Short-Term Wind Energy Production Forecasting and Target Plant Selection Based on Meteorological Data Using Empirical Mode Decomposition

İsrafil Karadöl
Department of Electrical-Electronics Engineering, Faculty of Engineering-Architecture, Kilis 7 Aralık University, Kilis, Turkey

Keywords: Wind Energy, XGBoost, EMD, İzmir.

Abstract: This study aims to identify the most suitable target wind power plant (WPP) for short-term wind energy production forecasting. Hourly meteorological data for 2022 from İzmir Province were processed using the Empirical Mode Decomposition (EMD) method to generate 56 Intrinsic Mode Function (IMF) signals, which were used as input variables for the XGBoost model. As output, production data from 52 different WPPs located within the same provincial boundaries were individually used as target variables. The model's performance was evaluated using R^2, MAE, and MSE metrics. The results indicated that while high prediction accuracy was achieved for some plants, the model's performance was limited for others. The best forecast accuracy was obtained using data from WPP35, whereas the poorest performance was observed with WPP7. These findings suggest that, despite being within the same province, differences in the geographical locations of meteorological stations and WPPs, as well as region-specific meteorological characteristics, can significantly affect prediction accuracy.

1 INTRODUCTION

Globally, energy demand is increasing day by day(Renewable Energy Agency, 2025). Meeting the increasing energy demand from traditional energy sources causes many environmental consequences. The depletion of traditional energy sources is also recognized as a major problem by energy producers(Karadöl & Şekkeli, 2022). For these reasons, states have turned to renewable energy sources as an alternative to traditional energy sources(Irena, 2025). The fact that renewable energy sources are environmentally friendly and sustainable is seen as a great advantage for governments. However, the fact that these sources have random generation characteristics is recognized as an economic and technical problem by electricity grid operators(Karadöl et al., 2021). To overcome this problem, forecasting energy production is of great importance.

When we examine the research in the field of short-term forecasting of wind energy, it is seen that many studies have been carried out with various machine learning based methods(Joseph et al., 2023; Liu et al., 2019; Mustaqeem et al., 2022; Shukla & Pasari, 2025).In addition to these studies, Li et al. proposed the XGBoost regression model optimized by Genetic Algorithm (GA) to solve the accuracy and speed problems in wind power forecasting(X. Li et al., 2023). Ma et al. proposed the XGBoost algorithm for very short-term wind power forecasting due to the variability of wind power(Ma et al., 2020).

There is no research in the literature on wind power generation forecasting for different targets at 1-hour horizon. To fill this gap in the literature, this study examines the forecasting performance of the XGBoost model for different target allocations at 1-hour horizon. The meteorological data and WPP generation data of Izmir province for the year 2022 are used to perform this investigation. All data sets used are hourly resolution and real-time data. All meteorological data obtained were decomposed into signals by EMD method. As a result of this process, 56 signals were generated. The generated signals were defined as input to the XGBoost model. With this model, 52 WPP productions were defined as targets in order to perform forecasts for a 1-hour period. For each target, the forecasting performance of the XGBoost model was analyzed. As a result of these examinations, the best facilities for 1-hour forecasts were determined.

Karadöl, İ.
Short-Term Wind Energy Production Forecasting and Target Plant Selection Based on Meteorological Data Using Empirical Mode Decomposition.
DOI: 10.5220/0014299300004848

In *Proceedings of the 2nd International Conference on Advances in Electrical, Electronics, Energy, and Computer Sciences (ICEEECS 2025)*, pages 164-168
ISBN: 978-989-758-783-2

2 MATERIAL METHOD

In this study, it is aimed to determine the most suitable target facility to realize wind power generation forecasting with the XGBoost model at 1-hour time horizon. For this purpose, firstly, meteorological measurement data and WPP generation data of Izmir province were obtained. The obtained meteorological data were preprocessed by EMD method and decomposed into IMF signals. As a result of this decomposition, 56×8760 signals were generated. These signals were used as input data for the XGBoost model. With the XGBoost model, 52 WPP productions were selected as separate targets in order to make WPP production forecasts with a time horizon of 1 hour. R^2, MAE, and MSE metrics were used to evaluate the WPP production forecasting performance according to the selected target plants. According to these metrics, the best target WPPs in Izmir province were determined for the 1-hour horizon.

