

Analysis of possibilities to improve quality of spatial wind speed forecasts for efficient forecasting of electric energy production in onshore wind farms in Poland

Paweł Piotrowski^{1,*}, *Dariusz Baczyński*¹, *Marcin Kopyt*^{1,2}, and *Karolina Szafranek*³

¹Warsaw University of Technology, Faculty of Electrical Engineering, Warsaw, Poland

²Globema sp. z o.o., ul. Wita Stwosza 22, 02-661 Warsaw, Poland

³Warsaw University, Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw, Poland

Abstract. The most important factor responsible for the quality of energy production forecasts in wind farms is the accurate wind speed forecast. An extensive statistical analysis of meteorological data (NWP) from 16 base nodes of the "300" grid in the "Łódź" area was made. The intention of the statistical analysis was to select potential explanatory variables for models predicting wind speed in the remaining 206 nodes of the grid's mesh. Next, tests of selected prognostic methods were performed in order to compare their effectiveness with bilinear method which is not computationally complex. It should be emphasized that the main problem in spatial wind speed forecasting is the very large number of nodes for which the forecasts are calculated. As a consequence, more advanced and computationally complex forecasting methods cannot be applied in practice due to too long calculations time and difficulties in huge amounts of data processing. Conclusions with proposals of preferred forecasting methods that could be used in practice were developed.

1 Introduction

Acquiring electrical energy from wind sector's dynamic development (onshore and offshore) is intertwined with the need for forecasting of wind speeds and total electrical energy production in the wind farms area as accurately as possible. Accurate forecasts enable more effective control of electrical power systems [1]. Costs minimization of conventional power plants production or costs minimization of energy bought in energy cluster area, in microgrid area or by prosumer can be given as examples of optimization objectives [2].

It is also worth pointing out, that accurate wind speed forecasts have an important part to play in choosing the best time for performing wind turbines maintenance and potential repairs [3].

* Corresponding author: pawel.piotrowski@ee.pw.edu.pl

Methods of wind energy production depend on the forecast horizon and available data. These methods can be divided into two separate parts: energy production prediction using statistical and regression models and/or meteorological forecast determination using meteorological data from Numerical Weather Prediction (NWP) models. There are plenty of studies on the advanced statistical analysis of historical power records in order to develop new methods of wind speed or power generation forecast systems [4-13]. Authors of [14] introduce and evaluate two hybrid forecasting models for wind speed and power generation (ARIMA–SVM and ARIMA–ANN, where ARIMA - autoregressive integrated moving average, ANN – artificial neural network, SVM - support vector machine). ARIMA prediction model is used for the linear component of a time series and a nonlinear prediction model for the nonlinear component. This approach has advantages comparing to single ARIMA, ANN, and SVM forecasting models. Wind power generation prediction was based on historical time series. For wind speed forecasting 2-year hourly dataset was retrieved from a wind observation site. Study [15], used past values of the wind speed and directions and their spatio-temporal correlations measured at numerous geographical locations to produce simple prediction model for the hourly mean wind speed and direction from 1 to 6 h ahead at multiple sites in the UK.

Use of precise meteorological forecasts for power generation determination is quite simple when analysing single or multiple locations. More complex analysis has to be done when large areas are taken into consideration, mainly because of the large amount of data and high cost of calculation.

Statistical analyses and spatial forecasts that were carried out concerned area with a codename „Łódź” for grid „300”. This area concerns environs of Łódź city with area size defined as a result of uniform deployment of circa 300 nodes on Poland’s territory (and adjacent parts of neighbouring countries). This way 196 computational meshes were obtained, one of which was „Łódź”. Schematic diagram of one mesh is shown in Fig. 1. One mesh consists of 206 nodes (pale blue color in Fig. 1) for which wind speed forecast are to be done (horizon from +1 h to +72 h). Base nodes A1..A4 and B1..B12 are points for which values of meteo variables forecasted values are known. These variables are wind speed, air pressure, wind azimuth and solar irradiance, each one of them with horizons spanning from +1 h to +72 h. Between neighbouring nodes Ax-Ax, Bx-Bx, Ax-Bx there are 12 nodes vertically and 13 nodes horizontally.

