

Wind Speed and Wind Power Forecasting using Statistical Models: AutoRegressive Moving Average (ARMA) and Artificial Neural Networks (ANN)

Pedro Gomes, Rui Castro

IST – Technical University of Lisbon, Portugal
Cie³ – Centre for Innovation in Electrical and Energy Engineering

Abstract

Wind power, i.e., electrical energy produced making use of the wind resource, is being nowadays constantly connected to the electrical system. This has a non-negligible impact, raising issues like network stability and security of the supply. An accurate forecast of the available wind energy for the forthcoming hours is crucial, so that proper planning and scheduling of the conventional generation units can be performed. Also, with the liberalization of the electrical markets worldwide, the wind power forecasting reveals itself critical to assure that the bids are placed with a minimum possible risk. This work addresses the issue of forecasting wind power with two statistical models, the Autoregressive Moving Average and Artificial Neural Networks. The basic theory and the respective application of these models to perform wind power prediction are presented. Furthermore, their forecasting ability is compared in three different case studies. At the end, some conclusions are drawn about the performance of both models regarding its forecasting capabilities when compared with reference models.

Keywords: Short-Term Wind Power Prediction, Autoregressive Moving Average, Artificial Neural Networks, Persistence.

1. Introduction

Wind power has gained a great importance in the electrical sector over the past years. The wind is nowadays a source of energy economically viable to explore in terms of electrical energy production, and thus competitive with other sources. It is also environmentally safe, since the corresponding process of energy production does not emit any pollutant gases. The growth witnessed in the installed capacity of wind power plants confirms these two aspects.

The last available figures (2011) account for around 200 GW of installed wind power worldwide, almost half of which is located in Europe. Since the beginning of this century, Portugal is also pursuing an ambitious plan on the promotion of renewable energy, in general, and on the wind power, in

particular. In April 2011, the Portuguese government issued the National Energy Strategy with the aim of further reduce the energy dependency from abroad. The target is to decrease this need from 83% in 2008 to 74% in 2020. This target is to be accomplished by the incorporation of 31% of renewables in the final energy consumption, while reducing this consumption by 20%, in relation to the projections in a BAU (Business As Usual) scenario.

It is expected that wind power plays a major role in these goals, as a capacity of about 7 GW is foreseen in the 2020 horizon, against 4,1 GW currently installed. In 2011, wind power accounted for 18% of the electrical energy consumption in Portugal. The recent developments in wind power in Portugal are depicted in Figure 1.

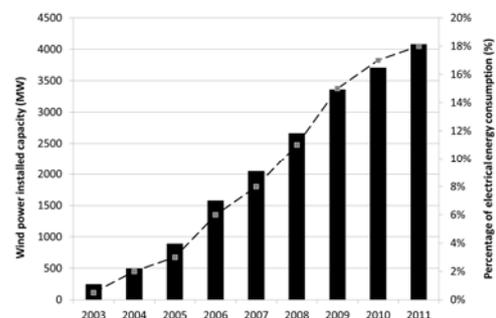


Figure 1. Wind power installed capacity in Portugal and percentage of electrical energy consumption. Source: REN – Portuguese Transmission System Operator.

The use of wind energy allows saving between 0,5 and 1 tons of greenhouse effect gas (GHG) per MWh that would be emitted to the atmosphere if natural gas or coal were used instead, respectively. This is a great contribution to the achievement of the Kyoto Protocol goal, the reduction GHG emissions by 5% of the 1990 levels over the five-year period 2008–2012. Regarding the economic benefits, in 2010, the wind power sector had a turnover of 40 billion Euros, and employed 670 000 people worldwide. It is also expected that this sector will offer one million jobs in 2012 [1].

The growing importance of wind power raises the issue of understanding its behavior and its impact in the electrical sector. Wind power production, being subject to the available wind, can only be controlled

within the margin of the possible production correspondent to the available wind, thus it has reduced control capacities. The extraction of energy from the wind should thus be maximized for economic and environmental reasons.

In comparison to the cost of the energy produced with a conventional coal or natural gas power plant, the cost of producing wind energy is still slightly higher. This cost is strongly related to the amount of wind energy produced: since the utilization of the wind is free, the more wind energy is produced the minor is its cost per MWh. Figure 2 shows how wind energy leveled costs reduce with the number of equivalent hours at full load (equivalent full load hours is the number of hours within one year during which the wind generator would have to run at full power in order to produce the energy delivered throughout that year).

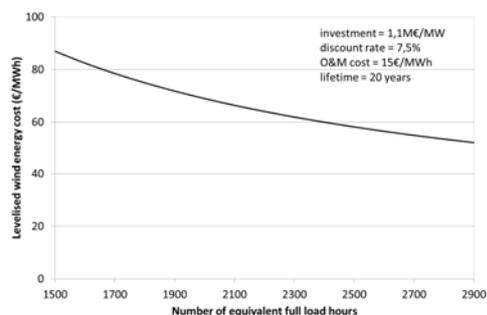


Figure 2. Wind energy leveled cost as a function of the number of equivalent full load hours.

