Forecasting of wind parks production by dynamic fuzzy models with optimal generalisation capacity

Georges Kariniotakis

Forecasting of Wind Parks Production by Dynamic Fuzzy Models with Optimal Generalisation Capacity.

George Kariniotakis, *Member IEEE*
Ecole des Mines de Paris,
Centre d'Energétique
B.P. No 207, 06904 Sophia-Antipolis, France.
Tel: +33-493957501, Fax: +33-493957535, kariniotakis@cenerg.cma.fr

ABSTRACT: On-line forecasting of the power output of wind farms is of major importance for a reliable and secure large-scale integration of wind power, especially under liberalized energy market environment. This paper presents such a prediction tool that receives on-line SCADA measurements, as well as numerical weather predictions as input, to predict the power production of wind parks 48 hours ahead. The prediction tool integrates models based on adaptive fuzzy-neural networks configured either for short-term or long-term forecasting. In each case, the model architecture is selected through non-linear optimization techniques. By this way the accuracy of the model on out of sample data (generalization) is optimized. The forecasting models are integrated in the MORE-CARE Energy Management Software (EMS) software developed in the frame of a European research project. In this EMS platform, wind forecasts and confidence intervals are used by economic dispatch and unit commitment functions. The paper presents detailed results on the performance of the developed models on a real wind farm using HIRLAM numerical weather predictions as input.

Keywords: Wind power, time-series forecasting, numerical weather predictions, on-line software, adaptive fuzzy-neural networks.

I. INTRODUCTION

Nowadays, wind park installations in Europe exceed 23 GW, while the motivated by the Kyoto Protocol targets of the E.U. for 12% energy demand covered by renewables by year 2010, are translated to 21% electricity generation by renewables. To achieve these targets, wind power in the Member States should arise up to 45-60 GW. Such a large-scale integration of wind power emerges the development of appropriate tools to assist the wind farm operators on their management task.

Of major importance are tools that forecast wind parks production for the next 24-48 hours. In a liberalised market environment, prediction tools enhance the competitiveness of wind power, since they reduce the penalties resulting from the wind resource intermittence.

Research on wind speed forecasting and, correspondingly the forecast of power output from a wind park, is actively pursued by several research centres in Europe.

Actually there are two main state-of-the art approaches; one based on physical modelling and a second one based on timeseries modelling.

The "physical" approach for wind power forecasting is based on a detailed description of the site (orography, roughness, obstacles), a description of the wind turbines (hub height, power curve, thrust curve) and a description of the wind plant. The main input is numerical weather predictions (NWP). Model output statistics are developed to account for systematic errors [1]. Weather predictions are however updated only a limited number of times per day by meteorological services. For this reason, the performance of these models is often satisfactory for rather longer (>6 hours ahead) than short-term horizons.

The alternative "timeseries", or statistical, approach includes typical linear models (ARMA, ARX etc) [2] and non-linear ones (i.e. neural networks, conditional parametric models, etc) [3,4]. These models aim to predict the future by capturing temporal and spatial dependencies in the data. The input to these models can be on-line SCADA data and numerical weather predictions (NWP). For look-ahead times more than ~10 hours, NWPs are indispensable for an acceptable performance, since they represent weather dynamics that cannot be modelled using only recent on-line data. For shorter horizons, up to ~10 hours ahead, timeseries models can be based exclusively on recent measurements; however even in this case, NWPs as explanatory input improves results. It is noted that the threshold of 10 hours is mentioned as an example rather than a rule, since it depends on the characteristics of a specific wind profile.

The models presented in this paper belong to the timeseries approach. In previous work, several types of models have been benchmarked on the wind power prediction problem [5,6,7]. Linear autoregressive models, radial basis functions, wavelet networks, feed-forward and recurrent neural networks, and finally adaptive fuzzy-neural network (F-NN) models were compared for the task of short-term prediction. Fuzzy neural networks, originally used here for wind forecasting, were found to outperform the other approaches in both short-term and long-term wind prediction.

