

Chapter 7

Short-Term Wind Speed Forecasting Model Using Hybrid Neural Networks and Wavelet Packet Decomposition



Adel Lakzadeh, Mohammad Hassani, Azim Heydari, Farshid Keynia, Daniele Groppi, and Davide Astiaso Garcia

Abstract Wind speed is one of the most vital, imperative meteorological parameters, thus the prediction of which is of fundamental importance in the studies related to energy management, building construction, damages caused by strong winds, aquatic needs of power plants, the prevalence and spread of diseases, snowmelt, and air pollution. Due to the discrete and nonlinear structure of wind speed, wind speed forecasting at regular intervals is a crucial problem. In this regard, a wide variety of prediction methods have been applied. So far, many activities have been done in order to make optimal use of renewable energy sources such as wind, which have led to the present diverse types of wind speed and strength measuring methods in the various geographical locations. In this paper, a novel forecasting model based on hybrid neural networks (HNNs) and wavelet packet decomposition (WPD) processor has been proposed to predict wind speed. Considering this scenario, the accuracy of the proposed method is compared with other wind speed prediction methods to ensure performance improvement.

Keyword Sequence of estimators · Combined neural networks · Wavelet packet decomposition · Wind speed prediction

A. Lakzadeh · M. Hassani
Islamic Azad University, Kerman Branch, Iran

A. Heydari · D. Groppi · D. Astiaso Garcia (✉)
Sapienza University of Rome, Rome, Italy
e-mail: davide.astiasogarcia@uniroma1.it

A. Heydari
e-mail: azim.heydari@uniroma1.it

D. Groppi
e-mail: daniele.groppi@uniroma1.it

A. Heydari · F. Keynia
Graduate University of Advanced Technology, Kerman, Iran
e-mail: f.keynia@kgut.ac.ir

7.1 Introduction

Energy formed a plethora of necessities of human life in the past and present, but it will also have a more paramount role in a future life; therefore, it is logical to be one of the human concerns (Blaabjerg and Liserre 2012). Fossil resources, the first and foremost energy source, are dwindling; moreover, their production costs, transmission, and distribution are enlarging. As a result, they may not be reliable sources of energy anymore and do not have ample potential to meet all human needs (Razmjoo et al. 2021). All these reasons have provoked researchers into offering new methods of supplying energy (Masters 2013). Of particular, new sources of supplying energy are solar, wind, wave, and water, but the problem faced up using these resources is random changes, such as changes in wind speed, water level, solar radiation intensity, which lead generated power to be partially unpredictable and unreliable. However, from any angle, the benefits of these clean energy sources outweigh the disadvantages. As has been mentioned earlier, wind is one of the most critical renewable sources of energy. Availability and permanency are two key factors leading researchers and industrial owners to replace fire with clean electrical energy (Erdem and Shi 2011; Jeon and Taylor 2012). This source of energy is currently one of the strategic products in the commercial energy markets. According to the collected data, about 2.3% of the world's electricity consumption is generated using wind energy (Torres et al. 2005; Zhao et al. 2012; Khalid and Savkin 2012). The use of wind power is growing in all parts of the world, and it has been predicted that by 2030 it will account for nearly 30% of world electricity production (Bhaskar et al. 2012; Salonen et al. 2011; Kusiak and Li 2010).

Thus, the use of advanced methods in maintaining the health of wind turbines, including predicting changes in wind speed, is an integral part of studies in this field.

Not only does a wind turbine failure increase maintenance costs, but also it reduces power generation. In order to generate effective power, the performance of the wind turbine must be checked by the status monitoring system. One of the critical performance parameters along power and step screw is wind speed, according to which Condition Monitoring System (CMS) results should be analyzed (Ackermann 2005; Heydari et al. 2021a). Some reasons representing the importance of wind speed prediction are the maintenance of wind turbine units, power generation schedule, and energy recovery and storage plan.

