IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 121, No. 2, April – June, 2017, pp. 137–160

Statistical correction of the wind energy forecast at the Hungarian Meteorological Service

Helga Tóth[1*], **Brigitta Brajnovits**[2], and **Balázs Renczes**[3]

[1]*Hungarian Meteorological Service,*
Kitaibel Pál u. 1; H-1024 Budapest, Hungary
E-mail: toth.h@met.hu,

[2] *MAVIR Hungarian Independent Transmission System Operator Company Ltd.*
Anikó u. 4, H-1031 Budapest, Hungary
E-mail: brajnovitsb@gmail.com

[3] *Budapest University of Technology and Economics,*
Department of Measurement and Information Systems,
Magyar tudósok körútja 2, H-1117 Budapest, Hungary
E-mail: renczes@mit.bme.hu

**Corresponding author*

(Manuscript received in final form, March 10, 2017)

Abstract— In order to efficiently integrate renewable energy sources – the production of which can be planned harder – in the energy grid, quite accurate forecasts are needed. Especially in Hungary, where energy storage is yet an unsolved problem. The limited ability of wind forecast means that the delivered power of the wind farms cannot be predicted with sufficient accuracy. This work focuses on the improvement of wind power forecast precision by using different statistical methods. In the first part of the paper, simple BIAS correction approaches and more complex ensemble based methods are applied to improve the power prediction for the whole country. The second part of the work focuses on enhancement of the wind power forecast of a single wind farm. While mostly only wind speed is taken into consideration for the calculation of the generated power, it is shown that air density is also an important factor in the equations. Autoregressive filtering is launched in order to show that wind speed and power forecasts can also be improved by this kind of statistical method.

Key-words: wind power, numerical weather prediction; statistical correction, quantile regression, analog ensembles, characteristic curves, fuzzy model, autoregressive filtering

Paper is based on the presentation delivered at Meteorological Scientific Days held on 19–20 Nov, 2015 entitled „Meteorological aspects of use of renewable energies”.

1. Introduction

Nowadays it is a crucial question to use renewable energy all over the world and to integrate it to the electricity network. Due to huge energy demand in the 20th century, the depletion of conventional energy sources (e.g., fossil oil, natural gas), and the increasing CO_2-emission, wind energy, as renewable energy source, has become of increasing importance. In the last decades, the world's wind energy utilization exploded. While in 1996 only 6 GW built-in capacity was in operation, by 2015 this value has exceeded 433 GW, and the wind energy industry is set to grow by another 60 GW in 2016 (*Global Wind Energy Council,* 2016). The European Union is also committed to use renewable energy sources. The latest revised RES-E directive in 2016 sets out the EU target of at least 27% for the renewables in the final energy consumption by 2030 (*European Commission*, 2016). In 2014, 195.8 Mtoe (Million of tonnes of oil equivalent) renewable primary energy was produced in the EU-28 area. The share of the energy sources were the following: biomass and waste 63.1%, hydropower 16.5%, wind 11.1%, geothermal 3.2%, solar 6.1% (*Eurostat*, 2016). The Hungarian production was 2.05 Mtoe for renewable primary energy, and the share of the sources are: Biomass and waste 89.2%, Geothermal 6.3%, Wind 2.8%, Hydropower 1.3%, Solar 0.5% (*Eurostat*, 2016). In spite of many advantages, the users of 'green energy' are also facing several unresolved problems. Critics of wind power argue that the intermittent energy production of the turbines has a negative effect on the security of the power system. The limited ability of the forecast of wind eventuates that the delivered power of wind farms cannot be forecasted with sufficient accuracy. As a consequence, to ensure secure system operation, the energy supplied by wind should only have a limited proportion in the total electricity production. In the last few decades, fluctuations could easily be regulated, but since the wind energy has gained higher proportion in the power system, these uncertainties cannot be ignored any longer.

In Hungary, currently there is no scope for wind power expansion. In 2006, the Hungarian Energy and Public Utility Regulatory Authority limited the wind power capacity to 330 MW. This value seemed to be sufficient that time, but 329 MW was built in by 2011, so practically no more installation is possible at the moment. The current political and regulatory environment in Hungary does not allow installing new wind power plants. For capacity expansion, a more precise predictability of wind power would be elemental. Nowadays, several methods are being investigated in order to ensure this issue, such as the energy storage. By means of this, in case of oversupply, a battery can be charged, and the stored energy can be used, when undersupply occurs (*Hartmann*, 2012). Another area is the prediction of the generated wind power. Statistical data mining approach was applied for wind farm power prediction (*Kusiak*, 2009). Another possible way is to use numerical weather prediction (NWP) models. However, NWP models forecast the wind speed and direction values, the system operator needs wind power data. To overcome this problem, neural networks are generally constructed (*Barbounis et. al*, 2006). Another approach is to calculate the characteristic curve between wind speed and wind power, by means of, e.g., the bin method or regression (*Llombart et. al,* 2005).

The production of wind energy strongly depends on weather, resulting that its planning is very difficult. It is an important task for the meteorologists to prepare the most accurate forecasts for the target users. The goal for the Hungarian Meteorological Service (OMSZ) is to adapt and develop algorithms connected to the calculation of power energy for wind, solar, and water resources. The operational usage is also important and expected.

This work focuses on the forecast of the wind power and on the improvement of its precision by using different methods. In Hungary, the power production has to be given for two days ahead, and for this horizon, meteorological forecasts have certainly to be used to achieve the desired accuracy. The aim of this work is to ensure the possibly most accurate forecast using meteorological data.

In the first part of this work, statistical correction of wind power forecast is shown, which is summarized for all wind power stations in Hungary for the Hungarian Transmission System Operator Ltd (MAVIR). Simple BIAS correction approaches (*Sweeney* and *Lynch*, 2011), like short term rolling trend (STT), short term rolling bias (STB), composite forecast (COM) are applied to correct the wind power predictions, without real access.

Quantile regression (QR) and analog ensemble (AnEns) methods (*Alessandrini et al.,* 2014) were also applied to gain probabilistic information from a deterministic model. The expectation was that not only additional information can be generated, but the final result would mean a more accurate prediction than that obtained with a single model run.

The second part of the paper focuses on improving of the wind power forecast of a single wind farm. It is shown that not only wind speed, but also air density is an important influencing parameter of wind power production. Air density has already been taken into consideration in former works (*Rizwan et. al*, 2012; *Farkas*, 2011), calculating the characteristic curve from two separate inputs. In this work, the wind speed is normalized with the air density as suggested in *IEC 61400-12-1* (2005), so the characteristic curve is determined from one input parameter, keeping the model simpler. Results can also be improved by applying statistical methods, as autoregressive modeling (*Collomb*, 2009).

2. Data and methods

The kinetic energy of the wind, can be calculated by E formula: $E = \frac{1}{2} m v^2$, where m is the air mass, v is the wind speed. Assuming a constant wind speed, the quantity of air flowing through an area A, which is perpendicular to the wind direction can be given by:

$$A = \frac{m}{\rho v t}, \tag{1}$$

where ρ denotes the air density, and t is the time. The power is the time derivative of kinetic energy:

$$P_0 = \frac{dE}{dt} = \frac{E}{t} = \frac{1}{2}\rho A v^3 . \tag{2}$$

For standard atmosphere (temperature of 273 K and pressure of 1013 hPa), the power can be calculated using the following equation:

$$\frac{P_0}{A} = 0.647 v^3. \tag{3}$$

From this equation, two consequences follow. First, wind has low power density compared to conventional energy sources. Thus, large turbine diameter is needed in order to get sufficiently high power density (*Hunyár et al.*, 2001). On the other hand, the equation is strongly non-linear, which may lead to significant error in the later prediction that is based on the wind speed. The first issue is further augmented by the Betz's law (*Betz*, 1966), i.e., the maximum amount of power supply P_w is:

$$(P_w)_{max} = \frac{16}{27} A \frac{\rho}{2} v^3 \tag{4}$$

However, real wind power plants cannot produce arbitrarily large power. On the other hand, under and above given wind speeds, no power can be obtained – these are called cut-in and cut-out speed, respectively. The characteristic curve is illustrated in *Fig. 1*.

Furthermore, wind speed varies with the height. This can be calculated by:

$$v_h = v_{10} \left(\frac{h}{h_{10}}\right)^n, \tag{5}$$

where v_{10} is the velocity at height $h_{10} = 10\,m$, n is the so-called Hellmann exponent, which is different above dry land and sea, and also depends on the topography.

Meteorological forecasts do not provide the value of wind power, which is the most important parameter for system operator. Their outputs are wind speed, temperature, relative humidity, etc. However, these data can be used to determine the produced wind power.