This paper presents an advanced wind power forecasting tool developed at Ecole des Mines de Paris. This tool has been integrated in the MORE-CARE Energy Management

System (EMS) developed in the frame of a EU project. It is installed at the islands of Crete and Madeira, where it optimizes the operation of these power systems, and also in Ireland, where it operates as a stand-alone wind-forecasting platform for 11 wind farms. The system provides optimal forecasts for a horizon up to 48-72 hours ahead.

II. DESCRIPTION OF THE PREDICTION MODEL.

Adaptive fuzzy-neural networks are applied here for both short-term (<10 hours) and long-term (1-48 hours) prediction.

The adaptivity property stands for the capacity of the model to fine-tune its parameters during on-line operation. This is an important requirement for a non-stationary process like wind speed or power. The model adaptivity compensates changes in the environment of the application that may happen during the lifetime of a wind farm. Such changes can be changes in the number of wind turbines (extension of the wind farm, maintenance or availability of the machines that is usually not available through SCADA), in the performance of the wind turbines due to aging, changes in the surrounding of the wind park (i.e. vegetation), or changes in the configuration of the model used to produce the NWPs.

The core F-NN model is generic and can be trained on appropriate input depending on the final use, which can be either short-term or long-term prediction.

A. Short-term models.

Short-term models receive historic values of wind power as input, as well as explanatory data, such as wind speed and direction, to predict wind power. The general form of a simple model with input only past values of power is:

$$\hat{P}(t+1) = f\left(P(t), P(t-1), \ldots, P(t-m)\right)$$

The generic fuzzy-neural function *f*(.) is described in Section III. Multi-step ahead forecasts are generated using the model in an iterative way. I.e., in order to produce a forecast for t+2, the forecast for t+1 is fed back as input to the model. This approach presents the drawback that does not permit to iterate explanatory input, since no forecasts can be available for such quantities. To handle this problem, models using the look-ahead time k as input variable can be considered.

An alternative approach is to develop multi-output models, or to tune a different model for each time-step. The implementation of this approach is complex and requires high development effort, which can be prohibitive in case of system with a large number of wind farms.

The short-term models are found to outperform persistence up to 20% according to the time-step [4,5,6]. Such predictions are adequate for small applications, for which NWPs are not available, e.g. in the case of islands [9]. In larger systems, timeseries models based on meteorological information, as the one presented below, outperform short-term models.

B. “Long-Term” models (1-48 hours ahead).

For “long-term” horizons up to 24-48 hours ahead, it is necessary to include numerical weather forecasts as explanatory input to the model in order to have an acceptable performance. NWPs include usually wind speed, direction and temperature at 10 m, as well as at several levels related to levels of atmospheric pressure. They can be provided for the geographical coordinates of the wind farm or for a grid of four points surrounding the farm. In the second case, the spatial resolution of the NWP model is of primary importance. Meteorological models with high resolution are often more accurate but require high computation time to produce forecasts, and as a consequence, they do not update frequently their output (i.e. 1-4 times per day). In contrast, forecasts from low-resolution NWP models are more frequently available.

The developed forecasting tool is able to operate with input from different NWP systems. In the frame of this study it was tested and gave satisfactory results with input from the SKIRON system for the case-study of Crete, and also from the HIRLAM NWP system for the case of Ireland. SKIRON forecasts were provided for a grid of 15x15 km (System B in Figure 1), while HIRLAM predictions were provided at the level of the wind farm (System A in Figure 1).

Forecasts are generated every hour for the next 2 days. At the moment of update, the most recent available NWPs are used as input to the model together with measurements of wind power. Eventually measurements of wind speed or direction can be used as input. Model configurations that do not include such online information as input were found to perform worse than persistence in look-ahead times up to 6 hours ahead. Wind power data are necessary for the on-line updating procedure, independently if they are used or not as input variables to the model. The general scheme of the model is shown in Figure 1.