2.1 Meteorological Data

For the 1-hour time horizon, empirical mode-disaggregated meteorological data sets are used to determine the best target facility for wind power generation forecasting. The meteorological data used in the study are hourly resolution and have a 1-year period (01.01.2022-31.12.2022). These data sets consist of solar radiation, humidity, cloudiness, air temperature, soil temperature, rainfall, wind direction, and wind speed. Some statistical properties of these meteorological parameters are given in Table 1.

Table 1: Statistical properties of meteorological parameters

Meteorological parameters	Mean	Standard Deviation	Maks. value	Min value	Number of data	Period
Cloudiness	2.34	2.56	8	0	8760	01.01.2022-31.12.2022
Humidity	58.59	17.08	99	13		
Radiation	211.19	300.86	1036.4	0		
Wind rota	192.1	100.55	360	1		
Wind speed	2.86	1.47	11.5	0		
Air temperature	18.9	8.31	38.6	-0.1		
Soil temperature	20.64	9.8	42.9	0.7		
Rainfall	0.053	0.57	24	0		

2.2 WPP Generation Data

In the research, 52 WPP production data located within the borders of Izmir province were used. These data are real-time and hourly resolution for the year 2022. The total installed capacity of the facilities from which the data were obtained is 1742 MW. Within this installed capacity, there are facilities with different installed capacities, with a minimum capacity of 3 MW and a maximum capacity of 226 MW. Co.

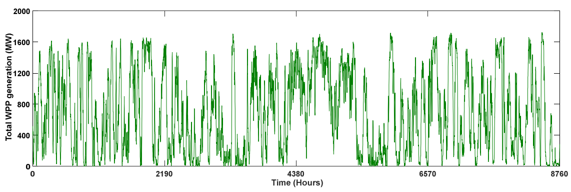

Figure 1: Total WPP production at 1 hour resolution

The total generation graph of these plants at hourly resolution is given in Figure 1. WPP generation data used in the study were obtained from Turkish Electricity Transmission and Distribution

2.3 Empirical Mode Decomposition (EMD)

The EMD method is based on the principle that nonlinear and non-stationary complex signals can be decomposed into sub-components with different oscillation characteristics(Jiang & Liu, 2023). Based on this approach, Huang et al. proposed that a complex signal can be decomposed into a residual signal with a finite number of intrinsic mode functions (IMFs) (Huang et al., 1998). This method has found wide application in many fields from engineering to climate science, from biomedical signal processing to energy systems, thanks to its structure suitable for time-frequency analysis.

The EMD method is based on the assumption that complex signals are composed of components that oscillate at various frequencies and amplitudes(Shukla & Pasari, 2025). According to this approach, the EMD decomposes a signal sequentially into subcomponents called Intrinsic Mode Functions (IMF) and a residual signal(Shang et al., 2022). IMF components reflect the short-term and high-frequency fluctuations of the signal, while the residual signal refers to the lower frequency and overall trend of the signal(Aladağ, 2023; Yuzgec et al., 2024). The mathematical expression of the EMD method is given in Equation 1(RajasundrapandiyanLeebanon et al., 2025).In this equation, Y_t is the input signal, N is the number of IMF signals, and R is the residual signal.