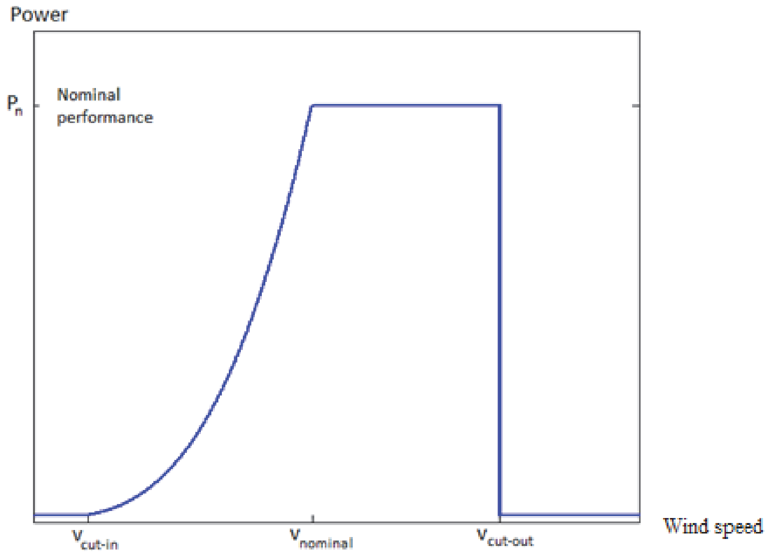

Fig. 1. Wind turbine performance as a function of wind speed (*Hunyár et al,* 2001).

The maximum obtainable power of wind turbines depends on several parameters (Eq. (4)). The most important of these is certainly the wind speed. Thus, mostly the wind power – wind speed (P-v) characteristic curve is given by manufacturers. However, according to Eq. (2), also air density plays an important role. This effect will be investigated later.

In practice, the shape of a characteristic curve is slightly different from that of *Fig. 1*, since it is sigmoid-shaped – curves for different turbines are given in *Fig. 2*. Furthermore, real power curves are significantly different from those of given by manufacturers, so determination of the real characteristic curve is crucial in order to give precise schedule.

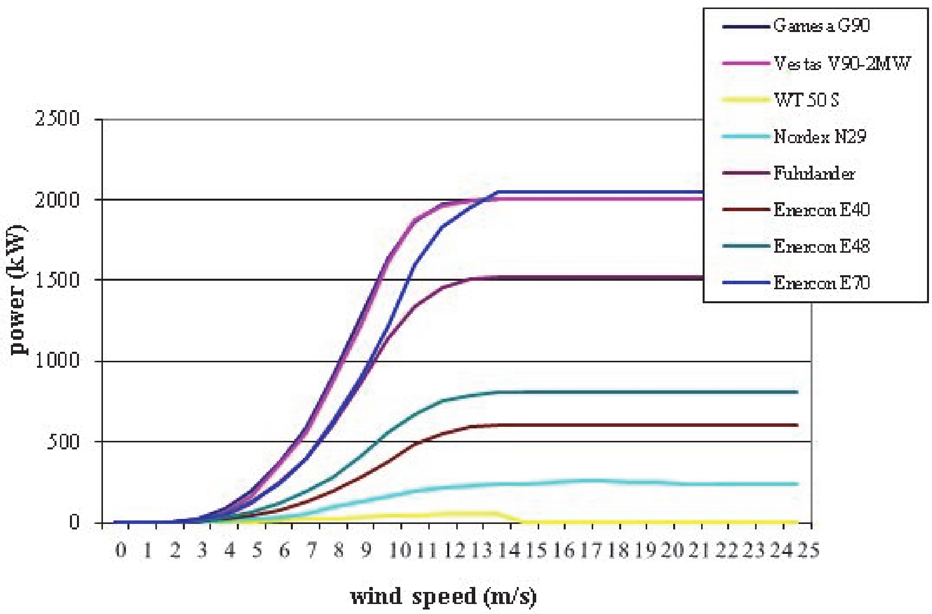

Fig. 2. Characteristic curves of different turbines.

To calculate power using the characteristic curves, wind speed forecast is needed on the hub height. The wind speed forecasts are provided by the AROME non-hydrostatic numerical weather prediction model (*Szintai et al.*, 2015) with 2.5 km horizontal and 15 minutes temporal resolution over the Carpathian domain (*Fig. 3*).

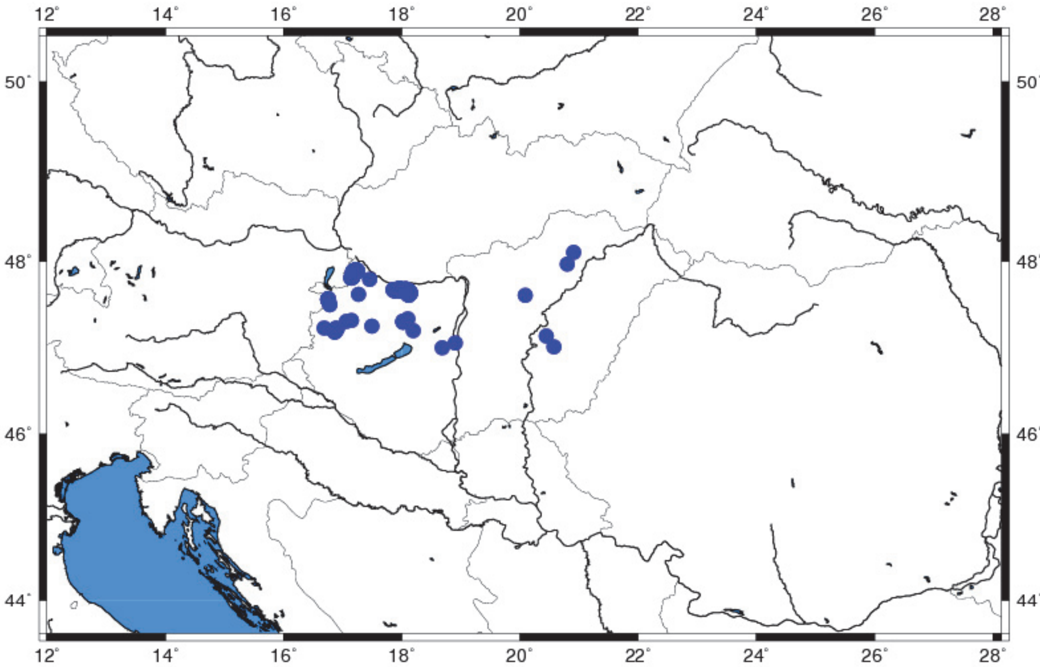

*Fig. 3***.** Domain of the AROME model as run operationally at OMSZ. Blue dots are wind power plants in Hungary.

The predicted wind speed is vertically interpolated to hub height, and horizontally to turbine position. From the individual wind speed values, power productions are calculated with a simple transformation of the characteristic curves. MAVIR gets the summarized power estimation of the 172 Hungarian tubes (*Fig. 3* and *Table 1*). Most of the wind power stations belong to 2 MW installed capacity and locate in the north western part of the country.

Forecasts are valid for +3–39 hours with 15 min temporal resolution, and are prepared 4 times a day (00 UTC, 06 UTC, 12 UTC, and 18 UTC). The system operator is able to use the latest wind power estimations for the planning in the next hours/day, and he/she is able to control the uncertainties of the wind power depending on the weather conditions (*Fig. 4*).