As illustrated in Figure 2, the costs range from approximately 70–85 €/MWh at poor wind sites, to about 50–55 €/MWh at rather windy sites, with an average of around 60 €/MWh at a wind site with average wind speeds.

The need to maximize wind power production and to carefully schedule the electrical energy production, taking into consideration the expected load and the available power plants, raises the subject of wind power prediction. Furthermore, in electricity markets, wind power prediction also helps to plan the optimal behavior of its agents.

Regardless all the advantages listed above, wind power has some related issues too. One of these issues, perhaps the most important, is the uncertainty associated with it, a problem shared with other renewable energies. This is an issue because, in a power system, the total amount of electricity that is provided at each instant has to match a varying load from the electricity consumer. To achieve this in a cost effective way, the power plants must be scheduled in advance according to increasing marginal operating costs. The variable production pattern of wind power changes the scheduling of the other production plants. This is always cost inefficient and sometimes technically unfeasible, leading to the necessity of using extra reserves,

generally expensive backup generators in low efficiency states, or the unscheduled use of the interconnections, which is normally very expensive.

These reasons justify that the electrical system cannot be totally or even mostly dependent on renewable energy plants, in order to be able to maintain the electrical balance between the supply and the load. Thus, conventional power plants are required to be active, no matter what growth is witnessed in the renewable energies. Nevertheless, an accurate forecast of the renewable energies contribution is mandatory to minimize the unreliability factors and therefore optimize the network planning and operation in a framework of high renewables penetration. As the wind power is the most important renewable energy utilized today, the wind power forecasting problem is of extreme importance.

2. Short-Term Wind Power Forecasting Models

Short-term wind power forecasting models belong to a subclass of the wind power time prediction. The time scales concerning short-term prediction are in the range of some days for the forecast horizon and from minutes to hours for the time-step. The purpose of these models is to predict the wind farm output. This kind of prediction is mainly oriented to system management and maintenance tasks scheduling purposes, but also to help in the planning of the agents in the electricity spot market (daily and intraday). Thus, it is of interest to system operators, electricity companies and wind farm developers [2].

Wind power forecasting models are mainly divided in two categories: physical models and statistical models.

Physical models try to use some physical considerations (terrain features, altitude, roughness, obstacles, etc.) as much as possible to reach the best estimate of local wind speed and output power, and often make use of a Numerical Weather Prediction (NWP) model output. NWP models consist in weather forecasting models that are able to approximately estimate the evolution of some variables of interest, such as temperature, wind speed, humidity and pressure, in the points of a mesh. The way each point influences the others is also taken into account for the evolution of the mesh state computation. The data obtained is usually treated using Model Output Statistics (MOS), in order to reduce the remaining error [3]. These models are not always used, because they require a meteorological service operational model that is expensive and might not be worth the cost [4].

Statistical models try to find relationships between a wealth of explanatory variables in order to perform that estimation. These variables are usually wind power and/or wind speed data measured on-

line, and may also include the NWP models outputs. This process usually occurs in two steps: the training and the prediction. The training consists in the optimization of the model in order to explain the given data as best as possible, adjusting its internal variables in order to minimize the error between the real historical data and the same data reproduced by the model. The prediction step is the completion of the forecast, making use of the model already trained.

The combination of physical and statistical techniques may result in different approaches for wind speed prediction. The three main blocks used in order to obtain the forecasting results are [5]:

1. Supervisory Control And Data Acquisition (SCADA) data, which is the on-line data.
2. Meteorological forecasts from the NWP model outputs.
3. Terrain information (height, terrain roughness, and others).

The combination of these input blocks to the forecasting models leads to five possible different approaches [5]:

- (1): short-term using only SCADA as input (forecast horizon up to 6 hours).
- (2): physical or statistical approaches (good performance in forecast horizons greater than 3 hours).
- (2)+(3): physical approach (good performance in forecast horizons greater than 3 hours).
- (1)+(2): statistical approach.
- (1)+(2)+(3): combined approach.

This work focuses on the first of the referred approaches, denoted as block (1), and addresses the use and comparison of two statistical wind speed forecasting models using only SCADA data. The two studied models are:

- Auto Regressive Moving Average (ARMA) model;
- Artificial Neural Networks (ANN) model.

The main objective of this paper is to compare two statistical models usually applied to wind power forecasting, the ARMA model and the ANN. The forecasting methodologies applied are defined and uniform criteria to adjust the required settings in both models are established. The ARMA models and ANN are employed to perform forecasts with the same historical data, in order to be compared. The quantities to be predicted are wind speed and wind power. Conclusions about their performance and application are drawn at the end.

3. Fundamentals on Statistical Forecasting Models

This section presents the fundamentals of the mathematical formulation of the forecasting models that are to be dealt with in this paper. Furthermore, the basis of the persistence model, commonly used as reference in wind forecasting studies, is given.

3.1. Persistence model

Persistence is the most commonly used reference model in wind power forecasting, and the great majority of comparisons are also made against this model. As such, this model is used hereafter as a baseline for comparison.

This forecasting model is usually applied to one hour ahead time series forecasts, equaling the forecast to the last known value of the time series. For N hour ahead forecast, the prediction is equaled to the N last known values of the series.

