TÉCNICO LISBOA

Development and implementation of a methodology for the optimization of wind power forecasting

Agata Mucha

Thesis to obtain the Master of Science Degree in

Energy Engineering and Management

Supervisors: Prof. António José Nunes de Almeida Sarmento

Dr. José Joaquim Marques Bação Cândido

Examination Committee

Chairperson: Prof. José Alberto Caiado Falcão de Campos

Supervisor: Prof. António José Nunes de Almeida Sarmento

Member of the Committee: Eng. Santiago Rubin Fernández

June 2018

Acknowledgements

I would first like to thank my thesis supervisors José Cândido, PhD and Prof. António José Nunes de Almeida Sarmento from WavEC for their valuable comments on this thesis. They allowed this paper to be my own work, but steered me in the right direction when needed.

I would also like to acknowledge Santiago Rubin Fernandez, Energy Forecasting Manager at EDP Renewables for all the information and data provided.

Further, I would like to convey my special thanks to InnoEnergy MSc SELECT. This thesis is the last element of the amazing journey that this program offered me.

Finally, I must express my very profound gratitude to my parents, sisters and to my partner for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Abstract

In this work, a methodology for the optimization of wind power forecasting is developed and implemented. The main idea is to combine three Wind Power Forecasting (WPF) models in order to minimize the forecast error. The available data include historical potential power and three WPF models' predictions from the case study Pestera wind farm in Romania, as well as relevant meteorological parameters from the same location. Two weather parameters that are most correlated with the WPF errors - wind speed and temperature - are used in weather class definition. Various approaches to combination weights are taken and later compared – the use of one vs two weather parameters, randomized vs chronological data partition. A performance comparison shows that a model based on the weighted combination method with the use of two weather parameters and the randomized data partition outperforms all individual WPF models as well as other combination approaches. The overall Normalized Mean Absolute Error decreased by over one percentage point, which stands for 10,1% improvement for the individually best performing model and 16,3% improvement in case of the weakest model. In monetary terms, WPF improvements cause a decrease in mispredictions in day-ahead energy markets and in the case of the Pestera wind fam, with the installed capacity of 90MW, allowed for imbalances penalties and opportunity cost savings of over € 57 000 in less than two months.

Keywords: wind power forecasting, combined forecasting, correlation of weather parameters and forecast errors

Neste trabalho, uma metodologia para a otimização da previsão de energia eólica é desenvolvida e implementada. A ideia principal é combinar três modelos de previsão de energia eólica (WPF) para minimizar o erro de previsão. Os dados disponíveis incluem a potência potencial histórica e três previsões dos modelos de WPF do estudo de caso do parque eólico Pestera na Roménia, bem como parâmetros meteorológicos relevantes do mesmo local. O dois parâmetros meteorológicos mais correlacionados com os erros de WPF - velocidade do vento e temperatura - são usados na definição das classes climáticas. Várias abordagens para pesos de combinação são tomadas e posteriormente comparadas - o uso de um vs dois parâmetros meteorológicos, partição de dados aleatória vs cronológica. Uma comparação de desempenho mostra que um modelo baseado no método de combinação ponderada com o uso de dois parâmetros meteorológicos e a partição de dados aleatória supera qualquer dos modelos individuais de WPF, bem como outras abordagens combinadas. O erro absoluto médio normalizado global diminuiu em mais de um ponto percentual, o que representa 10,1% de melhoria em relação ao modelo com melhor desempenho individual e 16,3% de melhoria no caso do modelo mais fraco. Em termos monetários, os melhoramentos de WPF refletem-se numa diminuição das estimativas incorrectas nos mercados de energia do dia seguinte e, no caso do parque de Pestera, com uma capacidade instalada de 90MW, permitiram uma redução das penalizações e dos respectivos custos de oportunidade de mais de € 57.000 num período inferior a dois meses.

Palavras-chave: previsão de produção de energia eólica, previsão combinada, correlação de parâmetros meteorológicos e erros de previsão

Table of content

List of tables

List of figures

Abbreviations

ANLG	Analog-based model
ANN	Artificial Neutral Networks
CENER	Centro Nacional de Energías Renovables
CWPF	Combined Wind Power Forecast
ECMWF	European Centre for Medium-Range Weather Forecasts
EDPR	Energias de Portugal Renewables
EU	European Union
HIRLAM	High Resolution Limited Area Model
IEA	International Energy Agency
KDE	Kernel Density Estimation model
MM5	Fifth-Generation Penn State/NCAR Mesoscale Model
NMAE	Normalized Mean Absolute Error
NWP	Numerical Weather Prediction
OLS	Ordinary Least Squares
SCADA	Supervision Control And Data Acquisition
USA	United States of America
WAsP	Wind Atlas Analysis and Application Program
WPF	Wind Power Forecast

1. Introduction

The following chapter provides the background of the thesis, as well as its scope and outline.

1.1. Background

Nowadays, all around the world, the power sector is going through a significant change, aiming at achieving a clean, sustainable and reliable electricity market. Power system transformation is a complex process that requires a lot of actions at multiple levels: creation of local, national and global policy, establishment of operational and planning practices, major investments in innovation as well as the use of already existing efficient, smart and environment-friendly technology options. Renewable energy is already contributing to this transition to a great and growing extent. According to the Renewables 2017 Global Status Report[1], wind energy covered 10.4% of EU demand in 2016, proving to play an important role in power supply. The Global Wind Energy Council in their latest Outlook 2016[2] presents the advanced scenario, where global wind installations are foreseen to reach 5,806 GW by 2050.

Together with the continuous increase of wind power penetration rates, there comes the challenge of integration of this variable source into the electricity grid. Wind power brings many challenges to the electricity market, particularly utility operators, generating companies and regulators. The uncertainty and variability of wind hinders power generation scheduling and dispatch decisions as well as it affects trading performance on the electricity markets. Wind power forecasting (WPF) is therefore identified as an essential tool to deal with the progressively growing global wind power installations. WPF can be of great use in multiple applications: generation and transmission maintenance planning, determination of operating reserve requirements, unit commitment, economic dispatch, energy storage optimization (e.g., pumped hydro storage), and energy trading[3].

1.2. Scope of the study

Many attempts have been made to improve WPF and therefore decrease the issues that the uncertain and discontinue nature of wind brings. Usually, a number of alternative predictions is considered. A single forecast may not always provide satisfactory results, while a combined forecast using several individual ones might yield better results. Ideally, the combined final stage forecast ought to represent an improvement as compared to the individual one, or at least be equal to the best performing prediction[4]. Combination of several forecasts takes

---

[1] Global Wind Energy Council, "Global Wind Energy Outlook 2016 - Opening up New Markets for Business," 2016, http://files.gwec.net/files/GlobalWindEnergyOutlook2016.

[2] REN21, "RENEWABLES 2017 GLOBAL STATUS REPORT," 2017, http://www.ren21.net/wp-content/uploads/2017/06/17-8399_GSR_2017_Full_Report_0621_Opt.pdf.

[3] C Monteiro et al., "Wind Power Forecasting," *Information Sciences* 11, no. 4 (2009): 762–767, https://doi.org/10.1016/B978-0-8155-2047-4.10007-9.

[4] Ismael Sánchez, "Adaptive Combination of Forecasts with Application to Wind Energy," *International Journal of Forecasting* 24, no. 4 (October 2008): 679–93, https://doi.org/10.1016/j.ijforecast.2008.08.008.

advantage of the fact that each prediction model exhibits strengths and weaknesses in different situations[5]. The concept of originally proposed in 1969 by Bates and Granger[6], where they proved that a combined set of forecasts can generate lower mean-square error than either of the forecasts individually.

In the literature, there is a number of works on the combination of Numerical Weather Prediction (NWP) models. However, the combination of Wind Power Forecast (WPF) models is, to date, a somehow less popular practice. This thesis intends to obtain an improved WPF by combining several models according to the innovative methodology proposed in Chapter 4. The methodology was developed using data from and applied to a case study to demonstrate its potential for wider application. The purpose of the following work is to address and answer the following research questions:

- What are the impacts of different weather parameters on WPFs errors?

Although the wind power greatly depends on the wind speed, other meteorological parameters influence the wind farm production: air pressure, temperature, humidity, wind direction, etc.. In order to identify the parameters that mostly impact power production (so as to reduce the range of parameters to be analyzed in the subsequent steps), the correlation between the meteorological parameters measured at the case study wind farm and the actual energy output will be initially studied.

- What is the performance of various WPF models in different weather classes?

The approach is to evaluate the performance of the several WPF models available for the wind farm for different weather classes. These classes will be defined by ranges of the relevant meteorological parameters identified in an initial analysis stage.

- What is the optimal combination of WPF models for each weather class to minimize the absolute error between the combined forecasts and the measurement?

The definition of the combination weights for the different weather classes will reflect the differences in the accuracy of the models, by providing extra weight to the most accurate models in each weather class. The way to approach the weight combination will result from the observation of the forecast errors for the different weather classes.

1.3. Outline

The thesis first provides an overview of the literature review on the topic of wind power in power systems, the resulting challenges and the importance of accurate wind power predictions, as well as on the wind power forecasting and their combination methods. The following chapter introduces the available data used later in the

[5] Ceyda Er Koksoy et al., "Improved Wind Power Forecasting Using Combination Methods," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2015), 1142–47, https://doi.org/10.1109/ICMLA.2015.60.
[6] J. M. Bates and C. W. J. Granger, "The Combination of Forecasts," *The OR Society* 20, no. 4 (1969): 451, https://doi.org/10.2307/3008764.

methodology described in a subsequent section. Finally, the results of three different approaches to Combined Wind Power Forecasting development are presented and summarized. The last section discusses the outcomes, demonstrates practical implications, names the encountered limitations and suggests future research steps.

2. Literature review

This section provides a brief overview of the relevant engineering issues of the integration of wind power into power systems, a review of commercial and operational WPF tools as well as a summary of hitherto WPF combination methods.

2.1. Wind power in power system

The constantly increasing share of wind energy installations brings about engineering challenges related to the fluctuating nature of wind and its integration into the existing power systems. In the summary of the International Energy Agency (IEA) Wind collaboration[7], the author presents the possible impacts of wind power on power system reliability and efficiency, divided into various timescales and how wide the impacts stretch. The study described in the cited paper addresses the regional and system wide issues (marked with red circle in Figure 1), as opposed to the local issue of grid connection, like distribution efficiency or power quality. As depicted in Figure 1, the regional and system wide impacts fall under three main categories: grid, adequacy and balancing. Below, all three types of impacts are presented in more detail.

2.1.1. Grid

The influence of the wind power on transmission is contingent on the location of wind power plants comparative to the load, and the relation between wind power production and load consumption. The power flow in the network is affected by the wind power, which can alter its direction as well as increase or reduce power losses and bottleneck situations. In order to maintain the adequacy and security of the transmission, grid reinforcements might be required. The costs of such reinforcements of the grid are highly dependent on the location of the wind power plants, relative to the load and grid infrastructure. Current advancements in technology allow for the power plant design meeting industry expectations, i.e. riding through voltage dips, supplying reactive power to the system, controlling terminal voltage and participating in SCADA (supervision control and data acquisition) system operation with output and ramp rate control.

[7] Hannele Holttinen, "Estimating the Impacts of Wind Power on Power Systems—summary of IEA Wind Collaboration Hannele Holttinen," *Environ. Res. Lett* 3 (2008): 25001–6, https://doi.org/10.1088/1748-9326/3/2/025001.

