Hybrid and Ensemble Methods of Two Days Ahead Forecasts of Electric Energy Production in a Small Wind Turbine

Paweł Piotrowski [1,*] **, Marcin Kopyt** [1] **, Dariusz Baczyński** [1] **, Sylwester Robak** [1] **and Tomasz Gulczyński** [2]

1. Institute of Electrical Power Engineering, Warsaw University of Technology, Koszykowa 75 Street, 00-661 Warszawa, Poland; marcin.kopyt@ee.pw.edu.pl (M.K.); dariusz.baczynski@ee.pw.edu.pl (D.B.); sylwester.robak@ee.pw.edu.pl (S.R.)
2. Globema Sp. z o. o., Wita Stwosza 22 Street, 02-661 Warsaw, Poland; tomasz.gulczynski@globema.pl
* Correspondence: pawel.piotrowski@ee.pw.edu.pl

Abstract: The ability to forecast electricity generation for a small wind turbine is important both on a larger scale where there are many such turbines (because it creates problems for networks managed by distribution system operators) and for prosumers to allow current energy consumption planning. It is also important for owners of small energy systems in order to optimize the use of various energy sources and facilitate energy storage. The research presented here addresses an original, rarely predicted 48 h forecasting horizon for small wind turbines. This topic has been rather underrepresented in research, especially in comparison with forecasts for large wind farms. Wind speed forecasts with a 48 h horizon are also rarely used as input data. We have analyzed the available data to identify potentially useful explanatory variables for forecasting models. Eight sets with increasing data amounts were created to analyze the influence of the types and amounts of data on forecast quality. Hybrid, ensemble and single methods are used for predictions, including machine learning (ML) solutions like long short-term memory (LSTM), multi-layer perceptron (MLP), support vector regression (SVR) and K-nearest neighbours regression (KNNR). Original hybrid methods, developed for research of specific implementations and ensemble methods based on hybrid methods' decreased errors of energy generation forecasts for small wind turbines in comparison with single methods. The "artificial neural network (ANN) type MLP as an integrator of ensemble based on hybrid methods" ensemble forecasting method incorporates an original combination of predictors. Predictions by this method have the lowest mean absolute error (MAE). In addition, this paper presents an original ensemble forecasting method, called "averaging ensemble based on hybrid methods without extreme forecasts". Predictions by this method have the lowest root mean square error (RMSE) error among all tested methods. LSTM, a deep neural network, is the best single method, MLP is the second best one, while SVR, KNNR and, especially, linear regression (LR) perform less well. We prove that lagged values of forecasted time series slightly increase the accuracy of predictions. The same applies to seasonal and daily variability markers. Our studies have also demonstrated that using the full set of available input data and the best proposed hybrid and ensemble methods yield the lowest error. The proposed hybrid and ensemble methods are also applicable to other short-time generation forecasting in renewable energy sources (RES), e.g., in photovoltaic (PV) systems or hydropower.

Keywords: wind energy; wind turbine; hybrid methods; ensemble methods; short-term forecasting; electric energy production; machine learning; deep neural network; swarm intelligence

1. Introduction

Renewable energy sources have become a very important element of energy mixes in many countries across our planet. The majority of green energy is produced in large hydropower stations, wind and solar farms. However, more and more energy has been produced every year by various types of prosumer sources. A prosumer is most often

perceived as a user of photovoltaic systems. Nevertheless, some prosumers use small wind turbines to produce electricity. This may be particularly appropriate where insolation is low or wind conditions are favourable. Of course, small wind turbines are not as convenient as PV systems for prosumers. Small wind turbines need more free space around them and can be problematic due to noise and vibration that can adversely affect people and cause structural damage. That is why wind turbines are rarely installed on buildings.

Forecasting electricity generation for a small wind turbine is important both on a larger scale where there are many such turbines (because it creates problems for networks owned by distribution system operators), but also for prosumers due to the current energy consumption planning. It is also important for owners of small energy systems as it facilitates the use of various energy carriers and optimizes energy storage.

The problem of forecasting electricity generation for small wind turbines seems to be as complex as forecasting for large wind farms. The following reasons can be mentioned:

- Small wind turbines have small rotor inertia, therefore, any change of wind speed instantly affects energy generation;
- Towers of small turbines are low and forecasting of wind speed at low altitudes is burdened with large errors;
- Operation of small turbines can be affected by surrounding obstacles and rough terrain, and energy generation varies due to vegetation or lying snow.

The research presented in this article concerns small prosumer wind turbines. No actual wind speed data were collected, which made data analysis difficult.

1.1. Related Works

In recent decades, renewable energy sources (RES) has become an important way to address environmental concerns. RES shares have been rising at different rates across countries, e.g., for the European Union (EU) the share has tripled to 18% since 1990, while for Poland it has increased six-fold [1]. Wind energy is and will be important part of this equation. Since wind sources with their intermittency could become destabilizing factors for power system energy balance, different solutions were proposed by researchers to ease the situation. One aspect would be to determine maximum wind power penetration level in view of frequency response adequacy [2]. Another aspect would be good quality of generation predictions to allow us to properly plan for and increase wind power capacity to above the threshold [3].

Recent papers addressing wind power forecasts could be broadly classified into 5 categories: papers focused on how to increase NWP accuracy [4–8], good-practice prediction guidelines [9–11], comparisons of accuracy across prediction models [12–15], hybrid and ensemble methods [16–27], and conventional methods improved by, among other things, preprocessing [28–35]. At this point, clear distinction should be made between hybrid, ensemble and improved models. Methods classified as hybrid ones consist mainly of a set of forecasting engine tools without considering input data improvement tools part of the hybrid. Improved models, on the other hand, either preprocess input data using a traditional non-hybrid model or slightly alter/add to the model. Unlike both approaches, ensemble methods use different models working in parallel.

Papers addressing the topic of wind forecasts improvement [4–6] have attempted to incorporate Doppler light detection and ranging (lidar) or sodar systems readings into verification or enhancement of wind forecasts. The presented improvement proved to be the most promising for strong, volatile winds. Other papers [7,8] propose different approaches, like Gaussian process regression [8] and combination of artificial neural networks, ensemble learning, and feature selection techniques [7].

Research in [9–11] focused on adequate dealing with common problems associated with forecasting and machine learning. Tawn et al., 2020 [9] presented an approach for dealing with missing data, both for operational work and training of models. Messner et al., 2020 [10] demonstrate forecasting verification methods more fitted to perspective of the

forecast user, while Sewdien et al., 2020 [11] assessed the influence of different parameters on artificial neural network prediction accuracy.

Comparative studies [12–15] provide analyses of the forecasting quality of different single models. Mishra et al., 2020 [12] contrast deep neural networks, namely DFFNN, DCNN, RNN, AM and LSTM, both with and without preprocessing by FFT and DWT. Shetty et al., 2020 [13] confront ANN and the surface response method (RSM) with more fundamental models like cubic spline interpolation, least squares, power curve and power equation. The authors of [14] have chosen more complex methods, comparing LR with six machine learning (ML) models: multi-layer perceptron (MLP), Bayesian neural network (BNN), random forests (RF), gradient boosting trees (GBT), KNNR and SVR, while [15] compares MLP, SVR and ANFIS models. Overall winners of comparisons were AM in [12], ANN in [13] and ANFIS in [15]. In [14], the best model heavily depends on the performance indicator and no statement about any model superiority was made.

Hybridization [16–21] and parallelization [22–27] of prediction models use data-refining and error compensation, respectively, as an approach to maximize prediction accuracy. The most common bases for hybrid models in recent literature are ANNs [17–19,21] due to their generalization ability, while the most common hybrid add ons would be single optimization methods [16,18,20,21]. With varying implementation, error reduction can be achieved in different ways. For that purpose, Zhen et al., 2020 [17] propose a Bidirectional Long Short Term Memory–Convolutional Neural Network (BiLSTM-CNN) model for better extraction of spatio-temporal features from data and Zhou et al., 2020 [19] describe a long short-term memory–seasonal autoregressive integrated moving average (LSTM-SARIMA) model for capturing seasonality.

Ensemble models proposed in literature vary by the methods used and how output is aggregated. Studies pertaining to both switchable [22] and output-aggregated models [23–27] have been described in recent research papers. As for methods, researchers frequently use variations of regression trees [22–25]. Another approach is proposed in [26]. The solution suggested there is based on parallelization of stacked autoencoders for advanced feature extraction.

Papers concerning improved models are more data-focused. Durán-Rosal et al., 2018 [28] propose an algorithm for optimal reduction of the size of time series and the authors of [29] investigate the influence of reduced numerical weather prediction (NWP) data on forecast quality. Research by Yu et al., 2020 [30] and Fan et al., 2020 [31] focuses on assimilating spatial data with Graph Networks, while Li et al., 2020 [32] present an adaptive time resolution method to deal with the error hidden in the data due to error averaging. Some of the papers are focused on enhancing performance of long short-term memory (LSTM) networks. Shadid et al., 2020 [33] propose preprocessing by wavelets, an approach popular in recent years, while Zhang et al., 2020 [34] provide modification to LSTM by constructing error following forget gate. Other contributions, focused largely on improvements of the existing methods—kernel density estimation (KDE) [35] and others.

1.2. Objective and Contribution

The main objectives of this paper can be summarized as follows:

- Conduct statistical analysis of time series of energy generated from a small wind turbine and potential explanatory variables. Select 8 input data sets to verify how the types and amounts of input data impact on forecasts quality;
- Verify the accuracy of forecasts conducted by single methods, hybrid methods and ensemble methods (18 methods in total);
- Develop and verify an original ensemble method: an averaging ensemble based on hybrid methods without extreme forecasts. Conduct an original selection of combinations of predictors for ensemble methods;
- Indicate the most efficient forecasting methods with no historical wind speed measurements available, but with wind speed forecasts available for up to 48 h ahead.

Below are listed selected contributions of this paper:

1. This research applies to forecasting of small turbine generation using wind speeds predicted for up to 48 h. This problem is understudies in research so far, especially in comparison with forecasting for large wind farms;
2. Development of an ensemble method called "artificial neural network (ANN) type MLP as an integrator of ensemble based on hybrid methods", which includes a combination of original predictors, and has been arrived at by testing different combinations of predictors. Predictions by this method have yielded the lowest mean absolute error (MAE) among the tested methods;
3. Development of an original method called "Averaging ensemble based on hybrid methods without extreme forecasts". Predictions by this method have yielded the lowest root mean square error (RMSE) among the tested methods;
4. Completion of an extensive scenario analysis taking into account different degrees of data availability and model complexities. In total, more than 100 models with different parameters/hyperparameters were tested to choose an optimal model for this complex predictive problem.

