

# Variable Selection for Wind Power Prediction using Particle Swarm Optimization

R. Jursa  
ISET e. V.  
Königstor 59, D-34119 Kassel,  
Germany  
Tel. +49(0)5617294253  
rjursa@iset.uni-kassel.de

## ABSTRACT

Wind energy has an increasing influence on the energy supply in many countries, but in contrast to conventional power plants it is a fluctuating energy source. For its integration in the electricity supply structure it is necessary to predict the wind power hours or days ahead. There are models based on physical, statistical or artificial intelligence approaches for the prediction of wind power.

In this paper a new short-term prediction method is described based on variable selection using particle swarm optimization and nearest neighbour search. As input variables for this prediction method weather data of a numerical weather prediction model and measured power data from wind farms of several locations in a spread area are used. Additionally a prediction model based on neural networks is described and the results of the new method are compared to the results of the neural network approach.

As a result we get a reduction of the prediction error by using the new prediction method. An additional error reduction is possible by using the mean model output of the neural network model and of the nearest neighbour search based prediction approach.

## Categories and Subject Descriptors

J.2 [Physical Science and Engineering]: Mathematics and statistics, engineering, earth and atmospheric science

**General Terms:** Algorithms

## Keywords

Variable selection, particle swarm optimization, wind power prediction, nearest neighbour search, neural network

## 1. INTRODUCTION

Wind power utilization for electricity generation will play an important role in the electricity supply in many countries in the future. In Germany it currently accounts for about 6% of the electricity supply. Its share is planned to increase to about 15% by 2020.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'07, July 7–11, 2007, London, England, United Kingdom.  
Copyright 2007 ACM 978-1-59593-697-4/07/0007...\$5.00.

Transmission system operators (TSO) have to balance the electricity production, so that the electricity production at any time matches the consumption or the load. Wind energy is a fluctuating and non-stationary power source, so that a certain amount of reserve power is needed for compensating these fluctuations. But this reserve power lowers the environmental advantage gained by using wind energy and raises the costs of conventional power production, due to the sub-optimal fuel consumption of conventional power plants using as reserve. In a study it is shown, that the using of wind power predictions lowers the costs for the integration of wind power in the electricity supply system and decreases the emissions of greenhouse gases significantly [1].

There are mainly two different approaches used for wind power prediction. One employs physical models of wind farms to determine the relationship between weather data from a numerical weather prediction model and the power output of the wind farm [2], [3], [4]. The physical approach yields the wind speeds in hub height of each wind turbine in the wind farm. The total power output of the wind farm is then calculated by using the power curves of the wind turbines. The other is the mathematical modelling approach in which statistical or artificial intelligence methods for example neural networks are used to find the relationship between weather prediction data and power output from historical data sets [5], [6], [7]. In this approach the influence of the orography, surface roughness and losses due to turbulence on the power output of the wind farm are empirical estimated with the advantage that also hidden physical parameters are included in the modeling.

In the last years there was strong research effort on the improvement of wind power predictions using meteorological forecast data from numerical weather prediction (NWP) systems. In the European project ANEMOS several prediction models were applied and compared for the prediction of selected wind farms located in areas with different characteristics [8]. Improvements of wind power predictions are obtained by the application of different NWP systems or mesoscale- and computational fluid dynamic models with higher grid resolutions in the area of the wind farms [9], [10]. In other recent studies strong improvements up to about 20% are obtained by using the data of different NWP models or ensemble models as input data for the wind power prediction models [11], [12], [13], [14], [15].

For the short-term prediction with prediction horizons of some hours ahead (1-6h) the persistence, i.e. the simple prediction assumption that the current power output of the wind farm will be unchanged in the next hours, is a popular benchmark test. So in

general the improvement of the prediction models compared to the persistence are specified.

In the most prediction models only local data are used, i.e. weather prediction data and measured power data for the location of the wind farm in question. Due to the space-time development of weather fronts over an extended area there are dependencies between weather and power data of different locations of wind farms. This can be interpreted as an information flux from several points of the NWP and several wind farm locations to the power data of a considered wind farm. Alexiadis [16] built an approach for short-term predictions with prediction horizons up to two hours ahead with a neural network model for certain wind situations. This approach uses additional wind speed data as input from two other locations with long distances apart to the considered wind farm location. For a single wind farm in Greece they got about 20% improvement of the prediction quality compared to the persistence error for one hour ahead. Taking only local wind speed data from the wind farm they got only about 6% improvement. But the considered wind farm is located in an area with only one prevailing wind direction of strong local winds. These special conditions are not directly comparable to the general situation of e.g. middle European countries of the northern hemisphere with strong winds from different directions. In a similar study about wind speed prediction Larson and Westrick [17] showed recently that the short-term forecast of the wind speed with prediction horizons of some hours for one location can be improved by using wind data from two other observation locations.