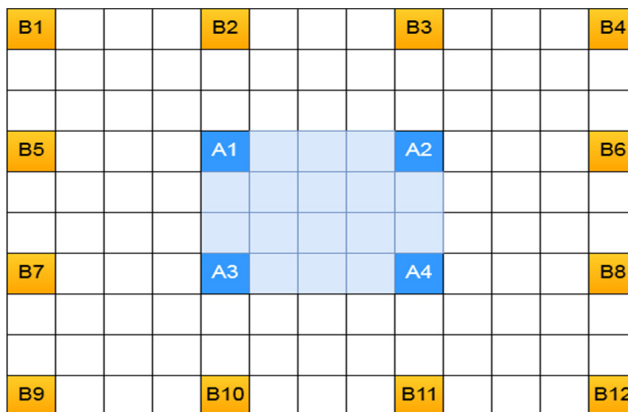


Fig. 1. Mesh of grid “300” - schematic diagram. Source: Own elaboration.

2 Statistical analysis of data

Statistical analysis of available data for the 300-nodes grid area (206 forecast grid nodes) for analysed area „Łódź” is performed. The Interdisciplinary Centre for Mathematical and Computational Modelling of the Warsaw University (ICM UW) provided the data (meteorological forecasts) for the scientific research. All analysed time series are from period of one year (1 hour values).

The Kolmogorov-Smirnov and Lilliefors tests show that the time series of wind speed forecasts for the analysed area do not have a normal distribution. The values of variance, standard deviation and coefficient of variation increase very significantly with the increase of the forecast horizon. The shape of the histograms also changes. With the increase in the forecast horizon, the value of kurtosis is becoming more and more negative. Probably forecasts with larger horizons used for research are less reliable as data because of disturbances in the algorithm that generates forecasts.

The variation of wind speed (variance) should be similar regardless of the horizon of the forecast, especially when the average wind speeds for different horizons are not significantly different. Table 1 shows selected statistical measures of time series of wind speed forecasts.

Table 1. Selected statistical measures of time series of wind speed forecasts. Source: Own elaboration.

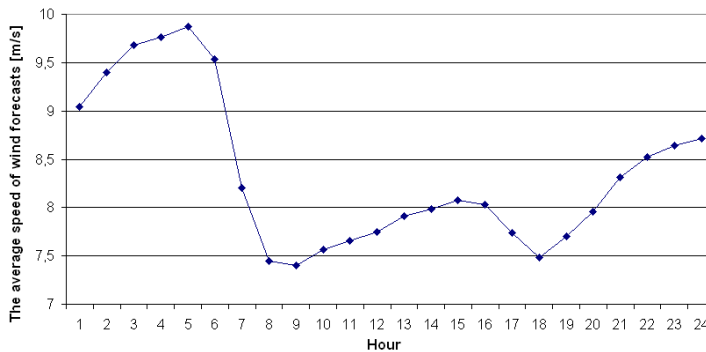
Statistical measures	Horizon of forecasts							
	+1 h	+3 h	+6 h	+12 h	+18 h	+24 h	+48 h	+72 h
Max speed [m/s]	15.00	14.96	16.07	15.92	18.74	16.98	14.95	16.92
Average speed [m/s]	8.35	10.01	10.63	11.92	8.30	10.14	7.79	7.63
Median [m/s]	7.83	10.18	11.03	8.34	8.31	9.99	8.62	7.84
Standard deviation [m/s]	1.89	2.38	2.89	3.44	3.19	3.19	3.41	4.11
Variation	3.58	5.70	8.37	11.87	10.17	10.15	11.65	16.92
Coefficient of variation [%]	22.66	23.79	27.22	28.85	38.42	31.41	43.77	53.92
Skewness	1.22	-0.32	-0.28	0.25	-0.08	0.21	-0.26	0.17
Kurtosis	1.05	-0.87	-1.03	-0.76	-0.35	-0.82	-1.15	-0.52

Wind speed forecasts from 16 base grid nodes (B1..B12, A1..A4) have very large values of Pearson linear correlation coefficient to wind speed forecasts from all 206 nodes. It is difficult to explain the increase in the value of Pearson's linear correlation coefficients for the larger horizons of wind speed forecasts. The vast majority of calculated linear correlation coefficients is statistically significant - significance at 5% level (excluding the length of the day). Table 2 shows the values of Pearson linear correlation coefficients between forecasts of wind speed from all 206 nodes and potential explanatory variables.