The research carried out and the methods developed here contribute significantly to the topic of small wind turbine generation prediction, especially for 48 h time horizon and lack of historical wind speed measurements, but with wind speed forecasts up to 48 h ahead available instead. In practice, it is impossible to obtain acceptable generation forecasts for 48 h horizon without using wind speed forecasts as input data. Nevertheless, using other data as additional explanatory variables have further reduced the error.

The remainder of this paper is organized is as follows: Section 2.1 presents statistical analysis of times series of small wind turbine hourly generation and wind speed forecasts. The process of input data selection for particular forecasting methods is described in Section 2.2, and Section 3 specifies the forecasting methods used in the paper. Evaluation criteria used for the assessment of forecasting quality are presented in Section 4, followed by many-sided analysis of the obtained results and their discussion in Section 5. Finally, the main conclusions of our studies are summarized in Section 6, and references are listed at the end of this paper.

2. Data

2.1. Statistical Analysis of Data

Data acquired from a small prosumer turbine located in the south of Poland were used for statistical analyses (exact location is confidential). The location is characterized by rather low wind speeds. Rated power of the turbine was 5 kW, maximum power 7.5 kW, and tower height 13 m. The turbine's cut-in speed was 2.5 m/s, and the cut-out speed was 25 m/s.

The collected data consisted of hourly wind turbine power output and wind speed forecasts with a horizon from 1 to 48 h. Measurements of actual wind speeds were not available. The data covered almost 2 years (22.5 months), from January of one year to 14 November of the following year, with exact date and time range being confidential. The time series consisted of 16,392 consecutive hours.

2.1.1. Statistical Analysis of Time Series of Hourly Wind Turbine Generations

Table 1 shows descriptive statistics for time series of hourly electric energy generated by the wind turbine considered here.

Table 1. Descriptive statistics for hourly electric energy generation.

Descriptive Statistics	Value
Mean	0.650 [kWh]
Standard deviation	0.899 [kWh]
Minimum	0.000 [kWh]
Maximum	6.684 [kWh]
Coefficient of variation	138.13%
The 25th percentile (lower quartile)	0.000 [kWh]
The 50th percentile (median)	0.364 [kWh]
The 75th (upper quartile)	0.835 [kWh]
The 90 percentile	1.676 [kWh]
Variance	0.808
Skewness	2.578
Kurtosis	8.208

The analysis of electric energy generation percentiles analysis shows that 0 values made up more than 30% of the time series samples. Usually, energy generation was within 0–1 kWh range (47.35% of the samples). Values closest to maximum power ranged from 6 to 7 kWh and made up for slightly more than 0.1% of samples. Percentage distribution of electric energy generation is shown in Figure 1.
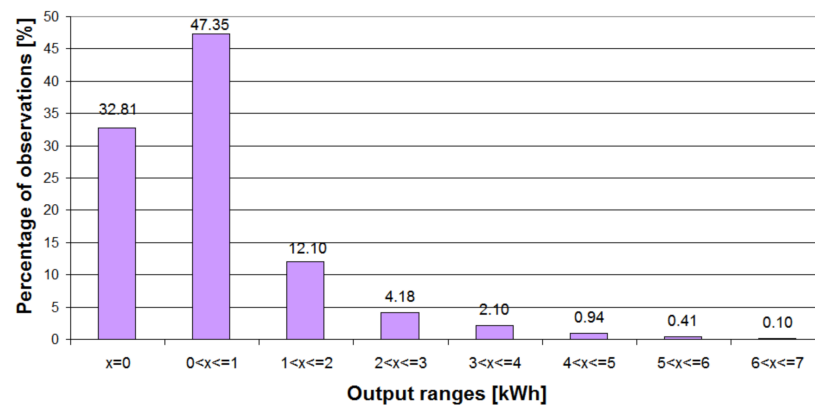
Figure 1. Percentage of time-series observations in particular generation ranges.

Autocorrelation coefficient (ACF) analysis of hourly generation time series shows slight daily periodicity, with increase of autocorrelation for number of backward periods equal to multiplicity of 24 h. Autocorrelation rapidly decreases for the consecutive hours of the first day. Details are presented in Figure 2. For the time series analyzed here, all autocorrelation coefficients are statistically significant (5% significance level) up to 7 days back (168 prior observations).
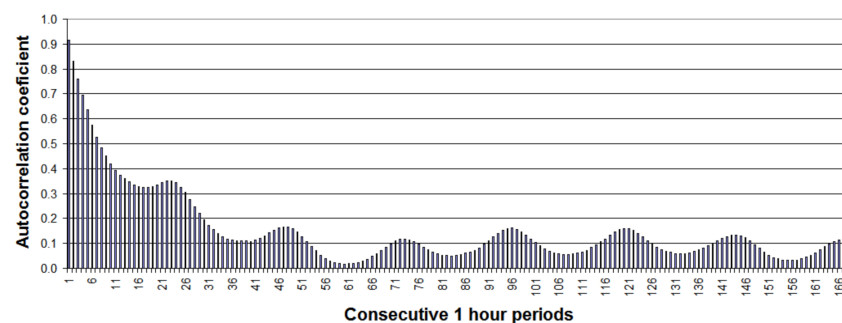
Figure 2. Autocorrelation function (ACF) of the wind turbine energy generation time series.

To check for daily variability of energy, hourly averages of generation were calculated for each hour of the day, based on the available 2-year dataset. Calculation results are shown in Figure 3. The analysis has demonstrated high daily variability of electric energy production. The ratio of generation for the period between 11:00 and 12:00 (maximum production) and generation for the period between 1:00 and 2:00 (minimum production) was 2.06.
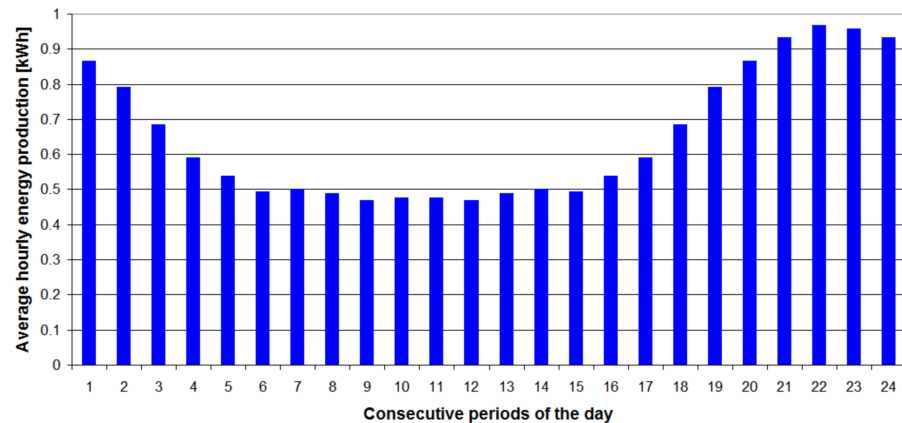

Figure 3. Daily variability of wind turbine energy generation.

To check for annual variability of energy, monthly averages of generation were calculated for each month of the year, based on the available 2-year dataset. Calculation results are shown in Figure 4.
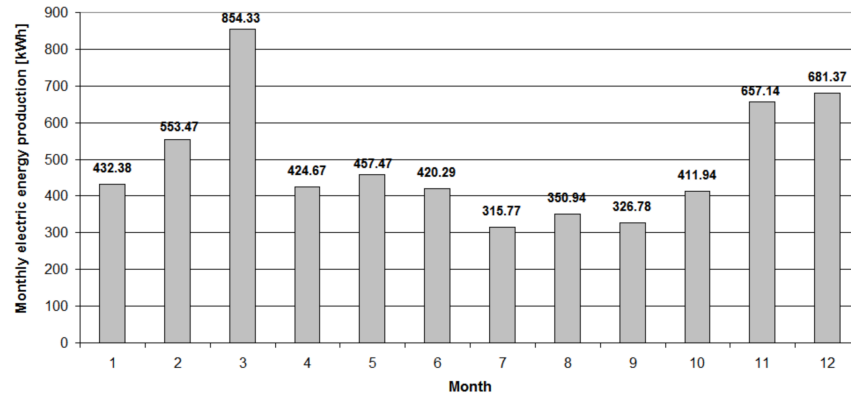

Figure 4. Monthly electric energy production—the mean over a 2-year period.

The analysis demonstrated strong seasonality of energy generation, with the highest production in winter months and lowest in summer ones. For both of the analyzed years, peak production was in March, which is an unusual, probably accidental phenomenon.

2.1.2. Statistical Analysis of Wind Speed Forecasts with Horizon from 1 to 48 h

Two days ahead wind speed forecasts were generated once per day, at 00:00 UTC. NWP unified model (UM) with 4 km × 4 km grid was used as the data source.

Wind speed forecasts for the period from *m-1* to *m* refer to mean forecasts for hours *m-1* and *m*. Descriptive statistics for the time series of forecasts with horizons from 1 to 24 h and forecasts with horizons from 25 to 48 h are shown in Table 2. For both horizons, means and variances were at the same level. For the 1–24 h horizon, mean, median, minimum and maximum wind speed forecasts were slightly higher.

Table 2. Descriptive statistics of wind speed forecasts with horizons from 1 to 24 h and 25 to 48 h.

Descriptive Statistics	Time Series of Wind Speed Forecasts for up to One Day Ahead (d + 1)	Time Series of Wind Speed Forecasts for the Second Day Ahead (d + 2)
Mean	3.592 [m/s]	3.626 [m/s]
Standard deviation	1.767 [m/s]	1.750 [m/s]
Minimum	0.088 [m/s]	0.228 [m/s]
Maximum	11.059 [m/s]	13.463 [m/s]
Coefficient of variation	49.18%	48.26%
The 25th percentile (lower quartile)	2.290 [m/s]	2.377 [m/s]
The 50th percentile (median)	3.284 [m/s]	3.341 [m/s]
The 75th (upper quartile)	4.672 [m/s]	4.641 [m/s]
The 90 percentile	6.089 [m/s]	6.023 [m/s]
Variance	3.121	3.063
Skewness	0.718	0.800
Kurtosis	0.277	0.697

2.2. Selection of Input Data for Particular Forecasting Methods

For input data selection, the Pearson linear correlation coefficient (R) was calculated between energy generation and potential explanatory variables. Analysis was made with omitting test-range data and results of analysis from chapter 2.1 were used for preliminary selection. Lagged inputs were selected by relying on ACF results and choosing 00:00 UTC of the following days as the starting point for 48 h generation forecasts, with horizons from 1 to 48 h. Table 3 describes results of R calculations and codes of input variables selected for use in explanatory data sets. All R values were statistically significant (5% significance level). R values of wind speed forecasts for the second day ahead were lower than for one day ahead, which suggests that wind speed forecasting accuracy decreases with increasing horizon. Wind speed was the most important input variable, with the highest R value. EN_L—coded input variables for different lag variants were given in a decomposed form. Separate results were given for horizons 1–24 h and 25–48 h for better visualization.