In this study we show a method to use this additional space-time information of a spread area to improve the wind power prediction. We use a prediction method based on nearest neighbour search (NNS) and select the input variables from weather and power data of different locations in the spread area. Taking all data from all locations as input variables leads to known disadvantages like the curse of dimensionality and a bad prediction performance by using irrelevant input variables for the NNS [18], [19]. So a method for selecting relevant variables and eliminating of irrelevant variables or features is required. There are many approaches in the area of feature selection, which can be used for prediction modeling [20]. Here an embedded method is studied for selecting the relevant input variables using particle swarm optimization (PSO) [21].

In Germany a neural network model is used operationally by TSO [12]. For the day-ahead prediction meteorological forecast data from the NWP system of the German weather service are used as input to the artificial neural networks (ANN). The simultaneous data of measured power output of wind farms are used to train the ANN, which are then able to calculate the power output from weather prediction data. Data from about 100 representative wind farms in Germany are available for this.

The introduced prediction method is evaluated on a test data set and compared to the results of the ANN based prediction model. The ANN gets local input data from the wind farm. For a better comparison of the performance of the variable selection algorithm with the ANN performance we train a second ANN getting the same variables as input data as the NNS.

We present in chapter 2 a brief description of the ANN model and a description of the NNS based approach using particle swarm optimization for variable selection. In chapter 3 we present the error measures for the evaluation of the models, in chapter 4 the results of example calculations with a discussion and at the end a summary.

## 2. PREDICTION METHODS

### 2.1 Artificial Neural Network

The artificial neural network consists of nonlinear functions, which are combined by a series of linear filters with weights [22]. In this study we use a neural network with three hidden layers and train the ANN by gradient descent with the back propagation algorithm. It minimizes the least square error  $E$  between the measured power  $P_n$  and the predicted power  $\hat{P}_n$  at time step  $n$  over a training data set with  $N$  data points:

$$E = \sum_{n=1}^N \left( \frac{P_n - \hat{P}_n}{P_{rated}} \right)^2 \quad (1)$$

where  $P_{rated}$  is the rated power of the wind farm. The scaling of the error by the rated power is the standard scaling for wind power predictions [23]. For the training of the ANN we use historical NWP data and historical measured power data for the same discrete time steps of the NWP data. As input variables of the ANN for the time  $t$  and the prediction horizon  $\hat{t}$  we use the NWP data for the location of the wind farm, namely 6 values of wind speed  $ws(t)$ ,  $ws(t + \hat{t})$ ,  $ws(t + 2\hat{t})$  for a height 10m and 100m above sea level, the sine and cosine of the wind direction  $wd(t + \hat{t})$  in 10m height, 3 values of power data  $p(t - 2\hat{t})$ ,  $p(t - \hat{t})$ ,  $p(t)$  and 4 values related to the time  $t + \hat{t}$ , i.e. the sine and cosine of the time with a period of one year and of one day. The same ANN is trained with the input variables, which are calculated by the PSO, denoted in the following by ANNpso. After the training the ANN provides the predicted power value.

### 2.2 Nearest Neighbour Search

This prediction method for the power output of single wind farms uses NNS. To build an input structure for NNS we construct time delay vectors from weather and measured power data. In detail, we construct from numerical weather prediction data and measured power data for every time  $t$  a time delay vector of several locations in a spread area, i.e. we take the wind speeds  $ws_i(t)$ , the wind directions  $wd_i(t)$  and the measured power  $p_i(t)$  of different locations  $i$  and build a common time delay vector  $\vec{S}(t)$  as a phase space reconstruction [22] of the weather and power data.

In this representation scheme of the input variables we can describe possible selections of the input variables by the construction parameters of the phase space reconstruction, i.e.:

- The embedding dimensions as the number of data points of each weather and power variable related to each location.
- The time shift as the difference between the actual time  $t$  and the time related to the weather and power variables.
- The time delay as the time difference between the data points of the weather and power variables.