$$Y_t = \sum_{i=1}^{N} IMF_i + R(t) \tag{1}$$

2.4 The Extreme Gradient Boosting (XGBoost) Model

XGBoost is a supervised machine learning algorithm based on decision trees, providing high accuracy and high computational efficiency(Kanji & Das, 2025). It is widely used in the literature due to its superior performance, especially in classification and regression problems (Hakkal & Lahcen, 2024; Joshi et al., 2024; M. Li et al., 2020; Rathore et al., 2023; Shi et al., 2021; Yelgeç & Bingöl, 2022; Zhang et al., 2024). While models in traditional boosting algorithms try to directly reduce the error values, in the extreme gradient boosting approach, the model focuses on the pseudo-residuals that are derivatives of these errors(Demirer et al., 2024).In other words, XGBoost is defined as an algorithm based on gradient boosting, which is faster, more accurate, and more robust to overlearning.

The gradient boosting method is based on incrementally training successive decision trees in a way that reduces the error rates in their previous models(Demirer et al., 2024; X. Li et al., 2022). Each new model focuses on the prediction errors of the previous model and tries to correct them(Yan et al., 2022). In this way, the generalization capability of the model is increased, and more accurate predictions are obtained.

In the literature analysis, it was found that the XGBoost model is a highly effective approach for modelling complex and dynamic systems such as wind power generation forecasting due to its high accuracy, flexibility, and computational efficiency(Guan et al., 2023; W. Li et al., 2020; Phan et al., 2021; Zheng & Wu, 2019).In this study, XGBoost is used as the main modelling tool for short-term power generation forecasts. The parameters of the model used are given in Table 2.

Table 2: XGBoost model parameters

Parameters	Value
n_estimators	400
Learning_rate	0.1
Max_depth	6
Subsample	0.8
Colsample_bytree	0.8
Random_state	42
Objective	reg:squarederror

3 RESULTS

3.1 Meteorological IMF Signals

The first stage of the research is the conversion of meteorological data into IMF signals by the EMD method. At this stage, each meteorological parameter was converted into 7 IMF signals. A total of 56 IMF signals were obtained from all meteorological parameters. IMF signals of wind speed data are given in Figure 2 as an example.

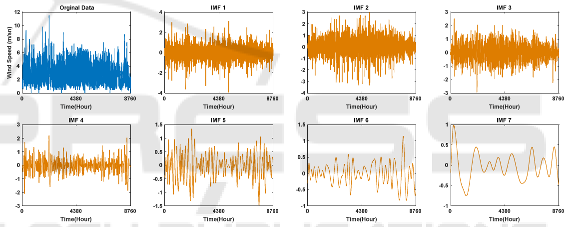

Figure 2: IMF signals of the wind speed data.

3.2 XGBoost Model WPP Forecasting Results

In this study, the XGBoost model is used to predict wind power plant (WPP) generation at 1-hour forecast horizon and to determine the most suitable target facility within the same province by using meteorological measurement data. In this context, meteorological data of Izmir province were transformed into Internal Mode Functions (IMF) using the Empirical Mode Decomposition (EMD) method, and these signals were used as input data to the XGBoost model.

To evaluate the performance of the model, 52 different WPP production data sets, each of which is treated as a different target variable, are used individually as the output of the model. For these targets, the forecast performance metrics obtained with the XGBoost model are presented in Table 3. When the table is analyzed, it is seen that for some power plants, forecasts with very high accuracy are obtained, while for some power plants, the performance of the model is poor. Forecast results with R^2 values of 0.90 and above are shown in bold in the table.

Table 3: WPP production forecast performance metrics over a 1-hour horizon.