For the analysed area there is quite a strong variation of wind speed forecasts in the period of the day. The "hour" (a number from 1 to 24) appears to be a significant explanatory variable in the forecasting model using neural network - the multilayer perceptron (MLP). A large night peak is visible from 9 pm to 7 am. Figure 2 shows changes in the average wind forecast speed over the 24-hour period for data from all horizons from +1 h to +72 h.

Table 2. Values of Pearson linear correlation coefficients between forecasts of wind speed from all 206 nodes and potential explanatory variables. Source: Own elaboration.

Potential explanatory variables	Pearson linear correlation coefficient	
	Horizon of the forecast	
	+1 h	+72 h
Wind speed forecast - node B1	0.743	0.855
Wind speed forecast - node B2	0.814	0.915
Wind speed forecast - node B3	0.808	0.880
Wind speed forecast - node B4	0.844	0.918
Wind speed forecast - node B5	0.836	0.915
Wind speed forecast - node B6	0.85	0.913
Wind speed forecast - node B7	0.855	0.928
Wind speed forecast - node B8	0.851	0.936
Wind speed forecast - node B9	0.814	0.875
Wind speed forecast - node B10	0.873	0.899
Wind speed forecast - node B11	0.875	0.902
Wind speed forecast - node B12	0.844	0.851
Wind speed forecast - node A1	0.829	0.970
Wind speed forecast - node A2	0.852	0.947
Wind speed forecast - node A3	0.891	0.950
Wind speed forecast - node A4	0.874	0.971
Air temperature forecast - node A1	0.725	0.289
Atmospheric pressure forecast - node A1	-0.712	-0.334
The length of the day	0.162	-0.137
Wind azimuth forecast - node A1	-0.240	-0.398
Wind azimuth forecast - node A2	-0.256	0.119
Wind azimuth forecast - node A3	-0.280	-0.651
Wind azimuth forecast - node A4	-0.283	-0.675

**Fig. 2.** Changes in the average wind speed forecast over the 24-hour period for data from all horizons from +1 h to +72 h for the analysed area. Source: Own elaboration.

As the last element of statistical analysis, it was checked if the values of linear correlation coefficients between wind speed forecasts in subsequent nodes between base nodes and wind speed forecasts in base nodes A1,...,A4, B1, ... B12 may be useful as potential explanatory variables in statistical forecasting models. Calculations of linear correlation coefficients were made both separately for each wind speed forecasts horizon (from +1 h to +72) and together for forecasts from all forecast horizons. The values of linear correlation coefficients are significantly different in each of the nodes for each horizon of forecasts. The part of the values of linear correlation coefficients are statistically

insignificant at the significance level of 5%. Therefore, it can be concluded that only the values of linear correlation coefficients calculated separately for each of the forecast horizon can be potentially valuable as input data to the statistical forecasting model. Furthermore, there is no significant relationship between the distance of a given node to the base node, and the calculated values of linear correlation coefficients between the given node and the base node. Also it is visible that the values of linear correlation coefficients are very similar in each of the nodes for each horizon of forecasts for correlation coefficients calculated for all horizons together from +1 h to +72 h (see Fig. 3). In this case all the values of linear correlation coefficients are statistically significant at the significance level of 5% but only slightly above the level of statistical significance.

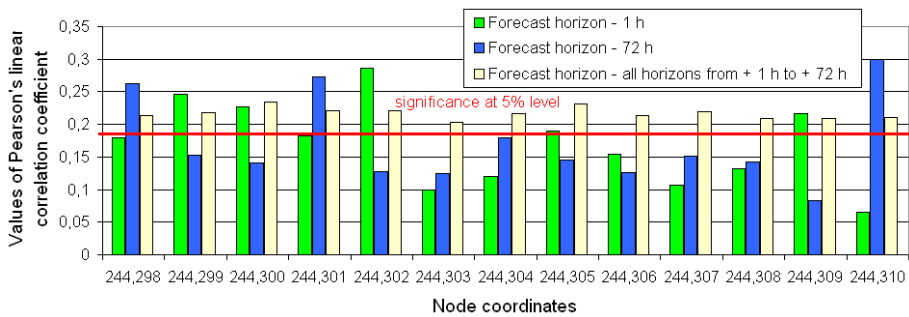


Fig. 3. The values of linear correlation coefficients between the base node A3 and subsequent nodes located to the right of it for the horizon +1 h, +72 h and all horizons together. Source: Own elaboration.