For a time series $\{Y_t\}$, given a historical set of data $H_t = \{Y_0, Y_1, \dots, Y_t\}$, the forecast of the forthcoming k values of $\{Y_t\}$ by a persistence process pp is:

$$pp_k(H_t) = \{Y_{t-k+1}, Y_{t-k+2}, \dots, Y_t\} \quad (1)$$

In particular, the forecast of the forthcoming value of $\{Y_t\}$ by a persistence process pp is

$$pp_1(H_t) = Y_t \quad (2)$$

3.2. Autoregressive moving average models

The Autoregressive Moving Average (ARMA) model is a useful and powerful tool to describe the dynamics of an individual time series. This model allows one to estimate the forthcoming value of an individual time series as a linear combination of values already observed. The computation of the coefficients of this linear combination, which are the parameters of the model, is based on the time series itself, so that each value of the series is explained by the linear combination of some of its prior values, in the best possible way. This computation corresponds to the training step of the model, and the estimation of the forthcoming value corresponds to the prediction step.

A mixed p th-order autoregressive process and q th-order moving average process, ARMA (p,q), is formally given by [6]:

$$Y_t = C + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3)$$

In equation (3), $\{Y_t\}$ is the time series to be described, C is an internal constant value of the process, $(\phi_1, \phi_2, \dots, \phi_p)$ are the AR process parameters, $(\theta_1, \theta_2, \dots, \theta_q)$ are the MA process parameters and $\{\varepsilon_t\}$ is a white noise process.

A sequence $\{\varepsilon_t\}$ is called a white noise sequence if each ε_t is a random variable with mean zero and covariance σ^2 and, for every $t, \tau \geq 0$ with $t \neq \tau$, ε_t and ε_τ are uncorrelated. Formally,

$$E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_\tau) = 0 \quad (4)$$

This process includes a component computed from the known information up to the instant $t-1$, which is the autoregressive process, and another component that represents the uncertainty of the mean of $\{Y_t\}$, which is the moving average process.

To estimate the parameters of the ARMA models, a least squares minimization is often employed. This mathematical procedure tries to find the best-fitting curve to a set of values of a time series by minimizing the sum of the squares of the residuals – the differences between the values of the time series and the same values reproduced by the model.

3.3 Artificial neural networks models

Artificial Neural Networks (ANN) are structures inspired by the natural intelligence and the human ability to adapt its way of thinking to complex problems, in its learning form experience and generalizing capabilities. These structures are nonlinear data-driven approaches, capable of performing nonlinear modeling without a priori knowledge about the relationships between input and output variables. Thus, they are a powerful general and flexible modeling tool for forecasting purposes [7].

ANN consist of many interconnected identical simple processing units – neurons or nodes – organized in layers – an input, an output and one or more hidden layers – and each of the neuron's connections has an adjustable weight factor associated. They can be classified as static or dynamic networks. The static networks outputs only depend on the current value of the inputs, while in dynamic networks the units are governed by differential or difference equations, thus exhibiting memory.

The weights associated with the connections are adjusted through a supervised or unsupervised learning process. In a supervised learning process the pairs (inputs, outputs) are presented to the network and the weights are adjusted in order to minimize a predefined error function. These outputs presented with the inputs are known to be the desired outputs to the correspondent inputs. In an unsupervised learning process only the inputs are presented to the

network, and this is trained to identify different classes of data.

ANN may also be classified as feedforward and recurrent. Feedforward networks, in which the topology graph does not contain any directed cycles, are those usually applied to wind speed and wind power prediction [8], [9], [10].

The multilayer perceptron is a static feedforward network typically trained with a supervised learning process and one of the most applied to wind forecasting. Its typical structure is shown in Figure 3. The network units are the white circles and the connections are the line segments. In the input/output layer there is one unit per input/output. As illustrated, the inputs of each layer are the outputs of the previous one, multiplied by the weight w of each connection.

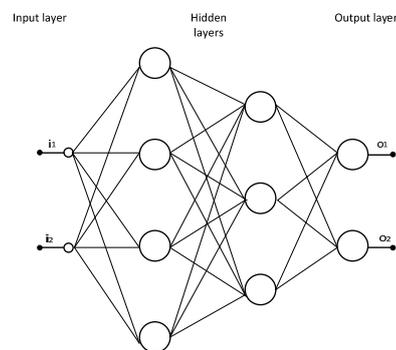


Figure 3. Structure of a multilayer perceptron network.

The output o of a node is the image of the weighted sum of all its inputs i by some activation function f , as illustrated in Figure 4 and formalized in equation (5). The activation functions employed in a multilayer perceptron can be linear, log-sigmoid or tan-sigmoid.

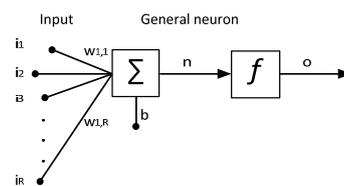


Figure 4. Basic operation of a multilayer perceptron network unit.

$$o = f \left(\sum_j w_j i_j + b \right) \quad (5)$$

During the network's training step the weights w and the constant value b are adjusted for each unit. To perform this, the backpropagation algorithm is employed.