Plant Number	MAE	MSE	R^2	Plant Number	MAE	MSE	R^2
WPP1	0.075	0.012	0.892	WPP27	**0.062**	**0.008**	**0.931**
WPP2	**0.078**	**0.013**	**0.907**	WPP28	**0.064**	**0.008**	**0.913**
WPP3	0.073	0.010	0.897	WPP29	**0.074**	**0.011**	**0.904**
WPP4	**0.065**	**0.009**	**0.934**	WPP30	**0.060**	**0.007**	**0.924**
WPP5	0.091	0.016	0.856	WPP31	**0.064**	**0.008**	**0.911**
WPP6	0.089	0.016	0.860	WPP32	**0.057**	**0.006**	**0.937**
WPP7	0.048	0.005	0.555	WPP33	**0.065**	**0.009**	**0.936**
WPP8	**0.070**	**0.010**	**0.916**	WPP34	**0.058**	**0.007**	**0.920**
WPP9	**0.063**	**0.007**	**0.933**	WPP35	**0.059**	**0.007**	**0.938**
WPP10	0.079	0.013	0.858	WPP36	0.067	0.009	0.892
WPP11	**0.069**	**0.010**	**0.924**	WPP37	0.100	0.020	0.867
WPP12	0.070	0.009	0.878	WPP38	0.081	0.013	0.889
WPP13	**0.068**	**0.009**	**0.922**	WPP39	**0.065**	**0.009**	**0.928**
WPP14	0.079	0.011	0.888	WPP40	0.059	0.007	0.897
WPP15	**0.062**	**0.008**	**0.906**	WPP41	**0.068**	**0.009**	**0.912**
WPP16	**0.061**	**0.007**	**0.936**	WPP42	**0.064**	**0.008**	**0.901**
WPP17	0.063	0.008	0.880	WPP43	**0.067**	**0.010**	**0.915**
WPP18	0.076	0.011	0.887	WPP44	**0.062**	**0.008**	**0.929**
WPP19	0.063	0.014	0.885	WPP45	**0.067**	**0.009**	**0.904**
WPP20	0.069	0.009	0.875	WPP46	**0.061**	**0.007**	**0.930**
WPP21	0.082	0.013	0.898	WPP47	**0.071**	**0.009**	**0.918**
WPP22	**0.060**	**0.007**	**0.934**	WPP48	**0.065**	**0.008**	**0.928**
WPP23	**0.067**	**0.009**	**0.920**	WPP49	**0.058**	**0.007**	**0.926**
WPP24	**0.072**	**0.010**	**0.917**	WPP50	0.079	0.012	0.894
WPP25	0.071	0.011	0.889	WPP51	**0.068**	**0.009**	**0.917**
WPP26	**0.067**	**0.009**	**0.923**	WPP52	**0.079**	**0.013**	**0.902**

According to the findings, the most accurate generation forecasts for the 1-hour forecast horizon were realized with the data from WPP35. On the other hand, the lowest forecast performance was obtained with the production data of plant WPP7.

4 CONCLUSIONS

As a result of the research, it was concluded that it is not always possible to use meteorological measurement data and all power plant production data within the same provincial borders together in forecasting studies. One of the main reasons for this situation is that meteorological measurement stations and wind power plants are located in different locations. Moreover, even if the measurement station and the power plant are located in the same province, the fact that some power plants have endemic meteorological conditions specific to their region is another reason for the differences in forecasting performance.

REFERENCES

Aladağ, E. (2023). Ampirik Mod Ayrıştırmasına Dayalı ARIMA Modeli Kullanılarak Van İli Hava Kirliliğinin Tahmini. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *28*(2), 495–509. https://doi.org/10.53433/yyufbed.1220578

Demirer, H., Ayci, D., Ozturk, S. A., Karabal, C., & Valles, J. A. (2024). A Machine Learning-Based Reference Baseline Consumption Model for Energy Efficiency Measurement and Verification. *Electrical-Electronics and Biomedical Engineering Conference, ELECO 2024 - Proceedings*, 1–5. https://doi.org/10.1109/ELECO64362.2024.10847193

Guan, S., Wang, Y., Liu, L., Gao, J., Xu, Z., & Kan, S. (2023). Ultra-short-term wind power prediction method combining financial technology feature engineering and XGBoost algorithm. *Heliyon*, *9*(6), e16938. https://doi.org/10.1016/j.heliyon.2023.e16938