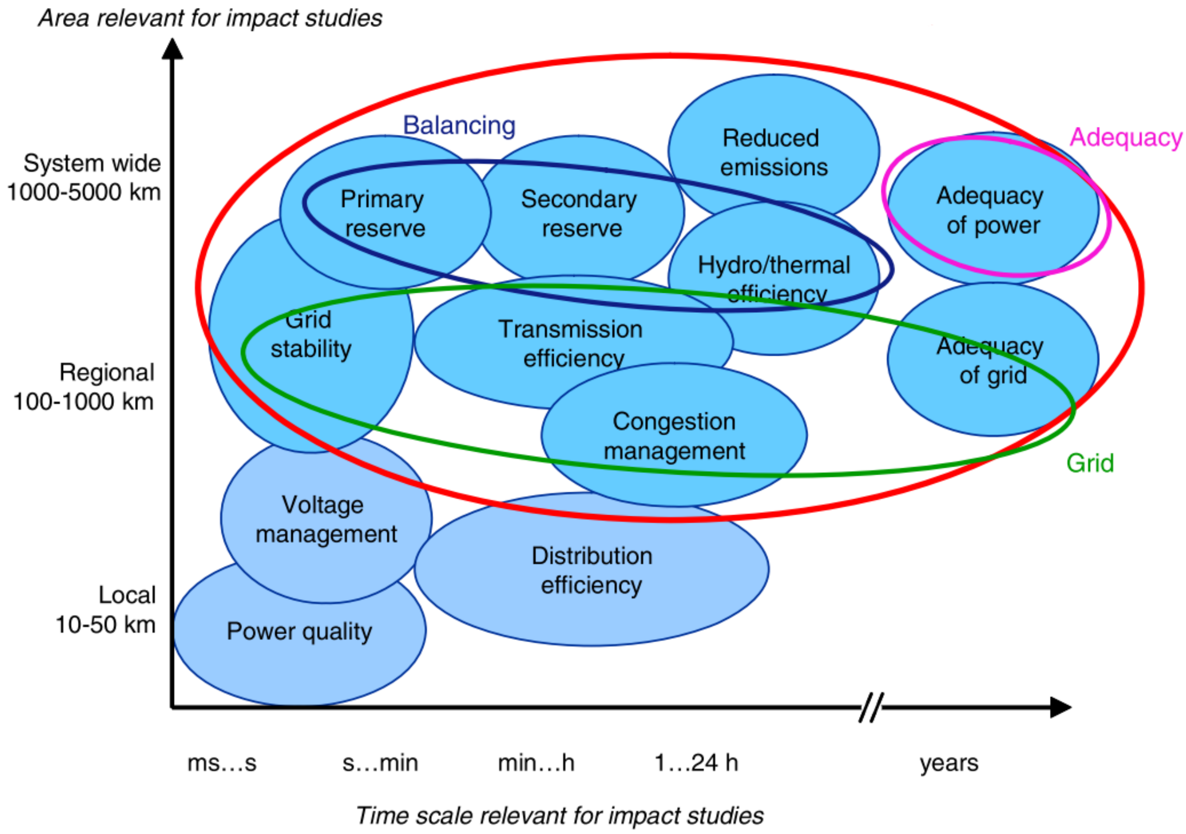

Figure 1 Impacts of wind power on power systems, divided into different timescales and width of area relevant for the studies[8]

2.1.2. Power system adequacy

The operators of the power system are obliged to maintain the system adequacy at a defined high level. Strictly speaking, the generation system has to be able to cover the peak demand, avoiding loss-of-load events, for a specified security of supply. The national regulations for the security of supply differ among the counties, ranging from 99 per cent security level, as in Germany, to 91 per cent, as in the United Kingdom[9].

The adequacy estimation is performed to assess the static conditions of the system, determining the total supply available during peak load situations in timescales of several years. The estimation takes into account the necessary generation capacity as well as the reliability data of various power sources. In the case of wind installations, the issue arises in proper assessment of wind power's aggregate capacity credit in the relevant peak load events. The geographical dispersion and interconnection play a major role in the estimations. According to the referenced study[10], wind power's contribution to the overall supply available in terms of a static condition of the system can be close to the average power produced by wind power at times of peak load. Conforming to the authors, this is true when the share of wind power is not high, and the capacity value of wind will decrease with

---

[8] Holttinen.
[9] Wind Energy The Facts, "Security of Supply and System Adequacy," accessed April 21, 2018, https://www.wind-energy-the-facts.org/security-of-supply-and-system-adequacy.html.
[10] Holttinen, "Estimating the Impacts of Wind Power on Power Systems—summary of IEA Wind Collaboration Hannele Holttinen."

the increase of wind power penetration. Yet, aggregation of large areas with wind power installations benefits the capacity credit of wind power.

2.1.3. Power system balancing

A real-time balance between electricity generation and consumption is essential for assuring system security. The biggest part of uncertainty lies in the unexpected failures of power plants or other elements of the system, possibly resulting in abrupt variations in power supply. According to the IEA report *The Power of Transformation - Wind, Sun and the Economics of Flexible Power Systems*[11], at shares above 2% to 3% in annual generation, wind power generation is also likely to cause an increase in supply-side variability and uncertainty.

According to the state-of-the-art report on studies carried out on various power systems[12], the increase in the balancing cost will be a function of the wind penetration level. However, as depicted in Figure 2, the impact of wind power's unpredictability and uncertainty varies for different countries/regions and power systems. The visible distinctions result from different methodologies of the studies in terms of timescales for prediction errors of wind power, costs for new reserve capacity investment as well as a size of balancing areas[13].

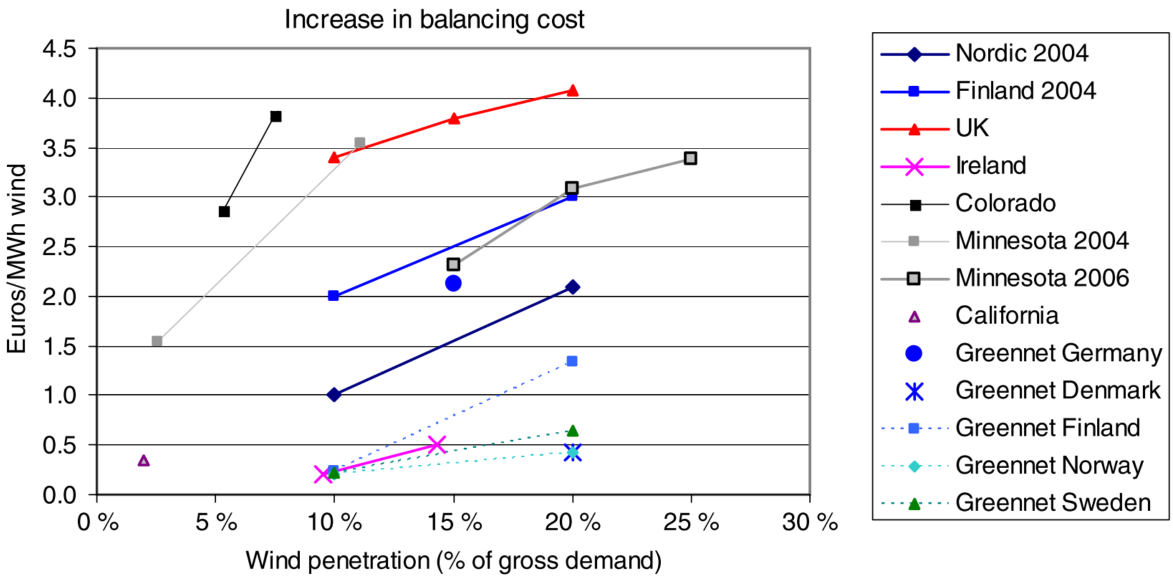

Figure 2 Results from estimates for the increase in balancing and operating costs due to wind power[14].

[11] International Energy Agency, "The Power of Transformation - Wind, Sun and the Economics of Flexible Power Systems," 2014, https://www.iea.org/publications/freepublications/publication/The_power_of_Transformation.pdf.
[12] Hannele Holttinen et al., "Design and Operation of Power Systems with Large Amounts of Wind Power State-of-the-Art Report," accessed April 21, 2018, http://www.vtt.fi/publications/index.jsp.
[13] Holttinen, "Estimating the Impacts of Wind Power on Power Systems—summary of IEA Wind Collaboration Hannele Holttinen."
[14] Holttinen et al., "Design and Operation of Power Systems with Large Amounts of Wind Power State-of-the-Art Report."

2.2. Wind Power Forecasting

As already pointed out, the primary barrier in wind power integration into the grid is its variability. Due to its dependence on the weather, wind power output can never be guaranteed at any time. Even though the addition of wind power into power systems bring benefits, such as reduction of emissions of electricity generation and reduction of operational costs due to the decreased amount of fuel used, the uncertain nature of wind biases utilities against using wind power. Accurate wind power forecasting, predicting power input from the wind farm into the grid, could lessen the network operating issues and therefore improve the perception of wind power[15].

Network operating issues are not the only challenges that precise wind power forecasting could influence in a highly positive manner. The unexpected fluctuations in wind power have also an impact on the price of wind-generated electricity. Even though some countries provide a higher, subsidized price to wind generators, the trend is shifting towards deregulated electricity markets. Under these conditions, the producers might be obliged to contract to provide firm power and later be penalized in case of over- or underproduction. Such situations negatively influence the value of the energy they sell, thus reducing the will to invest in wind energy. Accurate forecasts together with aggregation of power output of wind farms could increase the achieved price, resulting in increased feasibility of many sites and an improved attitude towards wind power[16].

2.2.1. Commercial and operational WPF tools

Research and development in the area of wind power forecasting has been focused on the creation of effective and reliable tools, and a number of approaches have been proposed. Mainly, the models take a physical or statistical approach, in some cases the combination of both[17]. The following subsection provides an overview of both approaches and presents examples of existing models.

Physical forecasting approach

The physical approach to wind power forecasting uses the detailed physical inputs in order to model the on-site conditions at the wind farm location. Figure 3 illustrates the basic operation of a model based on the physical approach.

The wind forecasts utilizing the physical approach take the output of the Numerical Weather Prediction (NWP) models, combine it with the on-site conditions and finally refine it with the use of downscaling methods. For the downscaling method, the detailed information about the wind farm is required, such as wind farm layout, wind turbine power curve, terrain orography, roughness, obstacles, etc.. The refined wind speed is then applied to the power curve to estimate the power production. Additionally, on-line data may be used to perform model output

15 Thomas Ackermann, *Wind Power in Power Systems*, 2005, https://simsee.org/simsee/curso2010/wind_power_in_power_systems.pdf.
16 Ackermann.
17 Jaesung Jung and Robert P Broadwater, "Current Status and Future Advances for Wind Speed and Power Forecasting," 2014, https://pdfs.semanticscholar.org/34e8/b8ceb3c68a7d0ee437b26eb9a757d619a814.pdf.

statistics to reduce the forecast error. In contrast to the statistical approach, the physical one does not need training input from historical data[18].