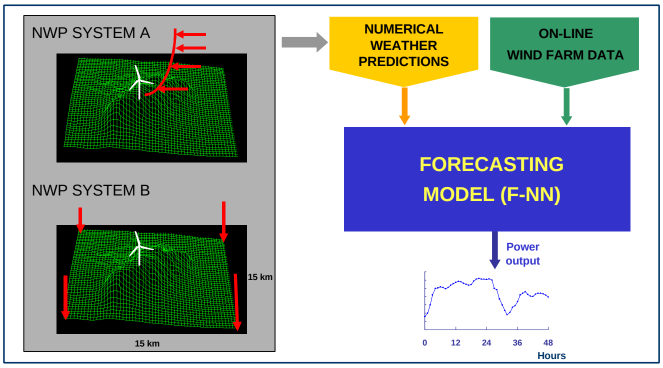

Figure 1 : *General scheme of the “long-term” prediction model with examples of two configurations of NWPs used as input (SKIRON,HIRLAM).*

The aim of the prediction model is to capture the relations between input (meteorological information, on-line data) and output (future total wind park power). Such mapping includes the following implicit relations:

- Temporal correlations between past and future data of the process (autoregressive aspect of the model).
- Conversion of wind speed (meteorological predictions) from the height or the atmospheric level of the NWP model to the hub height.

- Spatial projection of the meteorological wind speed forecasts from the NWP grid points (eg. 15x15 km) to the level of the wind farm.
- Correction of the wind park output for factors affecting the total production (i.e. array effects, effect of wind direction etc).

The advantage of a model like fuzzy neural networks compared to a physical one is that it permits to avoid all the above intermediate modeling steps. Moreover, its adaptive mode can compensate situations like the ones explained in the previous Section.

The above mapping relations introduce inaccuracy in the modeling procedure. Among the difficulties, one should add the error of weather forecasts, without neglecting the intermittent nature of wind itself. Wind speed is a non-stationary process both in the mean and in the variance. Wind power is nonlinear w.r.t. speed with a major difficulty in the area of cut-off speed, where prediction intervals can extend from maximum to zero wind power.

III. MODEL DEVELOPMENT AND GENERALIZATION.

A. General description of the fuzzy-neural network model.

The fuzzy model can be expressed in the form of rules of the type:

"IF $\underline{x}$ is A THEN y is B"

where $\underline{x}$, y are linguistic variables and A, B are fuzzy sets. In the case of time-series prediction rules may have the form:

$$R: \quad \text{IF} \quad x_1 \text{ is } A_1, \text{ and } \ldots, \text{ and } x_n \text{ is } A_n \quad \text{THEN} \quad y = g(x_1, \ldots, x_n)$$

where:

$x_1,\ldots,x_n$ are real-valued variables representing input variables of the system defined in the universes of discourse $X_1,\ldots,X_n$ respectively.

$A_1,\ldots,A_n$ are fuzzy sets.

y is variable of the consequence whose value is inferred. In the specific problem it represents future wind power $\left(\hat{P}(t+1), \hat{P}(t+2), \ldots\right)$.

$g(.)$ is a function that implies the value of y when $x_1,\ldots,x_n$ satisfy the premise. The function $g(.)$ in the consequent part of the rules may be a linear or a non-linear one or even a constant. In the case of a linear function the fuzzy rule-base takes the form:

$$\begin{array}{lllll} R^1: & \text{IF} & x_1 \text{ is } A_1^1, \ldots, \text{ and } x_n \text{ is } A_n^1 & \text{THEN} & y^1 = p_0^1 + p_1^1 x_1 + \ldots + p_n^1 x_n \\ \vdots & & & \vdots & \\ R^m: & \text{IF} & x_1 \text{ is } A_1^m, \ldots, \text{ and } x_n \text{ is } A_n^m & \text{THEN} & y^m = p_0^m + p_1^m x_1 + \ldots + p_n^m x_n \end{array}$$

Each rule gives an estimation of the output y_i according to the conditions defined by the fuzzy sets in the premises. In the context of timeseries prediction, each variable x_i in the premise corresponds to a past value of the process (i.e. power: $P(t)$, $P(t\text{-}1)\ldots$), or past values of explanatory input (i.e. wind speed: $WS(t)$, $WS(t\text{-}1)\ldots$) or meteorological forecasts ($WS^m(t+1)$, $WS^m(t+2)$, …).