In (Torres et al. 2005), the Autoregressive Moving Average (ARMA) model has been presented in order to predict the hourly average wind speed. In (Bhaskar et al. 2012), the adaptive and wavelet neural network model has been provided in order to predict wind speed based on NWP information. Zhang et al. developed a new combined intelligent forecasting model based on multivariate data secondary decomposition approach and deep learning method in order to predict wind speed (Zhang et al. 2021). Heydari et al. presented a novel combined short-term wind forecasting model using metaheuristic optimization algorithm and deep learning neural network (Heydari et al. 2021b). Neshat et al. (2021) provided a machine learning-based model

in order to predict short-term wind speed. The model is applied for the Lillgrund offshore wind farm in Sweden.

In this research, we have used the HNN method and WPD processor to predict wind speed. Three types of neural networks have been used in HNN neural network training based on three different algorithms which include Levenberg–Marquardt (LM), Broyden–Fletcher–Goldfarb–Shanno (BFGS), and Bayesian Regularization (BR) algorithms.

According to previous studies, with the proper selection of perceptron neural network training methods, the HNN neural network can be trained better than a single neural network. The best results for related neural networks are those that use the LM algorithm on the first step. LM is a fast training algorithm, so this algorithm should be selected as the first neural network in the hybrid neural network. On the second step of the HNN neural network, in order to identify better weights and biases in the search space, the BFGS algorithm was used to train the neural network, which caused the BFGS algorithm to start the training process from a suitable point, and was able to find the appropriate and minimal answers in the search space. In the proposed method, it is better to use the BR algorithm as the last neural network in order to adjust the parameters and obtain the maximum training efficiency.

7.2 Methods

7.2.1 Neural Network

In recent years, the use of artificial intelligence to solve complex and nonlinear problems such as classification and prediction has expanded. Artificial neural networks can identify the nonlinear relationships between variables through the training process (Yu 2010). In general, methods based on artificial intelligence are more accurate than conventional statistical methods.

One of the most well-known neural networks is the multilayer perceptron (MLP) neural network, which usually has a feedforward architecture. The MLP network consists of input and output layers and at least one hidden layer. Each of these layers has several neurons and includes processing units, and each unit is completely connected to the units of the next layer by weight (W_{ij}) (Alavi et al. 2010). Y output is obtained by transferring the sum of previous outputs and mapping by an activation function.

7.2.2 Wavelet Packet Decomposition

Wavelet packet analysis can be seen as a particular type of wavelet analysis. WPD is a classic signal processing method, which can break down the signal into suitable

components. The level of wavelet analysis can profoundly affect WPD performance. WPD consists of two types: continuous wavelet transform and discrete wavelet transform. In this method, the general stage divides the approximation coefficients into two parts. After dividing the approximation coefficient vector and the detail coefficient vector, we obtain both on a large scale. Missing information is taken between two consecutive approximations in detail coefficients. Then, the next step involves dividing the vector of the new coefficient. Sequential details are never reanalyzed. However, in the corresponding wavelet packet, each detail coefficient vector is also decomposed into two parts using the same approximation vector approach.

7.2.3 *Proposed Intelligent Forecasting Model*

In this study, a suitable combination of MLP neural networks is utilized to design the prediction model. The HNN prediction model can significantly improve the training capability of neural networks in modeling a complex process (including wind speed prediction). In addition, the use of several structures (parallel and consecutive) to combine neural networks with more capabilities is described in (Amjady et al. 2010).

This study proposes a new prediction model to wind speed forecasting based on neural networks and noise reduction data. Since wavelet transform works as a preprocessing method, the data are decomposed by using wavelet transform in the first step. In the following step, decomposed data feed into neural networks as input data.

Figure 7.1 shows the process of the wind speed prediction according to the proposed method. As shown in the figure, each level of the HNN prediction model has been arranged by particular types of neural networks. Since all neural networks have the same number of input data, layers, and neurons in the proposed method, they can use the weights and biases values of the previous. Therefore, each neural network improves the knowledge gained from the previous one. In addition, the second set of outcomes transmitted between neural networks is the predicted value of the target variable. The second and third neural networks also receive the initial predicted value of the target variable as input, which increases the neural network prediction accuracy.

