differently from what it was expected and, to filter those points, the DBSCAN clustering algorithm was used. The results obtained with this method can be seen in Figure 6. The proposed approach was able to correctly classify the samples more distant to the turbine's expected behavior as outliers. Those points were removed from the dataset and not used to adjust the wind power curve.

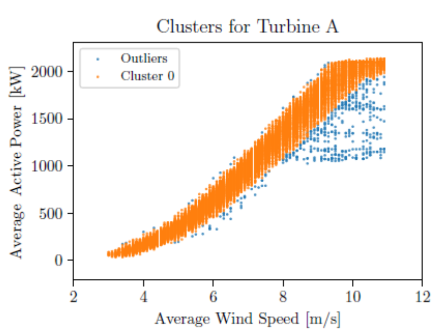

Figure 6: DBSCAN clusters for Turbine A

With the remaining samples, a 4^{th} degree polynomial was fitted for region II. Finally, with this polynomial, and considering an output power of zero for regions I and IV and the nominal power in region III, the wind power curve for this turbine was defined.

6.5 Power Estimation

The adjusted wind power curve was used to estimate the wind power production using the wind speeds measured with the SCADA and the ones of the proposed initial model and the NWP. The first was chosen so that the best possible result with the available data could be analyzed, as the error corresponding to this estimate is probably related to noise and sampling of the data. The choice for the initial model wind speed is due to its better results when compared to the adjusted model, which was not able to improve the initial results. Finally, the NWP wind speed was used to generate the power forecasts of the baseline scenario.

The results of the wind power forecast can be seen in Table 5. One can notice that the power estimates made with proposed model wind speed presented an improvement of 26.5% and 24.5% over the RMSE and MAE of the ones obtained using the NWP wind speed, moving closer to the results found using the SCADA wind speed. These findings validate the proposed methodology and show that having more accurate wind speed forecasts is crucial to improve the power predictions.

Table 5: Results of the wind power forecasts for Turbine A

Input	RMSE (kW)	MAE (kW)
SCADA Wind Speed	73.40	214.45
Proposed Model Wind Speed	**266.86**	**385.56**
NWP Wind Speed	363.18	510.90

In order to illustrate the achieved results, Figure 7 shows the true wind power output and the forecasts made with both the proposed model and NWP wind speeds. First, it is noticeable the better performance of the proposed model when compared to the baseline during most of the time. The main exception can be seen during nominal power production (from timestep 0 to 30 approximately). At those timesteps, the NWP seemed to perform better. A probable explanation to this is that the proposed model primarily adjusts the average wind speed and, therefore, the higher wind speeds turn out to be reduced and the power output do not correspond to the nominal power. However, in general, the proposed model tends to be more accurate and generates better estimates.

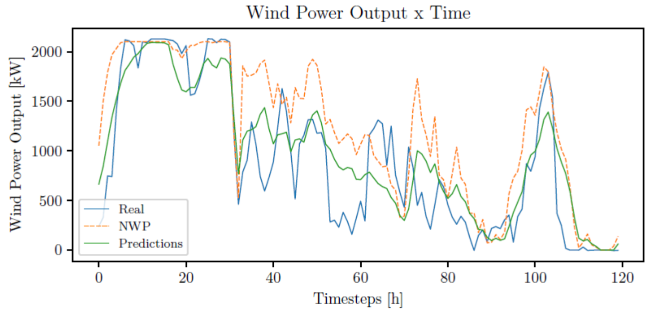

Figure 7: Wind Power forecasts for Turbine A

7 Conclusions

In this work, a FFNN to forecast electrical load demand, wind speed and power in different time scales by incorporating an error correction step following the initial models results was proposed.

With regard to the electrical load demand forecasts, several models were created to make predictions in three time horizons: 10-minutes, 1-hour and 12-hours ahead. However, due to space limits, only the results for the 1-hour ahead were presented here. The results demonstrated that the proposed initial model already outperformed the linear regression model (baseline). This difference was strengthened by the incorporation of the proposed error correction step, which resulted in a more accurate model.

For the wind speed forecasts, models for three turbines in a wind farm were created to predict two time horizons: 12-hours and 24-hours ahead. Also due to space limits, only the results for one of the turbines (Turbine A) in the 12-hour horizon were presented here. The results showed that the proposed initial model was able to slightly outperform the benchmark models, namely the NWP and an adjusted NWP. The findings suggest that the proposed model was able to find and correct error patterns other than the one caused by the difference between the NWP model height and the turbine hub height. However, in this case, the incorporation of the error correction step was not able to enhance the initial forecasts.

In addition to that, the DBSCAN clustering algorithm was used to identify samples that apparently did not have the expected wind speed and power relationship. The performance of this method was satisfactory and, after the removal of these samples, the wind power curve was approximated with a polynomial. With the adjusted curve and the wind speed estimated by the proposed model, wind power forecasts were made. The results accomplished with the proposed wind speed model were superior than those of the original NWP.