A linear function in the consequence is indeed an ARX (autoregressive with exogenous variables) model. It is clear that with the above definitions, the rule-base consists of an ensemble of "local" models. Local modeling is a desired property of the model, especially in the case of a non-stationary process such as wind generation.

Fuzzy sets in the premises are modeled here using Gaussian functions:

$$m_{A_j^i}(x_j) = \exp\left(-\left(\frac{x_j - a_j^i}{b_j^i}\right)^2\right)$$

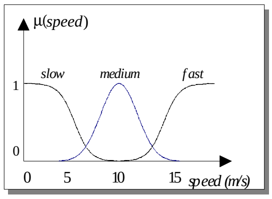

Figure 2: *Representation of fuzzy wind speeds. "Speed" is a linguistic variable with three terms "slow", "medium", and "fast" represented as fuzzy sets with the membership functions shown in the Figure.*

In the case of a linear function in the consequence, the model may be written analytically as following:

$$\hat{y} = \frac{\sum_{i=1}^{r}\left(p_0^i + \sum_{j=1}^{m} p_j^i x_j\right)\prod_{j=1}^{m} m_{A_j^i}(x_j)}{\sum_{i=1}^{r}\prod_{j=1}^{m} m_{A_j^i}(x_j)} \hat{=} \sum_{i=1}^{r} w^i \hat{y}^i$$

B. Learning and Generalization.

Model building is characterized by two phases: optimization of the model architecture and tuning of the model internal parameters (learning).

These two phases are driven by the requirement for good "generalization". Generalization is the capacity of the model to perform well when it predicts new data (data not used during the two phases of model development). Consequently, it is a primary requirement for the on-line use of a model.

The tuning of the model parameters is performed taking into account [8]:

- *Learning rules* based on stochastic gradient for tuning the parameters *a*, *b*, *p* of the model.
- Learning rules are appropriately developed to minimize simultaneously prediction error and the Information content of the model (max entropy). This acts as a *self-regularization* process that permits to avoid overfitting of the data.
- *Simulated annealing* is performed for controlling the evolution of the learning process through appropriate adaptation of the learning rate.
- *Early-stopping* is applied to the learning process in order to avoid overfitting.
- *Cross-validation* is applied to terminate learning. For this purpose, a subset of the data (validation set) is reserved.
- The cross-validation criterion is expressed as a weighted function of the performance of the model over the whole prediction horizon. By this way, generalization is optimized for multi-step ahead prediction.

The above process permits to tune optimally a model with a specific architecture. The architecture of a model is defined by the types of input variables and the number of fuzzy sets associated to each one. For each type of measured data it is needed to decide the number of past values to be used as input. When NWPs are considered as input it is necessary to select the relevant information (forecasts of wind speed, direction, etc) for the model.

This selection procedure, which is also similar to other types of models like neural networks, is a time consuming one due to the infinite number of combinations that can be evaluated. Often it is performed by trial-and-error, where several candidate configurations are tested. It is noted that the evaluation of each candidate model requires carrying out the above-described learning process.

In this work, the trial-and-error has been replaced by a fully automated process for model architecture optimization. The constrained nonlinear simplex ("Complex") optimization algorithm is used for this purpose. The algorithm has been modified for handling both discreet and continuous decision variables. The optimization process is based on the evaluation of the surface of the generalization function (defined as the performance of a model on the validation set) using a complex of points. Each point corresponds to a candidate model. The computational cost is high due to the necessity of the algorithm to tune each candidate model. However, in global, the automatic nature of the process permits to save considerable engineering time compared to the trial-and-error.

An alternative genetic algorithm approach did not present any advantages with respect to the simpler "Complex" algorithm. Genetic algorithms appeared to be less parsimonious w.r.t the number of models they need to test in order to converge compared to the Complex algorithm.
Each decision variable in Complex represents the number of fuzzy sets associated to each type of input data. In the special case, when the algorithm converges to zero-number of fuzzy sets for a specific type of data, then this input is excluded from the model as non-significant. By this way the algorithm performs input selection. When the number of fuzzy sets is converging to one, then the variable does not participate in the premises, but appears only in the function of the consequent part. Parsimony in the selection of input is critical to avoid overfitting by overparametrized models.