In the prediction models using wavelet transform as a noise reduction method, the process of noise reduction has three phases (see Fig. 7.2).

The first step of the noise reduction process is converting the raw data to decomposed layers by wavelet, called function analysis. Each part of the decomposed signal can be considered as a wavelet coefficient and a scale factor. By applying wavelet and scaling filters to the main time series at each step, wavelet and scaling coefficients are obtained, which are repeated in the form of the following pyramidal algorithm (Fig. 7.3).

In fact, in wavelet analysis, data are divided into two subsets of high-frequency and low-frequency data. The sparse-frequency data obtained by applying the father wavelet wave over the main series reflect the main features of the series. Much data

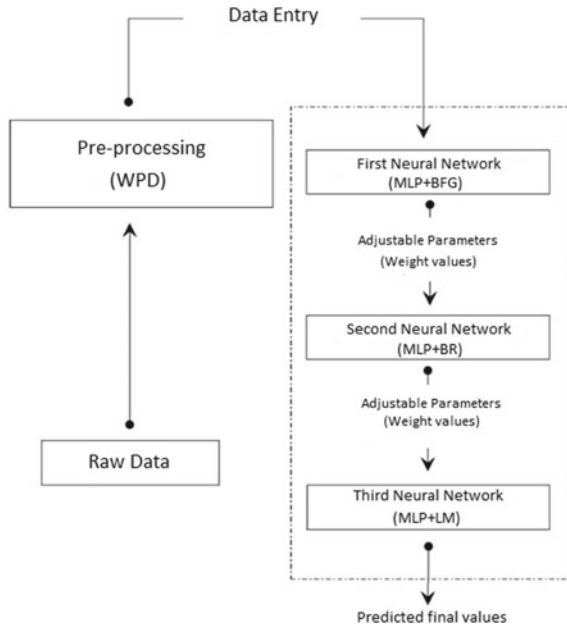


Fig. 7.1 Flowchart of the proposed method

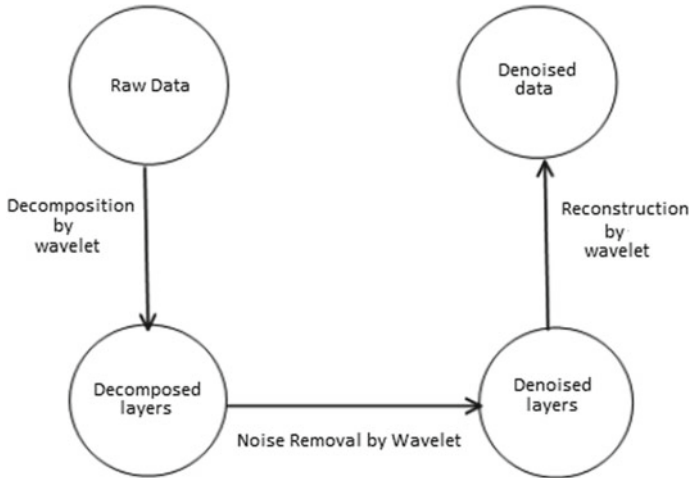


Fig. 7.2 Noise reduction by wavelet transform

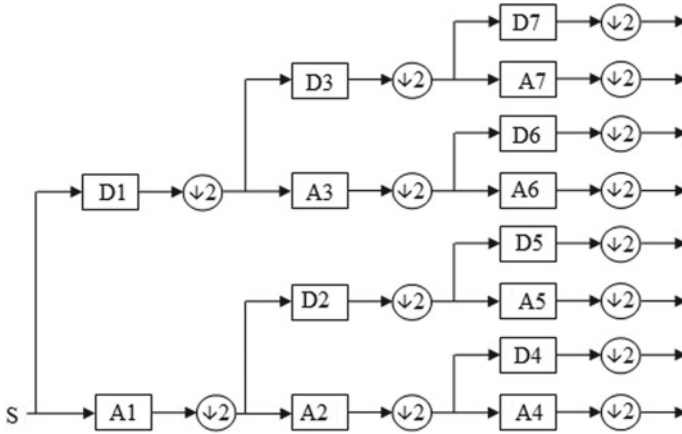


Fig. 7.3 Pyramidal algorithm for wavelet analysis, S: main data; $A[i]$ is the low-pass approximation coefficients, $D[i]$ is the high-pass detail coefficients