As a result of the statistical analysis, it was assumed that all the analysed potential explanatory variables (excluding the length of day) may be useful in classic statistical forecasting models and forecasting models using artificial MLP neural networks.

3 Initial tests of developed statistical models for spatial wind speed forecasting

The goal of the tests is to find a improved statistical forecasting method (not very computationally complex) better than the bilinear method. The bilinear method consists in subsequently executing linear interpolations in 2 orthogonal directions.

Two proposed original forecasting models were tested, using wind speed forecasts at selected base nodes and the values of linear correlation coefficients between selected base nodes and a given node for which wind speed forecast is calculated. Figure 4 shows the comparison of tested three forecasting methods.

For the „16-node model”, the wind speed forecast for single horizon in each of the n nodes (x,y coordinates) of the grid is calculated by

$$\hat{p}_{(x,y)} = \frac{\sum_{l=1}^{16} r_{A_l, x,y} \cdot v_{A_l}}{\sum_{l=1}^{16} r_{A_l, x,y}} \quad (1)$$

where $r_{A_l, x,y}$ is separately calculated (for a given single horizon of the forecast) Pearson's linear correlation coefficient between a node with x, y coordinates and a base node with

index l , v_{A_l} is forecast of wind speed in the base node A_l , and A_l is subsequent base node (B1,..B12, A1,..A4).

For the “2-node vertical / horizontal model”, the wind speed forecast for single horizon in each of the n nodes (x,y coordinates) of the grid between the two A-type base nodes with index k and l is calculated by

$$\hat{p}_{(x,y)} = \frac{r_{A_l,(x,y)} \cdot v_{A_l} + r_{A_k,(x,y)} \cdot v_{A_k}}{r_{A_l,(x,y)} + r_{A_k,(x,y)}} \quad (2)$$

where $r_{A_l,(x,y)}$ is separately calculated (for a given single horizon) Pearson's linear correlation coefficient between a node with x, y coordinates and a base node with index l (A type base nodes have indexes from 1 to 4), and v_{A_l} is forecast of wind speed in the base node A_l .

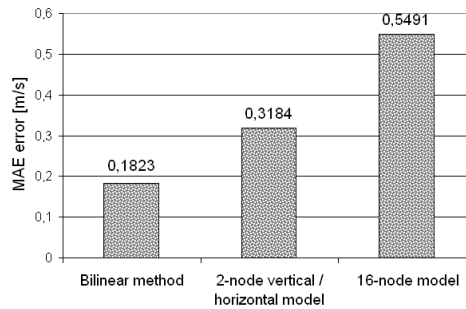


Fig. 4. The comparison of wind speed forecast errors for the horizon of +1 h for 13 nodes between the base nodes A3 and A4. Source: Own elaboration.

Finally, the suitability of multiple regression models for forecasting wind speed was verified. The test results showed that “multiple regression model - all horizon” is inappropriate and generates higher wind speed forecasting errors than the simple bilinear method. In the next step, multiple regression models were tested (a single model for a given node for only one horizon). Thus, 72 separate models were built for each node. For the „multiple regression model – single horizon”, the wind speed forecast for single k horizon in each of the n nodes (x,y coordinates) of the grid is calculated by

$$\hat{P}_{(x,y),t+k} = a + A_{1,t+k} \cdot w_{A1} + A_{2,t+k} \cdot w_{A2} + \dots + A_{4,t+k} \cdot w_{A4} + B_{1,t+k} \cdot w_{B1} + B_{2,t+k} \cdot w_{B2} + \dots + B_{12,t+k} \cdot w_{B12} \quad (3)$$

where $A_{1,t+k}..A_{4,t+k}$, $B_{1,t+k}..B_{12,t+k}$ are explanatory variables (wind speed forecasts in nodes A1..A4, B1..B12 in $t+k$ period), a is the constant (offset), $w_{A1}..w_{A4}$, $w_{B1}..w_{B12}$ are parameters, and k is forecast horizon.