The backpropagation algorithm is a gradient-based algorithm that seeks to minimize the prediction error of the network in the training data. For each

training sample (i_i, o_i) , the backpropagation algorithm tries to minimize the error:

$$E = \sum_i \|o_i - \hat{o}_i\|^2 \quad (6)$$

In equation (6), \hat{o}_i denotes the predicted network output, adjusting the weights w and constants b by gradient descent.

The basic algorithm can be summarized in the following seven steps [10]:

1. Initialize the network weights.
2. Present the first input vector, from the training data, to the network.
3. Propagate the input vector through the network to obtain the correspondent output.
4. Calculate the error signal by comparing the actual output to the desired (target) output.
5. Propagate the error signal back through the network.
6. Adjust weights to minimize the overall error (find an error surface's minimum).
7. Repeat steps 2 to 6 with the next input vector, until the overall error is satisfactorily small.

4. Application to Wind Speed and Wind Power Forecasting

The specific application of ARMA and ANN models to wind speed and wind power forecasting is addressed in this section.

4.1. ARMA models

Forecasting methodology. The employed ARMA methodology is performed in three steps:

1. Model definition: selection of the (p, q) order of the ARMA model, denoted as model structure.
2. Model training: estimation of the model parameters through a least squares minimization process.
3. Prediction of the forthcoming value: estimation of the forthcoming value of the time series.

If the estimation of more than one forthcoming value is required, one estimation at a time is performed, and from the first one on, the estimated values are used to perform the next.

Software tools. The simulations performed make use of the *GARCH* (Generalized Autoregressive Conditional Heteroskedasticity) *Toolbox* for use with Matlab® software [11]. Besides implementing the ARMA model, this toolbox also provides the possibility to implement the *Akaike and Bayesian Information Criteria* (AIC and BIC) and also the

lratiotest. These indices allow the evaluation of the performance of the model before the predictions are effectively done. Thus, it is possible to select the best between several predetermined models, although it is necessary to estimate the parameters of all these before the evaluation functions are employed.

Definition of the ARMA model structure. One of the main problems of ARMA models is the model structure definition, i.e., setting p and q . For the choice of p one may attend to the *Partial Autocorrelation Function* (PACF), which computes the autocorrelation between two elements of a time series separated by k , Y_t and Y_{t-k} , not considering the linear dependence on the intermediate elements. If it is zero for two elements, Y_t and Y_{t-k} , then Y_{t-k} does not have relevant information to help explain Y_t . Thus, the maximum value of p is given by the lag after which the PACF turns zero (or approximately zero), k_p . However, as the range of possible pairs (p, q) may be limited by the time available to perform the forecast, equation (7) has been used to set the maximum p , p_{max} . The range for the values of q is set in equation (8) and provides a reasonable trade-off between simulation time and model complexity.

$$p_{max} = \min\{k_p, p_{upper}\}, p_{upper} = 4 \quad (7)$$

$$q = \{1, 2, 4, 12, 24\} \quad (8)$$

4.2 ANN models

Forecasting methodology. The employed ANN methodology is performed in five steps:

1. Network architecture definition: number of layers, number of nodes and activation functions.
2. Data organization: in pairs (inputs, outputs) with the sizes established in step 1.
3. Data pre-processing: not always applied, but in some cases improves the performance of the network forecasting task.
4. Network training: estimation of the weights w and the constant values b , employing the backpropagation algorithm.
5. Prediction of the forthcoming value: by presenting the network an input vector with the past observations.

Software tools. The simulations performed make use of the *Neural Network Toolbox* for use with Matlab® [12]. This Toolbox allows one to implement both the multilayer perceptron ANN and the backpropagation algorithm.

Definition of the ANN model structure. One of the main issues in working with ANN models is the definition of the network architecture, i.e., the

number of hidden layers and nodes, the number of input and output nodes and the activation functions.

The hidden layers are responsible for capturing the data pattern and perform nonlinear mapping between input and output variables. Since one hidden layer is sufficient to approximate any complex nonlinear function with any desired accuracy [7], the used ANN models have one hidden layer. However, there is no consensus in the literature about the definition of the number of nodes in the hidden layer. Considering the simulations ran to test the implementation of the ANN model, the criterion established to address this issue is to compare the results obtained with three different networks, in which the number of hidden nodes is half the number of input nodes (rounded up if the number of input nodes is odd), equal to the number of input nodes or twice the number of input nodes.

The number of input nodes of the ANN employed corresponds to the number of past observations that contain relevant information to explain the value to be predicted. Thus, one may attend to the ARMA problem of setting p_{max} and set the number of input nodes equal to k_p , obtained with the PACF. One forecast is performed at a time, thus the network has one unit in the output layer.

The activation functions employed are the tan-sigmoid for the hidden layer and the linear for the output layer. The outputs of the tan-sigmoid activation function are confined to a specific interval: $d \in]-1, 1[$. When computing data that exceeds the values of this interval, it is necessary to make adjustments in order to obtain reasonable results from the network. The solution employed is to map the original data to the Standard Normal Distribution (with means equal to 0 and variances equal to 1) in the interval $I \in]-1, 1[$, process the mapped data and then transform the results back to the original interval.