Figure 3 shows an example of a run of the Complex algorithm. 115 candidate models are totally examined. The input selection is performed among past values of wind power and Hirlam wind speed, direction and temperature forecasts. The upper left figure shows the evolution of the Complex objective function. Each point in the figure corresponds to the "generalization" performance of a candidate model on the validation set. The rest of Figures show the number of fuzzy sets associated by the algorithm to some of the input variables

When the number of fuzzy sets for all variables is either one or zero then a single "rule" is obtained. The premise has no significance and the model corresponds to a simple linear function of the input variables. This limit case corresponds to the ARX class of models. Consequently, the optimization process can indeed exclude the use of a nonlinear fuzzy model and lead to a classical linear one. In this way, a selection between linear and nonlinear models is performed.

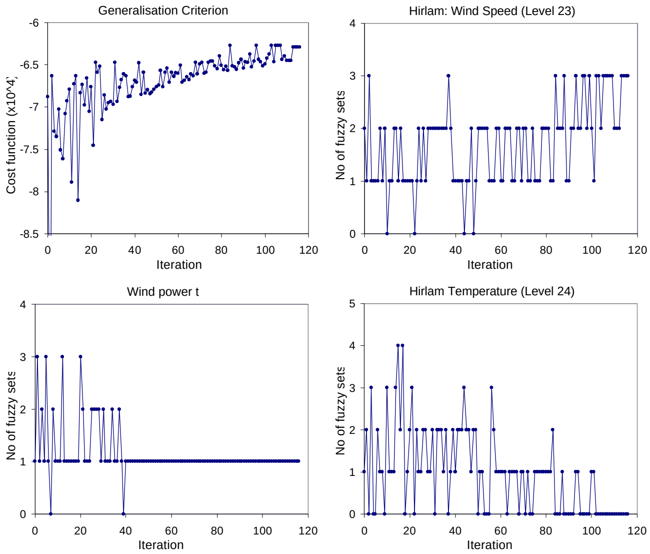

Figure 3: *Evolution of the algorithm for the model architecture optimization.*

IV. RESULTS

The case study of a real wind farm in Ireland is presented. This farm contains 20 wind turbines of 300 kW each and 1 turbine of 450 kW. Online data and Hirlam forecasts have been used covering the period between 5th February 2001 to 31st March 2002. The time series cover a period of 10000 hours from which 6600 were used for training (learning set), 1000 for cross- validation and 2400 (100 days) for testing the performance of the model. The results presented here are on the testing set.

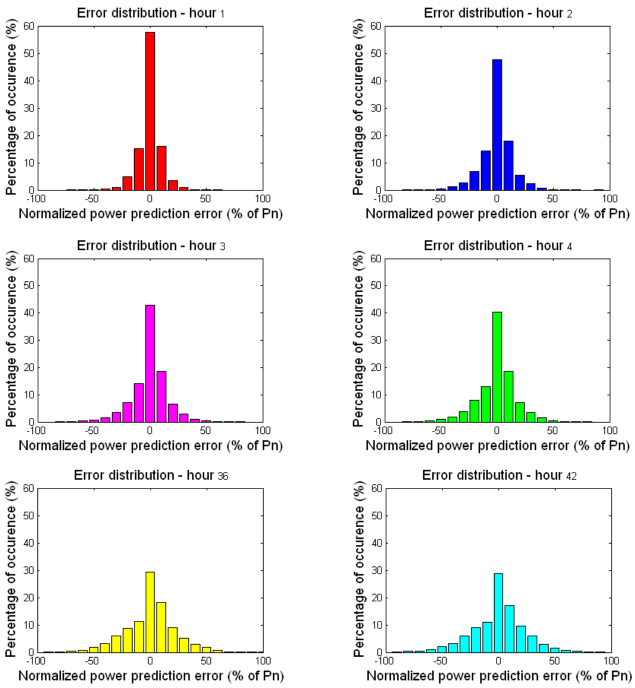

Figure 4: *Distributions of prediction errors for various look ahead times (+1h, +2h, +3h, +4h, +36h, +42h).*