Hakkal, S., & Lahcen, A. A. (2024). XGBoost To Enhance Learner Performance Prediction. *Computers and Education: Artificial Intelligence*, *7*(January), 100254. https://doi.org/10.1016/j.caeai.2024.100254

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Snin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Irena. (2025). Renewable capacity statistics 2025 International Renewable Energy Agency. In *Irena*. https://irena.org/-/media/Files/IRENA/Agency/Publication/2020/Mar/IRENA_RE_Capacity_Statistics_2020.pdf

Jiang, T., & Liu, Y. (2023). A short-term wind power prediction approach based on ensemble empirical mode decomposition and improved long short-term memory. *Computers and Electrical Engineering*, *110*(March), 108830. https://doi.org/10.1016/j.compeleceng.2023.108830

Joseph, L. P., Deo, R. C., Prasad, R., Salcedo-Sanz, S., Raj, N., & Soar, J. (2023). Near real-time wind speed forecast model with bidirectional LSTM networks. *Renewable Energy*, *204*(May 2022), 39–58. https://doi.org/10.1016/j.renene.2022.12.123

Joshi, A., Vishnu, C., Mohan, C. K., & Raman, B. (2024). Application of XGBoost model for early prediction of earthquake magnitude from waveform data. *Journal of Earth System Science*, *133*(1). https://doi.org/10.1007/s12040-023-02210-1

Kanji, S., & Das, S. (2025). Assessing groundwater potentialities and replenishment feasibility using machine learning and MCDM models considering hydro-geological aspects and water quality constituents. *Environmental Earth Sciences*, *84*(1), 1–18. https://doi.org/10.1007/s12665-024-11996-2

Karadöl, İ., & Şekkeli, M. (2022). Ramp Events in River Type Hydroelectric Power Plants: The Case of Turkey. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *34*(1), 361–375. https://doi.org/10.35234/fumbd.1029081

Karadöl, İ., Yıldız, C., & Şekkeli, M. (2021). Determining optimal spatial and temporal complementarity between wind and hydropower. *Energy*, *230*. https://doi.org/10.1016/j.energy.2021.120790

Li, M., Fu, X., & Li, D. (2020). Diabetes Prediction Based on XGBoost Algorithm. *IOP Conference Series: Materials Science and Engineering*, *768*(7). https://doi.org/10.1088/1757-899X/768/7/072093

Li, W., Peng, X., Cheng, K., Wang, H., Xu, Q., Wang, B., & Che, J. (2020). A Short-Term Regional Wind Power Prediction Method Based on XGBoost and Multi-stage Features Selection. *2020 IEEE Student Conference on Electric Machines and Systems, SCEMS 2020*, 614–618. https://doi.org/10.1109/SCEMS48876.2020.9352249

Li, X., Leung, F. H. F., Su, S., & Ling, S. H. (2022). Sleep Apnea Detection Using Multi-Error-Reduction Classification System with Multiple Bio-Signals. *Sensors*, *22*(15), 1–17. https://doi.org/10.3390/s22155560

Li, X., Wang, J., Geng, Z., Jin, Y., & Xu, J. (2023). Short-term Wind Power Prediction Method Based on Genetic Algorithm Optimized XGBoost Regression Model. *Journal of Physics: Conference Series*, *2527*(1). https://doi.org/10.1088/1742-6596/2527/1/012061

Liu, G., Zhou, J., Jia, B., He, F., Yang, Y., & Sun, N. (2019). Advance short-term wind energy quality assessment based on instantaneous standard deviation and variogram of wind speed by a hybrid method. *Applied Energy*, *238*(January), 643–667. https://doi.org/10.1016/j.apenergy.2019.01.105

Ma, Z., Chang, H., Sun, Z., Liu, F., Li, W., Zhao, D., & Chen, C. (2020). Very Short-Term Renewable Energy Power Prediction Using XGBoost Optimized by TPE Algorithm. *2020 4th International Conference on HVDC, HVDC 2020*, 1236–1241. https://doi.org/10.1109/HVDC50696.2020.9292870