So the possible selection of the input variables is transformed to the selection of time delay vectors with certain construction parameters inside the vector space of all possible time delay vectors. For the  $i$ -th location we have the following representation scheme  $\vec{S}_i(t)$ :

$$\begin{aligned} \vec{S}_i(t) = & (w_{i,ws}(t-s_{i,ws}), w_{i,ws}(t-(s_{i,ws} + \tau_{i,ws})), \dots, w_{i,ws}(t-(s_{i,ws} + (d_{i,ws} - 1)\tau_{i,ws})), \\ & w_{i,wd}(t-s_{i,wd}), w_{i,wd}(t-(s_{i,wd} + \tau_{i,wd})), \dots, w_{i,wd}(t-(s_{i,wd} + (d_{i,wd} - 1)\tau_{i,wd})), \\ & p_i(t-s_{i,p}), p_i(t-(s_{i,p} + \tau_{i,p})), \dots, p_i(t-(s_{i,p} + (d_{i,p} - 1)\tau_{i,p}))) \end{aligned} \quad (2)$$

In equation (2)  $\tau_{i,ws}$ ,  $s_{i,ws}$  and  $d_{i,ws}$  denote the time delay, the time shift and the embedding dimension at the  $i$ -th location for the wind speed values respectively. Analog  $\tau_{i,wd}$ ,  $s_{i,wd}$ ,  $d_{i,wd}$  are the construction parameters for the wind direction values and  $\tau_{i,p}$ ,  $s_{i,p}$ ,  $d_{i,p}$  that ones for the power values. The state representation of the whole system  $\vec{S}(t)$  with  $M$  locations is a common time delay vector built from all  $\vec{S}_i(t)$ :

$$\vec{S}(t) = (\vec{S}_1(t), \vec{S}_2(t), \dots, \vec{S}_M(t)) \quad (3)$$

From historical NWP data and historical measured power data we build for every hour the same time delay vectors as for the actual data. So it is possible to calculate the distance between an actual time delay vector and all historical time delay vectors.

Due to a non-linear relationship between wind speed and the power output of the wind farms we use in  $\vec{S}(t)$  also the quadratic and the cubic values of the wind speeds given for two different heights, namely for 10 and 100m above sea level. And to get periodic functions of the degree values of the wind directions with a step size of  $45^\circ$  we use the sine and cosine parameterization of the wind direction. All wind speed and power values are scaled to values between 0 and 1.

For the estimation of distances of the NNS the Euclidean distance [24] is used:

$$d(u, v) = \sqrt{\sum_{i=1}^D (u_i - v_i)^2} \quad (4)$$

In our case  $D$  is the dimension of the state vector  $\vec{S}(t)$  and the  $u_i$ ,  $v_i$  are its coordinates. Here the  $k$ -nearest neighbour search is used, which provides a certain number  $k$  of nearest neighbours [25]. For the wind power prediction of a considered wind farm we use the power data of this wind farm corresponding to the  $k$  nearest neighbours to the actual state  $\vec{S}(t)$ . These  $k$  power values and the power value of the actual time are the input of our prediction model, which is a linear combination of the nearest

neighbours weighted by their distance  $d$  and the actual power value:

$$\hat{P}(t + \hat{t}) = \sigma \frac{1}{\sum w_n} \sum_{\vec{S}_n \in U_{\vec{S}(t)}} w_n P(t_n + \hat{t}) + (1 - \sigma) P(t) \quad (5)$$

$$\text{with } U_{\vec{S}(t)} = \{\vec{S}_n : |\vec{S}_n - \vec{S}(t)| < \varepsilon\}, n \in \{1, \dots, k\} \quad (6)$$

where  $U_{\vec{S}(t)}$  is the  $\varepsilon$  neighbourhood of  $\vec{S}(t)$  within the  $k$  estimated neighbours are measured by the Euclidean distance measure  $d$ .  $P(t_n + \hat{t})$  are the corresponding power values of the neighbourhood state  $\vec{S}_n$  at time  $t_n$  plus the prediction horizon  $\hat{t}$ . For the weighting of the neighbours we use the bi-weight kernel-function [26]:

$$w_n = \left(1 - \frac{d_n^2}{d_k^2}\right)^2 \quad (7)$$

where  $d_k$  is the distance of the neighbour with the largest distance.  $\sigma$  is a weighting factor between 0 and 1, so that we use a weighted combination of the nearest neighbour based model and the persistence approach.

