Short time ahead wind power production forecast.

Alla Sapronova(1), Catherine, Meissner(2), Matteo Mana(2)

(1) Uni Research, Bergen, Norway (2) WindSim, Tonsberg, Norway

alla.sapronova@uni.no

Abstract. An accurate prediction of wind power output is crucial for efficient coordination of cooperative energy production from different sources. Long-term ahead (from 6 to 24 hours) prediction of wind power for onshore wind parks can be achieved by using coupling model that would bridge mesoscale weather prediction data and computational fluid dynamics (as presented in [1]). When a forecast for shorter time horizon (less than 1 hour) is anticipated, an accuracy of predictive model that utilizes hourly mesoscale weather data is decreasing as the higher frequency fluctuations of the wind speed are lost when data is averaged over an hour. It is observed in [2] that wind speeds can vary up to 50% in magnitude over a period of 5 min and the analysis of wind speed fluctuations over periods from minutes to hours [3] shows that higher frequency variations of wind speed and direction have to be taken into account for accurate short-term ahead energy production forecast. In this work a new model for accurate wind power output forecast for 5- to 30-minutes ahead is presented. The model is based on machine learning techniques and categorization approach [4] and uses historical park production time series and NWP to predict the total power production of the wind park.

[1] Introduction

Currently, wind energy plays a significant role in total electricity production in Europe and therefore it is vital to predict the wind energy output timely and accurate. The variability of the wind is a main challenge for obtaining an accurate forecast for minutes to hours ahead.

Many studies have been devoted to the improvements of wind forecasting techniques with number of models developed and launched. These models are based on physical, statistical or hybrid approaches and often using Numerical Weather Prediction (NWP) data at tens of kilometers resolution as a main component. But NWP based models often not providing satisfactory accuracy for site-specific and short horizon forecasts.

For site-specific forecast the mesoscale-microscale coupling model was proposed in [1]: there artificial intelligence methods were used to issue 1 to 3 hours ahead wind forecast from NWP data and CFD module was used to calculate the flow at finer scales [5]. The accuracy of coupling model [1] was found to be superior comparing to that based on polynomial fittings as well as ARMA models, although accuracy for shorter than one hour forecast remains unsatisfactory.

In this work, the model proposed in [1] is adapted to forecast wind power output for less than one hour ahead by utilizing power production time-series in addition to NWP data. Adapted model also uses a data categorization approach [4], when information obtained from categorization of a single variable (wind speed) is supplied as an input in addition to other variables.

[2] Model selection

The model for short-time ahead energy output prediction is based on time series analysis of historical park production and NWP data. Since the quality of forecast depends on ability of the model to predict the wind flow near the ground in the complex terrain, where roughness and complexity affect the flow at microscale, the model employs machine learning techniques as those are proven to be efficient for nonlinear multivariable functions approximation when explicit physical based models have limited application or not available.

In this work a combination of artificial neural networks (ANN) module and kernel (SVM) module were used. The modules are trained, tested and validated on historical data from several on-shore wind parks from Sweden and Norway. Data pre-processing included data cleaning, one- or 5-minutes time intervals averaging, and normalization.

To forecast the total wind park power production 5- to 30-minutes ahead two feed forward (FF) ANNs have been build: the first ANN was trained on time-series of multiple environmental variables (e.g. wind speed and direction, NWP data), and the second was trained on wind power production time-series. The root mean square percentage error (RMSPE) for validation data set is shown in Table 1.

Back propagation training method and Encog Workbench open source software package [7] has been used for ANNs training and validation.

Prediction time-window	NWP and historical power production time series	Power production time series
5 minutes ahead prediction	12.2	8.2
15 minutes ahead prediction	14.9	9.7
30 minutes ahead prediction	15.1	9.5

Table 1. ANN-module performances for short time ahead wind power forecast.

ANN that trained on only historical power production time series has been selected for further modification.

A double-module architecture has been suggested to improve the model's accuracy: one ANN has been used to issue a “coarse” prediction and the output from the first module has been submitted to the second machine learning unit (a “correcting” module) to issue more accurate forecast. With suggested double-module approach the RMSPE has been lowered 7.9 and 8.9 for 5 minutes ahead and 30 minutes ahead prediction respectively.

The “correcting” module architecture based on ANN has been compared to one based on kernel methods. The RMSPE for different “correcting” module architectures is shown in Table 2.

Prediction time-window	ANN module	SVM module
5 minutes ahead prediction	7.9	6.4
30 minutes ahead prediction	8.9	7.9

Table 2. Correcting-module performances for short time ahead wind power forecast.

The performance of double-module model has been compared to data categorization approach based model. This approach is described in [4] and suggests that the entire data-set is grouped into several discrete categories which allow identical category values to be treated in the same manner for non-equal continues numerical data. Previously it was shown [6] that the selection of methods for categorization is not critical, so in this work the wind speed was categorized. The information obtained from categorization of wind speed variable was supplied to ANN as input in addition to time series variables.

RMSPE on validation data set for all the above mentioned models is summarized in Table 3.

Single-module ANN based model [1]	Double-module, ANN based model	Double-module, ANN and SVM based model	Double-module model with data categorization approach
14.9	8.9	7.9	7.5

Table 3. RMSPE for 30 minutes ahead power output forecast for different models architectures.