Mustaqeem, Ishaq, M., & Kwon, S. (2022). A CNN-Assisted deep echo state network using multiple Time-Scale dynamic learning reservoirs for generating Short-Term solar energy forecasting. *Sustainable Energy Technologies and Assessments*, *52*(PC), 102275. https://doi.org/10.1016/j.seta.2022.102275

Phan, Q. T., Wu, Y. K., & Phan, Q. D. (2021). A hybrid wind power forecasting model with xgboost, data preprocessing considering different nwps. *Applied Sciences (Switzerland)*, *11*(3), 1–19. https://doi.org/10.3390/app11031100

RajasundrapandiyanLeebanon, T., Murugan, N. S. S., Kumaresan, K., & Jeyabose, A. (2025). Long-term solar radiation forecasting in India using EMD, EEMD, and advanced machine learning algorithms. In *Environmental Monitoring and Assessment* (Vol. 197, Issue 3). https://doi.org/10.1007/s10661-025-13738-8

Rathore, H., Meena, H. K., & Jain, P. (2023). Prediction of EV Energy consumption Using Random Forest And XGBoost. *Proceedings - 2nd International Conference on Power Electronics and Energy, ICPEE 2023*, 1–6. https://doi.org/10.1109/ICPEE54198.2023.10060798

Renewable Energy Agency, I. (2025). Renewable Capacity Statistics 2025. In *Irena*. www.irena.org

Shang, Y., Miao, L., Shan, Y., Gnyawali, K. R., Zhang, J., & Kattel, G. (2022). A Hybrid Ultra-Short-Term and Short-Term Wind Speed Forecasting Method Based on CEEMDAN and GA-BPNN. *Weather and Forecasting*, *37*(4), 415–428. https://doi.org/10.1175/WAF-D-21-0047.1

Shi, R., Xu, X., Li, J., & Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Applied Soft Computing*, *109*, 107538. https://doi.org/10.1016/j.asoc.2021.107538

Shukla, S., & Pasari, S. (2025). Short-term wind speed prediction with adaptive signal processing based hybrid statistical models. In *Energy Systems* (Issue 0123456789). Springer Berlin Heidelberg. https://doi.org/10.1007/s12667-025-00727-6

Yan, X., Wang, L., Zhu, J., Wang, S., Zhang, Q., & Xin, Y. (2022). Automatic Obstructive Sleep Apnea Detection Based on Respiratory Parameters in Physiological Signals. *2022 IEEE International Conference on Mechatronics and Automation, ICMA 2022*, 461–466. https://doi.org/10.1109/ICMA54519.2022.9856347

Yelgeç, M. A., & Bingöl, O. (2022). Ayrık dalgacık dönüşümü ve Xgboost ile rüzgâr gücü tahmini. *Uluslararası Teknolojik Bilimler Dergisi*, *14*(2), 58–65. https://doi.org/10.55974/utbd.1132336

Yuzgec, U., Dokur, E., & Balci, M. (2024). A novel hybrid model based on Empirical Mode Decomposition and Echo State Network for wind power forecasting. *Energy*, *300*(April), 131546. https://doi.org/10.1016/j.energy.2024.131546

Zhang, T., Zhang, X., Rubasinghe, O., Liu, Y., Chow, Y. H., Iu, H. H. C., & Fernando, T. (2024). Long-Term Energy and Peak Power Demand Forecasting Based on Sequential-XGBoost. *IEEE Transactions on Power Systems*, *39*(2), 3088–3104. https://doi.org/10.1109/TPWRS.2023.3289400

Zheng, H., & Wu, Y. (2019). A XGBoost Model with Weather Similarity Analysis and Feature Engineering for Short-Term Wind Power Forecasting. *Applied Sciences (Switzerland)*.