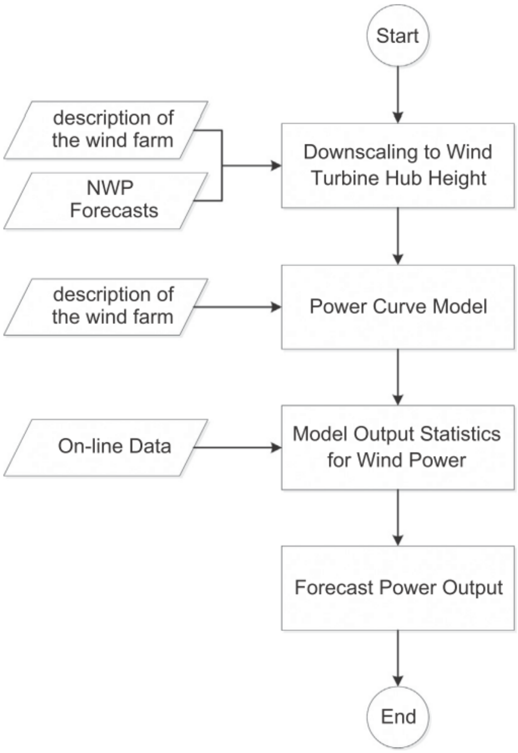

Figure 3 The physical approach to forecasting wind speed and power[19].

A great number of models basing on the physical approach has been implemented[20,21]. The *Prediktor* was developed by the Risoe National Laboratory in Denmark. It takes into account the local physical conditions in a form of NWP data from High Resolution Limited Area Model (HIRLAM) with the use of the Wind Atlas Analysis and Application Program (WAsP) as well as the PARK program. Further, the *Previento*, a model developed by the University of Oldenburg in Germany, with a similar physical approach as *Prediktor,* uses a different NWP forecast - from Lakelmodell of the German Weather Service. Then, the *LocalPred* was developed by CENER – National Renewable Energy Centre in Spain. It involves adaptive optimization of the NWP forecast, time series modelling, meso-scale modeling with the MM5 model[22]. and power curve modeling. Finally, the *eWind*, created by AWS TrueWind Inc. in the USA, has a similar physical approach to *Prediktor*, though it uses a high-resolution boundary layer model (ForeWind) as a numerical weather model to account for the local conditions.

---

[18] Jung and Broadwater.
[19] Jung and Broadwater.
[20] Junhui Huang, Case P. van Dam, and Henry Shiu, “Wind Energy Forecasting: A Review of State-of-the-Art and Recommendations for Better Forecasts. Final Report, Appendix B, California Renewable Energy Forecasting, Resource Data and Mapping,” 2010.
[21] Jung and Broadwater, “Current Status and Future Advances for Wind Speed and Power Forecasting.”
[22] National Center for Atmospheric Research Earth System Laboratory, “MM5 Community Model Homepage,” 2014, accessed May 11, 2018, http://www2.mmm.ucar.edu/mm5/overview.html.

Statistical forecasting approach

An alternative approach to the wind power forecasting is the statistical approach. Figure 4 illustrates the general form of the statistical model.

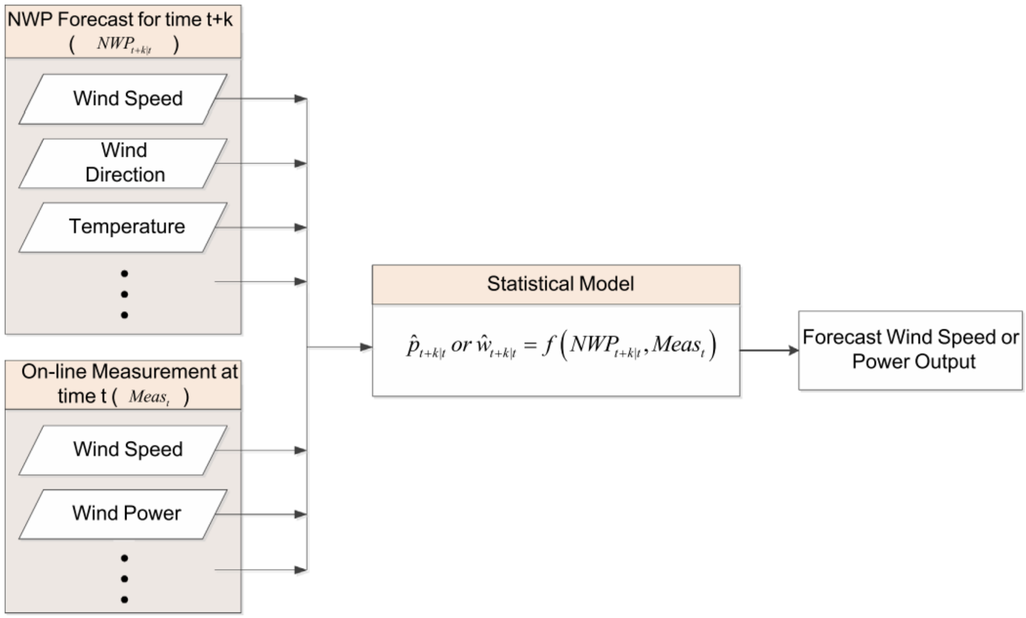

Figure 4 The statistical approach to forecasting wind speed and power[23].

In order to build the statistical model, the statistical approach uses historical data. This second approach uses NWP forecast for time *t + k* as well as on-line measurements at time *t* to predict the present values over the next few hours. In conventional statistical approach to wind power forecasting, a time series model is employed. As proposed by Box-Jenkins, in order to make a mathematical model, it is divided into four steps: model identification, model estimation, model diagnostics checking and forecasting. Although conventional statistical approach is relatively easy and inexpensive to build and use, it requires historical data to train the model[24].

There is a number of more advanced approaches, such as Artificial Neural Networks (ANN) approach, ANN-Fuzzy approach or different combinations of physical and statistical approaches. For more information on those methods, please refer to review papers by Jung and Broadwater[25] or Huang, Shiu and van Dam[26].

---

[23] Jung and Broadwater, "Current Status and Future Advances for Wind Speed and Power Forecasting."
[24] Jung and Broadwater.
[25] Jung and Broadwater.
[26] Huang, van Dam, and Shiu, "Wind Energy Forecasting: A Review of State-of-the-Art and Recommendations for Better Forecasts. Final Report, Appendix B, California Renewable Energy Forecasting, Resource Data and Mapping."

2.2.2. WPF combination methods

In the literature, a number of papers can be found on the improvement of WPF to eventually decrease the inherent error. Forecast errors of WPF models are at least partly uncorrelated, so that the combination of two forecasts can cut down overall errors even if one of the models is significantly poorer than the others. In fact, the combination of NWP with different meteorological forecasts resulted in improvements as high as 15%[27].

In the attempts to accurately predict the wind energy production, one has oftentimes access to a number of alternative meteorological forecasts or several available models for the transformation of meteorological forecasts into wind power forecasts. With the use of appropriate weights, the individual forecasts are multiplied and summed up, to end up with a final combined model, being a combination of NWP models. The idea was first proposed in 1969 by Bates and Granger[28], where they proved that a combined set of forecasts can generate lower mean-square error than either of the forecasts individually. From that point onwards, many methods have been suggested. In their *Review of Guidelines for the Use of Combined Forecasts*[29], Menezes, Bunn and Taylor selected seven well-established combining methods, being good representatives of varying degrees of sophistication. All the below-described methods adopt a linear formulation by which a vector, *f*, of *n* forecasts is linked through a transpose of a linear weighting vector, *w*, as $f_{combined}=w'f$.

1) *Simple average:* a simple and robust method, which has been repeatedly proven to outperform more complex combining schemes[30]. In the past, it has been a choice of many researchers, as outlined in Clemen's review[31].
2) *Outperformance:* in 1975, Bunn proposed an approach to forecast combination based on simplex of probabilities ($f_{combined}=p'f$, *p* being a simplex of probabilities), which can be evaluated in a Bayesian manner[32]. Each determined weight defines the probability that the respective forecast will perform the best (having the smaller absolute error) in the upcoming time.
3) *Optimal:* method introduced by previously mentioned Bates and Granger[33], where the linear weights are calculated in order to minimize the error variance of the combination. The vector of combining weights is calculated in accordance to the following equation:

$$w = \frac{S^{-1} \cdot e}{e' \cdot S^{-1} \cdot e} \tag{1}$$

[27] Henrik Aa. Nielsen et al., "Optimal Combination of Wind Power Forecasts," *Wind Energy* 10, no. 5 (September 2007): 471–82, https://doi.org/10.1002/we.237.
[28] Bates and Granger, "The Combination of Forecasts."
[29] Lilian M De Menezes, Derek W Bunn, and James W Taylor, "Review of Guidelines for the Use of Combined Forecasts," *European Journal of Operational Research* 120 (2000): 190–204, https://pdfs.semanticscholar.org/97c3/ea22e05eabda946695431e63d04bfbae5af4.pdf.
[30] Victor Richmond, R Jose, and Robert L Winkler, "Simple Robust Averages of Forecasts: Some Empirical Results," accessed April 22, 2018, https://doi.org/10.1016/j.ijforecast.2007.06.001.
[31] Robert T Clemen, "Combining Forecasts: A Review and Annotated," *International Journal of Forecasting* 5 (1989): 559–83, https://pdfs.semanticscholar.org/7117/9279738b91df0520061b351cb3e0124a411c.pdf.
[32] D. W. Bunn, "A Bayesian Approach to the Linear Combination of Forecasts," *Operational Research Quarterly (1970-1977)* 26, no. 2 (June 1975): 325, https://doi.org/10.2307/3008467.
[33] Bates and Granger, "The Combination of Forecasts."

where *e* is the *(n+1)* unit vector and *S* is the *(n x n)* covariance matrix of forecast errors. It was shown by Granger and Ramanathan[34] than the optimal method is equivalent to a least squares regression, with suppressed constants and weights constrained to sum to one.

4) *Optimal (adaptive) with independence assumption:* a method where the estimation of *S* in (1) is restricted to be diagonal, including only the individual forecast error variances.
5) *Optimal (adaptive) with restricted weights:* this method imposes an additional restriction on individual weights not to be outside the interval [0,1].
6) *Regression:* in this approach the forming forecasts are used as regressors in an ordinary least squares (OLS) regression with the inclusion of a constant.
7) *Regression with restricted weights:* the final method utilized the least squares regression with the inclusion of a constant and the restriction of sum of the weights equal to one.

The above-introduced methods present an overview of the existing extensive range of approaches, indicating the large number of attempts taken towards WPF improvements based on the combination of individual NWP models. For further details and implications of the methods please refer to the mentioned papers.

3. Available data

In order to develop and apply the methodology described below, data from the Pestera wind farm in Romania (44°12'36.0"N 28°03'00.0"E), were provided by EDP Renewables. The wind farm is composed of 30 operational units with the cumulative installed capacity of 90MW. Apart from the historical potential power from the farm (assuming 100% availability), data from the following three WPF models is used:

- Kernel Density Estimation model 1 (KDE150)
- Kernel Density Estimation model 2 (KDE250)
- Analog-based model (ANLG50)

The two KDE models are based on kernel density estimation, i.e. a non-parametric way to estimate the probability function of a random variable[35]. One of them uses 50 different variables from the European Centre for Medium-Range Weather Forecasts (ECMWF)[36], and the other one a selection of them. The analog model includes identifying past trajectories (i.e., analogs) similar to the current state (the initial condition), forecasting then the current state forward by assuming that the initial condition will develop forward according to a similar course as the identified past analogs[37].