Table 1. Most important parameters of wind farms in Hungary (*MEH*, 2015)

Type of wind turbine	Number of pieces	Nominal capacity of turbine [MW]	Rotor height [m]
Enercon E40	7	0.6	78
Enercon E48	5	0.8	75
Enercon E70	5	2.0	100
Fuhrlander MD77	2	1.5	100
Gamesa G90	91	2.0	100
Nordex N29	1	0.3	50
Repower MM 82	12	2.0	100
Vestas V27	1	0.23	32
Vestas V52	1	0.8	86
Vestas V90 – 1.8 MW	2	1.8	105
Vestas V90 – 2 MW	36	2.0	100–105
Vestas V90 – 3 MW	8	3.0	100
WT 50 S	1	0.05	24

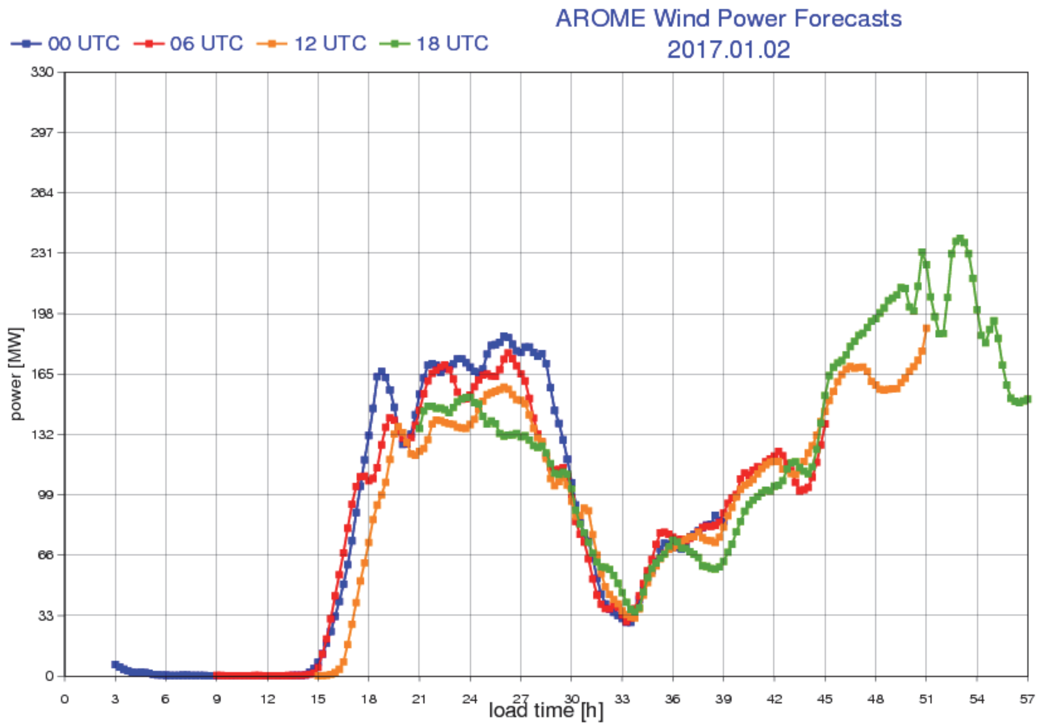

Fig. 4. AROME wind power forecast for Hungary (blue: 00 UTC, red: 06 UTC, orange: 12 UTC, green: 18 UTC forecasts).

MAVIR expects that the power forecasts should be as precise as it is possible, and the mean absolute error (MAPE) should not exceed 5%:

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{P_{me}-P_{mo}}{P_{me}}\right| * 100\ \%\ , \tag{6}$$

where P_{me} means the power measurement and P_{mo} denotes the modeled value. In real purposes this expectation is unrealistic, e.g., the value of *MAPE* was 53% in the summer of 2015. Mean relative error (*modeled-measured*) histogram in *Fig. 5* shows that the model significantly overestimates the wind power. In 30 % of the cases, the error reaches 100 %. This result seems scary, but most of the forecasted and measured data are relatively small, so the errors are not so large (*Fig. 6*).

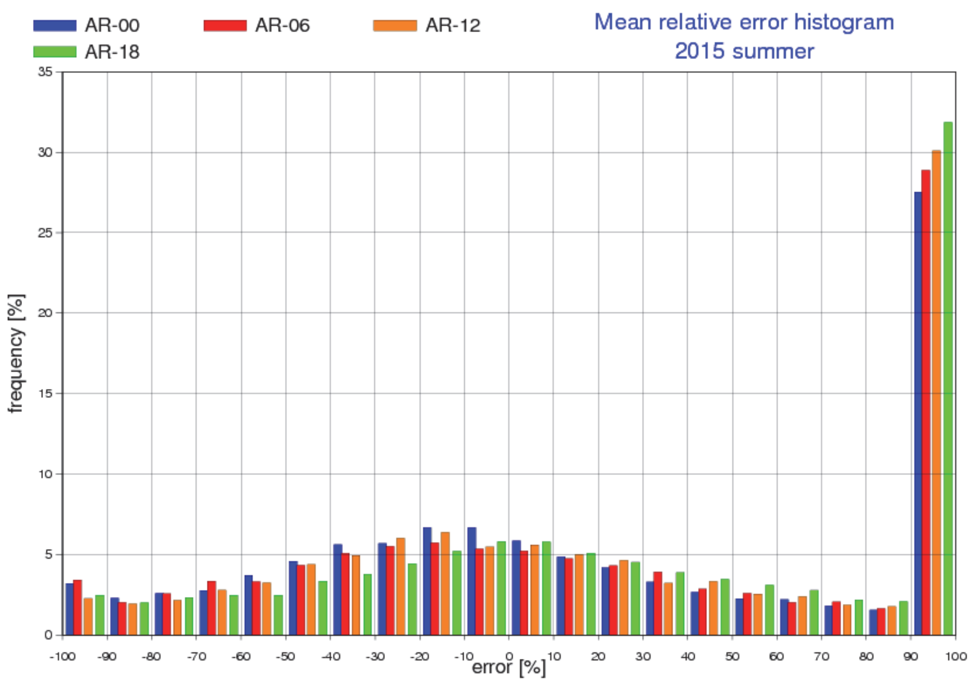

Fig. 5. Mean relative errors of AROME wind power forecast for Hungary for summer, 2015 (blue: 00 UTC, red: 06 UTC, orange: 12 UTC, green: 18 UTC forecasts).

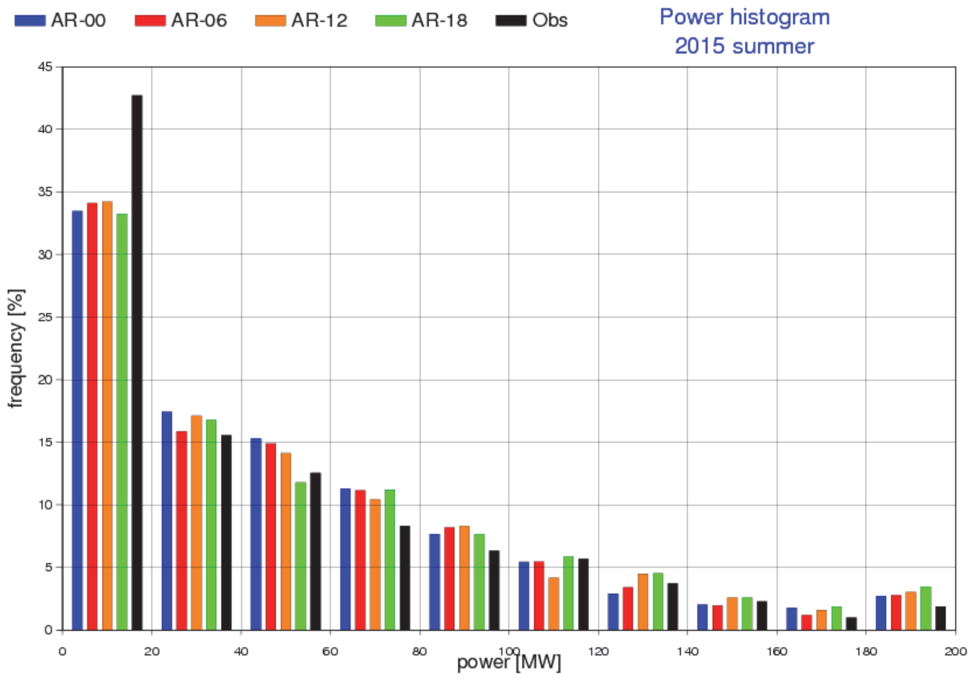

Fig. 6. Forecasted and measured power values for Hungary for summer, 2015 (blue: 00 UTC, red: 06 UTC, orange: 12 UTC, green: 18 UTC forecasts, black: measured).

Wind is a very fluctuating meteorological parameter. Thus, the power also changes rapidly due to the cubic relationship between the two variables (Eq. (2)). Unfortunately, this kind of issue affects also the results of BIAS corrections.

2.1. Statistical corrections of wind power forecasts

First of all, simple BIAS correction methods were applied on the summarized wind power forecast valid for the whole country. The power measurements are also available only for the whole country. This makes any kind of statistical correction work less accurate. Some model output statistics were provided with 100 days training period to improve the forecast skill.

Short-term rolling trend (STT) correction calculates the average errors for every time step over the previous 28 days (*Sweeney* and *Lynch,* 2011). Then the method corrects the raw forecast values with the computed ones at every time steps. Short-term rolling bias (STB) computes the average error of the previous 3 days, and corrects the original forecasts with this single value. Composite method (COM) is given by the sum of weighted STT, STB, and simple BIAS corrections.