Table 3. Pearson linear correlation coefficients between selected input variables and energy generation.

Description of Input Data	Code of Input Data	R
Indicator of annual seasonality—mean daily energy generation for given month	MONTH_I	0.263
Indicator of daily variability of electric energy production—mean hourly energy generation for given hour	DAY_I	0.195
Wind speed forecasts—horizons from 1 to 24 h	V_F	0.756
Wind speed forecasts—horizons from 25 to 48 h	V_F	0.717
Electric energy production forecast by physical method—the function of polynomial degree 3	POLY_F	0.690
Hourly energy generation lagged by 24 h, for 1–24 h horizons	EN_L	0.343
Hourly energy generation lagged by 48 h, for 25–48 h horizons	EN_L	0.166
Hourly energy generation lagged by 25 h, for 1–24 h horizons	EN_L-1	0.326
Hourly energy generation lagged by 49 h, for 25–48 h horizons	EN_L-1	0.161
Hourly energy generation lagged by 26 h, for 1–24 h horizons	EN_L-2	0.304
Hourly energy generation lagged by 50 h, for 25–48 h horizons	EN_L-2	0.157
Hourly energy generation lagged by 48 h, for 1–24 h horizons	EN_L-24	0.166
Hourly energy generation lagged by 72 h, for 25–48 h horizons	EN_L-24	0.117
Hourly energy generation lagged by 72 h, for 1–24 h horizons	EN_L-48	0.117
Hourly energy generation lagged by 96 h, for 25–48 h horizons	EN_L-48	0.161

Figure 5 describes correlation between per unit values of hourly generations of electric energy and wind speed forecasts. Data were normalized to the <0.1> range and sorted in ascending order by energy production. Figure 5 demonstrates that for some production data with the zero value, NWP-UM wind speed forecasts had non-zero values, with some wind speeds above cut-in speed of the analyzed wind turbine.
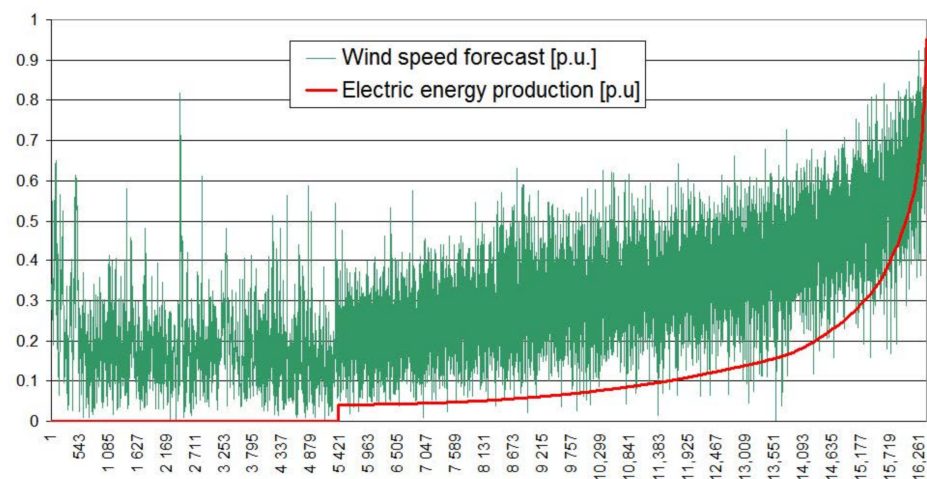
Figure 5. Correlation between hourly energy generations and wind speed forecasts.

A dispersion diagram for relationships between wind speed forecasts and actual energy generation [p.u.] is presented in Figure 6. The diagram demonstrates non-linear relationships between wind speed vector's module and production of electric energy. Data concentration is low, hence a wind turbine power curve typical of singular wind turbine is not clearly visible. This is mostly due to the differences between forecast and actual wind speeds. Extreme outliers were treated as unreliable data and were removed.
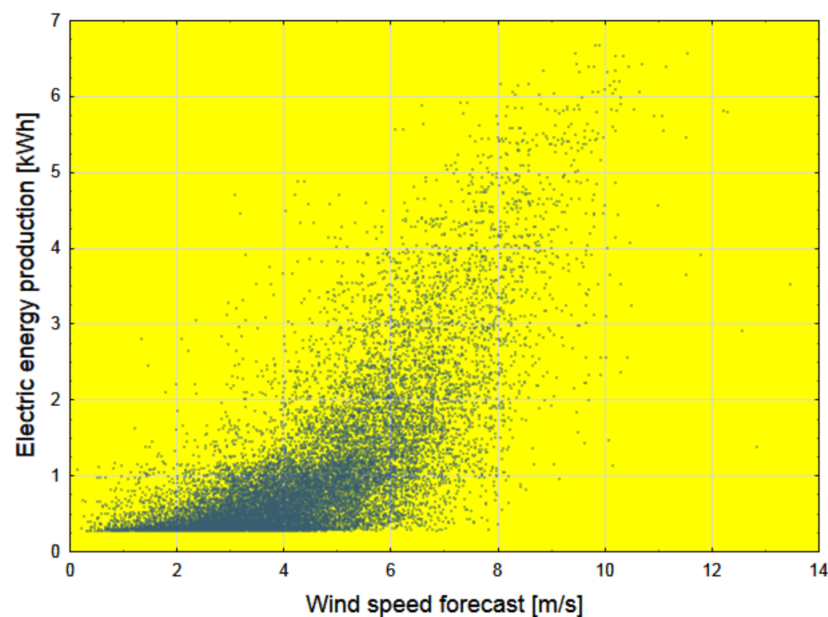
Figure 6. Relationship between wind speed forecasts [p.u.] and actual electricity production.

Eight different sets of input data with different information potential were proposed for the forecasts in order to analyze differences in the quality of forecasts between the sets. Table 4 presents sets of appropriate, selected sets of input data for the forecasting methods described in Section 3. Global sensitive analysis in the MLP model was used to eliminate unnecessary input data in the given set (from set 4 to set 8).

Table 4. Sets of input data selected for forecasting methods.

Name of Set	Description of Set	Codes of Input Data
Set 1	Time lag of forecasted time series	EN_L
Set 2	Selected time lags of forecasted time series	EN_L, EN_L-1, EN_L-2
Set 3	Selected time lags of forecasted time series	EN_L, EN_L-1, EN_L-2, EN_L-24, EN_L-48
Set 4	- Time lag of forecasted time series - Indicator of annual seasonality - Indictor of variability of daily electric energy production	EN_L, MONTH_I, DAY_I
Set 5	Wind speed forecast	V_F
Set 6	- Indictor of annual seasonality - Indictor of variability of daily electric energy production - Wind speed forecast	MONTH_I, DAY_I V_F
Set 7	- Time lag of forecasted time series - Indicator of annual seasonality - Indictor of variability of daily electric energy production - Wind speed forecast	EN_L, MONTH_I, DAY_I V_F
Set 8	- Time lag of forecasted time series - Indicator of annual seasonality - Indicator of variability of daily electric energy production - Wind speed forecast - Electric energy production forecast by physical model	EN_L, MONTH_I, DAY_I V_F POLY_F

3. Forecasting Methods

This section describes the methods employed in this paper. Forecasts are made using single methods based on time series only, other single methods using also additional exogenous variables including machine learning methods and the most advanced (complex)—ensemble methods. Below there is a brief description of the proposed methods. Naive model and naive smoothing model are benchmarks for the quality of other, more advanced forecasting methods.

Naive model (method code NAIVE). The naive model, which is simplest to implement, assumes that forecast generation values are identical to the actual energy generation values for the last known period being the multiplicity of 24. An unquestionable advantage of such model is simplicity and ability to take daily and seasonal variabilities of energy generation into account. Forecasts are calculated by the following Formula (1):

$$\hat{y}_t = y_{t-24 \cdot n} \tag{1}$$

where $\hat{y}_t$—forecast electric energy generated by wind turbine for hour t, $y_{t-24 \cdot n}$—energy generation for period lagged by $(t-24 \cdot n)$ from forecast period t, $n = 1$ for forecasting horizons from 1 to 24 h and $n = 2$ for horizons from 25 to 48 h.

Naive smoothing model (method code SMOOTHING). The naive model with smoothing assumes that forecast energy generation is the same as "smoothened" generation for the last known period being the multiplicity of 24. "Smoothened" production from the last known period is calculated as weighted average, with weight of production from $t-24 \cdot n$ equal to 0.5 and weights for productions from $t-24 \cdot n-1$ and $t-24 \cdot n+1$ and equal to 0.25.

$$\hat{y}_t = 0.5 \cdot y_{t-24 \cdot n} + 0.25 \cdot y_{t-24 \cdot n-1} + 0.25 \cdot y_{t-24 \cdot n+1} \tag{2}$$

where $\hat{y}_t$—energy generation forecast for hour t, $n = 1$ for horizons 1 to 23, and $n = 2$ for horizons 24 to 48.

Physical model (method code PHM). This forecasting model of generated hourly power is a function of forecasted wind speed. Form of function is the 4th order polynomial described by Formula (3). Catalogue data of wind turbine power curve were used to develop the function equation, with data points responding to powers for 3–25 m/s wind

speeds, with 1 m/s steps. Second, third, and fourth order of polynomials were tested and 4th order function was chosen due to the highest determination coefficient R^2, equal 0.9964.

$$\hat{y}_t(v) = 0.0005 \cdot v_t^4 - 0.0285 \cdot v_t^3 + 0.4992 \cdot v_t^2 - 2.5708 \cdot v_t + 4.3363 \tag{3}$$

where $\hat{y}_t(v_t)$—energy generation forecast for period t, and v_t—wind speed forecast for t.

Multiple Linear regression model (method code LR). It is a linear model that assumes a linear relationship between the input variables and the single output variable [36]. The input data are selected lags of the forecasted output variable and other input explanatory variables correlated to the output variable. The model is fitted using the least squares approach.

