

Experiments and Reference Models in Training Neural Networks for Short-term Wind Power Forecasting in Electricity Markets

Juan Méndez, Javier Lorenzo, and Mario Hernández

Institute of Intelligent Systems(SIANI)
Univ. Las Palmas de Gran Canaria
35017 Las Palmas, Canary Islands, Spain
Email: jmendez@iusiani.ulpgc.es

Abstract. Many published studies in wind power forecasting based on Neural Networks have provided performance factors based on error criteria. Based on the standard protocol for forecasting, the published results must provide improvement criteria over the persistence or references models of its same place. Persistence forecasting is the easier way of prediction in time series, but first order Wiener predictive filter is an enhancement of pure persistence model that have been adopted as the reference model for wind power forecasting. Pure enhanced persistence is simple but hard to beat in short-term prediction. This paper shows some experiments that have been performed by applying the standard protocols with Feed Forward and Recurrent Neural Networks architectures in the background of the requirements for Open Electricity Markets.

1 Introduction

The economical and environmental advantages of the use of renewable energy sources (RES) in the electricity generation connected to public grids has a counterpart in the instability and unpredictability introduced into the electric systems. The electric generation from the more important renewable sources, wind and solar power, is mainly concerning with the weather in a local geographic area. However, the weather is a chaotic system whose predictability is limited. In each national system, the Transmission System Operator(TSO) deals in the management of the electric system in the different control and planning levels.

The statistical approach for short-term wind prediction has been used due to the system complexity of whether and the chaotic fluctuations of wind speed. The statistical models such as ARMA, ARX and Box- Jenkins methods have been used historically for short-term wind forecasting up to few hours ahead [1–3]. Giebel [4] reports some of the statistical state of the art models and methods for wind power forecasting which have been developed and used, such as time series models for up to a few hours by means of statistical approaches and neural networks, as well as models based on Numerical Weather Prediction(NWP).

The Electric Authorities of countries had included the power forecasting in its Regulatory Norms which goal is to preserve the quality of the electricity

supply. The planning of an Electric System requires several levels related with different time scales as well as the whether forecasting. Very near short-term forecasting or nowcasting is the immediate prediction in time scale ranging from some minutes to several hours. Short-term forecasting concerns with time scale that ranges from one to three days, while middle-term forecasting covers from four days to several weeks time range.

The most interesting time scale in power predictions for using ANN is the nowcasting, which can be carried out by using the time series analysis. The short-term scale requires the cooperation between statistical and NWP tools, in regional and mesoscale weather models and cooperating with predictive systems as HIRLAN and MM5. The power forecasting for RES in Spanish Regulations is related to hourly periods of planning of the electricity market. All the power supplies and demands of the energy agents must be related to these hourly periods. The regulations for the short-term Spanish Electricity Market comprise two steps:

Short-term Forecasting. The RES producers, solar and wind farms, with power greater than 10 MW must provide 30 hours ahead the power forecasting for every hourly period of a full range of 24 hours.

Nowcasting. One hour ahead of each hourly period, corrections to the previous values can be sent to the Electricity Authority.

This means that in the nowcasting time scale, the computation of the predicted value must be carried out for the period covering two hours ahead. The second step can be carried out by using ANN, but the first requires the cooperation with NWP tools.

Artificial Neural Networks (ANN)[5] have been widely used for modeling and predictions in the field of renewable energy systems [6, 7] because they are able to handle noisy, incomplete data and non-linear problems to perform predictions and classifications [8–10]. Hippert et al [11] have addressed the construction, and evaluation, of their performance of very large ANNs in electric systems to forecast the load profile. Recurrent ANN [12] have been used as generalizations of predictive systems as ARMA. Also, they can be used to generalize linear predictive systems as Kalman filter [13]. Recurrent and recurrence in each layer, called multilayer recurrent, architectures have been also used in wind power prediction[14].