The estimation of the parameters of the “multiple regression model - single horizon” models was performed twice by the cross-validation method. At first, estimation of model parameters was made on the first half of available data (the other half of the data was the test range of the quality of forecasts). Then, the ranges of data were exchanged and the model parameters were re-estimated. The final errors of the wind speed forecasts (MAE error) were calculated as the mean values of the MAE error pairs from the test data ranges.

The multiple regression models for shorter forecast horizons have evidently lower MAE errors than the bilinear method. In particular, it is visible for the horizons of forecasts from +1 h to +6 h. Figure 5 shows the example of results for node coordinates (251,297).

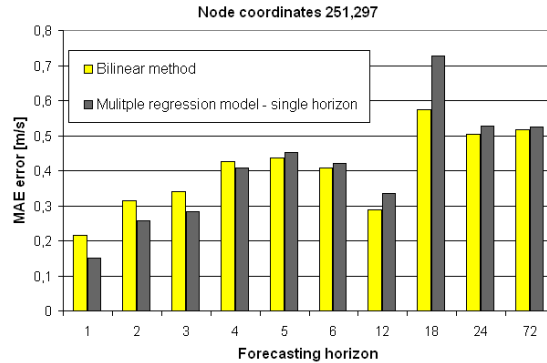


Fig. 5. The comparison of wind speed forecast MAE errors for the (251,297) node for chosen horizons of forecasts. Source: Own elaboration.

Multiple regression models with additional exogenous explanatory variables were also tested but unfortunately they were generating higher MAE forecast errors than models that use only wind speed forecasts in the base nodes as explanatory variables. Table 3 shows the comparison of forecasts for +1 h horizon for 12 nodes between the base node A3 and the base node A4.

Table 3. The comparison of forecasts for +1 h horizon for 12 nodes between the base node A3 and the base node A4. Source: Own elaboration.

Forecasting method	MAE error [m/s]
Bilinear method	0.6000
Multiple regression model – single horizon Explanatory variables: wind speed forecasts in base nodes A1..A4, B1..B12	0.5996
Multiple regression model – single horizon Explanatory variables: wind speed forecasts in base nodes A1..A4	0.6292
Multiple regression model – single horizon Explanatory variables : wind speed forecasts in A1..A4, B1..B12 (all 16 base nodes), air temperature forecast in node A1, atmospheric pressure forecast in node A1, the length of the day, wind azimuth forecasts in base nodes A1..A4	0.6336

It was also observed that for each tested forecasting method, the largest errors of forecasts occur for the +18 h forecast horizon. This occurrence is repeatable for each of the analysed nodes. The reason for this phenomenon could not be determined.

Finally, it could be concluded that multiple regression model should be unique for each node and each horizon (14832 models in total). It results from dissimilarities of times series features for each horizon and each node.

4 Propositions of effective solutions for spatial wind speed forecasting problem

For the purpose of problem solving („Łódź” area containing 206 nodes), multiple regression models (14832 models in total) and single neural network type MLP type generating forecasts for each node and each forecast horizon were proposed. Parameters of multiple regression models (3) were optimized by PSO algorithm. For the MLP learning purpose BP algorithm with adaptive learning coefficient, momentum technique, periodic

facts shuffling and periodic weight values disruption were used. Estimation, testing and validation were performed on the same periods for multiple regression models and MLP. During calculations MLP used 10% of total data, whereas multiple regression models used 100% of total data.

The problem is composed of $n=206$ nodes for which wind speed forecasts are to be done. For each node forecasts are done for horizons spanning from +1 h to +72 h. Number of explanatory data sets for MLP used for forecasting is considerable. It is equal to product of number of full data days, number of nodes, number of forecasts horizons ($365 \times 206 \times 72 = 5143680$). Consequently, there is a problem (mainly great temporal computational complexity) related to insufficient computational power of computer (64-bit processing) on which MLP learning is carried out. Hence, it seems appropriate to search for learning data sets number reduction methods with minimizing risk of lowering MLP forecasting accuracy at the same time.