---

[34] C.W.J. Granger and R. Ramanathan, "Improved Methods of Forecasting," *Journal of Forecasting 3*, 1984.
[35] Murray Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics* 27, no. 3 (September 1956): 832–37, https://doi.org/10.1214/aoms/1177728190.
[36] European Centre for Medium-Range Weather Forecasts ECMWF, "ECMWF | Advancing Global NWP through International Collaboration," accessed May 12, 2018, https://www.ecmwf.int/.
[37] Patrick L. McDermott & Christopher K. Wikle, "Analog Forecasting: A Flexible and Parsimonious Alternative for Nonlinear Prediction - Statistics Views," 2016, http://www.statisticsviews.com/details/feature/9652191/Analog-Forecasting-a-Flexible-and-Parsimonious-Alternative-for-Nonlinear-Predict.html.

In order to identify the parameters that mostly impact power production (so as to reduce the range of parameters to be analyzed in the subsequent steps), the meteoblue[38] global weather simulation archive is used. Historical weather simulation data, with a spatial resolution between 4 and 30 km , i.e. wind speed and direction, temperature, relative humidity as well as mean sea level pressure, are taken for the location of the Pestera wind farm (see Figure 5).

Figure 5 Distribution of wind turbines at the Pestera Wind Farm. Red location indicates the coordinates corresponding to the weather data.

Table 1 presents the available time periods of all the data types provided. Only a period where all the WPF models overlapped was taken for the analysis, i.e. from 12/04/2016 to 31/05/2017.

Table 1 Time periods of available data

Data type	**Starting from**	**Until**
Potential Power (from EDP)	01/01/2013	18/06/2017
EDPR_KDE250	12/04/2016	31/05/2017
EDPR_ANLG50	02/06/2015	31/05/2017
EDPR_KDE150	05/05/2015	31/05/2017
Historical weather simulation data	12/04/2016	31/05/2017

[38] Meteoblue, "Weather History+," accessed April 23, 2018, https://www.meteoblue.com/en/historyplus.

4. Methodology

In the following methodology, 85% of data will be used to set up the model and the remaining 15% to assess the quality of the combined WPFs. The proposed methodology is divided into three stages:

1) Analysis
2) Setup
3) Operation

Each step is described in detail below.

4.1. Analysis

WPFs error characterization

The first step of the analysis stage is the identification of the errors (i.e. deviations between forecasted and actual power generation) of the different WPF models for the case study wind farm. This includes the visual inspection of time-series (WPFs and actual wind power production) and scatter plots, as well as the computation of traditional error/evaluation parameters (normalized mean absolute error, correlation). This procedure aims at understanding the origin of the differences between forecasts and measurements, comparing the accuracy of the different WPF models.

Impact of different weather parameters on WPFs errors

The second step of the analysis stage consists in the identification of the most relevant weather parameters in terms of their impact on the WPFs' errors. For that purpose, the correlation between time series of weather parameters measured at the wind farm and time series of WPF errors are studied. The Pearson product-moment correlation coefficient, which provides a measure of the linear correlation between two variables, is used. This coefficient (ρ) is calculated as follows:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

where cov is the covariance of two variables and $\sigma_X \sigma_Y$ is the product of their standard deviations.

4.2. Setup

The proposed methodology is based on the conditional weighted combination of WPF models. This is a classical linear regression model for combined forecasting in which the time-varying combination weights are a function of relevant meteorological variables.

Weather classification

The approach is to evaluate the performance of the several WPF models available for the wind farm for different weather classes. These classes are defined by ranges of the relevant meteorological parameters identified in the analysis stage. The following aspects are taken into consideration when establishing the weather classes:

- Average behaviour/accuracy of the WPF models as a function of the meteorological parameters;
- Specific weather classification;
- Actual wind power generation of the farm as a function of the meteorological parameters.

The use of a single weather parameter will be compared to the use of the two most relevant weather parameters in the definition of weather classes.

Combination weights definition

It is assumed that the several WPF models perform differently for each weather class, i.e. some models may be more accurate for certain weather classes than others. The definition of the combination weights for the different weather classes will reflect these differences in the accuracy of the models, by providing extra weight to the most accurate models in each class. However, it may happen that some models consistently overestimate and others underestimate power production for given weather classes, thus suggesting that the combination of the two by appropriate weights provides a much more accurate estimate. A criterion to select the optimal combination of WPF models for each weather class is to minimize the least square error between the combined forecast and the measurements.

4.3. Operation

The operation (running and validation) of the Combined Wind Power Forecast (CWPF) model developed in the previous stages based on 85% of forecasted and measured data, uses as inputs:

- the remaining 15% of the WPF's results and meteorological parameters available
- the optimal combination of WPFs for each weather class, as identified in the model setup stage
- the actual power output of the wind farm

The outcome of this stage is a consensus, improved forecast of the power generation of the wind farm for the desired forecast horizon, as well as the resulting errors for the different weather classes and a conclusion about the suitability of the model.

5. Results

The subsequent chapter presents the outcomes of the methodology described in Chapter 4. It is divided in three parts, in accordance with the methodology's subchapters: analysis, setup and operation.

5.1. Analysis

In the first step of the analysis stage, the WPF errors were characterized. With the aim of visual inspection, the plot in Figure 6 illustrates a time-series of 48h of the actual potential power as well as the three WPF models. Then, Figure 7 and Figure 8 present scatter plots displaying the relation between the historical potential power (assuming 100% availability) and i) the considered weather parameters (i.e. temperature, relative humidity, mean see level pressure, wind speed and wind direction) (Figure 7) and ii) the models' forecasts (Figure 8).

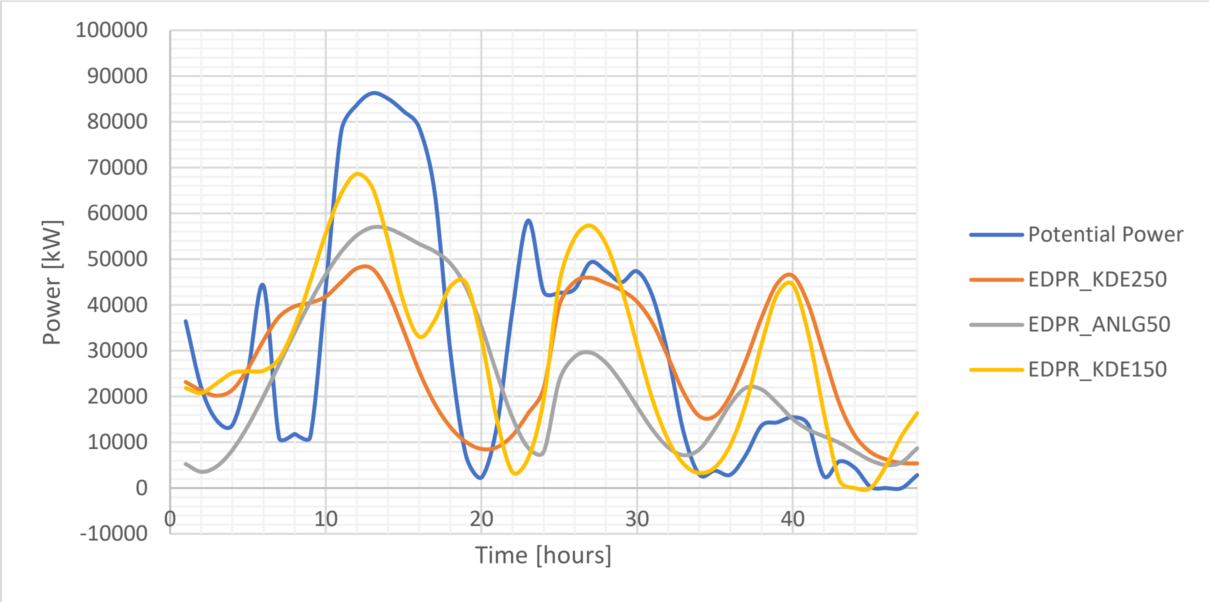

Figure 6 Time-series (48h) of Potential Power and the three EDPR WPF models

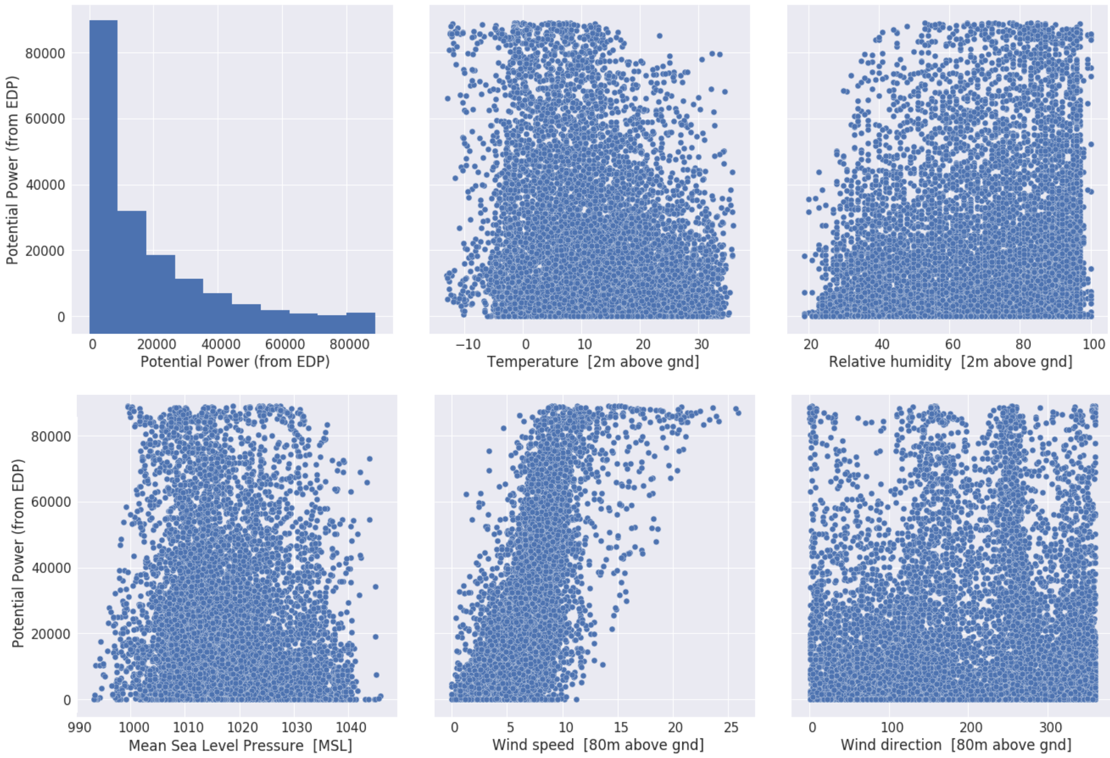

Figure 7 Scatter plots presenting relation between historical Potential Power and the weather parameters

Together with relation between the potential power and the models' forecast, Figure 8 also includes the regression line and its equation (*y=mx+b*), as well as the coefficient of determination R^2 (the square of the correlation coefficient *R*) for each case. The resulting value of R^2 indicates that the regression model accounts for around 63.0% of the variance, differing slightly between the models.