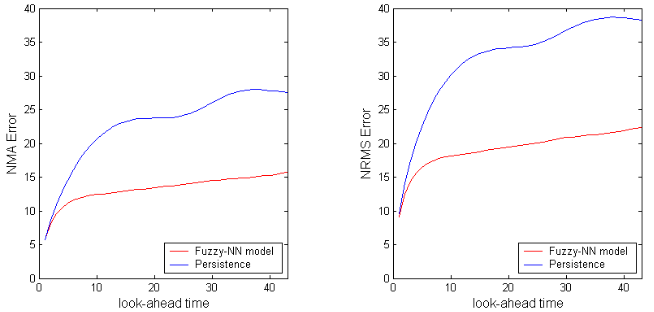

Figure 5: *Normalized Mean Absolute Error (%) and Root Mean Square Error (%) as a function of look ahead time. Comparison between Persistence and the advanced model. Normalization is made using the wind farms nominal power.*

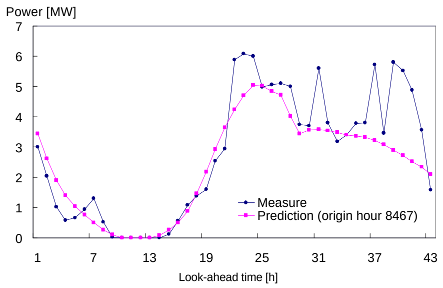

Figure 6: *Examples of a set of 42 hours ahead predictions of the total wind park production produced at different time origins (hours: 8467).*

Figure 4 shows the distribution of prediction errors for various look-ahead times. Figure 5 compares the performance of the fuzzy model with that of "Persistence" ("wind park production in the future is same as wind park production now", $\hat{P}(t+k)=P(t)$) on the testing set. Persistence is a simple method that is used as reference to evaluate advanced models. An advanced method is worth to implement for on-line operation only if it manages to beat Persistence. The F-NN model outperforms Persistence for all time steps. The Improvement or Skill of the F-NN model w.r.t. Persistence rises up to 55% according to the time step. The improvement for look-ahead time k is given by (i.e. for the NMAE criterion):

$$skill(k)=\frac{NMAE_p(k)-NMAE_{F-NN}(k)}{NMAE_p(k)}\cdot 100\%$$

An important issue is that this performance is positive for all time steps. This is one of the main benefits of this approach compared to other approaches like the "physical" one, which starts to have an improvement w.r.t. Persistence only after 3-5 hours ahead. A good performance in the short-term is required in several applications; i.e. in the case of a spot market where a bid for the wind production has to be made 1-3 hours in advance.

A practical problem that arises is related to the fact that there is a delay on the release of the NWPs by the meteo service. This delay can be in the order of 3-6 hours. One could argue that due to this delay the advantage of a statistical model over the first hours is lost since after 3-6 hours physical models start also to outperform persistence. This is not however true; the evaluation against Persistence is often done in a different way by statistical and physical models and thus the statistical criteria in the two cases are not directly comparable. Statistical models, as the ones presented here, produce new predictions every hour and operate with a sliding window scheme, while physical models usually produce predictions only when new NWPs arrive (i.e. every 6 hours). A fair comparison would require to simulate predictions by the physical model using a sliding window scheme and to compare with appropriate persistence forecasts.

In the short-term (1-6 hours) the inclusion of meteorological forecasts contributes to a better performance (twice as good) compared to that of the short-term models that do not consider meteorological forecasts as input. Figure 6 shows the performance of the model for the examined wind farm on a specific situation. This is a typical situation where for the first 24 hours the performance of the model is good, while it becomes lower for horizons more than 24 hours.