## 2.3 Selection of the Input Variables

In the state vector  $\vec{S}(t)$  we have several parameters. With them we can define the selection of the input variables of the prediction model. One part of these parameters are the construction parameters  $\tau_{i,ws}$ ,  $s_{i,ws}$ ,  $d_{i,ws}$ ,  $\tau_{i,wd}$ ,  $s_{i,wd}$ ,  $d_{i,wd}$ ,  $\tau_{i,p}$ ,  $s_{i,p}$  and  $d_{i,p}$  of the common time delay vector  $\vec{S}(t)$  and the other part are the number of estimated nearest neighbours  $k$  and the weighting factor  $\sigma$  of the prediction model equation (5).

### 2.3.1 Particle Swarm Optimization

To select the input variables we use particle swarm optimization (PSO). The PSO algorithm is a population-based non-linear optimization algorithm and the population is called swarm and each individual is called a particle and each particle has a position defined by coordinates in a certain search space. Each position of a particle yields a certain personal fitness value according to the optimization criteria. Each particle has also a velocity, which is used to change the coordinates at each iteration step of the optimization algorithm.

In our optimization problem the position of a particle corresponds to a certain selection of the input variables and the fitness values of the PSO is the prediction error of the NNS. At the beginning of the optimization the positions and velocities are initialized by assigning random values. At each optimization iteration  $j$  the velocities  $V_i^j$  and the positions  $X_i^j$  of each particle  $i$  are updated by the following rules:

$$\begin{aligned} V_i^{j+1} &= \omega V_i^j + c_1 r_1 (P_i^j - X_i^j) + c_2 r_2 (P_g^j - X_i^j) \\ X_i^{j+1} &= X_i^j + V_i^{j+1} \end{aligned} \quad (8)$$

where  $X_i^j$  is the current position of the  $i$ -th particle at the  $j$ -th iteration step. For  $\sigma$  we can use the real values provided by the algorithm, but for the construction parameters of  $\vec{S}(t)$  and the number of neighbours  $k$  we use a cast of the real values to integer values.  $P_i^j$  is the best previous position according to the best personal fitness value and  $P_g^j$  is the best global position corresponding to the best previous position achieved by all particles.  $r_1$  and  $r_2$  are random numbers between 0 and 1. The other values  $\omega$ ,  $c_1$  and  $c_2$  are optimization parameters of the PSO.

Additionally to this standard PSO a random change of the coordinates in each iteration step is introduced to avoid in a better way suboptimal local minima in the search space. For the changes of the coordinates corresponding to the coordinate classes of the embedding dimensions, the time delays and the time shifts an additional random change

$$X_i^{j+1} = X_i^{j+1} \pm 1 \quad (9)$$

is introduced in 5 percent of the cases of coordinates changes in each coordinate class. For the coordinate corresponding to the number of the nearest neighbours  $k$  and for the coordinate corresponding to the weighting factor  $\sigma$  we use in 10 percent of the cases

$$X_i^{j+1} = c(2r - 1)X_i^{j+1} \quad (10)$$

with  $r$  as a random number in the interval  $[0,1]$  and  $c$  as a scaling factor set here to 0.1.

### 3. ERROR MEASURES

For the evaluation of the performance of our prediction methods we use standard error measures [23] and we present the results scaled by the rated power  $P_{rated}$  of the wind farm. In addition to the mean power output MEAN of the wind farm

$$MEAN = \frac{1}{N} \sum_{n=1}^N \frac{P_n}{P_{rated}} \quad (11)$$

we calculate the BIAS

$$BIAS = \frac{1}{N_T} \sum_{n=1}^{N_T} \frac{P_n - \hat{P}_n}{P_{rated}} \quad (12)$$

and the RMSE

$$RMSE = \sqrt{\frac{1}{N_T} \sum_{n=1}^{N_T} \left( \frac{P_n - \hat{P}_n}{P_{rated}} \right)^2} \quad (13)$$

In the above equations  $P_n$  and  $\hat{P}_n$  are the measured and predicted power respectively for every time step  $n$ .  $N$  is the number of available data points and  $N_T$  the number of test data points, which are not used in the optimization procedure.