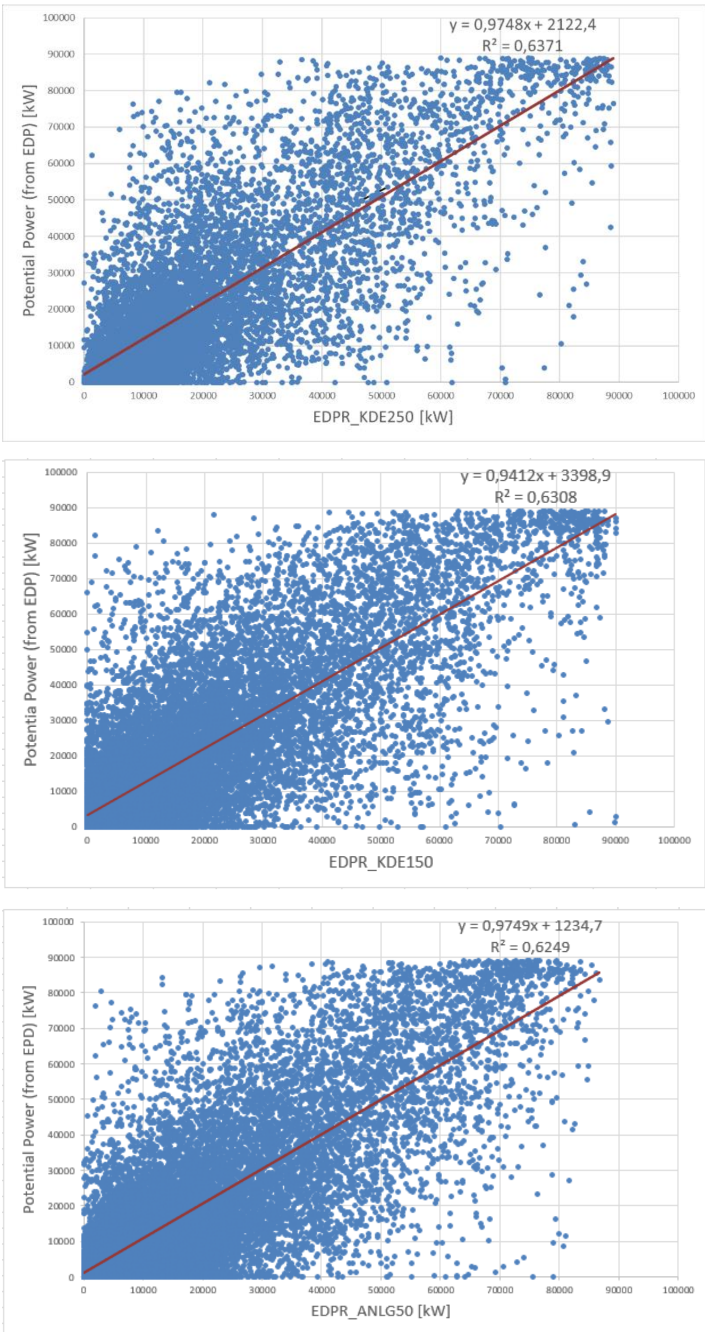

Figure 8 Scatter plots presenting relation between historical Potential Power and the three EDPR WPF models

In order to assess the accuracy of the available WPF models, the Normalized Mean Absolute Error was calculated. First, the MAE was determined with the use of the following formula:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (3)$$

As the name suggests, MAE is an average of the absolute errors, where y_i is the prediction, x_i is the true value and n is the number of readings. Then, MAE was normalized to the installed capacity of the wind farm. Table 2 summarizes the results of the NMAE calculations for each model separately.

Table 2 Normalized Mean Absolute Error of WPF models

WPF model	Normalized Mean Absolute Error [%]
NMAE_KDE250	10.504867
NMAE_KDE150	10.53881
NMAE_ANLG50	11.284863

From the results of the NMAE calculations, one can observe that all models exhibit similar errors of around 10%. Although the difference may not seem large and significant, in the case of wind power predictions, the consequences of mispredictions and resulting over- or under-productions can be costly. Therefore, even small decreases in WPF errors may bring visible positive outcomes.

The following step consisted in the calculation of the Pearson product-moment correlation coefficient according to equation (2). The correlation was calculated for the historical potential power, the WPF models forecasts as well as all considered weather parameters. The resulting matrix of correlations is presented in Figure 9.

The first thing to take into consideration is the correlation between the actual potential power and the WPF models. The KDE250 model, which has the lowest NMAE, has also the highest correlation with the potential power. Also, the models turn out to be rather highly correlated with each other. The reason for the high correlation of the KDE250 and KDE150 models is that both models are based on the Kernel Density Estimation method and use a great part of the same variables as input.

The results of the Pearson correlation calculations allow also for an identification of the parameters that mostly impact power production. As can be seen in Figure 9, the highest correlation between the actual potential power and the weather parameters is for the **wind speed** – equal to 0.71, followed by the **temperature**, corresponding to -0.27. A negative correlation indicates inverse proportionality. Additionally, a correlation between the WPF models' NMAE and the weather parameters was determined and is presented in Figure 10. In this case, also wind speed and temperature are the parameters that have the highest correlation, just as previously in the case of the correlation with the potential power and the individual WPF models forecasts. Figure 11 illustrates the correlation between the two most relevant weather parameters and the WPF errors. As a conclusion, for the further analysis, only wind speed and temperature will be considered.

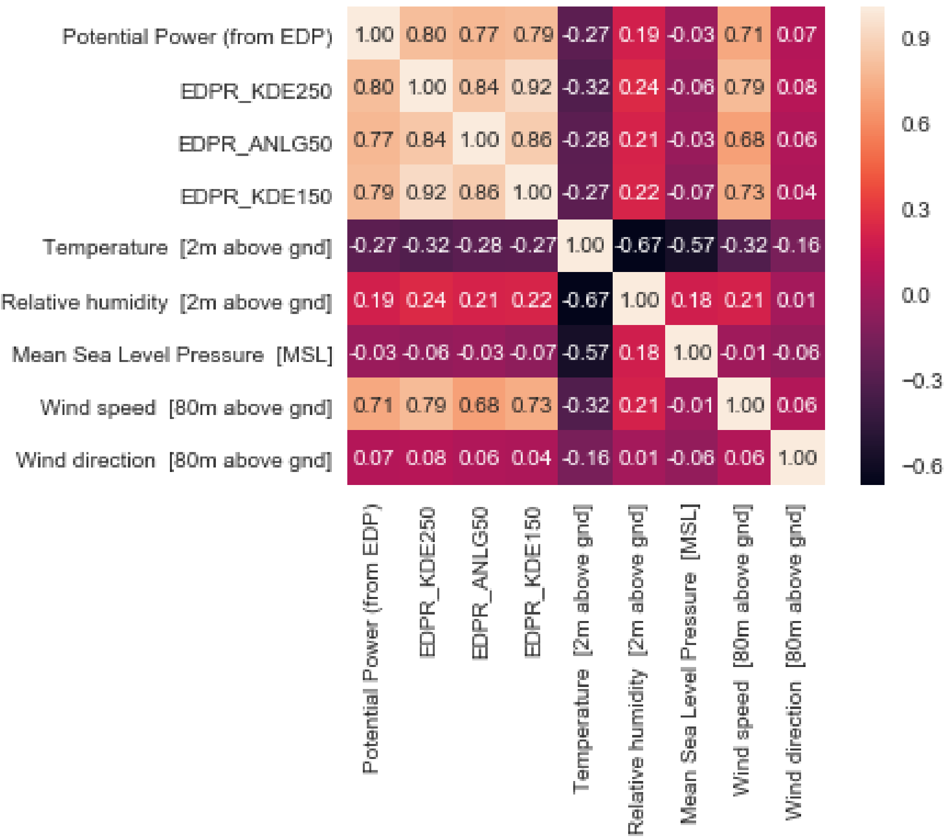

Figure 9 Matrix of Pearson product-moment correlations for potential power, individual forecasts and weather parameters

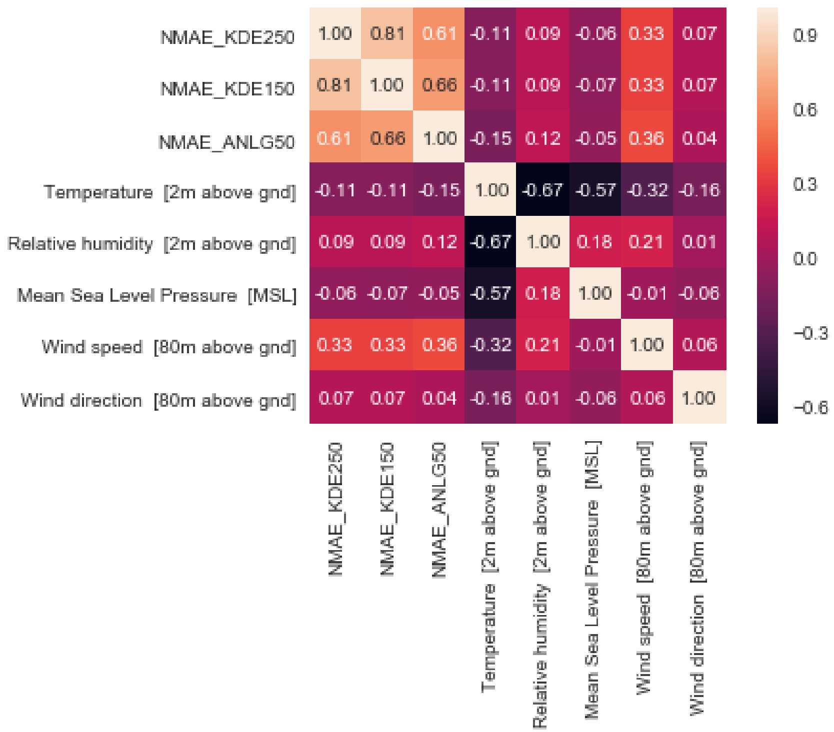

Figure 10 Matrix of Pearson product-moment correlations for WPF errors and weather parameters

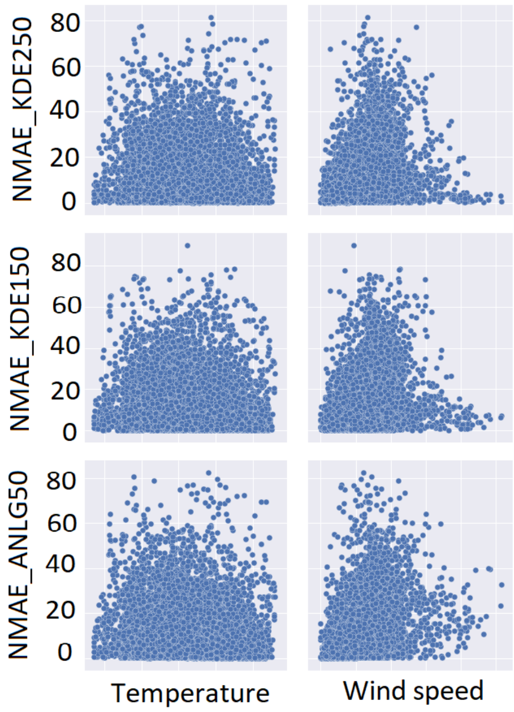

Figure 11 Scatter plots presenting the correlation between the WPF errors and the two chosen weather parameters: wind speed and temperature

5.2. Setup

As concluded in the analysis step of the followed methodology, in the subsequent setup stage, only two meteorological parameters - wind speed and temperature - were considered for the definition of different weather classes. In order to carry out a comparison between the use of one parameter and the use of two parameters, firstly only wind speed will be considered for the weather classes definition. Table 3 presents the 27 defined classes (divided evenly), together with the number of readings for both training (85% of data) and test (15% of data) models. In this case, the randomized partition of data was applied in order to arrive at a reasonable distribution of readings throughout all the weather classes.