Secondly, quantile regression (QR) and analog ensemble (AnEns) were applied to get probability information based on deterministic forecast using statistical approach. QR was provided to generate a set of the wind power predictors. 19-member ensemble forecasts were based on the AROME model from quantile 0.05 to 1 by step 0.05. QR can be carried out for a random variable *Y* as follows:

$$Q(\tau) = \beta_0(\tau) + \beta_1(\tau)x_1 + \beta_2(\tau)x_2 + \cdots + \beta_p(\tau)x_p, \tag{7}$$

where x_p are the *p* known regressors, $\beta_p(\tau)$ are the coefficients for the τth quantile and τ is the probability of finding values of *Y* below $Q(\tau)$. The coefficients for each quantile might be gained throughout a learning period. In our case, the total dataset is one-year-long (July 2012 – August 2013), therefore both the training period and the test period are half-year-long. Similarly to the work in *Alessandrini et al.* (2014), wind speed, wind direction, and forecast leadtime had been used as predictors, testing which can probably be neglected. Thus, four setups have been tried, in two of them forecast leadtime was used as predictor (wind speed and leadtime, then wind speed, wind direction, and leadtime as predictors). The coefficients for each quantile had been trained, and for the other two setups (wind speed, and wind speed and wind direction as predictors) not only each quantile but every leadtime had their own coefficients.

The analog ensemble technique is based on the work of *Alessandrini et al.* (2014). It was originally applied for ensemble forecasts, but can also be used with deterministic forecasts (*Delle Monache et al.*, 2012). The method, as its name suggests, searches for the closest *N* analogues to each forecast leadtime's prediction in the training period. These *N* members are the ensemble members for the given leadtime. In our case, the program searches for analogous wind speeds and wind directions in the training period, and extracts the corresponding power measurement. The general metric used to search for the best analogue for the current forecast is as follows:

$$\|F_t, A_{t'}\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{i,t-j} - A_{i,t'+j})^2}\ , \tag{8}$$

where F_t is the forecast that we search the $A_{t'}$ analogue for in the past at time t'. N_v is the number of physical variables and w_i are their weights, while σ_{fi} is the standard deviation of the past forecasts of the given variable for the selected location (*Alessandrini et al.*, 2014). As the number of ensemble members should be the square root of the number of training days, we chose the first 15 analogues, gaining a 15-member EPS.

2.2. *Corrections of a single wind farm data*

2.2.1. *Fitting the P-v curve with fuzzy model*

Measured data of two turbines (600 kW E40 and 2 MW E70) are available from May 2010 to June 2011, containing wind speed, wind direction, and wind power with the time resolution of 10 minutes. For fitting of the *P-v* characteristic curve, a cluster estimation method is applied (*Chiu*, 1994). This method can deal with more input parameters, so not only wind speed, but, e.g., wind direction can be handled. In this work, however, only wind speed is considered as input in order to keep our model as simple as possible.

The model assumes that in the investigated multi-dimensional function, different groups can be created that behave similarly. For wind turbines, at low speeds the output power is zero, and at high speeds (but lower than the cut-out speed) it equals to the nominal performance of the turbine. The center of these groups can be calculated, and these centers characterize the groups accurately. The applied fuzzy method implements the principle of *Sugeno* and *Tanaka*, 1992. In this case, the model will be able to describe a complex behavior with only a few rules. Using this approach, the characteristic curve fitting can be executed.

2.2.2. *Consideration of air density*

Most power forecasting methods take only wind speed into consideration. However, Eq. (4) clearly shows that the maximum obtainable performance is the linear function of air density. Certainly, the most important parameter is the wind speed, but only a few percent of improvement can be crucial for the system operator (who wants to ensure system stability) and the turbine operator (whose aim is to minimize the regulatory surcharge to be paid).

Fig. 7 shows the effect of air density on the produced power. Usually the higher the air density, the higher the generated power (at the same wind speed). However, there are some exceptions, since the power of the wind is determined by several parameters that were not investigated in this work, e.g., by the distribution of the wind speed in the 10 minutes interval.

The international standard of wind power measurements also deals with the effect of air density. For pitch regulated wind turbines it specifies the following normalization:

$$v_n = v_{10min} \sqrt[3]{\frac{\rho_{10min}}{\rho_0}}, \tag{9}$$

where v_n denotes the normalized wind speed, v_{10min} and ρ_{10min} are .... the measured wind speed and air density averaged over 10 minutes, while ρ_0 is the reference air density (*IEC 61400-12-1,* 2005).

Wind speed forecasts created by OMSZ have 15 minutes temporal resolution. Thus, these data had to be interpolated in order to ensure the same resolution for both datasets. It has to be mentioned that there were no air density data measurements, so the training can only be performed using the forecasted values. However, parameters determining air density (temperature, air pressure, humidity) can be forecasted more precisely than wind speed.

Due to the high computational demand, air density data were only determined for two weeks, one week in June 2010 and one week in February 2011.

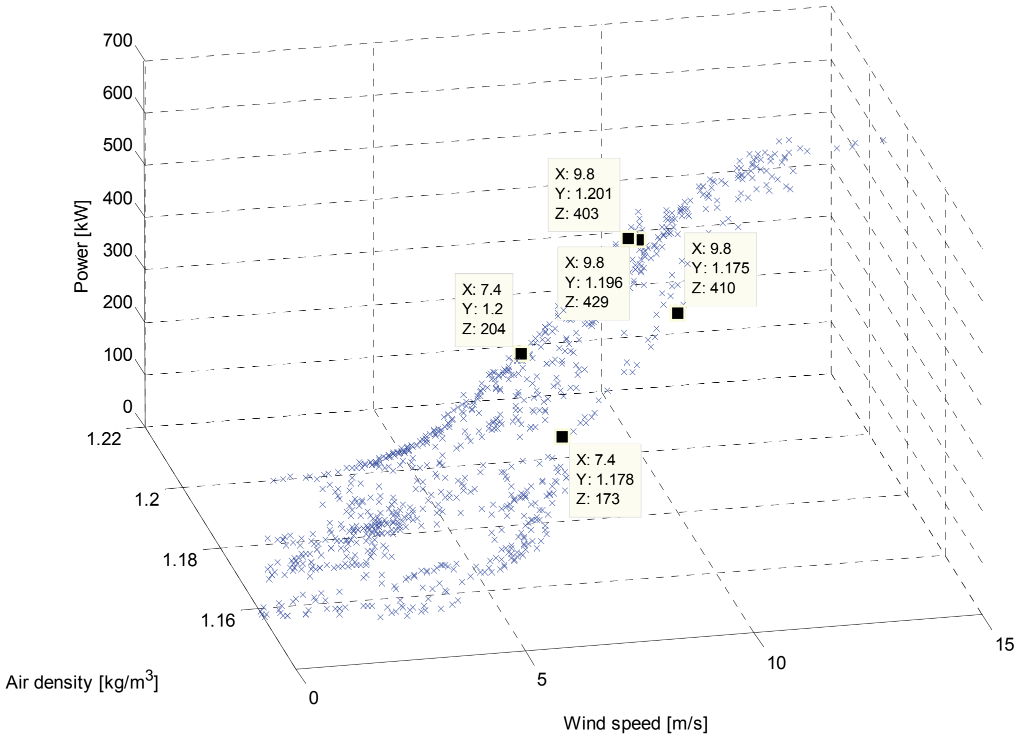

Fig. 7. The effect of air density on the generated power.

2.2.3. Autoregressive filtering of wind speed forecast

In this section, wind speed prediction is processed by autoregressive filtering. First of all, the definition of an autoregressive process is given. If X_n can be calculated as:

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \cdots + a_p X_{n-p} + \xi_n, \tag{10}$$

where a_i is real and ξ_n is white noise, then X is an autoregressive process of order p. The autoregressive process is stationary, i.e., a_i is constant in time. Consequently, it has to be declared that although the method can model atmospheric processes, it can only describe an 'average' behavior. Thus, it will be unable to follow the unique nature of the atmosphere at a given time. Nevertheless, the algorithm is expected to reduce the mean squared error.

To determine the coefficients a_i, the output is given as $y_n = -\sum_{i=1}^{p} a_i y_{n-i}$. This formula is similar to Eq. (10), the difference is the sign of the coefficients. This description is useful, because this way an IIR (infinite impulse response) filter can be created with a transfer function of:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \cdots + a_p z^{-p}}. \tag{11}$$

The other difference is that this description does not contain noise, so the output can only be approximated with an error. The coefficients have to be determined so that the mean squared error is minimal:

$$\sum_{n=-\infty}^{\infty} \left[y_n - \left(-\sum_{i=1}^{p} a_i y_{n-i} \right) \right]^2 = \sum_{n=-\infty}^{\infty} \left(y_n + \sum_{i=1}^{p} a_i y_{n-i} \right)^2 . \tag{12}$$

The error is minimal if we solve the following system of equations:

$$\forall j \in [1, k], \ \sum_{i=1}^{k} a_i R_{|j-i|}, = 0 \tag{13}$$

where R is the autocorrelation function (*Collomb,* 2009). This system of equations can be solved using the Levinson-Durbin algorithm that computes the coefficients recursively.