Many studies about the use of ANN in wind power have been preformed, but the criteria to evaluate their performance have been mainly based on error parameters. Based on more modern standard protocol for forecasting[15], the published results will provide improvement criteria over the persistence or references models of its same place. Persistence forecasting is a simple model that is intrinsic to the data, that is, it is a no algorithm approach. Any new proposed algorithm is so good or bad as how much is able to overtake the persistence. The use of ANN can provide a suitable procedure to beat it and other reference model based on the Wiener predictive filter. An application is presented applying the standard protocols with Feed Forward(FNN) and Recurrent Neural

Networks(RNN) architectures in the background of the requirements for Open Electricity Markets. The paper contains a Methods section where the persistence and reference models are defined as well as the used ANN architecture and training procedures. The Results section shows the collected results of some experiments carried out by using wind power data series, and Conclusion section summarizes the main contributions and remarkable results.

2 Methods

Persistence is the simplest model for forecasting. It is based on the assumption of a high inertia in the subjacent physical model. If $y(t)$ is the value at time t of a time series, in persistence model the predicted value for k times ahead is: $\hat{y}(t+k) = y(t)$. This kind of forecasting is really simple but can be very useful in practical, because it can be used as a reference model to compare different theoretical and practical applications. Any proposal of a new model or approach that requires some computational resource is required to have at least a better performance than this simple one. The level of improvement over this reference model must be a level of utility of the additional formal and computational cost. A high value in an error parameter, as MAE or RMSE, in a hardly predictable site can be a better result than a small value in a easily predictable site. However there are not a parameter to define what site has a hardly or easily predictable wind. A option is the use the own persistence as the reference to which compare the performance of proposed algorithms.

The pure persistence model can be overtaken by other model that involve persistence-like information. A reference model to compare different forecasting models has been proposed [16, 15]. It is more advanced because it includes very short-term information, as persistence, and long-term information. This proposed reference model is an extension of the pure persistence as a linear expression: $\hat{y}(t+k) = b + ay(t)$. A detailed analysis allows to show that is really the first order case of a more general linear predictive filter, as the Wiener filter with general expression:

$$\hat{y}(t+k) = B + \sum_{i=0}^m A_i y(t-i) \quad (1)$$

where coefficients A_i and B can be computed from the matrix containing the cross correlation between $y(t+k)$ and $y(t-i)$. The constant parameter is $B = (1 - \sum_{i=0}^m A_i) \bar{y}$, where \bar{y} is the large-term average value of $y(t)$. For the simplest case of first order filter: $\hat{y}(t+k) = B + A_0 y(t)$, the value of the coefficient is:

$$A_0 = \frac{\int [y(t+k) - \bar{y}][y(t) - \bar{y}] dt}{\int [y(t) - \bar{y}]^2 dt} \quad (2)$$

In an Electricity Market we have two kind of power values, the spot power $P(t)$ and its hourly average P_h . For the TSO, the spot power is very important

to assure the system stability at any time, but in the Electricity Market the hourly average is the required to RSE agents. The proposed reference model for wind power forecasting by Madsen [15], is applied for hourly average power in nowcasting as the required in the Spanish regulation as:

$$\hat{P}_{h+2} = A_0 P_h + (1 - A_0) \bar{P} \quad (3)$$

where A_0 and \bar{P} are parameters computed from large-term training information. This reference model, which we can call as improved persistence or Wiener persistence, is harder to beat because is based in the shortest-term information, P_h , and in the longest-term information, \bar{P} .

The basic theory for using ANN in prediction, its architectures and algorithms are in the area of adaptive and predictive linear filter[12]. The use of ANN has generated generalizations that has introduced improvements in the original linear models by allowing the construction of nonlinear predictive systems. The relationship between ANN, in special recurrent architectures, with linear predictive systems as ARMA allows nonlinear generalizations of previous statistical linear approaches.

A generalization of recurrent ANN is the multilayer recurrent[12,14]. In the wind power forecasting the problem can be formulated by using Feed Forward(FNN), without feedback, or Recurrent(RNN) ones:

$$\hat{P}_{h+2} = F [V_h, \dots, V_{h-n+1}, P_h, \dots, P_{h-m+1}] \quad (4)$$

The used training procedure was the Bayesian regularization [17,18] which updates the weight and bias values according to the Levenberg-Marquardt [19,20] optimization procedure. It uses as goal function a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well. The Bayesian regularization implementation that has been used is the implemented in the training function *trainbr* of the Neural Networks Toolbox of MATLAB[21]. The NARX architecture have been used for RNN with the same window size for input data, the wind speed, and feedback data, the wind power.