are also generated by applying the mother wavelet on the main series, often called noise. The main purpose of wavelet decomposition is to isolate the main features of the series from the noise.

7.3 Case Study

To evaluate the proposed model, 365 days of wind speed data from 1/3/2019 to 28/3/2020 were collected, all of which had a 5-min interval. Collected data consist of 12 samples per hour which mean 288 samples per day. On average, one data is extracted and used as the representative of that day. Collected data were divided into two categories: training and test. About 30% (120 days) of data are allocated for the test. Test data are used to evaluate the model performance.

7.4 Results and Discussion

The correlation of the obtained values by a model with the observed values (referred to as model evaluation or validation) is usually done by double comparisons between the simulated values and the observed values. There is the efficiency of the models and their validation, which can be determined by the Root Mean Square Deviation (RMSE), Mean Absolute Error (MAE), and Index of Agreement (IA).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \tag{7.1}$$

Table 7.1 Specifications of the artificial neural network

| Input method | Hidden layers | Neurons | Learning steps | Learning rate |
|----------------|---------------|---------|----------------|---------------|
| Neural network | 2 | 10 | 150 | 0.001 |

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \tag{7.2}$$

$$IA = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{7.3}$$

First, we normalized the data to [0–1] range. After applying WPD, we used a 2db wavelength to convert the wavelet. We took three wavelet steps from the data.

Wavelet steps: in the first step, we took the wavelet from the main data, which has two output signals including upper bound and lower bound. In comparison with the original signal, these two signals are completely brand-new ones.

In the next step, the first step instrument applies two times on the output signals of the previous step. As a result, we have eight signals after step two, and in the third, the neural networks were trained using 8% of data for the testing phase.

Each neural networks have two hidden layers with up to ten neurons in each layer. The characteristics of the artificial neural network are given in Table 7.1.

We trained neural networks with three various algorithms: BR, BFG, and LM (Table 7.2). Finally, we reversed the output network results from wavelet transform and measured the coefficients of determination (R^2), mean squared error (RMSE), absolute mean error (MAE), and agreement index (IA) using real data.

In the presented tables, the values related to different scales are calculated and displayed to examine the proposed combinations of networks and their performance differences according to the activation function. As shown in Table 7.3, using the TANSIG activation function, considering the order of placement of the networks, the lowest amount of RMSE was related to HNN2, which had the best performance with the proposed order. As can be seen, the best arrangement of the networks was 1-10-10-1 and had an equal value of 2.432. Considering the corresponding values of

Table 7.2 Different proposed HNNs for review in research

| | |
|--|--------|
| First MLP + LM, then MLP + BR, and finally MLP + BFG | HNN(1) |
| First MLP + LM, then MLP + BFG, and finally MLP + BR | HNN(2) |
| First MLP + BFG, then MLP + BR, and finally MLP + LM | HNN(3) |
| First MLP + BR, then MLP + BFG, and finally MLP + LM | HNN(4) |
| MLP + LM only | HNN(5) |