4.3. Time series filtering

The SCADA data utilized in the wind speed forecasting typically consists in a wind speed and wind power time series of hourly mean values. These often evidence sudden variations and sharp peaks. A model that tries to explain all these variations and peaks generalizes poorly, thus the explanations of forthcoming values becomes more difficult. One solution to this problem is to smooth the time series, before it is presented to the model to train it. For this purpose, the *Financial Time Series Toolbox* for use with Matlab® software [13] was used.

4.4. Performance evaluation

The performance evaluation of the forecasting models is done in three steps:

1. The elements to be forecasted are removed from the original time series, and the remaining series consists in the training series.
2. The removed elements are predicted employing one of the models described in section 3 – Persistence, ARMA and ANN.
3. The predictions are compared with the original values according to the evaluation criteria described below.

The evaluation criteria are the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Relative Error (MRE), which is employed to compare the results of forecasts with time series of different orders of magnitude (see equation (9)).

$$MAE(N) = \frac{1}{N} \sum_{t=1}^N |Y_t - Y_t^*|$$

$$RMSE(N) = \sqrt{\frac{\sum_{t=1}^N (Y_t - Y_t^*)^2}{N}} \quad (9)$$

$$MRE(N) = \frac{1}{N} \sum_{t=1}^N \left| \frac{Y_t - Y_t^*}{Y_t} \right|$$

4.5. Physical considerations

The forecasts performed should consider the physical limitations of the wind turbines. These limitations are only considered in the computation of the performance evaluation criteria and may be described as follows:

- For wind speeds below a certain value u_0 , the extraction of energy from the wind does not compensate, therefore the wind turbine is shut down.
- For wind speeds above the nominal wind speed u_{nom} the increase of the extraction of energy from the wind does not compensate due to the required robustness of the wind turbine, therefore the wind turbine is set to operate at its nominal wind power P_{nom} .
- For wind speeds above u_{max} , a characteristic value of each wind turbine considered dangerously high, the wind turbine is shut down for security reasons.

5. Case-Studies and Results

To evaluate the performance of ARMA and ANN models three different case-studies are presented: a one hour ahead wind speed forecast study, for several consecutive hours, a simulation to support bidding decisions in MIBEL, the electricity Iberian

market and finally a one hour ahead wind power forecast assessment.

5.1 Case-Study 1 (CS#1) – One hour ahead wind speed forecasts

A frequently employed case to test wind speed forecast models is to perform the forecast for the forthcoming hour for several consecutive hours [14], [15]. In this paper, the forecasts of five consecutive days are performed, using the methodologies outlined above. To obtain the forecasting results, wind speed historical data of one year is used. ARMA models are used in simulations denoted as *simulations 1* and ANN are used in simulations tagged as *simulations 2*.

Simulation 1 – ARMA. This forecasting simulation is performed as follows. Once the ranges for the possible values of p and q are set, the parameters of all models for all possible combinations (p,q) are estimated. Then, the best model according to AIC and the best model according to BIC are selected, and from these two, the best model is selected employing *lratiotest*. This selection of the best model occurs in all hours. This simulation is denoted as *simulation 1a*. The PACF for the wind speed time series employed led to the conclusion that $k_p=5$. Thus, $p_{max}=4$, according to equation (7).

A variation of this simulation considers that the model to be used is selected in the first hour and is employed in all hours. Therefore, only the first hour parameters of this model are estimated. This simulation is denoted as *simulation 1b*, and the particular parameters achieved are $p=1$ and $q=12$.

Simulation 2 – ANN. In this simulation, the ANN models are employed. Once the network structure is set, for each prediction the data is organized with all elements previous to the one to be forecasted and the network is created and trained. The prediction is performed presenting to the network as many immediately previous values to the one to be forecasted as the number of input nodes of the network. The number of input nodes is 5 ($k_p=5$) and the number of units in the hidden layer are 3 (*simulation 2a*), 5 (*simulation 2b*) and 10 (*simulation 2c*), according to Section 4.2.

The results of the simulations are presented in Table 1. It is worth to mention that the physical limitations of the wind turbine ($u_0=4\text{m/s}$, $u_{nom}=15\text{m/s}$ and $u_{max}=25\text{m/s}$) have been considered.

Table 1. Results for Case-Study 1.

Model	Simul.	MAE (m/s)	RMSE (m ² /s ²)	MRE	Proc. time (s)
Persistence		0,68	1,00	0,082	-
ARMA	1a	0,51	0,81	0,063	17367
ARMA	1b	0,50	0,78	0,060	590
ANN	2a	0,54	0,81	0,065	2772
ANN	2b	0,53	0,78	0,065	1993
ANN	2c	0,54	0,79	0,065	1680

The results evidence that, in general, the ARMA and ANN models perform better than the Persistence model. Between ARMA and ANN models, the former present the best results according to the MRE criterion. According to the MAE criteria, ARMA models also have a better performance than ANN models. It is important to take into consideration the processing time: the ARMA *simulation 1a* is the longest simulation, while the ARMA *simulation 1b* takes much less time, and is the shortest one. This *simulation 1b* also delivered the best overall results. As far as the ANN models are concerned, the variation that considers 5 hidden units (*simulation 2b*) achieved the best results, with a remarkable good result in the RMSE criterion, but the processing time was high.