K-Nearest Neighbours Regression (method code KNNR). This algorithm is a non-parametric method used for classification and regression [37]. The input consists of the k closest training examples in the feature space. In KNN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors. The main hyperparameter for tuning is the number of nearest neighbors. Distance metric is the second hyperparameter.

Support Vector Regression (method code SVR). SVM for regression of the Gaussian kernel (non-linear regression) transforms the classification task into regression by defining width ε tolerance region around the destination [38]. The learning task is reduced to the quadratic optimization problem and depends on few hyperparameters: regularization constant C, width parameter s of the Gaussian kernel and tolerance ε.

Artificial neural network, type MLP. It is a class of feedforward ANNs. MLP is a popular and effective non-linear or linear (depending on the type of activation function in hidden layer(s) and output layer) global approximator [29,36,39]. It consists of a single input layer, typically has one or two hidden layers, one output layer and uses the back-propagation algorithm for supervised learning. The main hyperparameter for tuning is the number of neurons in hidden layer(s). Two MLP models differing by the optimization algorithm are used for the forecasts.

- The BFGS method utilized for solving unconstrained non-linear optimization problems was chosen as a learning algorithm of a neural network (method code MLP$_{\text{BFGS}}$).
- The particle swarm optimization (PSO) method was used to determine MLP weights (method coded MLP$_{\text{PSO}}$). This hybrid combination of PSO and MLP has been implemented in an original computer program by one of the authors. The following tuning hyperparameters were investigated in our research: number of neurons in hidden layers (5, 8, 10, 20), number of particles in the swarm (20, 50, 100, 150), number of algorithm iterations (100, 1000), coefficients in the formula for particles movements (0.7, 1.0, 1.4), number of iterations between subsequent disturbances in the swarm (50, 100, 400), neighborhood width (2, 5, 10). The optimization concept is presented in Figures 7 and 8.
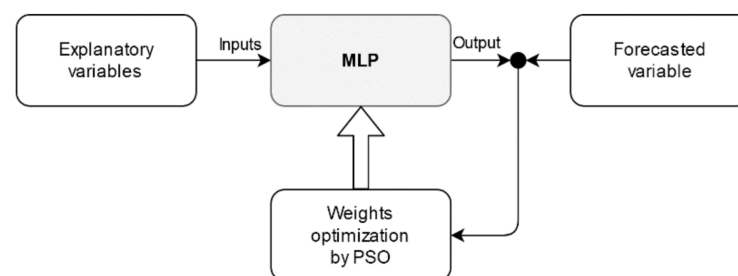

Figure 7. General concept of using particle swarm optimization (PSO) algorithm for multi-layer perceptron (MLP) training.
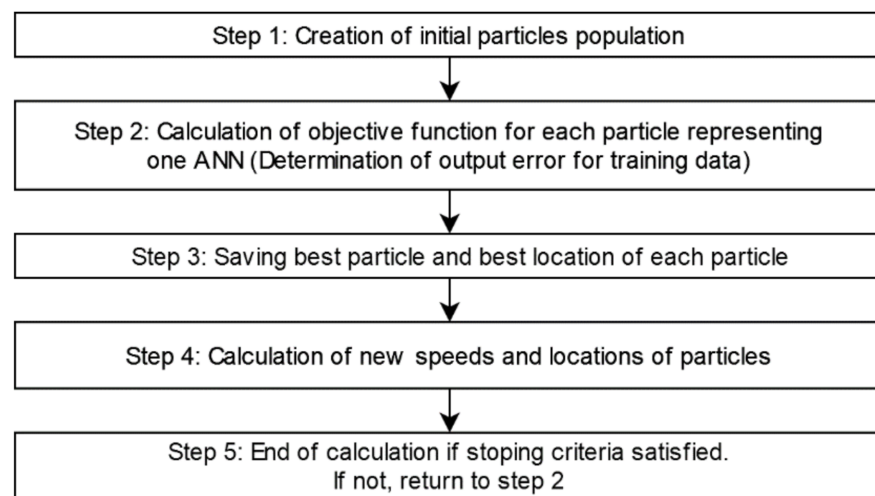
Figure 8. General algorithm of particle swarm optimization artificial neural network (PSO-ANN).

Deep neural network type LSTM (method code LSTM). It is a type of recurrent ANN. Internal modules different from the traditional RNNs allow LSTM to avoid problems with long-term dependencies [3,24], gradient explosion and vanishing [8,10]. Typically, a LSTM network consists of one input layer, one or two LSTM layers, and a dense output layer. Hidden LSTM layers consist of neurons for which input gate, forget gate and output gate are responsible for selective control of information [24,25]. Detailed information about algorithms incorporated in LSTM networks is presented in these papers. Between layers, dropout layers can be used to prevent model overfitting [3,8]. The principle of this mechanism for each node is to retain given node with a probability according to Bernoulli distribution, and dropping from network with complementary probability [40]. Settable hyperparameters for the LSTM model would be mainly the number of hidden layers and neurons in them and each layer activation function, but also batch size, number of training epochs and dropout degree. Besides those, various model optimizers like AdaGrad, RMSProp or ADAM can be used [3,10,24]. For optimal LSTM model selection, ReLU/ELU/PReLU/LeakyReLU/sigmoid/tanh activation functions were used for hidden layers, and sigmoid/linear/tanh/ ReLU were used for the output layer. The tested networks had one or two hidden layers with different combinations of neurons in them. Data shuffling and patience mechanisms were used. Networks were trained for 2000 epochs, with patience from 100 to 500, by the ADAM optimizer with lr = 0.001 and decay = 1×10^{-5}. To decrease computation time, batch size = 128 was used.

Hybrid methods. Hybrid methods are the physical model (PHM) and a single method connected in series. In these methods, information flows between two models. The first model receives forecast wind speeds (set 5) as input data. The following model receives not only output from the first model (generation forecast), but also endogenous and exogenous (set 7) input data. Both sets together form set 8. The concept of hybrid methods is described in Figure 9 and used for the following pairs of single methods connected in series:

- Physical model and multiple linear regression model (method code PHM+LR),
- Physical model and K-nearest neighbours regression (method code PHM+KNNR),
- Physical model support vector regression (method code PHM+SVR),
- Physical model and deep neural network type LSTM (method code PHM+LSTM),
- Physical model and artificial neural network type MLP (method codes: PHM+ML$_{BFGS}$ and PHM+MLP$_{PSO}$).
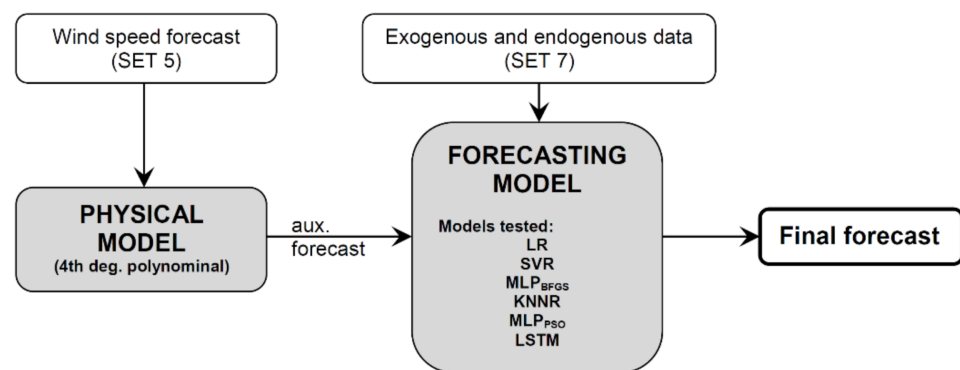
Figure 9. Diagram of series information processing in hybrid method.

Ensemble methods based on hybrid methods. This category of methods use more than one individual predictor and are supported by a simple or more complex integration system of individual forecasts. The general scheme of ensemble of predictors based on hybrid methods is presented in Figure 10. The simplest integration system is the weighted averaging of individual predictors and the most advanced integrator system is the ANN.
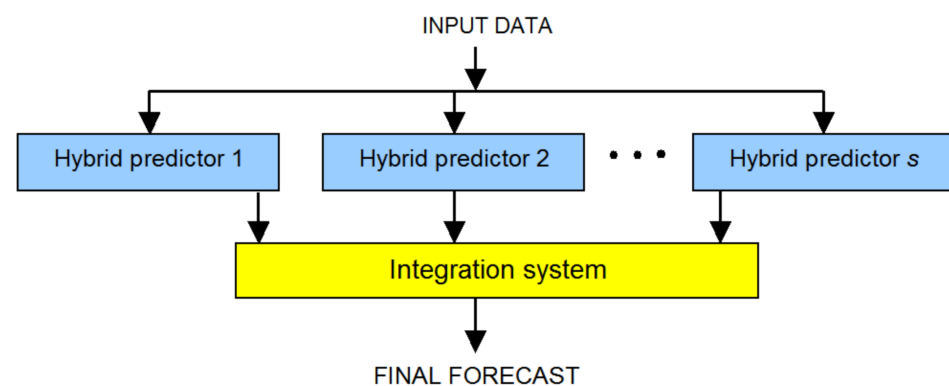

Figure 10. The general scheme of ensemble of predictors based on hybrid methods.

Four types of integration system were used for forecasts using "ensemble methods based on hybrid methods":

(1) Averaging ensemble based on hybrid methods (method code AVE_INT). It integrates the results of selected predictors into final verdict of the ensemble. The final forecast is defined as the average of the results produced by all *s* hybrid predictors organized in the ensemble [38]. The final prediction result is calculated by Formula (4). This formula uses stochastic distribution of predictive errors. The averaging of the forecast results is an established method of reducing the variance of forecast errors. Two main strategies of predictor choice were tested. An important condition for including the predictor into the ensemble is mutually independent operation and also similar levels of prediction error [38]. Hybrid predictors are included in the ensemble based on the smallest MAE errors on the validation subset and only predictors of different types.

$$\hat{y}_i = \frac{1}{s}\sum_{j=1}^{s} \hat{y}_i^j \qquad (4)$$

where, i is the prediction point, $\hat{y}_i$ is the final predicted value, $\hat{y}_i^j$ is predicted value by hybrid predictor number j, and s is the number of hybrid predictors in the ensemble.