To determine the coefficients of the autoregressive filter, a training set is needed to obtain the correlation coefficients. Furthermore, the order of the model is also to be determined. If the model is high-order, old data distort the output, while if too few number of coefficients are used, important details might not be taken into consideration. Analyzing the data, the model order was chosen so that data from the last 24 hours is used. Considering the 10 minutes resolution, this means that 144 coefficients were to be calculated. The training dataset was the wind speed measurements from April 2011.

3. Results and discussion

3.1. Statistical correction of wind power forecast

3.1.1. Simple methods

Test period was chosen from March 12, 2012 to July 31, 2012 to validate the simple lead time BIAS, STT, STB, and COM corrections. As it can be seen in *Fig. 8*, none of the methods have been able to improve the forecast at a given day significantly. In one case one method was more effective, while in another case the different one. For longer period, all kind of methods resulted in similar statistics (*MAPE, RMSE, BIAS*).

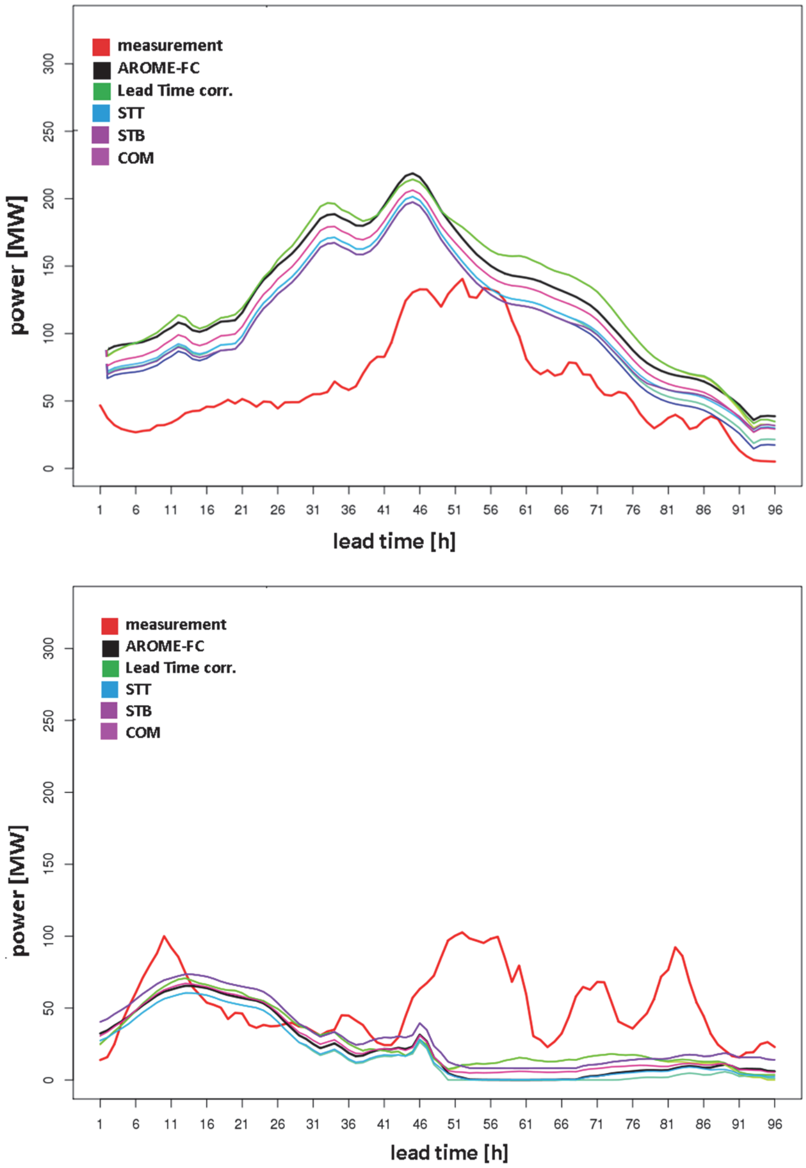

Fig. 8. Wind power measurement, forecast and BIAS corrected forecasts for some selected days.

3.1.2. Ensemble based methods

QR and AnEns were applied for the period of July 2012 – August 2013 based on AROME deterministic forecast. During the verification, BIAS and RMSE have been produced. This allows us to compare the results with those obtained by the higher resolution deterministic model, AROME, with the previously mentioned processes (QR and AnEns). These scores are shown for the EPS mean, median, and for AROME, as well as for the standard deviation of EPS members. In order to show, how good a probabilistic forecast is, several scores can be examined, some of them are (*Alessandrini et al.*, 2014):

- reliability diagram,
- rank histogram,
- spread Skill Plot,
- continuous ranked probability score (CRPS).

In the figures of BIAS (*Fig. 9*), basically three things can be observed:

1) the AROME's forecast is clearly better than the others,
2) the BIAS gets slightly larger when including leadtime as predictor,
3) the BIAS of AnEns's is concurrent with the BIAS of AROME (especially that of the median).

Of course that has to be set against RMSE, which may be seen in *Fig. 10*. We can state that:

1) There is no concurrence the AROME's forecast. The RMSE of AROME is about half of the others.
2) The RMSE of QR is slightly better, when wind speed and leadtime are used as predictors.
3) The correlation between the RMSE and the standard deviation of the eps members is high.
4) The RMSE of QR and AnEns medians are much worse than QR and AnEns means.

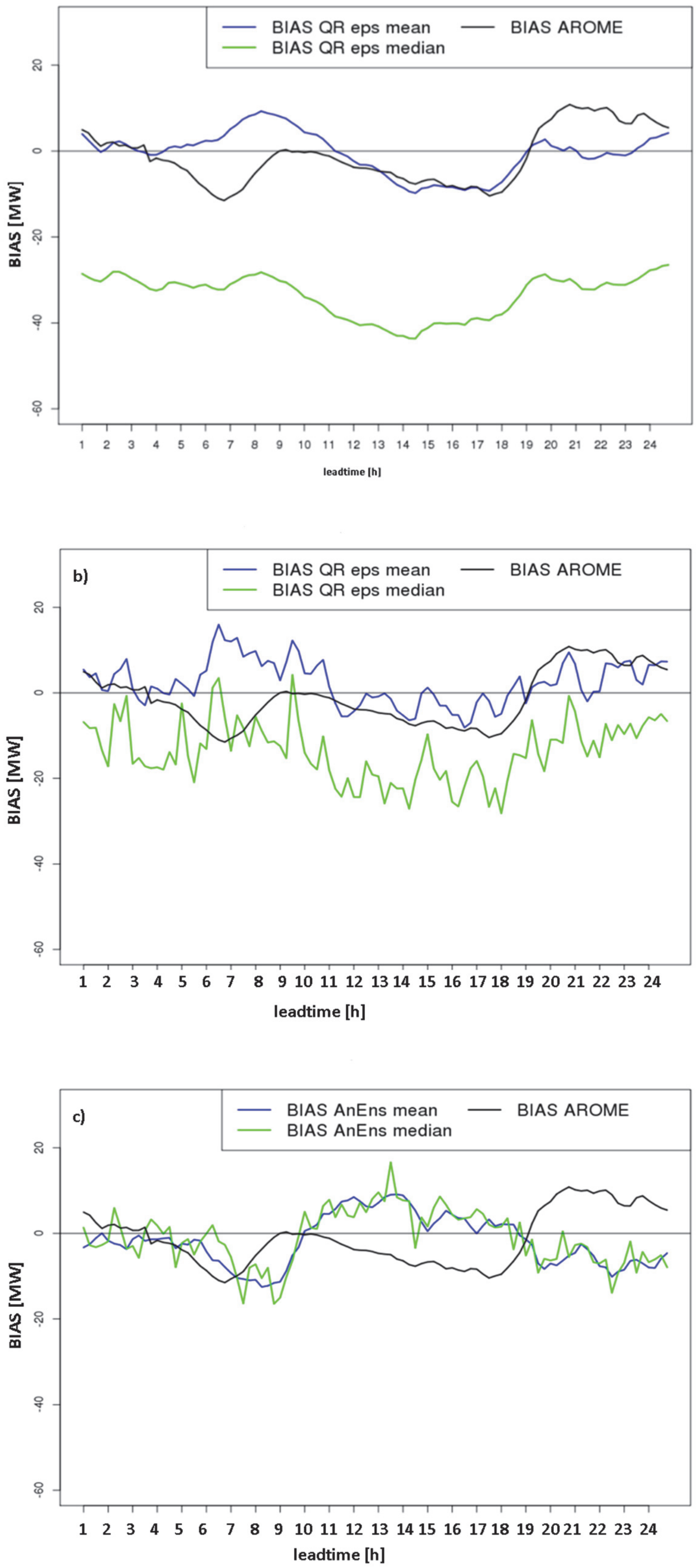

Fig. 9. a) The BIAS of the mean and median of the eps obtained by quantile regression with wind speed as predictor, b) the BIAS of the mean and median of the eps obtained by quantile regression with wind speed and leadtime as predictors, c) the BIAS of the mean and median of the eps obtained by analog ensemble and AROME as the function of leadtime.