Figure 5 presents a graphical sample of these results, for the ARMA *simulation 1b*.

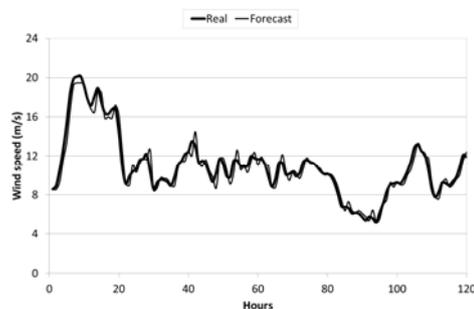


Figure 5. Sample of the results of Case-Study 1; ARMA simulation 1b.

5.2. Case-Study 2 (CS#2) – Wind speed forecasts to support the bidding decisions in MIBEL

The MIBEL (Electricity Iberian Market) is a liberalized market where producers and retailers may commercialize energy. The contracts may be celebrated in a bilateral forward trading, in the daily market or in the intraday market [16].

The bidding periods of the daily and intraday markets are shown in Figure 6. If one adopts the strategy of predicting each hour in the last opportunity to do it (in the intraday markets), six forecasts are required for the intraday market, from three to five hours, always with three hours in advance. Since ARMA and ANN models present a better performance in forecasting with few hours in

advance, the forecasts analyzed in this case-study follow this strategy.

The experience acquired with CS#1 showed that the best AIC models are always considered the best by the *Iratiotest* function. Thus, in this case-study only the AIC is considered to evaluate the performance of the model before the forecast is done. ARMA models are used in forecasting *simulation 3* and ANN are used in forecasting *simulation 4*.

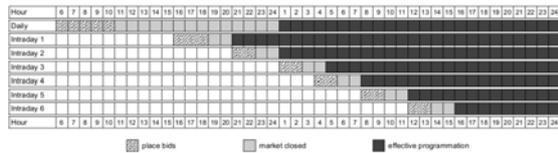


Figure 6. MIBEL bidding periods.

The objective of this CS#2 is to predict the hours of five consecutive days, in the last opportunity to do it, in order to bid in the intraday market, according to Figure 6. For the prediction of each block the historical data employed is the available data until the penultimate hour to place bids of the correspondent period. Figure 7 helps to understand the predictions realized in Case-Study 2.

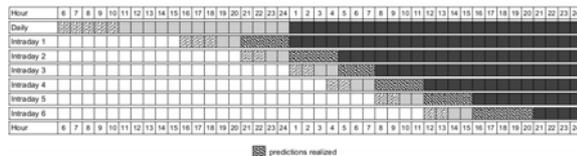


Figure 7. Predictions realized in CS#2.

Simulation 3 – ARMA. The ranges for the possible values of p and q are set at the beginning of the simulation. The parameters of all models (all possible combinations (p,q)) are estimated, the best model according to AIC is selected and its parameters are estimated at the beginning of each period.

This model is then utilized to perform the prediction of the next hour for all hours of the correspondent period. If the hour to forecast is one from the second to the last of the period, the forecasts already performed, from the first hour of that period, are included in the time series for input data of the model. Only the last hours of each period, according to Figure 7, are considered for the forecast. This simulation is denoted as *simulation 3a*.

A variation of *simulation 3a* consists in considering that, to perform the prediction of the next hour, the model structure is set at the beginning of each period, but the values of the parameters are set in all hours of each period. This simulation is tagged as *simulation 3b*.

Still another variation is to consider that both the model structure and the values of the parameters are set in all hours of each period, and then utilized to

perform the prediction of the next hour. This variation is denoted as *simulation 3c*.

Simulation 4 – ANN. Once the network structure is set, at the beginning of each period the available real data is organized, with all elements previous to that period, and the network is created and trained. The prediction of each hour of each period is performed presenting to the network as many immediately previous values to the one to be forecasted as the number of input nodes of the network.

Once again, as in CS#1, the number of input nodes is 5 and the possible number of units in the hidden layer are 3 (*simulation 4a*), 5 (*simulation 4b*) and 10 (*simulation 4c*) hidden units.

The same wind speed time series with one year of hourly means historical data utilized in CS#1 is employed, as well as the same physical considerations. The results of the simulations are presented in Table 2 and Figure 8 shows a graphical sample of these results, for ARMA *simulation 3c*, in the case.

Table 2. Results for Case-Study 2.

Model	Simul.	MAE (m/s)	RMSE (m ² /s ²)	MRE	Proc. time (s)
Persistence		2,12	2,67	0,248	-
ARMA	3a	1,86	2,27	0,212	4325
ARMA	3b	1,80	2,20	0,208	5957
ARMA	3c	1,80	2,21	0,207	30690
ANN	4a	1,90	2,34	0,220	101
ANN	4b	1,85	2,27	0,215	91
ANN	4c	1,95	2,38	0,223	89

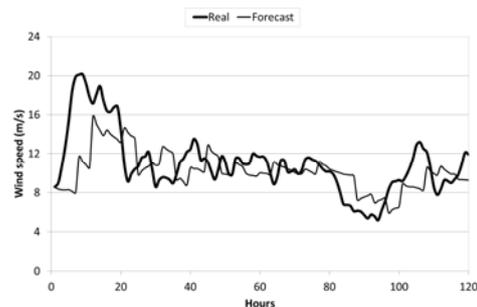


Figure 8. Sample of the results of Case-Study 2; ARMA simulation 3c.