Table 3 Weather classification for the case of a single parameter - wind speed

Weather class no.	Wind speed range [m/s]	No. of readings (train)	No. of readings (test)
0	0 to 1	154	18
1	1 to 2	409	46
2	2 to 3	684	76
3	3 to 4	870	97
4	4 to 5	1127	126
5	5 to 6	1240	138
6	6 to 7	1254	140
7	7 to 8	1135	127
8	8 to 9	819	91

9	9 to 10	472	53
10	10 to 11	228	26
11	11 to 12	144	17
12	12 to 13	63	8
13	13 to 14	34	4
14	14 to 15	36	4
15	15 to 16	25	3
16	16 to 17	20	3
17	17 to 18	16	2
18	18 to 19	9	2
19	19 to 20	12	2
20	20 to 21	14	2
21	21 to 22	7	1
22	22 to 23	2	1
23	23 to 24	2	1
24	24 to 25	0	0
25	25 to 26	1	1
26	26 to 27	0	0

It can be seen that most of the data are in the range between 3 and 8 m/s, therefore it can be expected that the model will be best-trained in these ranges. In the case of ranges where the number or readings is very small (for wind speed higher than 12 m/s), the model might not be sufficiently trained and may result in incorrect outcomes.

In the case of the use of two weather parameters, a matrix of wind speed and temperature ranges was defined. The choice of scopes of wind speed ranges was based on the number of readings available for particular values. E.g. for wind speeds lower than 4m/s or higher than 10m/s the amount of available data is smaller than for the range between 4 and 10m/s. In the case of the temperature, no such state has been identified, therefore the division is even. Again, the randomized partition of data was applied to obtain a sound distribution of readings throughout the weather classes. Concluding, the range distribution was adjusted to the available data readings and is presented in Table 4.

Table 4 Weather classification for the case of two parameters - wind speed and temperature

Weather class no.	**Wind speed range [m/s]**	**Temperature range [°C]**	**No. of readings (train)**	**No. of readings (test)**
1	0 to 4	-20 to 0	126	23
2	0 to 4	0 to 10	452	80
3	0 to 4	10 to 20	782	139
4	0 to 4	20 to 30	572	102
5	0 to 4	>30	97	18
6	4 to 6	-20 to 0	192	34

7	4 to 6	0 to 10	540	96
8	4 to 6	10 to 20	868	154
9	4 to 6	20 to 30	551	98
10	4 to 6	>30	133	24
11	6 to 8	-20 to 0	242	43
12	6 to 8	0 to 10	669	119
13	6 to 8	10 to 20	855	151
14	6 to 8	20 to 30	436	78
15	6 to 8	>30	90	16
16	8 to 10	-20 to 0	198	36
17	8 to 10	0 to 10	444	79
18	8 to 10	10 to 20	411	73
19	8 to 10	20 to 30	147	27
20	8 to 10	>30	31	6
21	10 to 14	-20 to 0	121	22
22	10 to 14	0 to 10	175	31
23	10 to 14	10 to 20	123	22
24	10 to 14	20 to 30	33	6
25	10 to 14	>30	0	0
26	14 to 20	-20 to 0	79	14
27	14 to 20	0 to 10	36	7
28	14 to 20	10 to 20	0	0
29	14 to 20	20 to 30	0	0
30	14 to 20	>30	0	0
31	>20	-20 to 0	28	6
32	>20	0 to 10	0	0
33	>20	10 to 20	0	0
34	>20	20 to 30	0	0
35	>20	>30	0	0

In order to arrive at the optimal combination of the available models, for each defined weather class (range), a linear regression was applied, minimizing the least square error between the forecast models and the historical potential power. For the purpose of model training, 85% of the available data was used, partitioned randomly.

Table 5 and Table 6 present the combination weights for each model in all the defined weather classes, for the use of one and two weather parameters, respectively.

Table 5 Combination weights for the three forecast models, for the case of a single weather parameter - wind speed

	Coefficients - combination weights		
Weather class no.	**EDPR_KDE250**	**EDPR_ANLG50**	**EDPR_KDE150**
0	0,44910363	0,3757933	-0,09471473
1	0,51649698	0,10991011	0,11639171
2	0,4957729	0,09430648	0,07307088
3	0,41374223	0,16098988	0,09152815
4	0,58089701	0,17130051	-0,06871616
5	0,19712213	0,23169844	0,21955051
6	0,21517291	0,42979076	0,16448834
7	0,11074775	0,35246816	0,39742383
8	0,25880625	0,42256584	0,21341866
9	0,40279594	0,43481312	0,14699079
10	0,29794714	0,41569215	0,22859398
11	0,24411227	0,60595313	0,13033295
12	0,7039078	-0,10522642	0,11650181
13	-0,57099588	0,38774845	0,68386296
14	-0,47076868	0,79854945	0,30529987
15	-0,35311967	0,68862602	0,67124733
16	-0,17317947	0,85194442	0,3244441
17	0,89479944	-0,05808955	-0,47984644
18	0,16688883	0,50015648	0,39905563
19	-2,67030711	0,51126328	0,49256932
20	-2,24152925	-0,38887268	0,9920203
21	0,50395969	0,12991208	-0,11992187
22	-0,00181814	0,01403662	0,00185257
23	0,00295562	-0,03567411	-0,00631467
24	-	-	-
25	0	0	0
26	-	-	-

Table 6 Combination weights for the three forecast models, for the case of two weather parameters - wind speed and temperature

	Coefficients – combination weights		
Weather class no.	**EDPR_KDE250**	**EDPR_ANLG50**	**EDPR_KDE150**
1	1,63125427	0,49241465	-0,98998371
2	0,56580834	0,01970049	0,03921237
3	0,58833195	0,17527112	0,03514658
4	0,31537379	0,08097527	0,15764513
5	0,02470215	0,83720558	0,15193093
6	0,73286087	0,51371715	-0,3405589

7	0,44632607	0,23523845	0,17342488
8	0,24426282	0,2441971	0,11463314
9	0,37453968	0,1450137	0,04781805
10	0,33305287	0,30206474	0,51605744
11	0,23902189	0,30979144	0,46633013
12	0,10146419	0,50079474	0,32029729
13	0,16561974	0,42444108	0,17927557
14	0,16681841	0,17822744	0,31504233
15	-0,67251716	-0,00399943	1,32085203
16	-0,01823331	0,10551435	0,76583699
17	0,39232301	0,38220083	0,24782445
18	0,28403239	0,51119824	0,13334355
19	0,14492652	0,39815433	0,12120874
20	0,06591901	0,03636813	0,05310182
21	0,16378829	-0,24283697	0,7397168
22	0,1029073	0,18364083	0,47246805
23	-0,03710915	0,59160275	0,22066688
24	0,36890899	0,49805537	-0,1565662
25	-	-	-
26	-0,14590724	0,30188719	0,64426621
27	0,72197155	0,13107439	0,26706955
28	-	-	-
29	-	-	-
30	-	-	-
31	-0,80012649	-0,14069735	0,616187
32	-	-	-
33	-	-	-
34	-	-	-
35	-	-	-

5.3. Operation

The running and validation of the Combined Wind Power Forecast obtained in the setup stage were performed with the use of the remaining 15% of the available data for the case study wind farm, as well as the selected meteorological parameters – wind speed and temperature for the corresponding period. In order to plot NMAE for the different weather classes, for each class the NMAE was calculated according to the formula:

$$NMAE_{class} = \frac{\sum_{i,j=1}^{n} |y_i - x_i| \cdot Power(j)}{Total\ Power} \quad (4)$$

where y_i is the forecast, x_i is the true value and Power(j) is the average power for the respective class. Figure 12 and Figure 13 illustrate the NMAE values for each weather class of both cases – use of one and two weather parameters, respectively (the weather classes in which there were no corresponding readings for the model to be trained were omitted in the graph, therefore there are only 24 classes in the first case and 27 in the second). The green line in the graphs illustrates the validated and tested Combined Wind Power Forecast and it is visible that, in general, it performs better than each of the individual forecasts in both cases.

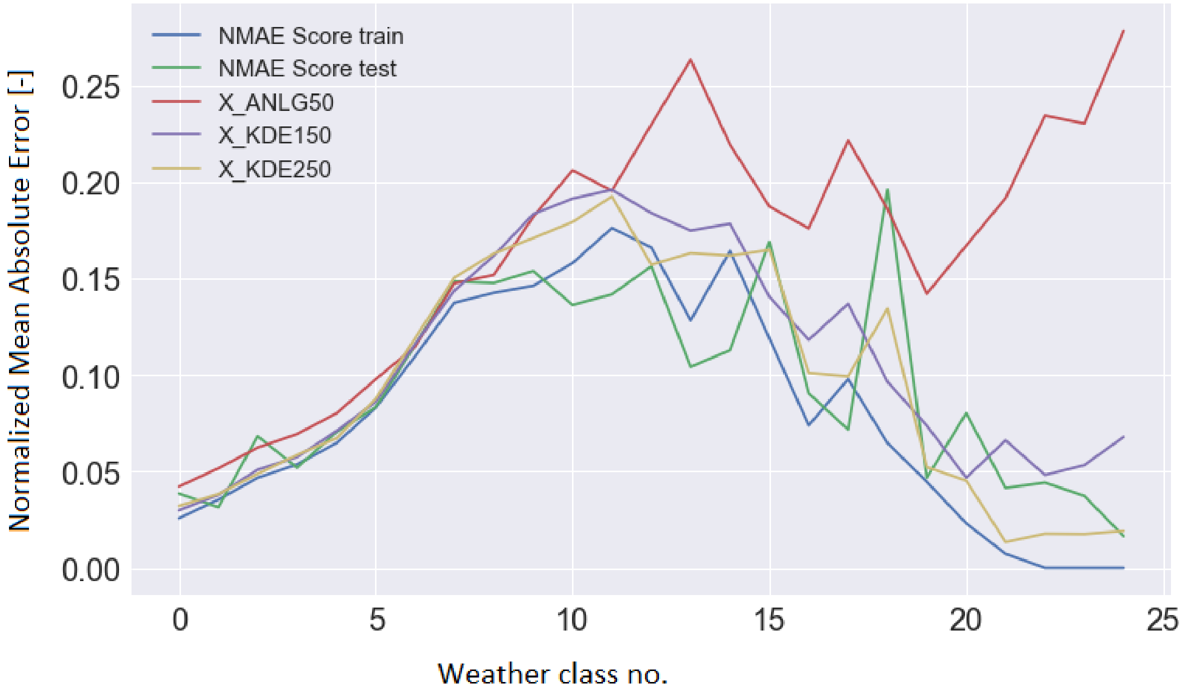

Figure 12 Case of a single weather parameter (wind speed): Normalized Mean Absolute Error of the trained (blue) and tested (green) Combined Wind Power Forecast, as well as of the three models individually (red, purple, yellow)

Weather class no.	0	1	2	3	4	5	6	7	8
No. of readings (train)	154	409	684	870	1127	1240	1254	1135	819
No. of readings (test)	18	46	76	97	126	138	140	127	91
Weather class no.	9	10	11	12	13	14	15	16	17
No. of readings (train)	472	228	144	63	34	36	25	20	16
No. of readings (test)	53	26	17	8	4	4	3	3	2
Weather class no.	18	19	20	21	22	23	24	25	26
No. of readings (train)	9	12	14	7	2	2	0	1	0
No. of readings (test)	2	2	2	1	1	1	0	1	0