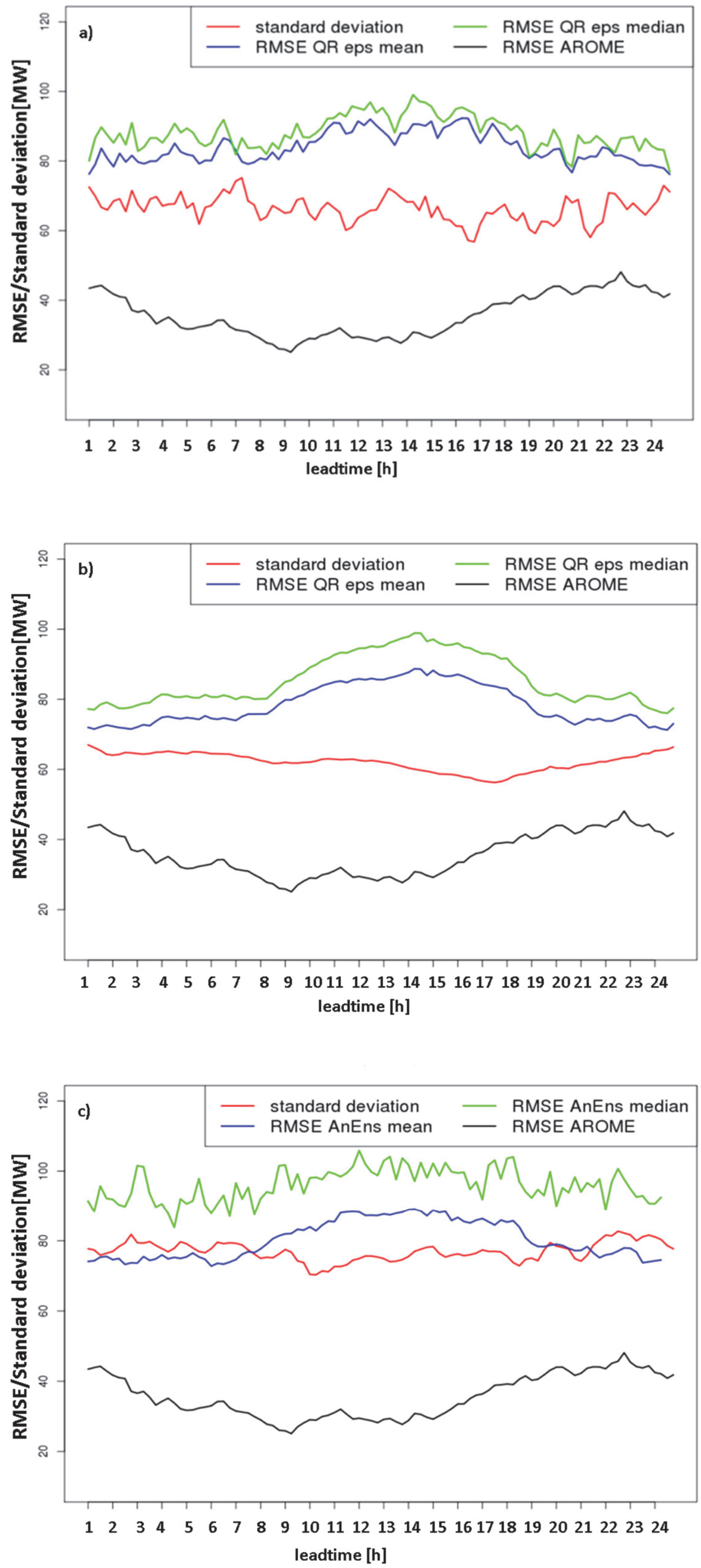

Fig. 10. a) The RMSE of the mean and median of the eps obtained by quantile regression with wind speed as predictor, b) the RMSE of the mean and median of the eps obtained by quantile regression with wind speed and leadtime as predictors, c) the RMSE of the mean and median of the eps obtained by analog ensemble and AROME as the function of leadtime.

In order to get more information about the probabilistic forecast, a number of indicators can be examined, e.g., the rank histogram (*Fig. 11*). This is a diagnostic tool to evaluate the spread of ensemble. The assumption is that all members are distributed with the same probability of occurrence of the observations within each bin. Usually the rank histograms show U-shape, so the spread of the members are not big enough, as it happens on the rank histogram of QR, when only the wind speed was used as predictor (*Fig. 11. a*). Using AnEns gives the best result, because it shows a little less equality between the bins (*Fig. 11. c*).

3.2. Corrections of a single wind farm data

3.2.1. Fitting the P-v curve with fuzzy model

The fuzzy model was trained in order to determine the relationship between the wind speed and the generated power. Measured data were available from May 2010 to June 2011. The first training set was a two-month-long data set (May and June 2010).

Fitting the function on all data points would obviously distort the result significantly. However, two simple rules can filter out outliers marked in *Fig. 12*. The first rule prescribes that above the cut-in speed, output values lower than a given value are ignored. For the second rule, it can be observed that at about 5 m/s, zero output was measured many times. A possible reason for that may be the power plant maintenance. This rule filters out data points, at which lower than 3.5 kW power was measured at higher than 3 m/s wind speed. The other rule takes into consideration that beyond the nominal wind speed, the output should be roughly constant. The saturation performance is 619 kW. Thus, according to this rule, above 14 m/s power supply, values lower than 610 kW are outliers.