(2) An ensemble based on hybrid methods with weight optimization for each predictor (method code W_OPT_INT). Each hybrid predictor in the ensemble has an individual weight. The optimal weights for each predictor in ensemble are calculated using the social

cognitive optimization (SCO) method on the validation data (forecasts from each predictor in the ensemble are input data) while minimizing MAE error of final forecasts (output data). SCO is a population-based metaheuristic optimization algorithm developed in 2002 [41]. This algorithm is based on the social cognitive theory. The key point of the ergodicity is the process of individual learning of a set of agents with their own memory and their social learning with the knowledge points in the social sharing library. The final prediction result is calculated by Formula (5). For w_s (weights) optimization by SCO, no limits were set for weight ranges or their sum.

$$\hat{y}_i = w_1 \cdot \hat{y}_i^1 + w_2 \cdot \hat{y}_i^2 +, \ldots, + w_s \cdot \hat{y}_i^s \tag{5}$$

where, i is the prediction point, $\hat{y}_i$ is the final predicted value, $\hat{y}_i^s$ is the predicted value by hybrid predictor number s, w_s is the weight associated with the forecast from the hybrid predictor with number s, and s is the number of hybrid predictors in the ensemble.

(3) **Averaging ensemble based on hybrid methods without extreme forecasts (method code MIN&MAX_SKIP).** This method assumes removing the minimum and maximum forecast from the set of hybrid predictors before each calculation of single final forecasts, being an average of forecasts from s hybrid predictors. For a 48 h horizon, removal is done 48 times for each forecast separately. Such a procedure should in theory decrease prediction errors (MAE and RMSE) and increase the value of R. An important condition for including the predictor into the ensemble is mutually independent operation and also similar levels of prediction error. The final prediction result is calculated by Formula (6).

$$\hat{y}_i = \frac{1}{s-2} \cdot \left(\sum_{k=1}^{s} \hat{y}_i^k - \min\left\{ \hat{y}_i^k \right\} - \max\left\{ \hat{y}_i^k \right\} \right) \tag{6}$$

where, i is the prediction point, $\hat{y}_i$ is the final predicted value, $\hat{y}_i^k$ is the predicted value by hybrid predictor number k, and s is the number of hybrid predictors in primary ensemble before the elimination of the outputs of predictors yielding extreme forecasts from the set of results.

(4) **ANN type MLP as a integrator of ensemble based on hybrid methods** (method code MLP_INT). It incorporates the results of selected predictors into final verdict of the ensemble using the MLP model. Finally, four hybrid predictors are chosen for the ensemble based on the smallest MAE errors on validation subset and predictors of differing type. The MLP integrator uses forecasts from individual hybrid predictors as input data, and the actual value of electric energy production from the wind turbine is the output. Dataset training is used for the training of the MLP integrator and validation dataset for MAE control, and the tuning of hyperparameters. Finally, the evaluation criteria are checked on a test data set. The general structure of MLP as an integrator of ensembles is presented in Figure 11.
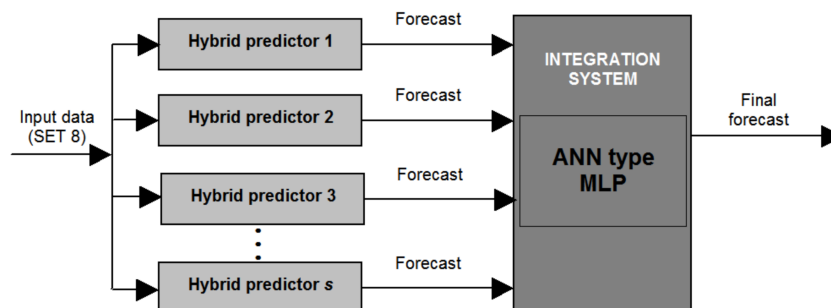

Figure 11. General structure of MLP as an integrator of ensemble of predictors based on hybrid predictors.

Different sets of explanatory variables with different information potentials were used for forecasts by single methods. The least information was contained in sets of variables that use only selected lagged variables of forecasted energy generation time series (set 1, 2, 3). The largest explanatory data set (set 8) was used for predictions by hybrid and ensemble methods. Table 5 shows tested input data sets for each method. One reason for organizing data into such sets was to verify the influence of the type and number of variables on forecast accuracy.

Table 5. Tested input data sets for each method.

The Method Code	Type of Method	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
NAIVE	Linear	+ *							
SMOOTHING	Linear		+						
PHM	Non-linear					+			
LR	Linear			+	+		+	+	
SVR	Non-linear			+	+		+	+	
MLP$_{BFGS}$	Non-linear			+	+		+	+	
MLP$_{PSO}$	Non-linear			+	+		+	+	
LSTM	Non-linear			+	+		+	+	
PHM+LR	Non-linear								+
PHM+SVR	Non-linear								+
PHM+MLP$_{BFGS}$	Non-linear								+
PHM+KNNR	Non-linear								+
PHM+MLP$_{PSO}$	Non-linear								+
PHM+LSTM	Non-linear								+
AVE_INT [PHM+LSTM, PHM+SVR PHM+MLP$_{PSO}$, PHM+KNNR]	Non-linear								+
AVE_INT [PHM+LSTM, PHM+KNNR, PHM+MLP$_{PSO}$]	Non-linear								+
W_OPT_INT [PHM+LSTM, PHM+KNNR, PHM+MLP$_{PSO,}$]	Non-linear								+
W_OPT_INT [PHM+LSTM, PHM+SVR, PHM+MLP$_{PSO}$, PHM+KNNR,]	Non-linear								+
MIN&MAX_SKIP [PHM+MLP$_{PSO}$, PHM+KNNR, PHM+LSTM, PHM+SVR]	Non-linear								+
MIN&MAX_SKIP [PHM+MLP$_{PSO}$, PHM+LR, PHM+LSTM, PHM+SVR, PHM+KNNR]	Non-linear								+
MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR]	Non-linear								+

Note: —* tested input data set.

4. Evaluation Criteria

For the performance tests of the methods, six evaluation criteria (measures of errors) are used, including RMSE, MAE, Pearson linear correlation coefficient (R), mean bias error (MBE), 75th percentile of the absolute errors (AE) and 99th percentile of the AE.

Mean absolute error is calculated by Formula (7)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

In the process of forecasting of wind turbine electric energy production, changes of RMSE and MAE follow the same trend, and the smaller the two errors, the more accurate the prediction results. MAE is related to the first order of error moment while RMSE is related to the second order.

Root mean square error, which is sensitive to large errors, is calculated by Formula (8):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

where, $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of prediction points.

Pearson linear correlation coefficient of the observed and predicted data is calculated by Formula (9):

$$R = \frac{C_{y\hat{y}}}{std(y) \cdot std(\hat{y})} \tag{9}$$

where, $C_{y\hat{y}}$ is the covariance between the actually observed and predicted data and std denotes standard deviation of the appropriate variable.

The bigger the error R value (range from -1 to 1), the more accurate the prediction results.

Mean bias error (MBE) captures the average bias in the prediction and is calculated by Formula (10):

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{10}$$

The value of single *i*-th Absolute Error (AE) needed for calculation of percentiles of AE errors is calculated by Formula (11):

$$AE_i = |y_i - \hat{y}_i| \tag{11}$$

The 75th percentile (PCTL75AE) is the value of AE error below which there are 75% of all AE errors, and it indicates very well the density of AE errors.

Similarly, the 99th percentile (PCTL99AE) is the value of AE error below which there are 99% of all AE errors, and it indicates very well the level of the biggest AE errors.

Measures of errors (MAE, RMSE) are basic measures to evaluate the accuracy of proposed models, while others measures (R, MBE, PCTL75 and PCTL99) are auxiliary.

5. Results and Discussion

Predictions were conducted sequentially, from single methods with a limited number of input variables to hybrid methods and ensemble methods using all selected input variables (set 7, 8). Such a procedure allows us to observe differences in the quality of results depending on the complexity of particular methods and the range of input variables used.

Data available from the 2-year period were divided into training, validation and test sets. Eighty-five percent of the first year's data were used as the training data set and the remaining 15% constituted the validation set.

Data for both data sets were chosen at random from the first year's data set. The second year's data set constituted the test set used for one-time final evaluation of the quality of specific prediction methods on data containing all seasons.

Table 6 provides the classification of forecasts by the range of input variables used, with class number increasing with the range of data used. It should be noted that no wind speed measurements were acquired, so they could not be used in the forecasting process, which is not unusual for small wind turbines. On the other hand, wind speed forecasts can be purchased, and they are the most important explanatory variable.

Table 6. Tested input data sets for each class forecast.

Class No.	Description of Input Data	Input Data Sets (Depend on Method)
Class 1	Only selected previously observed value of forecasted time series	set 1, set 2, set 3
Class 2	Selected previously observed values of forecasted time series and indicators of annual seasonality and variability of daily electric energy production	set 4
Class 3	Wind speed forecast and indicators of annual seasonality and variability of the daily electric energy production (without forecasted time series data)	set 5, set 6
Class 4	Selected previously observed values of forecasted time series, indicators of annual seasonality and variability of daily electric energy production and wind speed forecast	set 7
Class 5	Selected previously observed values of forecasted time series, indicators of annual seasonality and variability of the daily electric energy production, wind speed forecast and electric energy production forecast by physical model	set 8

Forecasts of Class 1 have practical use where the wind turbine has been installed relatively recently and wind speed forecasts are unavailable. NAIVE and SMOOTHING models in particular do not need collecting energy production time series from date times more than 48 h backwards. Class 1 forecasts are predictions based only on time series of forecasted process of energy production, hence their accuracy is low. Class 2 forecasts can be used if seasonal and daily variability markers have been calculated (if the time series covers at least 1 year) and wind speed forecasts are not used due to data acquisition cost. Tests of class 3 use predictions of wind speed and try to check whether lagged energy production can be included in inputs without decreasing the accuracy of forecasting. These tests verify if collecting energy production time series is reasonable. Class 4 represents forecasts using all available input data, while class 5 additionally uses electric energy production forecast as an output from physical model and input to another single method, which makes it a hybrid structure composed of two single methods connected in series.

The performance indicators on the test subset are presented separately for each class in Tables 7–12. Results are sorted in descending order by MAE values. Best result for each quality measure is bold-faced. A detailed description of the results of hyperparameters tuning for tested hybrid and ensemble methods (Tables 11 and 12) is provided in Table A1 of Appendix A.