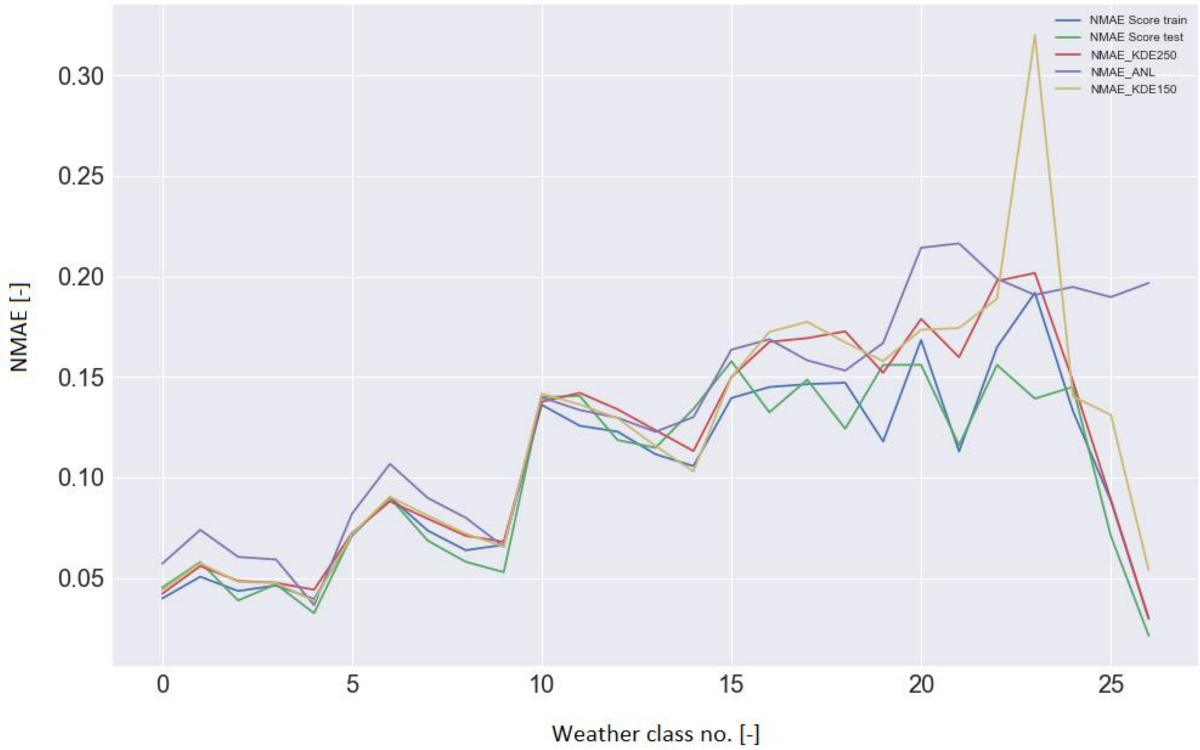

Figure 13 Case of two weather parameters (wind speed and temperature): Normalized Mean Absolute Error of the trained (blue) and tested (green) Combined Wind Power Forecast, as well as of the three models individually (red, purple, yellow)

Table 7 Case of two weather parameters (wind speed and temperature): Number of readings used in the train and test models

Weather class no.	1	2	3	4	5	6	7	8	9	10	11	12
No. of readings (train)	126	452	782	572	97	192	540	868	551	133	242	669
No. of readings (test)	23	80	139	102	18	34	96	154	98	24	43	119
Weather class no.	13	14	15	16	17	18	19	20	21	22	23	24
No. of readings (train)	855	436	90	198	444	411	147	31	121	175	123	33
No. of readings (test)	151	78	16	36	79	73	27	6	22	31	22	6
Weather class no.	25	26	27	28	29	30	31	32	33	34	35	
No. of readings (train)	0	79	36	0	0	0	28	0	0	0	0	
No. of readings (test)	0	14	7	0	0	0	6	0	0	0	0	

Below each graph illustrating the NMAE of the individual and combined forecasts for both cases, the respective table with the number of readings used in the train and test models is again presented (**Błąd! Nie można odnaleźć źródła odwołania.** and Table 7). They are here particularly important to properly analyze the results presented in the graphs (again, the weather classes in which there were no corresponding readings for the model to be trained were omitted in the graph, therefore there are only 24 classes in the first case and 27 in the second). When analyzing the two graphs, in both cases it is visible that the combined model performed the best in the weather classes for which the number of available readings was the highest. In the case of weather classes with a few readings only, the performance of the Combined Wind Power Forecast is considerably poorer.

Chronological data partition

In order to present a selected cut of the time-series, illustrating the individual WPF models together with the combined one, the data partition in the train/test models had to be changed from randomized to chronological. The reason is that the combined model can only be presented with the use of the 15% test data. In the case of the randomized partition, the data would not present the continuous time-series, but rather a random distribution of points.

Table 8 Chronological data partition: Weather classification for the case of two weather parameters - wind speed and temperature

Weather class no.	Wind speed range [m/s]	Temperature range [°C]	No. of readings (train)	No. of readings (test)
1	0 to 4	-20 to 0	149	0
2	0 to 4	0 to 10	498	34
3	0 to 4	10 to 20	751	170
4	0 to 4	20 to 30	628	46
5	0 to 4	>30	115	0
6	4 to 6	-20 to 0	226	0
7	4 to 6	0 to 10	583	53
8	4 to 6	10 to 20	802	220
9	4 to 6	20 to 30	557	92
10	4 to 6	>30	157	0
11	6 to 8	-20 to 0	285	0
12	6 to 8	0 to 10	704	84
13	6 to 8	10 to 20	829	177
14	6 to 8	20 to 30	429	85
15	6 to 8	>30	106	0
16	8 to 10	-20 to 0	234	0
17	8 to 10	0 to 10	451	72
18	8 to 10	10 to 20	438	46
19	8 to 10	20 to 30	159	15
20	8 to 10	>30	37	0
21	10 to 14	-20 to 0	141	2
22	10 to 14	0 to 10	167	39
23	10 to 14	10 to 20	124	21
24	10 to 14	20 to 30	38	1
25	10 to 14	>30	0	0
26	14 to 20	-20 to 0	93	0
27	14 to 20	0 to 10	41	2

28	14 to 20	10 to 20	1	0
29	14 to 20	20 to 30	0	0
30	14 to 20	>30	0	0
31	>20	-20 to 0	34	0
32	>20	0 to 10	0	0
33	>20	10 to 20	0	0
34	>20	20 to 30	0	0
35	>20	>30	0	0

Table 8 presents the weather classes for the case of chronological data partition, considering the use of two weather parameters. The weather classification applied here was exactly the same as in the similar case of the randomized data partition. It is clearly visible that the chronological partition method is not really applicable here, as the distribution of readings throughout the weather classes is constrained due to the little data available: if data from several years were used, the model could be trained with a certain number of years and then tested for the following years. Here, the 15% of data are all from the similar weather season and therefore there are many weather classes without any readings (marked in blue in Table 8). However, this distribution allows for a graphical illustration of a time-series. Figure 14 presents the NMAE of the combined model as well as of the individual WPF models, for a randomly selected continuous 72h of operation. Is it notable that, even for a casually chosen period of time, the combined model in general performs better than the singular models.

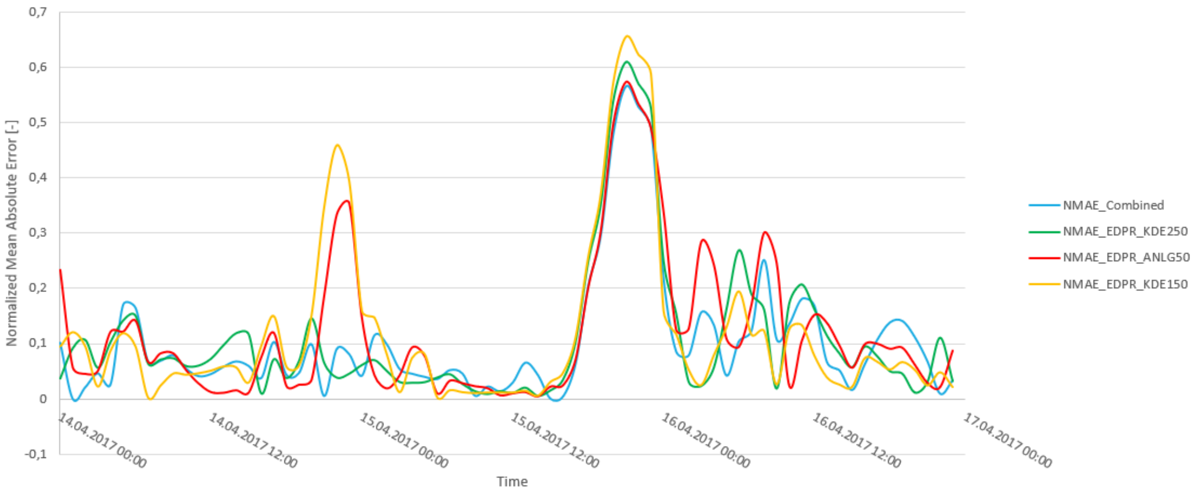

Figure 14 Time-series illustration of the NMAE of the combined model and of the individual WPF models, for a randomly selected continuous 72h of operation

6. Discussion

6.1. Summary of results

Table 9 summarizes the NMAE of the three different approaches followed to arrive at the Combined Wind Power Forecast:

1) Randomized data partition with the use of one weather parameter only – wind speed
2) Randomized data partition with the use of two weather parameters – wind speed and temperature
3) Chronological data partition with the use of two weather parameters – wind speed and temperature

All three approaches are then compared to the individual WPF models and the improvement over the individually best performing model (KDE250) as well as over the worst performing model (ANLG50) is determined. It is visible that in the best case (2), the overall NMAE has decreased by over one percentage point, which stands for 10,1% improvement in comparison with the individually best performing model (KDE250) and 16,3% improvement in the case of the weakest model (ANLG50).

Table 9 Summary of Combined Wind Power Forecast improvements

	WPF model	**Normalized Mean Absolute Error [%]**	**Improvement over individually best performing model (KDE250)**	**Improvement over individually worst performing model (ANLG50)**
Individual forecasts	NMAE_KDE250	10,505		
	NMAE_KDE150	10,539		
	NMAE_ANLG50	11,285		
1) 1 weather parameter (randomized data partition)	NMAE_Combined (test)	10,022	4,59%	11,19%
2) 2 weather parameters (randomized data partition)	NMAE_Combined (test)	9,442	10,11%	16,33%
3) 2 weather parameters (chronological data partition)	NMAE_Combined (test)	9,519	9,38%	15,64%

6.2. Implications

From the analysis of the results of the NMAE calculations, one can claim that the difference between the combined and the individual models does not seem large and significant. However, in the case of wind power predictions, the consequences of mispredictions and resulting over- or under-productions can be very costly. Wind generation imbalances penalties in day-ahead energy markets are usually settled using the current market price[39]. Therefore, even small decreases in WPF errors may bring visible positive monetary benefits.

In order to estimate the possible savings from the improved Combined Wind Power Forecast, the results from the model with the chronological data partition were used. In this case, the test data (15% of the available data) covered the period from 12.04.2017 - 16:00:00 until 31.05.2017 - 23:00:00, with hourly readings.