3 Results

We have used a wind data series acquired in Gran Canaria Island(Spain). The wind speed series comprise about 33 days data from a meteorological tower in time steps of one minute. Wind power series are obtained from the wind speed at 40 meters high and from a power transfer function with 5 and 12.5 m/sec cut-off values. Relative values about the nominal values, $P(t)/P_n$, are used in the power series.

The data set was split in two subset, the train and test. The train data is 2/3 of the global data. The standard protocol for performance evaluation suggested by Madsen[15] was used. It includes the definition of the Evaluation

Table 1. Comparative results for two hours ahead prediction by using several RNN configurations trained with Bayesian regularization. All Evaluation Criterion and their improvements over the reference model are in percent(%) normalize to the nominal power. The mean and standard deviation, $\mu \pm \sigma$, values are provided for 25 training trials

	Pers. Ref.	RNN1	RNN2	RNN3	RNN4	RNN5
Delay		(2:3)2	(2:5)4	(2:7)6	(2:7)6	(2:7)6
Hidden Nodes		80	40	10	40	60
BIAS	0.6 0.9	0.5 ± 0.1	0.3 ± 0.1	0.1 ± 0.3	0.3 ± 0.4	0.3 ± 0.1
MAE	14.5 15.3	15.5 ± 0.2	15.3 ± 0.1	15.7 ± 0.5	15.3 ± 0.2	15.3 ± 0.1
RMSE	23.7 22.3	22.3 ± 0.3	21.6 ± 0.1	22.5 ± 1.2	21.5 ± 0.1	21.6 ± 0.1
SDE	23.7 22.3	22.4 ± 0.3	21.6 ± 0.1	22.5 ± 1.2	21.6 ± 0.1	21.6 ± 0.1
Imp_MAE		-0.4 ± 1.2	1.1 ± 0.5	-2.5 ± 3.3	0.6 ± 1.0	0.4 ± 0.9
Imp_RMSE		0.0 ± 1.2	3.3 ± 0.3	-1.0 ± 5.3	3.3 ± 0.6	3.2 ± 0.6
Imp_SDE		-0.1 ± 1.2	3.2 ± 0.3	-1.1 ± 5.3	3.2 ± 0.6	3.1 ± 0.6

Table 2. Comparative results by using several FNN networks configurations. Additional data are the same as in Table 1

	FNN1	FNN2	FNN3	FNN4	FNN5	FNN6
Delay	(2:4)3	(2:4)3	(2:6)5	(2:6)5	(2:11)10	(2:11)10
Hidden Nodes	3	6	5	10	10	20
BIAS	3.0 ± 1.8	4.0 ± 2.5	1.4 ± 0.3	1.4 ± 0.9	2.4 ± 2.4	3.2 ± 3.1
MAE	16.2 ± 1.0	16.8 ± 1.2	15.7 ± 0.4	16.0 ± 0.7	16.7 ± 1.3	17.4 ± 1.7
RMSE	22.7 ± 0.4	22.9 ± 0.6	22.2 ± 0.4	22.4 ± 0.5	22.6 ± 0.8	23.4 ± 1.3
SDE	22.5 ± 0.2	22.5 ± 0.3	22.2 ± 0.3	22.4 ± 0.5	22.5 ± 0.6	22.0 ± 1.1
Imp_MAE	-4.8 ± 6.5	-8.6 ± 8.0	-1.3 ± 2.7	-3.1 ± 4.8	-7.4 ± 8.0	-12.3 ± 10.7
Imp_RMSE	-2.1 ± 2.0	-2.9 ± 3.0	2.7 ± 1.6	-0.6 ± 2.3	-1.5 ± 3.4	-4.7 ± 5.8
Imp_SDE	-1.0 ± 0.9	-0.8 ± 1.2	0.4 ± 1.5	-0.4 ± 2.1	-0.5 ± 2.5	-2.9 ± 5.0

Criteria(EC) BIAS, MAE, RMSE and SDE, and also the improvement over the reference model which are computed in percent value as:

$$Imp_{ref,EC}(\%) = 100 \frac{EC_{ref} - EC}{EC_{ref}} \quad (5)$$

Many training procedures of ANN use optimization procedures that run from initial random states. The optimization tries to reach a minimum value of some goal function, but the reached value and the trained network depend on the initial random state. In the practice, that means that the performance of a trained ANN has some random degree. To reduce the uncertainty in the results, we provide the mean and the standard deviation obtained from 25 training trials as: $\mu \pm \sigma$. Following the suggestion of Zang[22] that users should pay more attention to selecting the number of input nodes, we have cross correlated the power with itself and correlated it with the wind speed and concluded that the highest values are for offsets until the range of 4-6 hours back. It means that the size of the more useful data window must be around this range.