First, reduction methods possibilities were studied. "Random selection from blocks" method was proposed for data reduction effects testing. In this method k -learning data sets are split into inseparable blocks. One block consists of 206 data sets (number of nodes) containing information about given m -th day's (where m is number from 1 to 365) weather forecast with $+s$ horizon (where s is number from 1 to 72). Therefore, number of blocks equals 26280 (365×72) and each block contains 206 data sets. Objective of method is uniform selection of blocks, characterized by large, medium and small variation and large, medium and small mean. Due to that, it could be expected that blocks selected for MLP learning would be the most representative (carrying the most information). For 100-times reduction of learning data sets 1% of them was chosen from created 26280 sets. For each block (out of 26280), containing 206 information about forecasted wind speed in 206 forecasting nodes, variation and mean value was computed. Blocks were then sorted in ascending order in regard to variations corresponding to blocks and thereafter 0.5% of blocks was chosen by the means of uniform selecting every 200th block, starting with block of the smallest variation. The next step was sorting blocks in ascending order, this time in regard to means corresponding to blocks. Subsequently 0.5% of blocks was chosen using the same method as before, but with the use of means. Algorithm of block choosing prevented blocks repetitions.

To determine which percentage of blocks chosen from all data would be representative, MLP generalization ability tests were performed for number of blocks equal to 1/2/5/10% of total block number containing learning data. Tests were carried out for 26-input MLP (16 wind speeds forecasts from nodes A1..A4 and B1..B12, 2 geographical coordinates of output node, 4 wind directions forecasts from nodes A1..A4, number of hour of forecast, length of day, atmospheric pressure and air temperature forecasts for nodes A1). Data was divided into learning, testing and validation sets. These sets had similar wind speed mean values. Structure of MLP was composed of 26 inputs, 1 output, and 60, 40 and 30 neurons respectively in next hidden layers. MLP constructed this way includes 5190 weights in total –circa 45 times less of modifiable parameters when compared to all 14832 multiple regression models (237312 parameters in total). Huge data volumes (several dozen MB) processed during MLP learning and, consequently, high computational complexity induced creation of original computer program written in C++ language. MLP learning when done in MATLAB/OCTAVE type software would take more than 700 h according to estimations.

Quality test of MLP using reduced 100-times number of sets shown quality deterioration so great, that method was deemed as ineffective. The most probable explanation, is that each of 14832 (206×72) sub-models of MLP had insufficient number of data sets- in practice few data sets for single sub-model. 10% blocks turned out to be the most appropriate and they facilitated achieving suitably low error on test data. It is worth

mentioning that size of learning data files exceeded 30 MB, whereas validation data set (100% blocks) had a volume of circa 3 GB. Table 4 shows the comparison of forecasts for all horizons and all 206 nodes.

Table 4. The comparison of forecasts for all horizons and all 206 nodes. Source: Own elaboration.

Forecasting method	MAE error [m/s]	Percentiles of MAE errors [m/s]		
		50%	95%	100%
14832 multiple regression models – single horizon	0.471	0.307	1.454	26.266
Bilinear method	0.489	0.324	1.482	27.034
Neural network type MLP	0.533	0.375	1.551	28.388

5 Conclusions

Most accurate spatial forecasts were achieved by multiple regression models (14832 models in total). Accuracy differences for all of three methods were however minute and at the level of few percents. Slightly greater values of MAE error for MLP stems from considerably bigger number of parameters in multiple regression models and from explicit problem decomposition (each node and each horizon had its separate model). Decomposition was really significant, because wind speed forecasts in 16 base nodes were considerably different in regard to variation and mean for respective forecasts horizons. Importantly, change of MLP structure (number of hidden layers and neurons in layers) did not improve results. Aside from decomposition and greater number of parameters, cause of achieving better results by multiple regression models was using full volume of data by multiple regression models.