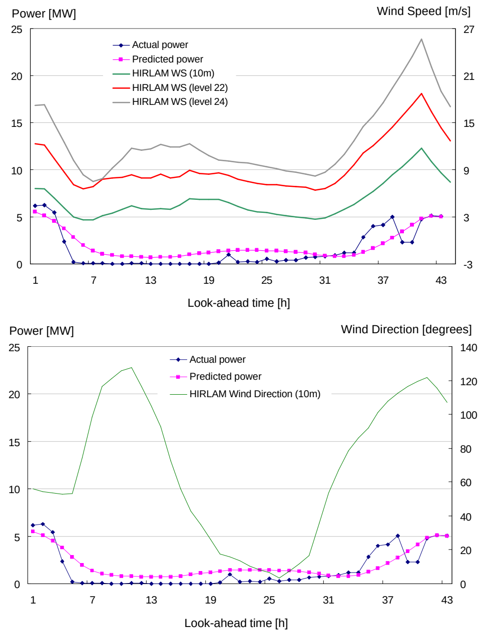

Figure 7: *Role of wind direction for adapting predictions.*

In Figure 7, it is demonstrated how the model exploits explanatory input information to improve its performance. As shown in the upper figure, the model predicts the same level of power (~5 MW) for the look-ahead times 1-3 and 37-43 although wind speed in the second case is higher. On the other hand, for look-ahead times 6-13, the predicted power decreases despite the speed increase. This means that during the two periods of 7-13 and 37-43, the model predicts less power than what would be expected by a direct consideration of Hirlam speed. This is in fact due to the influence of the wind direction on the wind farm production (in both cases wind direction is around 120 degrees).

Figure 8 shows a comparison between the performance of the fuzzy model and a simple model based on the conversion of the Hirlam wind speed at 10 m to the hub height and then to power using the machine power curve. The simple model under-predicts significantly power during the first hours. In contrast the fuzzy model predicts a higher power for the first lower peak of the Hirlam wind speed and captures the shape of the real curve.

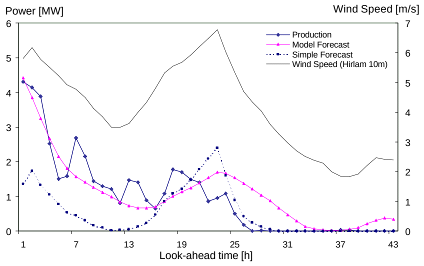

Figure 8: Comparison of predictions obtained by the F-NN models and a simple model that converts Hirlam forecasts at 10 m to the hub height.

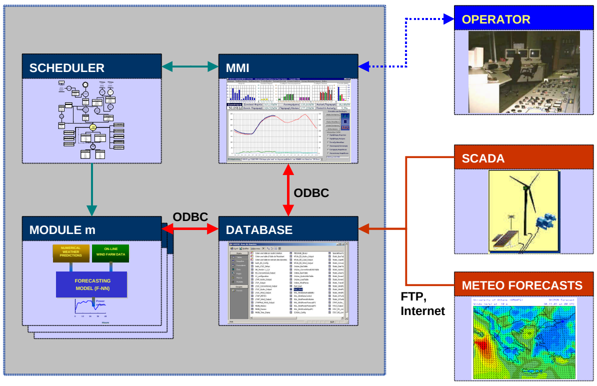

Figure 9: *Architecture of the forecasting software.*

V. CONCLUSIONS

The paper presents an adaptive fuzzy-neural network model developed for the prediction of the power output of a wind farm. The architecture of the model is optimized using the nonlinear simplex algorithm. The paper presents evaluation results on a wind farm in Ireland. Online and Hirlam data of one year were used for this purpose. The performance of the model is found to outperform both persistence as well as simple methods. The prediction module was developed in C++ using ODBC functionality and integrated in the More-Care EMS software - Figure 9. Several methods were tested and implemented for estimating the uncertainty of the predictions [10]. The software is installed for on-line operation at the islands of Crete and Madeira, where its performance is currently under evaluation. Since December 2002 it is also installed for on-line operation in Ireland for predicting the output of 11 wind farms. Figure 10 shows a display of the platform as configured to run in Ireland.

The above work permitted to assess the performance of advanced prediction techniques for wind power forecasting. In order to achieve higher levels of accuracy, substantial R&D efforts are necessary. In the frame of the European Project Anemos (ENK5-CT-2002-00665) which started in October 2002 (http://anemos.cma.fr) several fields of research in the area will be addressed for both onshore and offshore wind power prediction.