Table 7. Performance measures for single methods with input data class 1.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
LSTM	Set 3	**0.4841**	**0.8029**	0.1142	0.3632	2.7985	0.2416
MLP$_{PSO}$	Set 3	0.4842	0.8139	0.2052	**0.0954**	**0.5252**	**0.2455**
SVR	Set 3	0.4975	0.8173	0.1643	0.1151	0.8584	0.1587
LR	Set 3	0.5183	0.8556	0.3166	0.2923	1.0914	0.2301
MLP$_{BFGS}$	Set 3	0.5304	0.9105	0.3410	0.4291	2.9634	0.2335
SMOOTHING	Set 2	0.6244	0.9807	0.0323	0.4064	3.3131	0.2260
NAIVE	Set 1	0.6336	1.0026	**0.0322**	0.3997	3.6138	0.2193
Average measures for single methods with input data class 1		0.5389	0.8834	0.1723	0.3002	2.1663	0.2221

Notes: the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Table 8. Performance measures for single methods with input data class 2.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
LSTM	Set 4	**0.4788**	**0.7911**	0.1184	0.3491	2.6987	0.3323
LR	Set 4	0.4802	0.7927	0.1823	**0.1408**	0.7133	0.3175
MLP$_{PSO}$	Set 4	0.4823	0.7983	0.2224	0.1815	**0.5256**	0.3417
SVR	Set 4	0.5050	0.8280	0.2832	0.2109	0.9387	0.2776
MLP$_{BFGS}$	Set 4	0.5204	0.8912	**0.0322**	0.4114	2.9430	**0.3505**
Average performance measures for single methods with input data class 2		0.4934	0.8203	0.1677	0.2587	1.5639	0.3239

Notes: —the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Table 9. Performance measures for single methods with input data class 3.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
LSTM	Set 6	**0.2860**	0.4812	0.1189	0.3520	2.7899	0.8176
MLP$_{PSO}$	Set 6	0.2902	0.4711	0.0891	0.3693	2.7694	**0.8208**
MLP$_{BFGS}$	Set 6	0.2961	**0.4688**	**0.0328**	0.4041	2.8710	0.8177
SVR	Set 6	0.3299	0.5052	0.1448	**0.2796**	2.5782	0.8064
LR	Set 6	0.3562	0.5981	0.1054	0.4612	**1.2190**	0.7068
PHM	Set 5	0.4389	0.6155	−0.0662	0.4691	2.7801	0.6540
Average performance measures for single methods with input data class 3		0.3329	0.5233	0.0708	0.3892	2.5013	0.7706

Notes: the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Table 10. Performance measures for single methods with input data class 4.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
LSTM	Set 7	**0.2854**	0.4805	0.1155	0.3514	2.7834	0.8188
MLP$_{PSO}$	Set 7	0.2909	0.4691	0.0942	0.3717	2.6131	**0.8270**
MLP$_{BFGS}$	Set 7	0.2929	**0.4594**	**0.0293**	0.4140	2.8527	0.8253
SVR	Set 7	0.3412	0.5177	0.1474	**0.2988**	2.5490	0.7937
LR	Set 7	0.3531	0.6055	0.1068	0.4512	**1.2243**	0.7145
Average performance measures for single methods with input data class 4		0.3127	0.5064	0.0986	0.3774	2.4045	0.7958

Notes:—the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Table 11. Performance measures for hybrid methods with input data class 5.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
PHM+LSTM	Set 8	**0.2823**	0.4702	0.0830	0.3661	2.8905	0.8202
PHM+MLP$_{PSO}$	Set 8	0.2868	0.4727	0.0943	0.3668	2.8300	0.8209
PHM+MLP$_{BFGS}$	Set 8	0.2914	**0.4595**	**0.0329**	0.4062	2.8762	**0.8260**
PHM+KNNR	Set 8	0.2996	0.4746	0.0456	0.3763	2.6014	0.8148
PHM+SVR	Set 8	0.3239	0.4979	0.1294	**0.2989**	2.5510	0.8088
PHM+LR	Set 8	0.3380	0.5457	0.0984	0.3899	**1.9396**	0.7677
Average performance measures for hybrid methods with input data class 5		0.3037	0.4868	0.0806	0.3674	2.6148	0.8097

Notes: the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Table 12. Performance measures for ensemble methods based on hybrid methods with input data class 5.

The Method Code	Input Data Set	MAE [kWh]	RMSE [kWh]	MBE [kWh]	PCTL75AE [kWh]	PCTL99AE [kWh]	R
MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR]	Set 8	**0.2816**	0.4679	0.0855	0.3650	2.7511	0.8206
W_OPT_INT [PHM+LSTM, PHM+KNNR, PHM+MLP$_{PSO}$]	Set 8	0.2819	0.4681	0.0973	0.3560	2.7173	0.8205
AVE_INT [PHM+LSTM, PHM+KNNR, PHM+MLP$_{PSO}$]	Set 8	0.2825	0.4701	0.0972	0.3616	2.7871	0.8199
MIN&MAX_SKIP [PHM+MLP$_{PSO}$, PHM+KNNR, PHM+LSTM, PHM+SVR]	Set 8	0.2858	**0.4575**	0.0844	**0.3487**	**2.6534**	**0.8259**
W_OPT_INT [PHM+LSTM, PHM+SVR, PHM+MLP$_{PSO}$, PHM+KNNR,]	Set 8	0.2860	0.4696	0.0911	0.3563	2.8441	0.8228
AVE_INT [PHM+LSTM, PHM+SVR, PHM+MLP$_{PSO}$, PHM+KNNR]	Set 8	0.2861	0.4708	0.0929	0.3572	2.8479	0.8221
MIN&MAX_SKIP [PHM+MLP$_{PSO}$, PHM+LR, PHM+LSTM, PHM+SVR, PHM+KNNR]	Set 8	0.2909	0.4677	**0.0752**	0.3488	2.6536	0.8258
Average performance measures for ensemble methods based on hybrid methods with input data class 5		0.2850	0.4688	0.0891	0.3562	2.7506	0.8224

Notes: the best fitting results for each fitting measure is printed in bold in blue. The worst fitting result is printed in red.

Based on the results from Table 7, it is possible to draw the following partial conclusions which concern proposed single methods with input data class 1:

- Regarding MAE, qualitative differences between the best two methods (LSTM and MLP$_{PSO}$) are very small;
- The two clearly worst methods in terms of MAE are NAIVE (the reference method) and SMOOTHING;
- Regarding performance measures for the MLP model, PSO is clearly superior to BFGS as a weight optimization method;
- Values of R are very small and similar for all seven methods, which clearly indicates that single methods with class 1 input data are of little value for the forecasting of wind turbine generation.

Based on the results from Table 8, the following partial conclusions can be drawn concerning the proposed single methods with input data class 2:

- Regarding MAE, qualitative differences between the three best methods (LSTM, LR and MLP$_{PSO}$) are very small; linear method LR has surprisingly ranked second-best,
- The two clearly worst methods in terms of MAE are SVR and MLP$_{BFGS}$,
- Regarding performance measures, for the MLP model, PSO is clearly superior to BFGS as a weight optimization method,
- Values of R are very small and similar for all five methods, which clearly indicates that single methods with class 2 input data are of little value for the forecasting of wind turbine generation.

Based on the results from Table 9, the following partial conclusions can be drawn, regarding the proposed single methods with input data class 3:

- MAE, RMSE and R have demonstrated clear improvement as compared to the results of Class 2. The reason is that using for Class 3 as an input wind speed forecasts the most important explanatory variable.
- Regarding MAE, qualitative differences between the best three methods (LSTM, MLP$_{PSO}$ and MLP$_{BFGS}$) are very small,
- In terms of MAE value, PHM is clearly the worst method. This is due to using only wind speed forecasts as input data (set 5),

- The rank of linear method LR by MAE has notably decreased. It is the second worst method in the ranking, so it can be concluded that it is of little value as compared to non-linear methods when wind speed forecasts are included in input data.

Based on the results from Table 10, the following partial conclusions can be drawn regarding the proposed single methods with input data class 4:

- Performance measures have slightly improved as compared to the results of Class 3. This is due to additional use of selected, previously observed values of forecasted time series;
- Taking into account MAE value, qualitative differences among the three best methods (LSTM, MLP$_{PSO}$ and MLP$_{BFGS}$) are very small;
- MLP$_{BFGS}$ method deserves attention as its RMSE is not only the lowest, but also visibly less than for the LSTM method with the lowest MAE;
- The two clearly worst methods in terms of MAE value are SVR and LR. In particular, linear method LR seems to be of little value.

Based on the results from Table 11, the following partial conclusions can be drawn regarding the proposed hybrid methods with input data class 5:

- Performance measures slightly improved in comparison with results of Class 4. This is due to using a hybrid method—the energy generation forecast from PHM model is additional input data for another, more advanced model;
- Taking into account MAE value, qualitative differences among the two best methods (PHM+LSTM and PHM+MLPPSO) are very small,
- MLPBFGS method deserves attention as its RMSE is the lowest,
- The two clearly worst methods in terms of MAE value are PHM+SVR and PHM+LR. In particular, linear method LR seems to be of little value.

Based on the results from Table 12, the following partial conclusions can be drawn regarding the proposed ensemble methods based on hybrid methods with input data class 5:

- Performance measures have slightly improved as compared to the results of hybrid methods with input data of Class 5. This is due to the use of different integration systems to achieve the final forecast with the use of selected hybrid methods;
- Taking into account MAE value, qualitative differences among ensemble methods based on hybrid methods are quite small except the last method MIN&MAX_SKIP [PHM+MLPPSO, PHM+LR, PHM+LSTM, PHM+SVR, PHM+KNNR] using 5 hybrid methods;
- Special attention should be paid to MIN&MAX_SKIP [PHM+MLPPSO, PHM+KNNR, PHM+LSTM, PHM+SVR] using 4 instead of 5 hybrid methods (PHM+LR removed from ensemble). MAE of this method is similar to the value for the best method in the MLP_INT [PHM+LSTM, PHM+MLPPSO, PHM+KNNR] ranking. At the same time, it features the lowest RMSE, PCTL75AE, PCTL99AE and the highest value of R.

Table 13 and Figure 12 provide a collective set of results from Tables 7–12, made by choosing best MAE result from each table. The same, but with RMSE, applies to Table 14 and Figure 13. Moreover, Tables 13 and 14 contain percentage differences between various methods in relation to the method with the best value of the given error measure. A percentage value with the positive sign means a difference in favor of the given method.

Table 13. Collective set of best results with respect to MAE for each class of method.