Table 7.3 Comparison of the proposed method with RMSE

| Activation function | Network sequence | HNN(1) | HNN(2) | HNN(3) | HNN(4) | HNN(5) |
|---------------------|------------------|--------|--------|--------|--------|--------|
| Tansig | 1-3-3-1 | 3.822 | 2.801 | 5.989 | 4.448 | 4.182 |
| | 1-5-5-1 | 3.551 | 2.799 | 6.062 | 4.430 | 4.224 |
| | 1-10-10-1 | 3.101 | 2.432 | 5.745 | 4.210 | 4.099 |
| | 1-15-15-1 | 3.220 | 2.501 | 5.920 | 4.220 | 4.120 |
| Logsig | 1-3-3-1 | 3.534 | 2.821 | 5.997 | 4.966 | 4.228 |
| | 1-5-5-1 | 3.618 | 3.027 | 6.066 | 4.413 | 4.210 |
| | 1-10-10-1 | 3.321 | 2.643 | 5.832 | 4.401 | 4.110 |
| | 1-15-15-1 | 3.401 | 2.765 | 5.899 | 4.409 | 4.102 |

the LOGSIG function, the same result is true and this network had the lowest value of RMSE in the same order. The corresponding value for this arrangement was 2.643.

Table 7.4 shows the MAE scales. The figures obtained from these calculations show the same results as the scale in Table 7.3. The corresponding MAE values for the HNN2 network with the TANSIG activation function are the lowest at 1.743. Also, using the LOGSIG activation function has the best efficiency and an equal value to 1.899 compared to other suggested items. As a result, the best-case scenario is HNN2 with the order 1-10-10-1 and using the TANSIG function.

The calculation of the determination coefficient values is given in Table 7.5. As expected from the review of the previous tables, the highest value for this scale was equal to 0.863 and was related to HNN2 with the order of 1-10-10-1 and the TANSIG function. In the continuation of the values, it can be seen that in the calculations related to the LOGSIG activation function, the corresponding values in the coefficient of determination for the two sequences 1-10-10-1 and 1-15-15-1 were similar and equal to 0.853. The lowest value was related to HNN3 with TANSIG activation function and the proposed order was 1-3-3-1.

Table 7.6 examines the values corresponding to IA. The values calculated for this parameter are in line with the previous findings. The highest value is related to HNN2

Table 7.4 Comparison of the proposed method with the MAE

| Activation function | Network sequence | HNN(1) | HNN(2) | HNN(3) | HNN(4) | HNN(5) |
|---------------------|------------------|--------|--------|--------|--------|--------|
| Tansig | 1-3-3-1 | 2.221 | 1.945 | 3.131 | 2.908 | 2.593 |
| | 1-5-5-1 | 2.135 | 1.944 | 3.273 | 2.955 | 2.395 |
| | 1-10-10-1 | 2.111 | 1.743 | 3.023 | 2.899 | 2.274 |
| | 1-15-15-1 | 2.104 | 1.867 | 3.100 | 2.903 | 2.145 |
| Logsig | 1-3-3-1 | 2.151 | 1.913 | 3.303 | 2.965 | 2.409 |
| | 1-5-5-1 | 2.110 | 1.982 | 3.361 | 2.937 | 2.452 |
| | 1-10-10-1 | 2.132 | 1.899 | 3.198 | 2.910 | 2.346 |
| | 1-15-15-1 | 2.134 | 1.901 | 3.143 | 2.910 | 2.345 |

Table 7.5 Comparison of the proposed method with the R2

| Activation function | Network sequence | HNN(1) | HNN(2) | HNN(3) | HNN(4) | HNN(5) |
|---------------------|------------------|--------|--------|--------|--------|--------|
| Tansig | 1-3-3-1 | 0.798 | 0.805 | 0.594 | 0.696 | 0.755 |
| | 1-5-5-1 | 0.749 | 0.804 | 0.693 | 0.721 | 0.753 |
| | 1-10-10-1 | 0.789 | 0.863 | 0.701 | 0.756 | 0.769 |
| | 1-15-15-1 | 0.788 | 0.834 | 0.699 | 0.765 | 0.771 |
| Logsig | 1-3-3-1 | 0.731 | 0.843 | 0.645 | 0.698 | 0.735 |
| | 1-5-5-1 | 0.716 | 0.821 | 0.690 | 0.681 | 0.718 |
| | 1-10-10-1 | 0.728 | 0.853 | 0.687 | 0.702 | 0.727 |
| | 1-15-15-1 | 0.734 | 0.853 | 0.698 | 0.703 | 0.733 |