VI. REFERENCES

[1] Focken, U., Matthias, L., Waldl, H.P., “Reduction of wind power prediction error by spatial smoothing effects”, 2001 European Wind Energy Conference, Copenhagen, Denmark, July 2001, pp. 822-825.

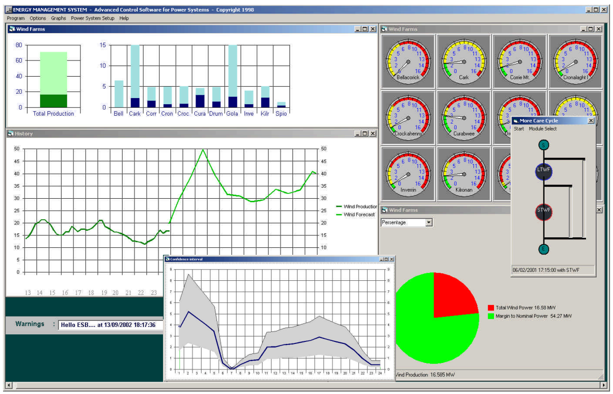

Figure 10: *The user interface of the forecasting platform.*

[2] Fukuda, H. et al, “The development of a wind velocity prediction method based on a Data-Mining type Auto-Regressive model”, 2001 European Wind Energy Conference, Copenhagen, Denmark, July 2001, pp. 741-744.

[3] Nielsen, T.S., Madsen, H., Nielsen, H.A., Landberg., L., Giebel, G., “Prediction of regional wind power”, 2002 Global Wind Power Conference, Paris, France, April 2002, CD proceedings.

[4] Kariniotakis G., Stavrakakis G. Nogaret E., “Wind power forecasting using advanced neural network models”, IEEE Trans. on Energy Conversion, Vol. 11, No. 4, Dec. 1996, pp. 762-767.

[5] Kariniotakis, G., Nogaret, E., Stavrakakis, G., “Advanced Short-Term Forecasting of Wind Power Production”, 1997 European Wind Energy Conference, Dublin, Ireland, 1997, pp. 751-754.

[6] Dutton A.G., Kariniotakis G., Halliday J.A., Nogaret E., “Load and Wind Power Forecasting Methods for the Optimal Management of Isolated Power Systems with High Wind Penetration”, Wind Engineering, Vol. 23, No 2, 2000.

[7] Kariniotakis G., Nogaret E., Dutton A., Halliday J. Androutsos A., "Evaluation of advanced wind power and load forecasting methods for the optimal management of isolated power systems.” Proceedings of the 1999 European Wind Energy Conf., Nice, France, 1-5 March 1999.

[8] Kariniotakis G., “Contribution to the development of an advanced control system for the optimal management of autonomous wind-diesel systems”, PhD thesis, Ecole Nationale Supérieure des Mines de Paris, Centre d’Energétique, December 1996.

[9] N. Hatziargyriou et al, “The CARE system overview: advanced control advice for power systems with large-scale integration of renewable energy sources”, Wind Engineering, Vol. 23, No 2, 2000.

[10] Pinson P., Kariniotakis G., “Wind Power Forecasting using Fuzzy Neural Networks Enhanced with On-line Prediction Risk Assessment.”, Proceed. of the IEEE Bologna Power Tech Conference 2003, June 23-26, Bologna, Italy.

ACKNOWLEDGMENT

This work has been performed in the frame of the MORE-CARE project (ERK5-CT1999-00019) supported by the European Commission. The authors would like to thank ESB National Grid for making available the data for the case studies.

BIOGRAPHY

George Kariniotakis was born in Athens, Greece. He received his production and management engineering and MSc degrees from the Technical University of Crete, Greece and his PhD degree from Ecole des Mines de Paris in 1996. He is currently with the Center of Energy Studies of Ecole des Mines de Paris as a scientific project manager. He is a member of IEEE. His research interests include renewable energies, distributed generation and artificial intelligence.