Class of Method	Method Code	MAE [kWh]	Difference [%]
The ensemble method based on hybrid methods with input data class 5	MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR]	0.2816	-
The hybrid method with input data class 5	PHM+LSTM	0.2823	0.25
The single method with input data class 4	LSTM	0.2854	1.35
The single method with input data class 3	LSTM	0.2860	1.56
The single method with input data class 2	LSTM	0.4788	70.03
The single method with input data class 1	LSTM	0.4841	71.91
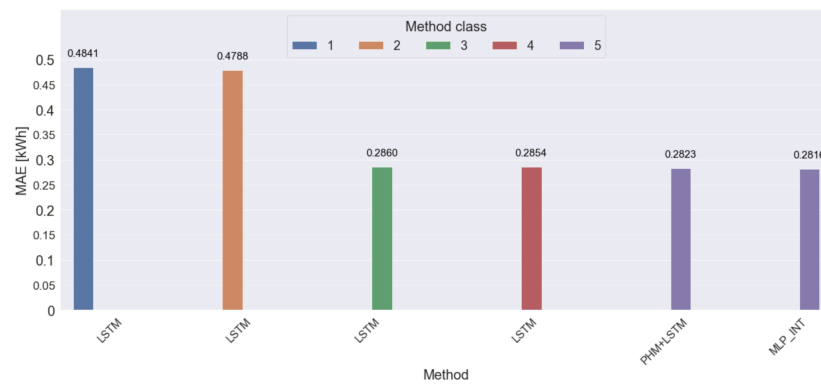
Figure 12. Collective set of best results with respect to MAE for each class of method.

Table 14. Collective set of best results with respect to RMSE for each class of method.

| Class of Method	Method Code	RMSE [kWh]	Difference [%]
The ensemble method based on hybrid methods with input data class 5	MIN&MAX_SKIP [PHM+MLP$_{PSO}$, PHM+KNNR, PHM+LSTM, PHM+SVR]	0.4575	-
The single method with input data class 4	MLP$_{BFGS}$	0.4594	0.42
The hybrid method with input data class 5	PHM+MLP$_{BFGS}$	0.4595	0.44
The single method with input data class 3	MLP$_{BFGS}$	0.4688	2.47
The single method with input data class 2	LSTM	0.7911	72.92
The single method with input data class 1	LSTM	0.8029	75.50
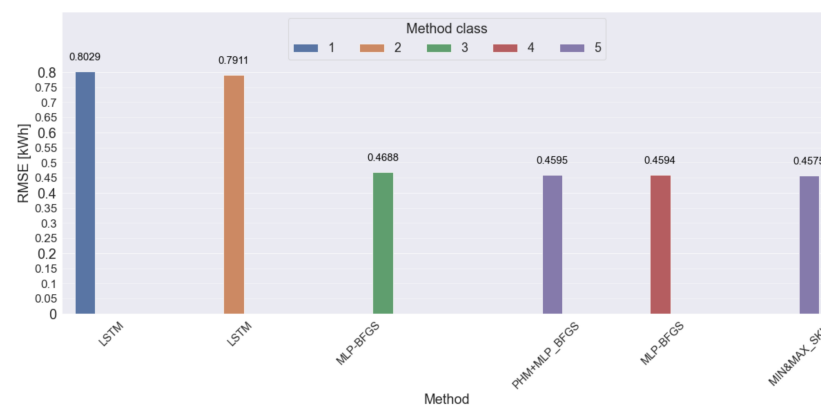
Figure 13. Collective set of best results with respect to MAE for each class of method.

Figure 14 provides a forecast from 1 to 48 h made by using method MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR] for 2 following days of an autumn month (October). The forecast curve smoothing effect is clearly visible. Figure 15 demonstrates a forecast from 1 to 48 h made by using the above method for two following days of a spring month (April). For the spring forecast, the curve smoothing effect is slightly less visible and prediction is less accurate, particularly for the second day, which is normal, because the accuracy of wind speed forecasts used as input data decreases with increasing time horizon.
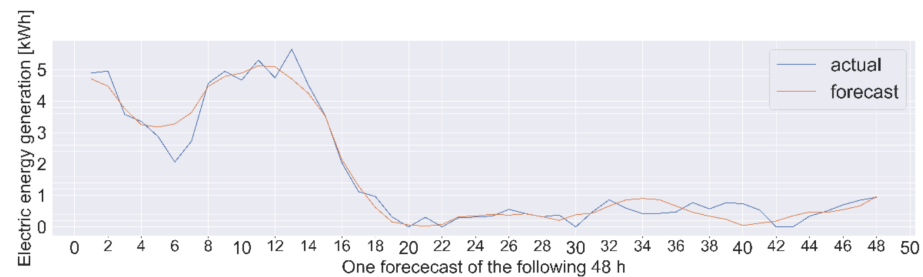
Figure 14. Forecast of electric energy generation from 1 to 48 h made by MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR] method for two following days of an autumn month (October).
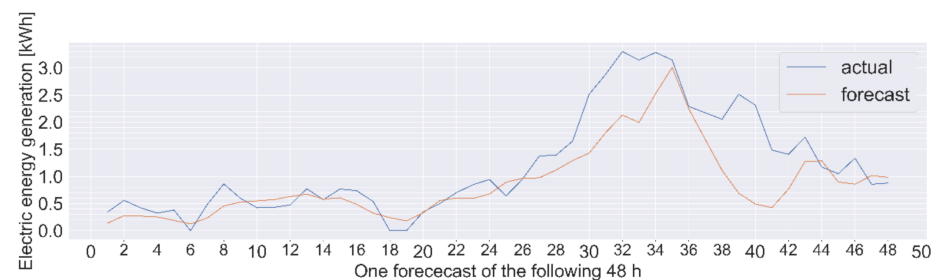

Figure 15. Forecast of electric energy generation from 1 to 48 h made by MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$, PHM+KNNR] method for two following days of the spring month (April).

6. Conclusions

The results of the study can be summarized as follows:

(1)　Original hybrid methods and ensemble methods based on hybrid methods, developed for researching specific implementations, have reduced errors of energy generation forecasts for a small wind turbine as compared to single methods.

(2)　The best integration system for ensemble methods based on hybrid methods for accuracy measures MAE, R, PCTL75AE and PCTL99AE is a new, original integrator developed for predictions called "averaging ensemble based on hybrid methods without extreme forecasts" (method code MIN&MAX_SKIP) with 3 hybrid methods in the ensemble. This method is notable for its simplicity, especially in contrast with MLP integrator which requires tuning parameters and hyperparameters.

(3)　The best integration system in ensemble methods based on hybrid methods for accuracy measure MAE is the MLP integrator.

(4)　"An ensemble based on hybrid methods with weight optimization for each predictor" performs better than the method with equal weights for each predictor.

(5)　Our research has demonstrated the merits of using ensemble methods based on hybrid methods instead of hybrid methods. Especially, high accuracy gain was achieved as compared to single methods.

(6)　Deep neural network LSTM is the best single method, MLP is the second best, while using SVR, KNNR and especially LR is less favorable.

(7)　An increase in valuable information in input data (from class 1 to class 5) decreases prediction errors. In particular, wind speed forecasts are the most important input

data. Using lagged values of forecast time series proved to slightly increase prediction accuracy. The same applies to using seasonal and daily variability markers.

(8)   If lagged values of actual wind velocities can be used as additional input data, the quality of forecasts should slightly improve.

(9)   More research is needed to verify, among other things, the following:

- Is prediction accuracy affected by using forecasts from more than one source?
- Does using greater amount of input data, especially wind speed forecasts from periods directly neighboring the forecast period, affect prediction accuracy?
- Will the proposed, original method of "averaging ensemble based on hybrid methods without extreme forecasts" be equally good for different RES predictions from different locations and 1 to 72 h ahead horizons?

Abbreviations

The following abbreviations are used in this manuscript:

ACF	Autocorrelation function
AE	Absolute Error
AM	Attention Mechanism
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ARIMA-FFNN	Autoregressive Integrated Moving Average—Feedforward Neural Network
BFGS	Broyden–Fletcher–Goldfarb–Shanno algorithm
BiLSTM-CNN	Bidirectional Long Short Term Memory–Convolutional Neural Network
BNN	Bayesian Neural Network
DCNN	Deep Convolutional neural network
DFFNN	Deep Feed Forward Neural Network
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
GBT	Gradient Boosting Trees
KDE	Kernel Density Estimation
KNNR	K-Nearest Neighbours Regression
LR	Linear Regression
LSTM	Long Short-Term Memory
LSTM-SARIMA	Long Short-Term Memory–Seasonal Autoregressive Integrated Moving Average
MAE	Mean Absolute Error
MBE	Mean Bias Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
NWP	Numerical Weather Prediction

PCTL75AE	The 75th percentile
PCTL99AE	The 99th percentile
p.u.	Per unit
PV	Photovoltaic System
R	Pearson linear correlation coefficient
R^2	Determination coefficient
RES	Renewable Energy System
RF	Random Forests
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RSM	Response Surface Method
SVR	Support Vector Regression
UM	Unified Model

Appendix A

Table A1. The results of hyperparameters tuning for hybrid methods and ensemble methods based on hybrid methods.

Method Code	Description of Method, the Name and the Range of Values of Hyperparameters Tuning and Selected Values
PHM+LSTM	The number of hidden layers:1–2, selected:2, the number of neurons in hidden layer: 3–15, selected: 5–5, the activation function in hidden layer: ReLU/ELU/PReLU/LeakyReLU/sigmoid/tanh, selected ReLU, the activation function in output layer: ReLU/sigmoid/tanh/linear, selected tanh, learning algorithm: ADAM, lr = 0.001, decay = 1×10^{-5}, epochs: 2000, patience: 100–500, selected: 100, batch size: 128; shuffle:True.
PHM+SVR	Regression SVM: Type-1, Type 2, selected: Type-1, kernel type: Gaussian (RBF), the width parameter σ: 0.333, the regularization constant C, range: 1–20 (step 1), selected: 3, the tolerance ε, range: 0.05–2 (step 0.05), selected: 0.1.
PHM+KNNR	Distance metrics: Euclidean, Manhattan, Minkowski, selected: Euclidean, the number of nearest neighbors k, range: 1–200, selected: 42.
PHM+MLP$_{PSO}$	The number of neurons in hidden layers: 5–20, selected 2 hidden layers 10–8, learning algorithm: PSO, the activation function in hidden layer: linear, hyperbolic tangent, selected: hyperbolic tangent, the activation function in output layer: hyperbolic tangent. The number of particles in swarm (20, 50, 100, 150), selected 50, number of algorithm iterations (100, 1000) selected 50, values of coefficients in formula for particles movements (0.7, 1.0, 1.4), selected 1.4 for first coef and 0.7 for the rest, number of iterations between subsequent disturbances in swarm (50, 100, 400), selected 400, neighborhood width (2, 5, 10), selected 5.
PHM+MLP$_{BFGS}$	The number of neurons in hidden layer: 5–20, selected: 13, learning algorithm: BFGS, the activation function in hidden layer: linear, hyperbolic tangent, selected: hyperbolic tangent, the activation function in output layer: linear.
MLP_INT [PHM+LSTM, PHM+MLP$_{PSO}$,PHM+KNNR]	The number of neurons in hidden layer: 3–10, selected: 4, learning algorithm: BFGS, the activation function in hidden layer: linear, hyperbolic tangent, selected: hyperbolic tangent, the activation function in output layer: linear.