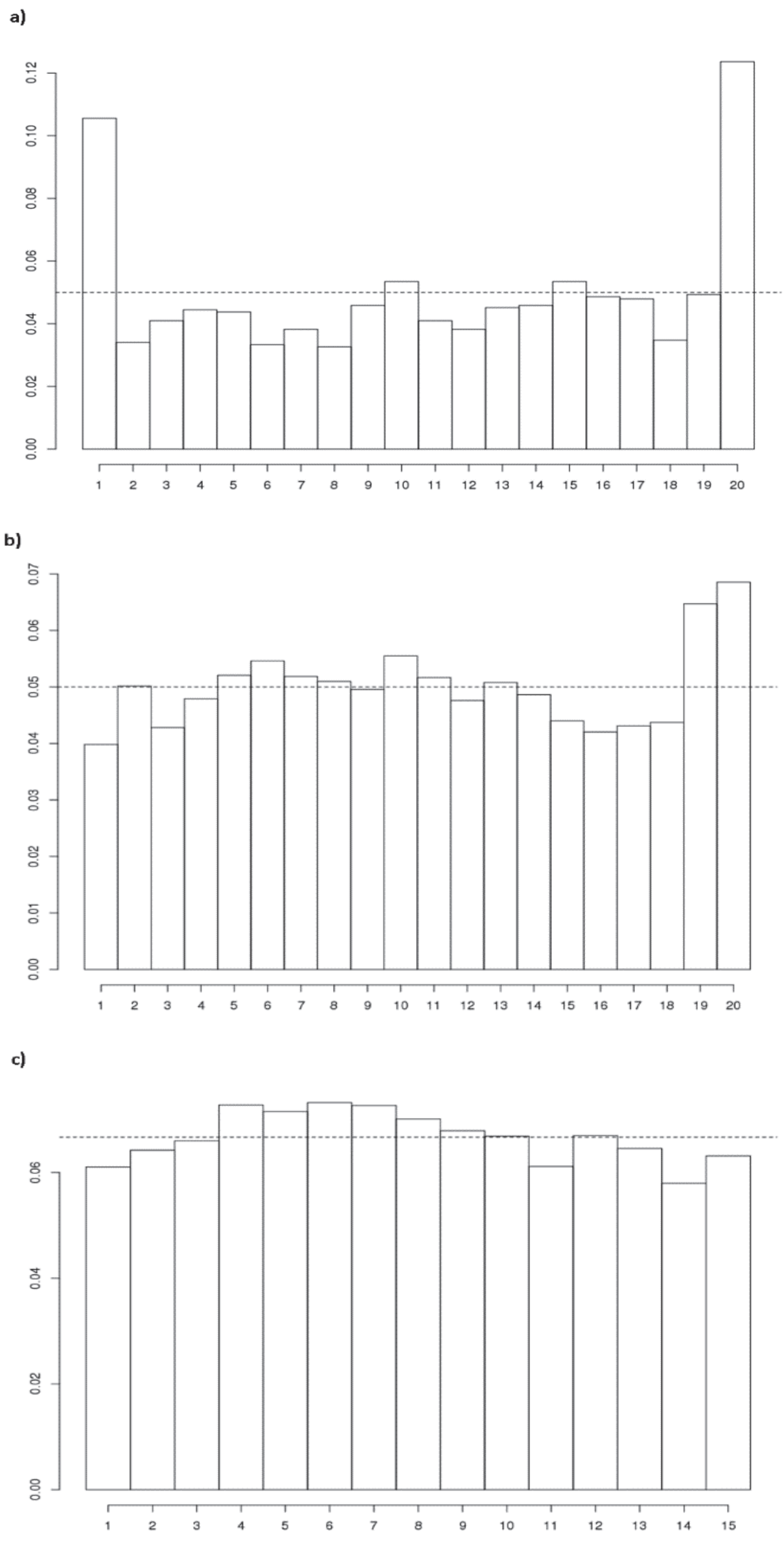

Fig. 11. Rank histograms of QR in case a) when wind speed and b) when wind speed and leadtime were used as predictors. c) Rank histogram of AnEns.

However, *Fig. 12* shows that these rules do not eliminate all of the outliers. In order to filter these points out, a preliminary fit is needed without data points filtered out previously. After that, data points that differ from the model output with more than 10% are to be eliminated. The resulting fitted curve now follows the desired sigmoid-shape.

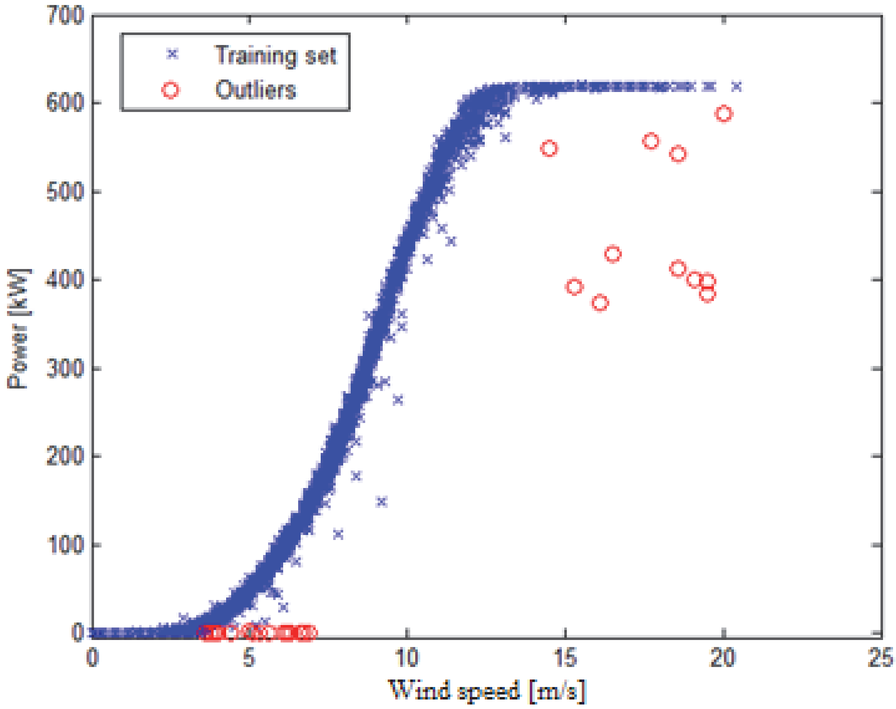

Fig. 12. Training set and the outliers (May and June, 2010).

This fit can be executed for data of different seasons. For both winter and summer, two months were selected as training set. It became evident that the resulting curves are significantly different (see *Fig. 13*). The difference at the two investigated wind speeds is about 5% in both cases. This is problematic, because characteristic curves are mostly considered as static. Furthermore, catalogue data differ from both curves, resulting that they cannot be used for precise forecasting.

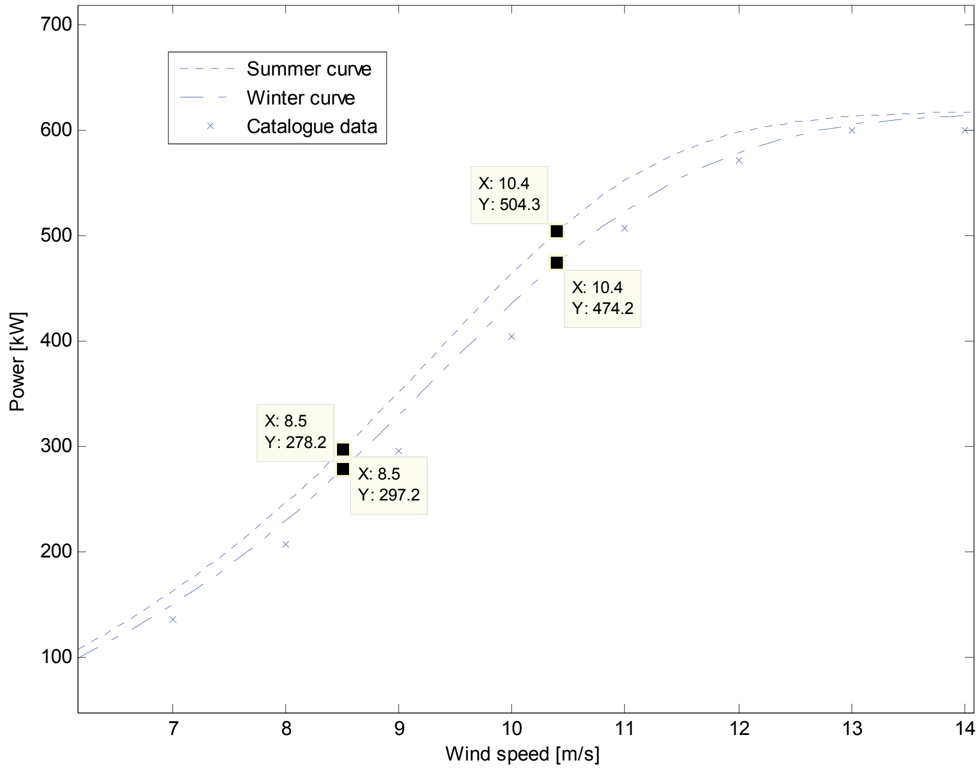

Fig. 13. Fitted curves in summer and winter (2010 and 2011), and catalogue data.

3.2.2. Consideration of air density

Curve fitting, using the normalized wind speed data of Section 2.2.2 showed that at low and high wind speeds, there was no significant difference to observe. Contrarily, at medium speeds, the difference is about the 5% of Section 3.2.1. It is obvious again that catalogue data cannot be used for precise forecasting. The difference between the two unnormalized curve is 108.5 kW^2, while for normalized curves this value is only 21.6 kW^2 (*Fig. 14*), which means that the mean squared erroro (MSE) is reduced to the fifth of its original value, where MSE is:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(P_{summer,i} - P_{winter,i}\right)^2. \qquad (14)$$

According to these results, it seems that one of the error sources that cause the difference between the curves is the air density.