The findings of this research suggest that the proposed methodology can be beneficial for the energy sector, as it provided enhanced forecasts with ML models that work with limited amount of data. When data is scarce, it is essential to extract as much information as possible from it. The analysis of the errors showed that sometimes there is more information in the data than the models are able to find and that including an error correction step can improve the forecasts.

REFERENCES

[1] IPCC. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2014. ISSN 00220248. URL https://www.ipcc.ch/report/ar5/syr/.

[2] IEA. Secure Energy Transitions in the Power Sector, 2021. URL https://www.iea.org/reports/secure-energy-transitions-in-the-power-sector.

[3] IEA. Power systems in transition, 2020. URL https://www.iea.org/reports/power-systems-in-transition.

[4] T. Ahmad and H. Chen. A review on machine learning forecasting growth trends and their real-time applications in different energy systems. Sustainable Cities and Society, 54 (October 2019):102010, 2020. ISSN 22106707. doi: 10.1016/j.scs.2019.102010. URL https://doi.org/10.1016/j.scs. 2019.102010.

[5] M. Q. Raza and A. Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews, 50:1352–1372, 2015. ISSN 18790690. doi: 10.1016/j.rser.2015.04.065. URL http://dx.doi.org/10.1016/j.rser.2015.04.065.

[6] A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal, pages 210–229, 1959.

[7] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016. URL http://www.deeplearningbook.org.

[8] C. M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

[9] A. Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly Media. ISBN 9781491962299.

[10] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan. Load forecasting, dynamic pricing and DSM in smart grid: A review. Renewable and Sustainable Energy Reviews, 54:1311–1322, 2016. ISSN 18790690. doi: 10.1016/j.rser.2015.10.117. URL http://dx.doi.org/10.1016/j.rser.2015.10.117.

[11] H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems, 16(1):44–55, 2001. ISSN 08858950.doi: 10.1109/59.910780.

[12] D. Solyali. A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. Sustainability (Switzerland), 12(9), 2020. ISSN 20711050. doi: 10.3390/SU12093612.

[13] K. B. Debnath and M. Mourshed. Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88(August 2016):297–325, 2018. ISSN 18790690. doi: 10.1016/j.rser.2018.02.002. URL https://doi.org/10.1016/j.rser.2018.02.002.

[14] C. Kuster, Y. Rezgui, and M. Mourshed. Electrical load forecasting models: A critical systematic review. Sustainable Cities and Society, 35(August):257–270, 2017. ISSN 22106707. doi: 10.1016/j.scs.2017.08.009. URL http://dx.doi.org/10.1016/j.scs.2017.08.009.

[15] IRENA. Renewable Power Generation Costs in 2020, 2021. ISSN 1476-4687.

[16] A. Tascikaraoglu and M. Uzunoglu. A review of combined approaches for prediction of short-term wind speed and power. Renewable and Sustainable Energy Reviews, 34:243–254, 2014. ISSN 13640321. doi: 10.1016/j.rser.2014.03.033. URL http://dx.doi.org/10.1016/j.rser.2014.03.033.

[17] S. S. Soman, H. Zareipour, S. Member, O. Malik, and L. Fellow. A Review of Wind Power and Wind Speed Forecasting Methods With Different Time Horizons. pages 1–8, 2010

[18] W.-Y. Chang. A Literature Review of Wind Forecasting Methods. Journal of Power and Energy Engineering, 02(04):161–168, 2014. ISSN 2327-588X. doi: 10.4236/jpee.2014.24023.

[19] M. S. Nazir, F. Alturise, S. Alshmrany, H. M. Nazir, M. Bilal, A. N. Abdalla, P. Sanjeevikumar, and Z. M. Ali. Wind generation forecasting methods and proliferation of artificial neural network: A review of five years research trend. Sustainability (Switzerland), 12(9), 2020. ISSN 20711050. doi:10.3390/su12093778.

[20] J. Wang, Y. Song, F. Liu, and R. Hou. Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. Renewable and Sustainable Energy Reviews, 60:960–981, 2016. ISSN 18790690. doi: 10.1016/j.rser.2016.01.114. URL http://dx.doi.org/10.1016/j.rser.2016.01.114.

[21] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian. A Critical Review of Wind Power Forecasting Methods-Past, Present and Future. Energies, 13(15):1–24, 2020. ISSN 19961073. doi: 10.3390/en13153764.

[22] I. Okumus and A. Dinler. Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Conversion and Management, 123:362–371, 2016. ISSN 01968904. doi:10.1016/j.enconman.2016.06.053. URL http://dx.doi.org/10.1016/j.enconman.2016.06.053.

[23] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pages 1–15, 2015.

[24] T. Pinto, H. Morais, and J. M. Corchado. Adaptive entropy-based learning with dynamic artificial neural network. Neurocomputing, 338:432–440, 2019. ISSN 18728286. doi: 10.1016/j.neucom.2018.09.092. URL https://doi.org/10.1016/j.neucom.2018.09.092.

[25] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pages 226–231, Portland, OR, 1996. AAAI Press.

[26] H. P. Kriegel, P. Kr¨oger, J. Sander, and A. Zimek. Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3):231–240, 2011. ISSN 19424795. doi:10.1002/widm.30.