References

1. Brodny, J.; Tutak, M.; Saki, S.A. Forecasting the Structure of Energy Production from Renewable Energy Sources and Biofuels in Poland. *Energies* **2020**, *13*, 2539. [CrossRef]
2. Masood, N.A.; Yan, R.; Saha, T.K. A new tool to estimate maximum wind power penetration level: In perspective of frequency response adequacy. *Appl. Energy* **2015**, *154*, 209–220. [CrossRef]
3. Wood, D. Grand Challenges in Wind Energy Research. *Front. Energy Res.* **2020**, *8*. [CrossRef]
4. Hämäläinen, K.; Saltikoff, E.; Hyvärinen, O. Assessment of Probabilistic Wind Forecasts at 100 m Above Ground Level Using Doppler Lidar and Weather Radar Wind Profiles. *Mon. Weather Rev.* **2020**, *8*. [CrossRef]
5. Wilczak, J.M.; Olson, J.B.; Djalalova, I.; Bianco, L.; Berg, L.K.; Shaw, W.J.; Coulter, R.L.; Eckman, R.M.; Freedman, J.; Finley, C.; et al. Data assimilation impact of in situ and remote sensing meteorological observations on wind power forecasts during the first Wind Forecast Improvement Project (WFIP). *Wind Energy* **2019**, *22*, 932–944. [CrossRef]
6. Theuer, F.; van Dooren, M.F.; von Bremen, L.; Kühn, M. Minute-scale power forecast of offshore wind turbines using long-range single-Doppler lidar measurements. *Wind Energ. Sci.* **2020**, *5*, 1449–1468. [CrossRef]

7. Papazek, P.; Schicker, I.; Plant, C.; Kann, A.; Wang, Y. Feature selection, ensemble learning, and artificial neural networks for short-range wind speed forecasts. *Meteorol. Z.* **2020**, *29*, 307–322. [CrossRef]

8. Cai, H.; Jia, X.; Feng, J.; Li, W.; Hsu, Y.M.; Lee, J. Gaussian Process Regression for numerical wind speed prediction enhancement. *Renew. Energy* **2020**, *146*, 2112–2123. [CrossRef]

9. Tawn, R.; Browell, J.; Dinwoodie, I. Missing data in wind farm time series: Properties and effect on forecasts. *Electr. Power Syst. Res.* **2020**, *189*, 106640. [CrossRef]

10. Messner, J.W.; Pinson, P.; Browell, J.; Bjerregård, M.B.; Schicker, I. Evaluation of wind power forecasts—An up-to-date view. *Wind Energy* **2020**, *23*, 1461–1481. [CrossRef]

11. Sewdien, V.N.; Preece, R.; Rueda Torres, J.L.; Rakhshani, E.; van der Meijden, M. Assessment of critical parameters for artificial neural networks based short-term wind generation forecasting. *Renew. Energy* **2020**, *161*, 878–892. [CrossRef]

12. Mishra, S.; Bordin, C.; Taharaguchi, K.; Palu, I. Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. *Energy Rep.* **2020**, *6*, 273–286. [CrossRef]

13. Shetty, R.P.; Sathyabhama, A.; Pai, P.S. Comparison of modeling methods for wind power prediction: A critical study. *Front. Energy* **2020**, *14*, 347–358. [CrossRef]

14. Spiliotis, E.; Petropoulos, F.; Nikolopoulos, K. The Impact of Imperfect Weather Forecasts on Wind. Power Forecasting Performance: Evidence from Two Wind Farms in Greece. *Energies* **2020**, *13*, 1880. [CrossRef]

15. Baptista, D.; Carvalho, J.P.; Morgado-Dias, F. Comparing different solutions for forecasting the energy production of a wind farm. *Neural Comput. Appl.* **2020**, *32*, 15825–15833. [CrossRef]

16. Hu, J.; Tang, J.; Lin, Y. A novel wind power probabilistic forecasting approach based on joint quantile regression and multi-objective optimization. *Renew. Energy* **2020**, *149*, 141–164. [CrossRef]

17. Zhen, H.; Niu, D.; Yu, M.; Wang, K.; Liang, Y.; Xu, X. A Hybrid Deep Learning Model and Comparison for Wind Power Forecasting Considering Temporal-Spatial Feature Extraction. *Sustainability* **2020**, *12*, 9490. [CrossRef]

18. Ma, Y.-J.; Zhai, M.-Y. A Dual-Step Integrated Machine Learning Model for24 h-Ahead Wind Energy Generation Prediction Based on Actual Measurement Data and Environmental Factors. *Appl. Sci.* **2019**, *9*, 2125. [CrossRef]

19. Zhou, B.; Liu, C.; Li, J.; Sun, B.; Yang, J. A Hybrid Method for Ultrashort-Term Wind Power Prediction considering Meteorological Features and Seasonal Information. *Math. Probl. Eng.* **2020**. [CrossRef]

20. Chen, K.-S.; Lin, K.-P.; Yan, J.-X.; Hsieh, W.-L. Renewable Power Output Forecasting Using Least-Squares Support Vector Regression and Google Data. *Sustainability* **2019**, *11*, 3009. [CrossRef]

21. Ding, J.; Chen, G.; Yuan, K. Short-Term Wind Power Prediction Based on Improved Grey Wolf Optimization Algorithm for Extreme Learning Machine. *Processes* **2020**, *8*, 109. [CrossRef]

22. Ouyang, T.; Huang, H.; He, Y.; Tang, Z. Chaotic wind power time series prediction via switching data-driven modes. *Renew. Energy* **2020**, *145*, 270–281. [CrossRef]

23. Bracale, A.; Caramia, P.; Carpinelli, G.; De Falco, P. Day-ahead probabilistic wind power forecasting based on ranking and combining NWPs. *Int. Trans. Electr. Energy Syst.* **2020**, *30*, 12325. [CrossRef]

24. Sun, M.; Feng, C.; Zhang, J. Multi-distribution ensemble probabilistic wind power forecasting. *Renew. Energy* **2020**, *148*, 135–149. [CrossRef]

25. Banik, R.; Das, P.; Ray, S.; Biswas, A. Wind power generation probabilistic modeling using ensemble learning techniques. *Mater. Today-Proc.* **2020**, *26*, 2157–2162. [CrossRef]

26. Chen, J.; Zhu, Q.; Li, H.; Lin, Z.; Shi, D.; Li, Y.; Duan, X.; Liu, Y. Learning Heterogeneous Features Jointly: A Deep End-to-End Framework for Multi-Step Short-Term Wind Power Prediction. *IEEE Trans. Sustain. Energy* **2020**, *11*. [CrossRef]

27. Shahid, F.; Khan, A.; Zameer, A.; Arshad, J.; Safdar, K. Wind power prediction using a three stage genetic ensemble and auxiliary predictor. *Appl. Soft Comput.* **2020**, *90*, 106151. [CrossRef]

28. Durán-Rosal, A.M.; Gutiérrez, P.A.; Salcedo-Sanz, S. A statistically-driven Coral Reef Optimization algorithm for optimal size reduction of time series. *Appl. Soft Comp.* **2018**, *63*, 139–153. [CrossRef]

29. Piotrowski, P.; Baczyński, D.; Kopyt, M.; Szafranek, K.; Helt, P.; Gulczyński, T. Analysis of forecasted meteorological data (NWP) for efficient spatial forecasting of wind power generation. *Electr. Power Syst. Res.* **2019**, *175*, 105891. [CrossRef]

30. Yu, M.; Zhang, Z.; Li, X.; Yu, J.; Gao, J.; Liu, Z.; You, B.; Zheng, X.; Yu, R. Superposition Graph Neural Network for offshore wind power prediction. *Future Gener. Comput. Syst.* **2020**, *113*, 145–157. [CrossRef]

31. Fan, H.; Zhang, X.; Mei, S.; Chen, K.; Chen, X. M2GSNet: Multi-Modal Multi-Task Graph Spatiotemporal Network for Ultra-Short-Term Wind Farm Cluster Power Prediction. *Appl. Sci.* **2020**, *10*, 7915. [CrossRef]

32. Li, L.; Li, Y.; Zhou, B.; Wu, Q.; Shen, X.; Liu, H.; Gong, Z. An adaptive time-resolution method for ultra-short-term wind power prediction. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105814. [CrossRef]

33. Shahid, F.; Zameer, A.; Mehmood, A.; Raja, M.A.Z. A novel wavenets long short term memory paradigm for wind power prediction. *Appl. Energy* **2020**, *269*, 115098. [CrossRef]

34. Zhang, P.; Li, C.; Peng, C.; Tian, J. Ultra-Short-Term Prediction of Wind Power Based on Error Following Forget Gate-Based Long Short-Term Memory. *Energies* **2020**, *13*, 5400. [CrossRef]

35. Guan, J.; Lin, J.; Guan, J.J.; Mokaramian, E. A novel probabilistic short-term wind energy forecasting model based on an improved kernel density estimation. *Int. J. Hydrog. Energy* **2020**, *45*, 23791–23808. [CrossRef]

36. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019.

37. Dudek, G.; Pełka, P. Forecasting monthly electricity demand using k nearest neighbor method. *Przegląd Elektrotechniczny* **2017**, *93*, 62–65.

38. Osowski, S.; Siwek, K. Local dynamic integration of ensemble in prediction of time series. *Bull. Pol. Ac. Tech* **2019**, *67*, 517–525.

39. Dudek, G. Multilayer Perceptron for Short-Term Load Forecasting: From Global to Local Approach. *Neural Comput. Appl.* **2019**, *32*, 3695–3707. [CrossRef]

40. Piotrowski, A.P.; Napiorkowski, J.J.; Piotrowska, A.E. Impact of deep learning-based dropout on shallow neural networks applied to stream temperature modelling. *Earth-Sci. Rev.* **2020**, *201*, 103076. [CrossRef]

41. Xie, X.F.; Zhang, W.J.; Yang, Z.L. Social cognitive optimization for nonlinear programming problems. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Beijing, China, 4–5 November 2002; pp. 779–783.