Tables 1 and 2 contain the results for several configurations of RNN and FNN respectively. Table 1 contains also the error values for the persistence and reference model. The computation of the reference model data was performed by using the train set, its parameters are: $A_0 = 0.82$ and $\bar{P} = 0.68$. The reported results are related to architectures including one hidden layer. The experiments have shown that more layers increases the computational cost and have no better performance. In both tables, the delays are taken in relation to the prediction time; they are represented as: $(h_1 : h_2)w$, where $w = h_2 - h_1 + 1$ is size of the time window. In all cases $h_1 = 2$ to met the regulations. Remark that the values of BIAS and MAE are related to the first moment of the error, therefore they are related to the generated power, but the values of RMSE and SDE are related to the second order moment and the variance of the error.

All the tested RNN architectures perform better on BIAS values, such as significantly reduce the level in relation to the reference model and the persistence. It means that the feedback of RNN architectures systematically corrects the biased offset in the prediction. The FNN architectures without such feedback are systematically biased. The inclusion of innovation filters can be needed for the FNN case but is no necessary for the RNN one. However, in MAE criterium the persistence value is not beaten neither reference nor any tested ANN architecture. The variance of the error provided by RMSE and SDE criteria are outperformed by some RNN architectures in relation to persistence, reference model and FNN. The range of parameters that provide better results are around values 4 and 6 for windows size, and around 40 for hidden nodes. The use of narrow windows or lower number of hidden nodes performs worse. There are not tradeoff between reducing the window size and increasing the hidden nodes as shows on the RNN1 case. The increasing of hidden nodes does not performs much better as is shown in RNN6 case. The FNN architectures are more unstable, eg. the FNN3 have a good improvement of 2.7 in mean value in the RMSE criterium, but has a big standard deviation value of 1.6. It is unstable if compared with the RNN2 case with 3.3 value in mean and 0.3 value in standard deviation.

Figure 1 shows the comparative performance in several hours ahead for the RMSE criterium. The included models are the persistence, the reference model the RNN2 and the FNN3 cases. It is shown that the reference model performs much better that the persistence and both ANN cases outperform the reference model. Also it is shown that the relative efficiency of the predictive models of ANN in relation to persistence increases when increases the ahead hours.

4 Conclusions

The short-term forecasting of wind power for Electricity Markets requires two kind of time scales prediction. The first requires detailed prediction for 1-2 days ahead, which needs the cooperation of some tools of NWP. The second is for the time scale of few hours ahead, which can be carried out by using time series

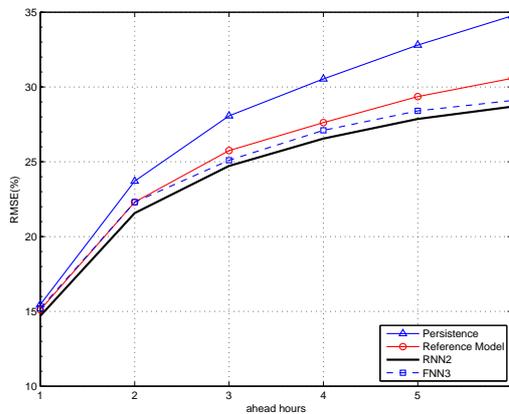


Fig. 1. Comparative RMSE of several models in the very short-term prediction

analysis. In this time scale, ANN can be applied successfully for wind power forecasting useful in Open Electricity Markets.