Computational cost was obviously the lowest for bilinear method. Usage of many multiple regression models or the single MLP means approximately similar cost that is a couple orders of magnitude greater than cost of bilinear method. Due to this cost, bilinear method should be preferred for typical calculations. MLP or multiple regression models should be used in special cases for chosen locations. These could include very important locations where lowest possible error is needed and locations with non-typical terrain orography (e.g. places with big difference of heights, lakes, etc.) where bilinear method forecasts accuracy clearly deteriorates.

The article was based on the results of the project "Spatial forecasting of energy generation from renewable energy sources including its impact on loads in network nodes", co-financed by the European Union through The National Center for Research and Development under the Operational Programme Smart Growth 2014-2020.

The Interdisciplinary Centre for Mathematical and Computational Modelling of the Warsaw University (ICM UW) provided the data (meteorological forecast) for the scientific research.

References

1. D. Baczyński, K. Księżyk, M. Parol, P. Piotrowski, J. Wasilewski, T. Wójtowicz, *Mikrosieci niskiego napięcia. Praca zbiorowa pod redakcją M. Parola*, (OWPW, Warszawa, 2013)
2. M. Sabat, D. Baczyński, K. Szafranek, Analiza szeregów czasowych produkcji energii ze źródeł odnawialnych pod kątem niezależności energetycznej wybranego obszaru, *Przegląd Elektrotechniczny* **93**, 11-15 (2017)

3. P. Piotrowski, M. Polewaczyk, R. Raczkowski, S. Robak, Problem narażenia pracowników morskich stacji elektroenergetycznych najwyższych napięcia na szkodliwe czynniki oraz działania minimalizujące ryzyko zagrożeń, *Medycyna Pracy* **67**, 51-72 (2016)
4. J. Tastu, P. Pinson, E. Kotwa, H. Madsen, H. Nielsen, Spatio-temporal analysis and modeling of short-term wind power forecast errors, *Wind Energy* **14**, 43-60 (2011)
5. R. Girard, D. Allard, Spatio-temporal propagation of wind power prediction errors, *Wind Energy* **16**, 999-1012 (2013)
6. M. He, L. Yang, J. Zhang, V. Vittal, A Spatio-Temporal Analysis Approach for Short-Term Forecast of Wind Farm Generation, *IEEE Transaction on power systems* **29**, 4 (2014)
7. T. Popławski, K. Dąsał, J. Łyp, P. Szelaąg, Zastosowanie modeli ARMA do przewidywania mocy i energii pozyskiwanej z wiatru, *Polityka Energetyczna* **13**, 385-400 (2010)
8. T. Popławski, Problematyka prognoz generacji wiatrowej w KSE, *Przegląd Elektrotechniczny* **90**, 119-122 (2014)
9. P. Piotrowski, D. Baczyński, Prognozowanie dobowej produkcji energii elektrycznej przez turbinę wiatrową z horyzontem 1 doby, *Przegląd Elektrotechniczny* **90**, 113-117 (2014)
10. J. Wasilewski, D. Baczyński, Short-term electric energy production forecasting at wind power plants in pareto-optimality context, *Renewable and Sustainable Energy Reviews* **69**, 177-187 (2017)
11. P. Piotrowski, Analiza statystyczna danych mających wpływ na produkcję energii elektrycznej przez farmę wiatrową oraz przykładowe prognozy krótkoterminowe, *Przegląd Elektrotechniczny* **88**, 161-164 (2012)
12. D. Baczyński, M. Kopyt, Analiza jakości prognoz produkcji energii w elektrowniach wiatrowych opracowanych na podstawie danych meteo z modeli UM i COAMPS, *Przegląd Elektrotechniczny* **93**, 171-175 (2017)
13. P. Piotrowski, *Prognozowanie w elektroenergetyce w różnych horyzontach czasowych*, Prace Naukowe Politechniki Warszawskiej „Elektryka” **144** (OWPW, Warszawa, 2013)
14. J. Shi, J. Guo, S. Zheng, Evaluation of hybrid forecasting approaches for wind speed and power generation time series, *Renewable and Sustainable Energy Reviews* **16**, 3471-3480 (2012)
15. S. Weiss, D. Hill, D. Infield, Short-term spatio-temporal prediction of wind speed and direction Jethro Dowell, *Wind Energy* **17**, 1945-1955 (2014)