[3] Results

The suggested double-module model comprise of (1) ANN, that uses park power production time-series to predict power output 5 to 30 minutes ahead, and (2) SVM, that utilizes output from ANN as one of the inputs along with NWP data and time-series of registered wind speed, wind direction, nacelle speed, nacelle position, wind power production. Schematic description of the model is shown in Figure 1.

Figure 1. Double-module model: "coarse"-predicting module is using longer time-series of a single variable; the output from "coarse" module is utilized by "correcting" module along with shorter time-series of multiple variables.

ANN module performance was tested on 1 and 5 minutes averaged data. ANN architecture and performance for various time-series windows and prediction horizons are shown in Table 4.

Prediction horizon	5 minutes			10 minutes			20 minutes			30 minutes		
Time-series window interval, minutes	10	20	30	20	30	40	30	40	60	40	60	90

Number of hidden neurons	10 to 30	20 to 60	30 to 90	20 to 60	30 to 90	40 to 100	30 to 90	40 to 100	60 to 100	40 to 100	60 to 100	90 to 100
Validation error (RMSPE)	14.9	9.7	7.9	15.1	12.2	8.1	18.1	12.4	8.6	16.2	8.9	8.9

Table 4. ANN module architectures and performances for short time ahead wind power forecast.

Output from ANN along with other inputs has been submitted to SVM to predict the power output for the same time ahead. With double-module approach RMSPE for 10 minutes ahead forecast has been lowered from 7.9 to 5.9. SVM module performance is shown in Table 5.

Prediction horizon	5 minutes				10 minutes				30 minutes					
Power production time-series length, minutes	5	5	10	10	5	5	10	10	5	5	10	10	20	20
NWP data used	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Wind speed and direction time-series, minutes	5	5	5	5	5	5	5	5	10	10	10	10	10	10
Validation error (RMSPE)	6.6	6.7	5.9	5.9	6.9	6.8	6.1	6.2	9.1	9.4	8.6	9.2	8.2	8.5

Table 5. SVM module performances for short time ahead wind power forecast.

SVM module performance is not significantly affected by luck of NWP data for very short term ahead prediction, yet for longer prediction horizons, like 30 minutes ahead, using NWP data as input helps to lower RMSPE from 8.5 to 8.2.

Finally, the model have been modified to employ categorization data (as described in [4]). The wind speed data were categorized and supplied as one of input variables to SVM module. The model that is using categorization approach shows better performance, as described in Table 6.

Prediction horizon	5 minutes		10 minutes		30 minutes		
Power production time-series length, minutes	5	10	5	10	5	10	20
Wind speed and direction time-series, minutes	5	5	5	5	10	10	10
RMSPE for model without categorization approach	6.7	5.9	6.8	6.1	9.1	8.6	8.2
RMSPE for model with categorization approach	4.5	4.6	5.1	5.1	6.5	6.2	6.2

Table 6. SVM module performances for short time ahead wind power forecast with and without categorization approach.

[4] Conclusions

In this work a new model to forecast wind power production 5 to 30 minutes ahead is presented.

The model uses NWP, historical wind power production data and employs machine learning methods and categorization approach.

The best performance is shown for combination of ANN and kernel method for double-module model.

The model shows that use of NWP data is not significantly improve the accuracy for very short time ahead forecast (5-10 minutes ahead), yet is valuable for 30 (and longer) minutes ahead forecasts.

If categorization of an input variable is used, the better accuracy of the forecast can be achieved even with smaller number input variables. It is observed that model's generalization arises from the model's ability to find similarity in the training data that usually consists of continuous numeric data. Since numbers are rarely exactly the same from one example to the next, the model can fail in selecting the margins for identical properties. In this case, the generalization can be improved by classification. As shown in [5] the choice of methods for categorization is irrelevant to the generalization improvement therefore the generalization approach can be adapted for various input parameters.

[5] Aknowledgement

The work is sponsored by the Norwegian Research Council, project ENERGIX, 2015-2016.

References

[1] Sapronova, A., Meissner, C., Mana, M.: Mesoscale-microscale coupled model based on artificial neural network techniques for wind power forecast. Poster at EWEA Offshore 2014, PO ID 385 (2014)

[2] A. Sapronova, NORCOWE report: NORCOWE-RR-C-14-WP2-018 (2014)

[3] C.L.Vincent, P.Pinson, and G.Giebela, Wind fluctuations over the North Sea, Int. J. Climatol. 31: 1584–1595 (2011)

[4] A.Sapronova, C.Meissner, M.Mana, Improving an Accuracy of ANN-Based Mesoscale-Microscale Coupling Model by Data Categorization: With Application to Wind Forecast for Offshore and Complex Terrain Onshore Wind Farms, Lecture Notes in Computer Science, 2014, Volume 8817, pp.61-66

[5] M.Mana, Short-term Forecasting of wind energy production using CFD simulations., Poster at EWEA 2013 (2013)

[6] Li, L., Pratap, A., Lin, H.-T., Abu-Mostafa, Y.S.: Improving generalization by data categorization. In: Jorge, A.M., Torgo, L., Brazdil, P.B., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 157–168. Springer, Heidelberg (2005)

[7] http://www.heatonresearch.com/wiki/Encog_Workbench