Table 7.6 Comparison of the proposed method with IA

| Activation function | Network sequence | HNN(1) | HNN(2) | HNN(3) | HNN(4) | HNN(5) |
|---------------------|------------------|--------|--------|--------|--------|--------|
| Tansig | 1-3-3-1 | 0.916 | 0.954 | 0.892 | 0.893 | 0.915 |
| | 1-5-5-1 | 0.912 | 0.954 | 0.906 | 0.909 | 0.904 |
| | 1-10-10-1 | 0.943 | 0.994 | 0.918 | 0.911 | 0.921 |
| | 1-15-15-1 | 0.940 | 0.992 | 0.911 | 0.908 | 0.922 |
| Logsig | 1-3-3-1 | 0.912 | 0.954 | 0.896 | 0.908 | 0.900 |
| | 1-5-5-1 | 0.915 | 0.946 | 0.906 | 0.904 | 0.897 |
| | 1-10-10-1 | 0.921 | 0.967 | 0.910 | 0.909 | 0.919 |
| | 1-15-15-1 | 0.919 | 0.966 | 0.911 | 0.908 | 0.918 |

with TANSIG activation function and order 1-10-10-1 and shows a value equal to 0.994. After that, the best and highest value of this index is related to the same network with the order of 1-15-15-1 and has a difference of 0.002. The lowest value in the calculations of this scale was similarly related to HNN3 with other TANSIG activation functions and the proposed order of 1-3-3-1.

Comparison of the output results of various neural network arrangements with error measurement criteria shows that the proposed model has the acceptable ability to predict wind speed. As shown in Tables 7.3, 7.4, 7.5, and 7.6 the most effective ANN model in terms of performance is the arrangement (1-10-1), sigmoid tangent function, and HNN hybrid algorithm (1) on Kerman time series data.

Different criteria were used to measure the model: coefficient of determination (R2), squared mean square error (RMSE), absolute mean error (MAE), and agreement index (IA). All the results showed that the proposed method performed very well. In particular, the excellent performance of this method in long-range forecasting will be a clear vision for the Condition Monitoring System (CMS) to prevent sudden events and damage to wind turbines and ensure the health of the turbine. This reduces turbine shutdown and reduces costs and increases production capacity.

7.5 Conclusions

At first, wind speeds are predicted for future using a new combined intelligent model.

For this purpose, three feedforward networks have been used. This type of wind speed prediction can be very efficient in the Iranian wind energy industry, and this research in this field is not similar in the world. The proposed model was evaluated with various network models, functions, and different numbers of neurons to extract optimal network structure. Furthermore, training and validation data were used to evaluate and compare diverse network structures' performance. The proposed model also was trained with training algorithms BFGS, LM, and BR. As the results show, with the right Perceptron neural network training methods, the HNN neural network can be trained better than a single neural network. As been explained before, the best results for class neural networks are those using the first-class LM algorithm. Training a Perceptron neural network with an LM algorithm leads training error to reduce quickly since the LM algorithm is reasonably fast. Therefore, it is logical to select as the training algorithm of the first level neural network in the hybrid model.

Acknowledgements Sapienza University of Rome (DPDTA—Horizon2020 project “GIFT” Geographical Islands FlexibiliTY Grant agreement ID: 824410) is kindly acknowledged for providing financial support for this work.