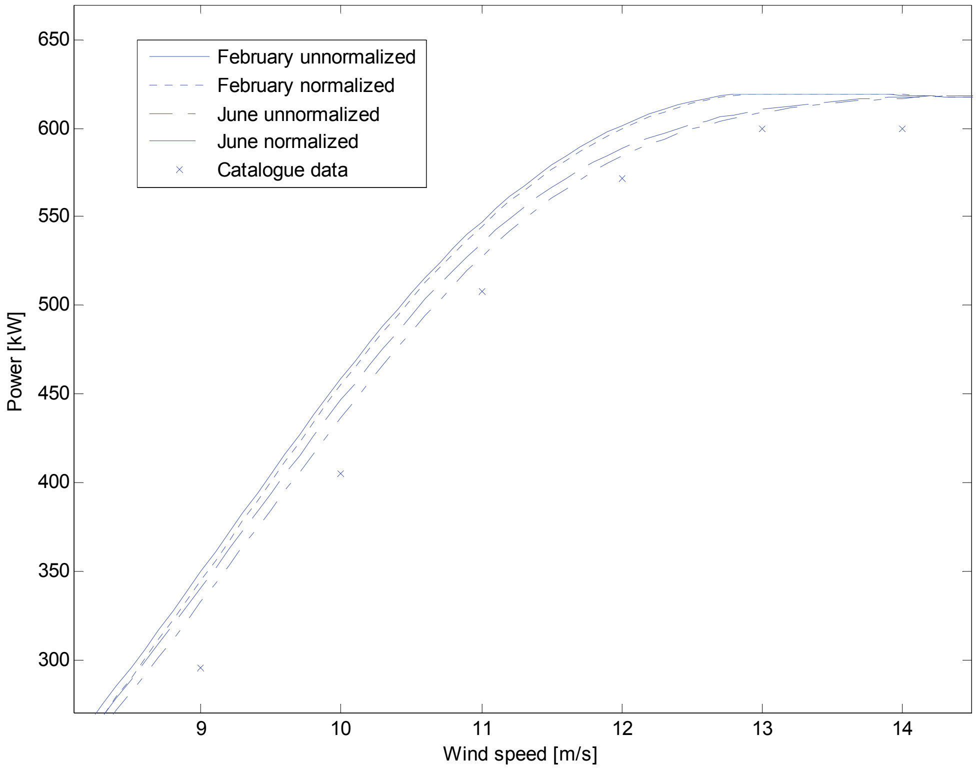

Fig. 14. Fitted power curves and catalogue data.

3.2.3. Autoregressive filtering of wind speed forecast

The IIR-filtering on the data given by the OMSZ was performed with the coefficients calculated with the method described in Section 2.2.3. *Table 2* contains the results, showing that the MSE of both wind speed and wind power forecasts have been improved by 3–4%. This means that if the filtered data were used as forecast instead of the 'raw' meteorological data, these improvements could have been achieved.

Table 1. MSE-improvements using autoregressive filtering

MSE-improvement	June, 2010	February, 2011	April, 2011
Wind speed	3.732%	3.958%	3.015%
Wind power	3.891%	4.387%	3.007%

4. Summary and conclusions

The professionals at MAVIR are responsible for the smooth energy supply of the whole country. Daily plans of the available energies are created for every day with 15 minutes time resolution by the system operator. One segment of this system is the prediction of the generated wind power. The task of meteorologists is to provide the most accurate forecast for the operators and the wind farm owners. The aim of this work was to improve the wind power forecasts.

In this paper, two types of approach were discussed. In the first part, statistical correction of the summarized wind power forecast was shown, which is valid for the whole country. Simple BIAS corrections and more comprehensive ensemble-based methods were applied without real access. None of the methods were able to improve the forecast significantly. Some improvements might be possible in a given day, but the absolute errors became worse for the whole period.

In the second part, we concentrated on the correction of the wind power forecast of a single farm. Investigations showed that the forecasts of the OMSZ follow the real production data well. It was also an important result that the air density should also be taken into account for the calculation of wind power. Namely, this is the parameter that may cause the difference between the characteristic curves in different months.

Another result of this work was that autoregressive filtering may also increase the precision of power forecasting. Using this method, several percent of improvement was achieved, which is significant considering the 5% expectation of the system operator.

Here it was not shown, but it was proven that using meteorological forecasts, much more precise scheduling can be made than by using predictions without physical considerations (data of the previous year, on average production). This would be advantageous for both the system operator and the turbine operators.

This work confirms that the wind energy forecast can be improved in two ways. The first one is to develop the meteorological forecast, so it is very important to create more complex and finer resolution (in horizontal and vertical) NWP models. The second one is using statistical corrections for the individual wind turbines and wind farms, and the sum of corrected values are able to produce more accurate wind power for the operators. This second step seems not realistic, because it is almost impossible to get measurements from all the 172 wind farms, which are necessary for the statistical corrections.

Acknowledgements: QR and AnEns were done in the framework of COST Action ES1002 WIRE. The authors would like to thank to the COST STSM host, *Stefano Alessandrini*, for his advices and scripts, to *Simone Sperati* (RSE Spa.) for his loads of help. The 'Improvement of the wind power forecast of a single wind farm' part of the paper is based on MSc thesis 'Study of Wind Turbine Production Forecasting Methods', which was defended at the Budapest University of Technology and Economics in 2013. The authors would like to thank *Professor István Kollár, Dr. Bálint Hartmann,* and *Roger Randriamampianina* for their help in writing the thesis. Remarks of the anonymous reviewer which lead to an improvement of the manuscript are greatly acknowledged.

References

European Commission, 2016: Proposal for a directive of the European Parliament and of the Council on the promotion of the use of energy from renewable sources (recast). ec.europa.eu/energy/sites/ener/files/documents/1_en_act_part1_v7_1.pdf (accessed on 8 March 2016)

Eurostat, 2016: ec.europa.eu/eurostat/statistics-explained/index.php/Renewable_energy_statistics (accessed on 8 March 2016)

GWEC, 2016: Global wind report. Annual market update 2015. www.gwec.net/wp-content/uploads/vip/GWEC-Global-Wind-2015-Report_April-2016_22_04.pdf_(accessed on 8 March 2016)

IEC 61400-12-1, 2005: Wind turbines – Part 12-1: Power performance measurements of electricity producing wind turbines.

MEKH, 2015: Beszámoló a kötelező átvételi rendszer 2015. évi alakulásáról. www.mekh.hu/download/1/6a/20000/kotelezo_atvetel_beszamolo_2015.pdf (accessed on 8 March 2016) (In Hungarian)

Alessandrini, S., Monache, L.D., Sperati, S., and *Niesen, J.N.,* 2014: A novel application of an analog ensemble for short-term wind power forecasting. *Renew. Energy 76*, pp: 768–781.

Barbounis ,T. G. and *J.B. Theocharis, J. B.,* 2006.: Long-term wind speed and power forecasting using local recurrent neural network models, IEEE Transactions on Energy Conversion, 21, 273–284.

Betz, A. 1966: Introduction to the Theory of Flow Machines. (D. G. Randall, Trans.) Oxford: Pergamon Press.

Chiu, S.L., 1994: Fuzzy model identification based on cluster estimation*. Intelligent Fuzzy Syst.* 2, 267–278.

Collomb, C., 2009: Linear prediction and Levinson-Durbin Algorithm, www.emptyloop.com/technotes/a%20tutorial%20on%20linear%20prediction%20and%20levinson-durbin.pdf (accessed on 8 March 2016)

Farkas, Z., 2011: Considering Air Density in Wind Power Production, *Study*, Eötvös University arxiv.org/pdf/1103.2198v1.pdf (accessed on 8 March 2016)

Hartmann, B., 2012: Szélerőművek integrálása a villamosenergia-rendszerbe, PhD Thesis Budapest University of Technology and Economics, pp. 110. repozitorium.omikk.bme.hu/bitstream /handle/10890/1198/tezis_eng.pdf?sequence=3&isAllowed=y (accessed on 8 March 2016) (in Hungarian)

Hunyár, M., Schmidt, I., Veszprémi K., and *Vincze Gy.,* 2001: A megújuló és környezetbarát energetika villamos gépei és szabályozásuk, Műegyetemi Kiadó, Budapest, 2001

Kusiak, A., Zheng, H., and *Song, Z.,* 2009: Wind farm power prediction: a data-mining approach. *Wind Energ. 12*, 275–293.

Llombart, A., Watson, S.J., Llombart, D., J.M. Fandos, J.M., 2005: Power Curve Characterization I: improving the bin method. Proceedings of the International Conference of Renewable Energies and Power Quality (ICREPQ).

Delle Monache, L., Eckel, F.A., Rife, D.L., Nagarajan, B. and *Searight, K.,* 2012: Probabilistic Weather Prediction with an Analog Ensemble. *Mont.Weather Rev. 141,* 3498–3516.

Rizwan, M., Saini, S., and *Singh, U.,* 2012: Prediction of wind energy using intelligent approach. *Power* Electronics (IICPE), 2012 IEEE 5th India International Conference on. IEEE

Sugeno, M. and *Tanaka, H.,* 1992: Successive identification of a fuzzy model and its applications to prediction of a complex system, *Fuzzy Sets Syst. 42*, 315–334.

Sweeney, C. and *Lynch, P.,* 2011: Adaptive post-processing of short-term wind forecasts for energy applications. *Wind Energ. 14*, 317–325.

Szintai, B., Szűcs, M., Randriamampianina, R., and *Kullmann, L.*, 2015: Application of the AROME non-hydrostatic model at the Hungarian Meteorological Service: physical parameterizations and ensemble forecasting. *Időjárás 119*, 241–265.