As in CS#1, ARMA and ANN models reveal a better performance than Persistence. Once again, in general, ARMA models present better results in comparison to ANN, according to all criteria. The ARMA *simulation 3c*, the longest one, presents the best results, as far as the MRE criterion is concerned, but was overtaken by *simulation 3b* in the MAE and RMSE criterion.

In what concerns the variations of the ANN related simulations, the simulation that considers 5 hidden units (*simulation 4b*) obtained the best

performance. Despite the slightly worse results obtained with ANN models, their processing time is significantly minor than the ARMA's.

5.3. Case-Study 3 (CS#3) – One hour ahead wind power forecasts

Wind speed forecasts have been performed up to now. Electrical system operators are in charge of managing the power flow in the grid, thus, from their point of view, wind power forecasts are more important than wind speed forecasts.

This Case-Study concerns direct wind power forecasts. The methodology used was similar to the one related to the wind speed forecast in CS#1, but this time a wind power historical data of one year was used instead.

The computation of the PACF for this wind power time series returns $k_p=4$, thus $p_{max}=4$ in the ARMA model, according to equation (7). This simulation is denoted as *simulation 5a*.

As in the wind speed forecast case, a variation of this simulation was considered, in which the model to be used is selected in the first hour and is employed in all hours. Therefore, only the first hour parameters of this model were estimated. This simulation is denoted as *simulation 5b*, and the particular parameters achieved are $p=1$ and $q=24$.

As far as the ANN models are concerned, the number of input nodes is 4 and the possible number of units in the hidden layer are 2 (*simulation 6a*), 4 (*simulation 6b*) and 8 (*simulation 6c*). The results of the simulations are presented in Table 3.

Table 3. Results for Case-Study 3.

Model	Simul.	MAE (kW)	RMSE (kW ²)	MRE	Proc. time (s)
Persistence		131,18	167,68	0,351	-
ARMA	5a	108,45	138,54	0,267	18040
ARMA	5b	108,26	138,23	0,266	2349
ANN	6a	121,48	154,84	0,311	2609
ANN	6b	122,33	155,89	0,312	2071
ANN	6c	119,43	152,88	0,306	1664

Once again, the results evidence that in general the ARMA and ANN models show a better performance than the Persistence model. ARMA models deliver better results than ANN, according to all evaluation criteria.

The processing time of the ARMA *simulation 5a* is again the biggest, while the remaining simulations present similar processing times. The simulation that delivers the best results is again the ARMA *simulation 5b*, with model ARMA (1,24) employed in all hours, the best AIC model in this situation.

Figure 9 presents a selected graphical sample of these results, concerning the ARMA *simulation 5b*, this time.

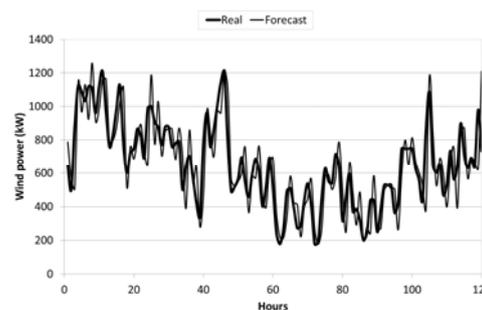


Figure 9. Sample of the results of Case-Study 3; ARMA simulation 5b.

5.4 Discussion

The first conclusion to outline from the results obtained is the worsening of the results from CS#1 to CS#2. In the former, the predictions were performed with one hour in advance, while in the latter they were performed with four to nine hours in advance, as Figure 7 evidences. This worsening is thus expected since short-term wind power forecasting models present a better performance for forecasting horizons of a few hours.

The performance of forecasting wind speed (CS#1) and wind power (CS#3) may be compared with the help of the MRE criterion, since it gives relative values of the error. In general, the forecasts for wind power time series are worse than for wind speed time series. This is justified by the type of the time series to predict: as one may see in Figures 5 and 9, the wind power time series utilized presents more variations, thus it is more difficult for the models to explain the historical data and therefore more difficult to estimate the forthcoming values.

Comparing ARMA models to ANN models, one may generally conclude that they both present similar results, with a slight better performance for the ARMA models and a better processing time for the ANN models. The processing time may in fact be an important decision factor. Depending on the application, one might need to opt for a faster forecasting model. To bid offers in the MIBEL, one has less than an hour to perform a forecast that includes the historical data up to the last hour and before the bidding period ends.