References

- Blaabjerg F, Liserre M (2012) Power electronics converters for Wind turbine systems. *IEEE Trans Ind Appl* 48(2):708–719. <https://doi.org/10.1109/TIA.2011.2181290>
- Razmjoo A, Rezaei M, Mirjalili S, Majidi Nezhad M, Piras G (2021) Development of sustainable energy use with attention to fruitful policy. *Sustainability* 13(24):13840. <https://doi.org/10.3390/su132413840>
- Masters GM (2013) *Renewable and efficient electric power systems*. John Wiley & Sons
- Erdem E, Shi J (2011) ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl Energy* 88(4):1405–1414. <https://doi.org/10.1016/j.apenergy.2010.10.031>
- Jeon J, Taylor JW (2012) Using conditional kernel density estimation for wind power density forecasting using conditional kernel density estimation for wind power density forecasting. *J Am Stat Assoc* 107(497):66–79. <https://doi.org/10.1080/01621459.2011.643745>
- Torres JL, De BM, De FA, Garcı A (2005) Forecast of hourly average wind speed with ARMA models in navarre (Spain). *Sol Energy* 79(1):65–77. <https://doi.org/10.1016/j.solener.2004.09.013>
- Zhao P, Wang J, Xia J, Dai Y, Sheng Y, Yue J (2012) Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China. *Renewable Energy* 43:234–241. <https://doi.org/10.1016/j.renene.2011.11.051>
- Khalid M, Savkin AV (2012) A method for short-term wind power prediction with multiple observation points. *IEEE Trans Power Syst* 27(2):579–586. <https://doi.org/10.1109/TPWRS.2011.2160295>
- Bhaskar K, Member S, Singh SN, Member S (2012) AWNN-assisted wind power forecasting using feed-forward neural network. *IEEE Trans Sustain Energy* 3(2):306–315. <https://doi.org/10.1109/TSTE.2011.2182215>

- Salonen K, Niemelä S, Fortelius C (2011) Application of radar wind observations for low-level NWP wind forecast validation. *J Appl Meteorol Climatol* 50(6):1362–1371. <https://doi.org/10.1175/2010JAMC2652.1>
- Kusiak A, Li W (2010) Estimation of wind speed: a data-driven approach. *J Wind Eng Ind Aerodyn* 98(10–11):559–567. <https://doi.org/10.1016/j.jweia.2010.04.010>
- Ackermann T (2005) *Wind power in power systems*. vol 200(5). New York, Wiley
- Heydari A, Garcia DA, Fekih A, Keynia F, Tjernberg LB, De Santoli L (2021a) A hybrid intelligent model for the condition monitoring and diagnostics of wind turbines gearbox. *IEEE Access* 9:89878–89890. <https://doi.org/10.1109/ACCESS.2021.3090434>
- Zhang S, Chen Y, Xiao J, Zhang W, Feng R (2021) Hybrid wind speed forecasting model based on multivariate data secondary decomposition approach and deep learning algorithm with attention mechanism. *Renewable Energy* 174:688–704. <https://doi.org/10.1016/j.renene.2021.04.091>
- Heydari A, Nezhad MM, Neshat M, Garcia DA, Keynia F, De Santoli L, Tjernberg LB (2021b) A combined fuzzy gmdh neural network and grey wolf optimization application for wind turbine power production forecasting considering scada data. *Energies* 14(12):3459. <https://doi.org/10.3390/en14123459>
- Neshat M, Majidi Nezhad M, Abbasnejad E, Mirjalili S, Bertling Tjernberg L, Astiaso Garcia D, Alexander B, Wagner M (2021) A deep learning-based evolutionary model for short-term wind speed forecasting: a case study of the Lillgrund offshore wind farm. *Energy Convers Manage* 236:114002. <https://doi.org/10.1016/j.enconman.2021.114002>
- Yu S (2010) Actual experience on the short-term wind power forecasting at Penghu-from an Island perspective. In: 2010 International conference on power system technology, IEEE Press, Zhejiang, China, pp 1–8. <https://doi.org/10.1109/POWERCON.2010.5666619>
- Alavi AH, Gandomi AH, Mollahassani A, Heshmati AA, Rashed A (2010) Modeling of maximum dry density and optimum moisture content of stabilized soil using artificial neural networks. *J Plant Nutr Soil Sci* 173(3):368–379. <https://doi.org/10.1002/jpln.200800233>
- Amjady N, Daraeepour A, Keynia F (2010) Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. *IET Gener Transm Distrib* 4(3):432–444

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