This study has used the standard protocols to evaluate the performance of forecasting procedures that some authors have introduced. We have compared the results according these protocol. We have shown that the new reference model, based on the first order Wiener filter, perform better in variance criteria as RMSE and SDE, but it is worse in first order moment as BIAS and MAE. Some ANN architectures, as Recurrent and Feed Forward, have been tested. The main conclusion is that Recurrent architectures have better performance in first and second order statistical moments and can beat the reference model in the range of nowcasting useful in the Electricity Market.

Acknowledgment

This research was partially supported by the Spanish Ministry of Science and Innovation, grant TIN2008-06068. We acknowledge to the Instituto Tecnológico de Canarias (ITC) for the use of the wind speed series used in this paper that were acquired in its facilities in Pozo Izquierdo, Gran Canaria Island.

References

1. Nielsen, H.A., Madsen, H.: Wind power prediction using arx models and neural networks. In: Proceedings of the Fifteenth IASTED International Conference on Modelling, Identification and Control. (1996) 310313
2. Landberg, L., Giebel, G., Nielsen, H.A., Nielsen, T., Madsen, H.: Short-term prediction an overview. *Wind Energy* **6** (2003) 273280

3. Nielsen, T.S., Madsen, H., Nielsen, H.A., Pinson, P., Kariniotakis, G., Siebert, N., Marti, I., Lange, M., Focken, U., von Bremen, L., Louka, P., Kallos, G., Galanis, G.: Short-term wind power forecasting using advanced statistical methods. In: European Wind Energy Conference & Exhibition 2006. (2006)
4. Giebel, G.: EC Project ANEMOS: The state-of-the-art in short-term prediction of wind power. a literature overview. Technical report, Risø National Laboratory (2003)
5. Haykin, S.: Neural Networks. Prentice Hall (1999)
6. Kalogirou, S.A.: Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Reviews* **5** (2001) 373401
7. Li, S., Wunsch, D.C., OHair, E., Giesselmann, M.G.: Neural network for wind power generation with compressing function. In: IEEE International Conference on Neural Network. Volume 1. (1997) 115–120
8. Alexiadis, M., Dokopoulos, P., H.S.Sahsamanoglou, Manousaridis, I.: Short-term forecasting of wind speed and related electric power. *Solar Energy* **63**(1) (1998) 61–68
9. Zhang, G., Patuwo, B.E., Hu, M.Y.: Forecasting with artificial neural networks: The state of the art. *Int. Jorunal of Forecasting* **14** (1998) 35–62
10. Kandil, N., Wamkeue, R., Saad, M., George, S.: An efficient approach for short term load forecasting using artificial neural networks. *Electrical Power and Energy Systems* **28** (2006) 525–530
11. Hippert, H.S., Bunn, D.W., Souza, R.C.: Large neural networks for electricity load forecasting: Are they overfitted? *Int. Journal of Forecasting* **21** (2005) 425–434
12. Mandic, D.P., Chambers, J.A.: *Recurrent Neural Networks for Prediction*. John Wiley and Sons Ltd (2001)
13. Haykin, S.: *Kalman Filtering and Neural Networks*. John Wiley and Sons, Inc (2001)
14. Li, S.: Wind power prediction using recurrent multilayer perceptron neural network. *IEEE Power Engineering Society General Meeting* **4** (2003) 225–230
15. Madsen, H.: EC Project ANEMOS: A protocol for standardizing the performance evaluation of short-term wind power prediction models. Technical report, Technical University of Denmark (2004)
16. Nielsen, T.S., Joensen, A., Madsen, H., Landberg, L., Giebel, G.: A new reference model for wind power forecasting. *Wind Energy* **1** (1998) 29–34
17. MacKay, D.J.C.: A practical framework for backpropagation networks. *Neural Computation* **4**(3) (1992) 448–472
18. Foresee, F.D., Hagan, M.: Gauss-newton approximation to Bayesian regularization. In: *Proceedings of the 1997 International Joint Conference on Neural Networks*. (1997) 19301935
19. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. *The Quartely of Applied Mathematics* **2** (1944) 164–168
20. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11** (1963) 431–441
21. Demuth, H., Beale, M., Hagan, M.: *Neural Network Toolbox 6, User's Guide*. The MathWorks, Inc. (2008)
22. Zhang, G.P., Patuwo, B.E., Hu, M.Y.: A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research* **28** (2001) 381–396