Another occurrence that is interesting to outline is the fact that in the first case study, with predictions with one hour in advance, the best results were obtained with a simulation for which the best model is selected in the first hour and trained in all the remaining. This allows one to save much time, since the choice of the best model implies the training of all possible models, and the training is the most time consuming task of the simulation. However, in the MIBEL case study, with predictions of several hours in advance, the best results were obtained with a simulation for which the selection of the best model is performed in all hours. This difference allows one

to draw the conclusion that the increase in forecasting difficulty also increases the importance of the ability of the simulation routine to adapt to different forecasting requirements.

About the best number of units in the hidden layer of the ANN models there is one fact that should be evidenced. There were three possible numbers of hidden units, according to the criteria set in Section 4.2. Both in CS#1 and CS#2, the number of hidden units that delivered the best results was the number of units in the input layer. However, to conclude that this choice delivers better results, more simulations should be ran.

6. Conclusions

This paper intends to present the basics on forecasting wind speed with ARMA and ANN, define similar criteria to adjust the required settings in both models and compare their performance under similar forecasting conditions. However, due to the inherent differences between the models, conclusions from the performance comparison should be drawn with care.

A general conclusion that may be drawn from the obtained results is that both ARMA and ANN do perform better than the reference persistence model. In what relates to the comparison between ARMA and ANN, one may conclude that, in general, ARMA models achieve slightly better forecasts, but they are more time consuming than the ANN models.

The predicted wind speed is much more accurate when the forecasts are performed with an hour in advance than when they are done with four to nine hours in advance, as it is the case of the MIBEL case. This worsening is thus expected, since short-term wind power forecasting models present a better performance for forecasting horizons of a few hours. The forecasts performed for the MIBEL case are not acceptable for any model employed, since all of them present high errors. This is a known limitation of the statistical forecasting models.

This work focuses on the forecast of wind speed with ARMA and ANN models, employing only wind speed hourly means time series. However, ARMA models allow one to utilize them with an exogenous time series and ANN models may have as input several kinds of data. This opens the possibility to train the models not only with wind speed data, but also with other variables of interest, such as wind power, temperature, atmospheric pressure and humidity. The combination of these data, as long as it is measured simultaneously in the same site, might conduct to better results, since the explanation of the input wind speed becomes more complete, thus the forecast is expected to be more accurate.

7. References

- [1] World Wind Energy Association (2011), *World Wind Energy Report 2010*, Germany.
- [2] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, (2008), "A Review on the Young History of the Wind Power Short-Term Prediction," *Renewable and Sustainable Energy Reviews*, vol.12, no.2.
- [3] G. Giebel, R. Brownsword, and G. Kariniotakis, (2003), "The State-of-the-Art in Short-Term Prediction of Wind Power – A Literature Overview," Project ANEMOS, Tech. Rep.
- [4] L. Landberg, G. Giebel, H. Nielsen, T. Nielsen, and H. Madsen, (2003), "Short-Term Prediction – An Overview," *Wind Energy*, vol.6, no.3.
- [5] G. Giebel, G. Kariniotakis, and R. Brownsword, (2003), "State-of-the-Art on Methods and Software Tools for Short-Term Prediction of Wind Energy Production," in European Wind Energy Conference & Exhibition – EWEC2003, Madrid.
- [6] J. Hamilton, (1994), *Time Series Analysis*, Princeton University Press.
- [7] G. Zhang, B. Patuwo, and M. Hu, (1998), "Forecasting with Artificial Neural Networks: The State of the Art," *International Journal of Forecasting*, vol.14, no.1.
- [8] M. Mohandes, S. Rehman, and T. Halawani, (1998), "A Neural Networks Approach for Wind Speed Prediction," *Renewable Energy*, vol.13, no.3.
- [9] P. Fonte, J.C. Quadrado, and R. Castro, (2006), "Short-Term Wind Speed Prediction and Wind Power Generation Output," in International Conference on Knowledge Engineering and Decision Support – ICKEDS'06, Lisbon.
- [10] M. Gardner and S. Dorling, (1998), "Artificial Neural Networks (The Multilayer Perceptron) – A Review of Applications in the Atmospheric Sciences," *Atmospheric Environment*, vol.32, no.14-15.
- [11] Garch Toolbox – For Use With Matlab, (1999), 1st ed., The MathWorks Inc.
- [12] H. Demuth, M. Beale, and M. Hagan, (2009), *Neural Network Toolbox*, 6th ed., The MathWorks Inc.
- [13] Financial Time Series Toolbox – For Use With Matlab, (2000), 1st ed., The MathWorks Inc.
- [14] D. Faria, R. Castro, C. Philippart, and A. Gusmão, (2009), "Wavelets Pre-Filtering in Wind Speed Prediction," in IEEE Second International Conference on Power Engineering, Energy and Electrical Drives – POWERENG2009, Costa de Caparica.
- [15] M. Milligan, M. Schwartz, and Y. Wan, (2003), "Statistical Wind Power Forecasting Models: Results for US Wind Farms," in Wind Power Conference – WINDPOWER, Austin.
- [16] T. Simão, (2009), *Pricing Renewable Energy in a Competitive Electricity Market*, MSc thesis, Instituto Superior Técnico.
- [17] P. Gomes, (2010), *Comparison of Wind Power Forecasting Models*, MSc thesis, Instituto Superior Técnico.