For comparison to the persistence we calculate also the improvement, i.e. the normalized difference of the RMSE of the persistence and the RMSE of the presented prediction models:

$$IMP = \frac{RMSE_{persistence} - RMSE_{model}}{RMSE_{persistence}} \quad (14)$$

Beside the model outputs of the single models ANN and NNS we use also the mean model output of both models as prediction values.

## 4. RESULTS FOR SINGLE WIND FARMS AND DISCUSSION

First prediction calculations for single wind farms were done with power output measurements of wind farms in the German E.ON control zone and NWP prediction data from the German weather service.

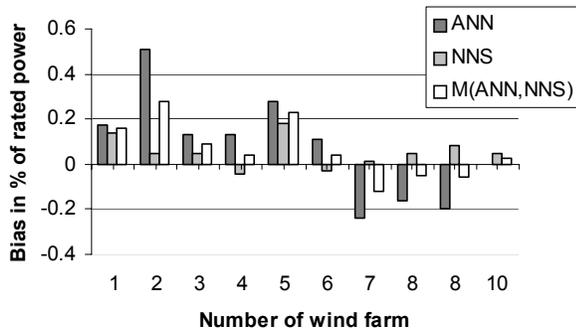
**Table 1. Number of test data and mean power output of the used wind farms for evaluation**

WF	Number of test data points	Mean power output in % of rated power
1	9311	19.07
2	9234	21.48
3	9077	16.5
4	7454	21.54
5	7232	34.63
6	8117	33.89
7	6169	13.83
8	9138	20.32
9	9357	13.2
10	9311	12.55

For the evaluation of the prediction we choose by random test data points of the whole data set with approximately 25000 data points. The remaining data points are used for the training of the ANN prediction model and the optimization procedure of the NNS based prediction model. Data from February 2004 to December 2006 have been used for modeling and testing. Table 1 shows the number of test data points for each wind farm (WF), which vary due to the different amount of available data points.

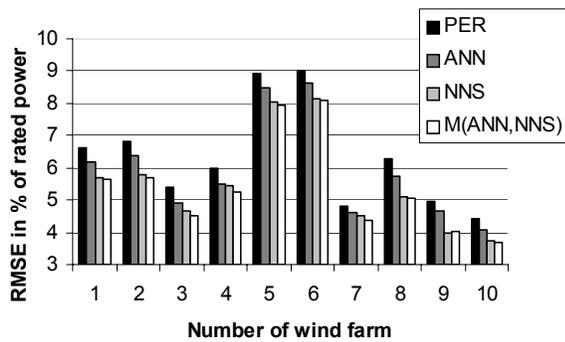
In the next figures we show the error measures as defined in chapter 3.

From figures 1, 2 and 3 it can be seen that the NNS-model provides smaller prediction errors than the ANN-model, but for two wind farms (4 and 7) the RMSE-values of the NNS are only slightly better than those of the ANN. One can also see that it is possible to reduce the RMSE additionally by using the mean value of both models for the power prediction.

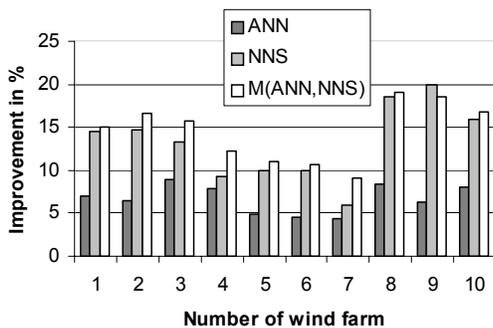


**Figure 1.** Bias of the predictions of the two prediction models ANN and NNS and the mean model output M(ANN,NNS).

Regarding the RMSE-values, we can see that the prediction error of the mean model output M(ANN,NNS) from ANN and NNS is in general smaller than the mean of the RMSE-values of both individual models. Only for wind farm 9 it would be a disadvantage to use the mean model output, because for this wind farm the prediction performance of ANN and NNS has the largest difference.



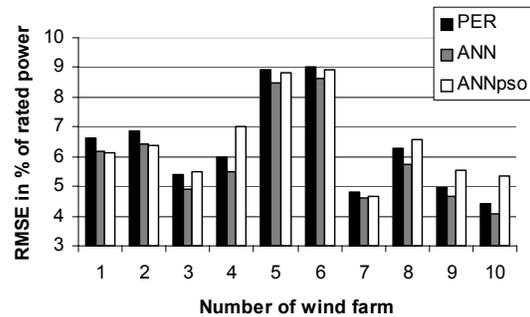
**Figure 2.** Comparison of the RMSE values of the persistence (PER), the ANN-model, the NNS-model and the mean of the ANN- and NNS-model.