Firstly, the difference in the over-estimated production and the under-estimated production resulting from using the improved combined WPF, in comparison to the individual WPF models, was determined.

Table 10 Combined Wind Power Forecast - Resulting change in under- and over production

	Change in underproduction [MWh]:	**Change in overproduction [MWh]:**
KDE250	571,79	-2156,13
KDE150	797,05	-2243,68
ANLG50	-1530,57	-1360,13

Table 10 presents the differences in mispredictions, comparing the Combined Wind Power Forecast with the individual WPF models. What is interesting to notice is that underproduction increased in comparison with the first two models (KDE250 and KDE150), however overproduction decreased by a significantly higher amount. In other words, the combined model compensates the overestimation trend of the individual forecasts, but has a higher tendency to underestimate (except for the case of the ANLG50 model). The compensations in the overestimation are larger than the underestimations, hence the overall error is smaller. It is important to further analyze this outcome, taking into account the actual monetary value of these differences in mispredictions.

To arrive at a reasonable assessment of the cost difference, the historical data of the electricity market price in Romania for April and May 2017 were used (see Table 11)[40]. If the generated wind power is less than what it was scheduled for (underproduction), the wind power producer incurs an imbalance penalty. This penalty is determined as the wind power producer is obliged to purchase the shortfall from the real-time market, which usually has higher prices[41]. Therefore, possible penalties were assumed, depending on the misprediction type. In

---

[39] C. Brunetto and G. Tina, “Wind Generation Imbalances Penalties in Day-Ahead Energy Markets: The Italian Case,” *Electric Power Systems Research* 81, no. 7 (2011): 1446–55, https://doi.org/10.1016/j.epsr.2011.02.009.
[40] Andreea Paul, “Prețurile La Energie Electrică În România Ating Recorduri Istorice În August 2017 – INACO,” 2017, https://inaco.ro/preturile-la-energie-electrica-in-romania-ating-recorduri-istorice-in-august-2017/.
[41] Desmond W H Cai, Sachin Adlakha, and K Mani Chandy, “Optimal Contract for Wind Power in Day-Ahead Electricity Markets,” accessed May 10, 2018, https://pdfs.semanticscholar.org/99a2/604f9eafd599b5e70fa1b6cedd42fb8116cf.pdf.

the case of underproduction, the penalty of 130% of electricity market price was adopted for every MWh of imbalance, while for overproduction the opportunity cost was equated to the market price.

Table 11 Historical data of electricity market price in Romania[42]

Month	Price [€/MWh]
April 2017	38,03
May 2017	42,54

Table 12 presents the results of the estimation of the wind generation imbalances penalties. It is evident that the improved Combined Wind Power Forecast model brings monetary benefits. For the case study Pestera wind farm in Romania, having the installed capacity of 90MW, the calculations resulted in savings of over € 57 000 in less than two months.

Table 12 Summary of the estimations of wind generation imbalances penalties

WPF model	Cumulated penalty for the period of 12/04/2017 - 31/05/2017
KDE250	€ 504 031,16
KDE150	€ 495 431,11
ANLG50	€ 581 507,77
Combined	€ 446 488,16

6.3. Conclusions

The following are the main conclusions that can be withdrawn:

- Weighted combination as a function of two relevant weather parameters resulted in greater improvements, in comparison to the use of one parameter.
- Randomized data partition provided better data distribution over defined weather classes and therefore better results than chronological partition.
- Larger scope of historical training data could result in better-trained model and consequent superior improvements.
- The improved Combined Wind Power Forecast model brings monetary benefits. The calculations for the given case study wind farm resulted in savings of over € 57 000 in less than two months.

[42] Andreea Paul, "Prețurile La Energie Electrică În România Ating Recorduri Istorice În August 2017 – INACO."

6.4. Limitations

In the course of the study, certain limitations were acknowledged. First of all, the sample size of the analyzed data was narrow, especially for the occasionally occurring high wind speeds. This limitation caused the trained model to be worse-trained in certain weather classes and the improvements for those classes were poor or unreliable. Also, a large number of years of data would allow to properly test the chronological partition. Secondly, two of the available individual WPF models – KDE250 and KDE150 - demonstrated high correlation between one another. The reason for that was that both models were based on the same physical Kernel Density Estimation method. The combination of WPF models with lower correlation could result in better improvements. Finally, the meteorological parameters used in the analysis were taken from one point location, chosen based on the numerical mean of the wind turbines' locations distributed over a certain area. This averaging could have caused slight miscalculations. Besides, the use of weather historical measurements instead of historical simulation data could lead to potential further improvements of training the model.

6.5. Future research suggestions

As for the future research, it is suggested to follow the methodology with the use of a larger sample size, in order to arrive at well-trained models for each of the defined weather classes. Additionally, it is advised to verify the methodology with the use of less correlated WPF models. Moreover, after seeing the improvement resulting from passing from a single weather parameter to two in weather class definition, even if seemingly the second had no significant correlation with the forecast errors, there is the potential for improvement by using a third weather parameter in the definition of the weather classes, or even a forth.

The results presented in this paper prove the high potential of Combined Wind Power Forecast models. The improvements over the individual WPF models, brought by their combination, are evident and therefore further research is highly recommended.

References

Ackermann, Thomas. *Wind Power in Power Systems*, 2005. https://simsee.org/simsee/curso2010/wind_power_in_power_systems.pdf.

Andreea Paul. “Prețurile La Energie Electrică În România Ating Recorduri Istorice În August 2017 – INACO,” 2017. https://inaco.ro/preturile-la-energie-electrica-in-romania-ating-recorduri-istorice-in-august-2017/.

Bates, J. M., and C. W. J. Granger. “The Combination of Forecasts.” *The OR Society* 20, no. 4 (1969): 451. https://doi.org/10.2307/3008764.

Brunetto, C., and G. Tina. “Wind Generation Imbalances Penalties in Day-Ahead Energy Markets: The Italian Case.” *Electric Power Systems Research* 81, no. 7 (2011): 1446–55. https://doi.org/10.1016/j.epsr.2011.02.009.

Bunn, D. W. “A Bayesian Approach to the Linear Combination of Forecasts.” *Operational Research Quarterly (1970-1977)* 26, no. 2 (June 1975): 325. https://doi.org/10.2307/3008467.

Cai, Desmond W H, Sachin Adlakha, and K Mani Chandy. “Optimal Contract for Wind Power in Day-Ahead Electricity Markets.” Accessed May 10, 2018. https://pdfs.semanticscholar.org/99a2/604f9eafd599b5e70fa1b6cedd42fb8116cf.pdf.

Clemen, Robert T. “Combining Forecasts: A Review and Annotated.” *International Journal of Forecasting* 5 (1989): 559–83. https://pdfs.semanticscholar.org/7117/9279738b91df0520061b351cb3e0124a411c.pdf.

European Centre for Medium-Range Weather Forecasts ECMWF. “ECMWF | Advancing Global NWP through International Collaboration.” Accessed May 12, 2018. https://www.ecmwf.int/.

Global Wind Energy Council. “Global Wind Energy Outlook 2016 - Opening up New Markets for Business,” 2016. http://files.gwec.net/files/GlobalWindEnergyOutlook2016.

Granger, C.W.J., and R. Ramanathan. “Improved Methods of Forecasting.” *Journal of Forecasting 3*, 1984.

Holttinen, Hannele. “Estimating the Impacts of Wind Power on Power Systems—summary of IEA Wind Collaboration Hannele Holttinen.” *Environ. Res. Lett* 3 (2008): 25001–6. https://doi.org/10.1088/1748-9326/3/2/025001.

Holttinen, Hannele, Bettina Lemström, Finland Peter Meibom, and Henrik Bindner. “Design and Operation of Power Systems with Large Amounts of Wind Power State-of-the-Art Report.” Accessed April 21, 2018. http://www.vtt.fi/publications/index.jsp.

Huang, Junhui, Case P. van Dam, and Henry Shiu. “Wind Energy Forecasting: A Review of State-of-the-Art and Recommendations for Better Forecasts. Final Report, Appendix B, California Renewable Energy Forecasting, Resource Data and Mapping,” 2010.

International Energy Agency. “The Power of Transformation - Wind, Sun and the Economics of Flexible Power

Systems," 2014. https://www.iea.org/publications/freepublications/publication/The_power_of_Transformation.pdf.

Jung, Jaesung, and Robert P Broadwater. "Current Status and Future Advances for Wind Speed and Power Forecasting," 2014. https://pdfs.semanticscholar.org/34e8/b8ceb3c68a7d0ee437b26eb9a757d619a814.pdf.

Koksoy, Ceyda Er, Mehmet Baris Ozkan, Serkan Buhan, Turan Demirci, Yusuf Arslan, Aysenur Birturk, and Pinar Karagoz. "Improved Wind Power Forecasting Using Combination Methods." In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 1142–47. IEEE, 2015. https://doi.org/10.1109/ICMLA.2015.60.

Menezes, Lilian M De, Derek W Bunn, and James W Taylor. "Review of Guidelines for the Use of Combined Forecasts." *European Journal of Operational Research* 120 (2000): 190–204. https://pdfs.semanticscholar.org/97c3/ea22e05eabda946695431e63d04bfbae5af4.pdf.

Meteoblue. "Weather History+." Accessed April 23, 2018. https://www.meteoblue.com/en/historyplus.

Monteiro, C, R Bessa, V Miranda, a Botterud, J Wang, and G Conzelmann. "Wind Power Forecasting." *Information Sciences* 11, no. 4 (2009): 762–767. https://doi.org/10.1016/B978-0-8155-2047-4.10007-9.

National Center for Atmospheric Research Earth System Laboratory. "MM5 Community Model Homepage." 2014. Accessed May 11, 2018. http://www2.mmm.ucar.edu/mm5/overview.html.

Nielsen, Henrik Aa., Torben S. Nielsen, Henrik Madsen, Maria J. San Isidro Pindado, and Ignacio Marti. "Optimal Combination of Wind Power Forecasts." *Wind Energy* 10, no. 5 (September 2007): 471–82. https://doi.org/10.1002/we.237.

Patrick L. McDermott & Christopher K. Wikle. "Analog Forecasting: A Flexible and Parsimonious Alternative for Nonlinear Prediction - Statistics Views," 2016. http://www.statisticsviews.com/details/feature/9652191/Analog-Forecasting-a-Flexible-and-Parsimonious-Alternative-for-Nonlinear-Predict.html.

REN21. "RENEWABLES 2017 GLOBAL STATUS REPORT," 2017. http://www.ren21.net/wp-content/uploads/2017/06/17-8399_GSR_2017_Full_Report_0621_Opt.pdf.

Richmond, Victor, R Jose, and Robert L Winkler. "Simple Robust Averages of Forecasts: Some Empirical Results." Accessed April 22, 2018. https://doi.org/10.1016/j.ijforecast.2007.06.001.

Rosenblatt, Murray. "Remarks on Some Nonparametric Estimates of a Density Function." *The Annals of Mathematical Statistics* 27, no. 3 (September 1956): 832–37. https://doi.org/10.1214/aoms/1177728190.

Sánchez, Ismael. "Adaptive Combination of Forecasts with Application to Wind Energy." *International Journal of Forecasting* 24, no. 4 (October 2008): 679–93. https://doi.org/10.1016/j.ijforecast.2008.08.008.

Wind Energy The Facts. “Security of Supply and System Adequacy.” Accessed April 21, 2018. https://www.wind-energy-the-facts.org/security-of-supply-and-system-adequacy.html.