**Figure 3.** Comparison of the improvement of the ANN-model, NNS-model and the mean of the ANN- and NNS-model.

In figure 4 the comparison between the ANN and ANNpso is presented. One can see that nearly for all wind farms the RMSE-values of ANNpso are higher than those of ANN. So the input variables selected by the PSO for the NNS should not be used for

the ANN. And we can conclude that the performance of the selection of the input variables by the PSO depends on the underlying prediction model, which is in our case the NNS.



**Figure 4.** Comparison of the RMSE values of the persistence (PER), the ANN-model and the ANNpso-model.

In table 2 we show the mean improvement over all 10 considered wind farms. As one can see the highest improvement can be obtained by using the mean model output of ANN and NNS, which is more than two times higher than the improvement of the ANN-model. The improvement of the NNS-model alone is only slightly worse for the most wind farms than this mean model output, due to the fact that the performance of the NNS-model is much better for the most wind farms than that one of the ANN-model.

**Table 2.** Mean improvement (IMP) of the ANN-model, of the NNS-model and of the mean model output of ANN and NNS.

IMP of ANN	IMP of NNS	IMP of M(ANN,NNS)
6.7%	13.2%	14.5%

## 5. SUMMARY

To summarize, we presented a new method for short-term wind power prediction based on particle swarm optimization and nearest neighbour search. The particle swarm optimization is used for variable selection of weather and measured power data of several locations in a spread area for the nearest neighbour search. We calculated with this approach wind power predictions for 10 wind farms in Germany and compared the results with an operationally used prediction method based on neural networks.

We can conclude, that the wind power prediction error can be reduced by using the presented variable selection method. In comparison to the neural network model we get for some wind farms a strong reduction of the prediction error. The mean improvement of the prediction error in comparison to the persistence of the nearest neighbour search is 13.2%, while that one of the neural network is only 6.7%. An additional reduction of the prediction error, at least for the RMSE-values, is possible by using the mean model output of both presented models, i.e. of the neural network and the nearest neighbor search. We get then

an improvement of 14.5%. Using the selected variables also as input data for the neural network yields to higher RMSE-values in comparison to those obtained with a neural network with manual selected local input data of the considered wind farm. The performance of the selection of input variables depends on the prediction model incorporated in the optimization algorithm.

In future work we want to use the particle swarm optimization for selecting the input variables with the neural network as the underlying prediction model. As additional input for the prediction models we want to use measured wind speed and wind direction data.

## 6. REFERENCES

- [1] Doherty, R. and O'Malley, M.: System operation with a significant wind power penetration. In *Proceedings of the IEEE Power Eng. Soc. General Meeting*, Denver, Colorado (USA), volume 1, 1002-1007, 2004.
- [2] Beyer, H.G., D. Heinemann, H. Mellinshoff, K. Mönich, and H.-P. Waldl: Forecast of Regional Power Output of Wind Turbines. In *Proceedings of the European Wind Energy Conference*, Nice, France, 1-5 March 1999, pp. 1070-1073.
- [3] G. Giebel, J. Badger, I. Martí Perez, P. Louka, G. Kallos, A. M. Palomares, C. Lac, G. Descombes: Short-term Forecasting Using Advanced Physical Modelling – The Results of the Anemos Project, Results from mesoscale, microscale and CFD modeling. In *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March 2006.
- [4] Lange, M., Focken, U.: *Physical Approach to Short-Term Wind Power Prediction*, Springer 2006.
- [5] Giebel, G., Landberg, L., Kariniotakis, G., Brownsword, R.: The State-Of-The-Art on Methods and Software Tools for Short-Term Prediction of Wind Energy Production. In *Proceedings of the European Wind Energy Association Conference*, Madrid, Spain, 16-19 July 2003.
- [6] Nielsen T.S., Madsen, H., Nielsen, H. Aa., 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 *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March 2006.
- [7] Sánchez, I.: Short-term prediction of wind energy production. *International Journal of Forecasting*, 22 (2006), 43 – 56.
- [8] Kariniotakis G., et al: Next generation short-term forecasting of wind power – overview of the ANEMOS project. In *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March 2006.
- [9] Kariniotakis, G., Martí, I., Casas, D., Pinson, P., Nielsen, T.S., Madsen, H., Giebel, G., Usaola, J., Sanchez, I., Palomares, A.M., Brownsword, R., Tambke, J., Focken, U., Lange, M., Louka, P., Kallos, G., Lac, C., Sideratos, G., Descombes, G.: What performance can be expected by short-term wind power prediction models depending on site characteristics? In *Proceedings of the European Wind Energy Conference*, London, UK, 22 February - 25 November 2004.
- [10] Martí, I., Kariniotakis, G., Pinson, P., Sanchez, I., Nielsen, T.S., Madsen, H., Giebel, G., Usaola, J., Palomares, A.M., Brownsword, R., Tambke, J., Focken, U., Lange, M., Sideratos, G., Descombes, G.: Evaluation of Advanced Wind Power Forecasting Models – Results of the Anemos Project. In *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March, 2006.
- [11] von Bremen, L.: Optimal linkage of different NWP models with neural networks for offshore wind power predictions. *Sixth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Delft, Netherlands, 26 –28 October 2006.
- [12] Lange, B., Rohrig, K., Ernst, B., Schlögl, F., Cali, Ü., Jursa, R., Moradi, J.: Wind power prediction in Germany – Recent advances and future challenges. In *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March, 2006.
- [13] Lange, M., Focken U., Meyer, R., Denhardt, M., Ernst, B., Berster, F.: Optimal combination of different numerical weather models for improved wind power predictions. *Sixth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms*, Delft, Netherlands, 26 –28 October 2006.
- [14] Gregor, G. (ed.), Badger, J., Landberg, L., Nielsen, H. Aa., Nielsen, T.S., Madsen, H., Sattler, K., Feddersen, H., Vedel, H., Tøfting, J., Kruse, L., Voulund, L.: *Wind Power Predictions using Ensembles*. Risø National Laboratory, September 2005. (available at [www.risoe.dk/rispubl/VEA/ris-r-1527.htm](http://www.risoe.dk/rispubl/VEA/ris-r-1527.htm))
- [15] Lang, S., Möhrle, C., Jørgensen, J., Gallachóir, B. O., McKeogh, E.: Application of a Multi-Scheme Ensemble Prediction System for wind power forecasting in Ireland and comparison with validation results from Denmark and Germany. In *Proceedings of the European Wind Energy Conference*, Athens, Greece, 27 February - 2 March 2006.
- [16] Alexiadis, M.C., Dokopoulos, P.S., Sahsamanglou, H.S.: Wind speed and power forecasting based on spatial correlation models. *IEEE Transactions on Energy Conversion*, 14, 836 – 842, 1999.
- [17] Larson, K. and Westrick, K.: Short-term forecasting using off-site observations. *Wind Energy*, 9, 5-62, 2006.
- [18] Kantz, H. and Schreiber, T.: *Nonlinear Time Series Analysis*. Cambridge Univ. Press, 1997.
- [19] Blum, A., and Langley, P.: Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2), 1997, 245-271.
- [20] Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A. (Eds.) *Feature Extraction: Foundations and Applications*. Series: Studies in Fuzziness and Soft Computing, Vol. 207, Berlin, Springer, 2006.

- [21] Eberhart, R. C. and Kennedy, J.: A new optimizer using particle swarm theory. *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, (Nagoya, Japan), pp. 39-43, IEEE Service Center, 1995.
- [22] Gershenfeld, N.: *The Nature of Mathematical Modeling*. Cambridge, Cambridge University Press, 1999.
- [23] Madsen, H., Kariniotakis, G., Nielsen, H. Aa., Nielsen, T.S., Pinson, P.: A Protokoll for standardizing the performance evaluation of short-term wind power prediction models. *Report of the ANEMOS Project*, 2004. (available at <http://anemos.cma.fr>)
- [24] Gimeno, J.M., Bejar, J., Sanchez-Marre, M.: Nearest-Neighbours for Time Series. *Applied Intelligence*, 20 (1) 2004, 21-35.
- [25] Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*. Springer, 2001.
- [26] McNames, J.: A nearest trajectory strategy for time series prediction. *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*, Katholieke Universiteit Leuven, Belgium, 112-128, July, 1998.