




Article

Evaluation of Weather Information for Short-Term Wind Power Forecasting with Various Types of Models

Ju-Yeol Ryu ^{1,*}, Bora Lee ^{2,*}, Sungho Park ¹, Seonghyeon Hwang ¹, Hyemin Park ¹, Changhyeong Lee ¹ and Dohyeon Kwon ¹

¹ Institute for Advanced Engineering, Yongin 17180, Republic of Korea

² Institute of Health & Environment, Seoul National University, Seoul 08826, Republic of Korea

* Correspondence: juyeol.ryu@iae.re.kr (J.-Y.R.); mintbora@sun.ac.kr (B.L.)

Abstract: The rising share of renewable energy in the energy mix brings with it new challenges such as power curtailment and lack of reliable large-scale energy grid. The forecasting of wind power generation for provision of flexibility, defined as the ability to absorb and manage fluctuations in the demand and supply by storing energy at times of surplus and releasing it when needed, is important. In this study, short-term forecasting models of wind power generation were developed using the conventional time-series method and hybrid models using support vector regression (SVR) based on rolling origin recalibration. For the application of the methodology, the meteorological database from Korea Meteorological Administration and actual operating data of a wind power turbine (2.3 MW) from 1 January to 31 December 2015 were used. The results showed that the proposed SVR model has higher forecasting accuracy than the existing time-series methods. In addition, the conventional time-series model has high accuracy under proper curation of wind turbine operation data. Therefore, the analysis results reveal that data curation and weather information are as important as the model for wind power forecasting.

Keywords: wind power forecasting; time-series model; linear regression; support vector regression; rolling origin



Citation: Ryu, J.-Y.; Lee, B.; Park, S.; Hwang, S.; Park, H.; Lee, C.; Kwon, D. Evaluation of Weather Information for Short-Term Wind Power Forecasting with Various Types of Models. *Energies* **2022**, *15*, 9403. <https://doi.org/10.3390/en15249403>

Academic Editor: Davide Astolfi

Received: 14 November 2022

Accepted: 6 December 2022

Published: 12 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of renewable energy enables sustainable development by protecting the environment and addressing the depletion of natural resources; hence, it has continuously attracted the attention of researchers. Furthermore, numerous countries have rapidly increased the share of renewable energy to practice carbon neutrality, thus emphasising the importance of energy production [1,2] and power grid connection using solar and wind power [3]. The International Renewable Energy Agency projects the share of electricity generation using renewable energy to reach 61% by 2050, and solar and wind power generation capacity to continuously grow to 8828 GW and 6044 GW, respectively [4]. As various countries are expanding the share of solar and wind power generation, many research institutes are studying and various models to forecast renewable energy power generation. Developing forecasting models for renewable energy power generation not only makes it possible to review the business feasibility of specific regions, but it can also help power generation companies to reduce costs and secure economic feasibility by facilitating the review of management and operation plans beforehand. Moreover, the integrations between forecasting model and power system such as thermal storage and combined heat and power (CHP) are important for high usability of renewable energy and planning in power grid system [5]. Wind speed is the most important factor in wind power generation, and forecasting models for wind speed are continuously being developed. Initial forecasting models that combined several models had mean absolute percentage errors of 15.06–17.13% [6], whereas that of recent models has improved to approximately 2% [7].

However, forecasting wind power generation requires the consideration of various environmental variables, as the forecasting power can be improved by using various weather data other than wind speed and direction. In actual wind power forecasting models, it is difficult to forecast the output because the wind speed and direction constantly fluctuate, and this variability causes numerous problems in power systems [8]. Accordingly, it is essential to develop a sophisticated model for minute-based forecasting rather than hourly forecasting of the output. To solve this problem, researchers have proposed various forecasting models for wind power generation, and the forecasting power is being continuously improved [9–11]. Representative statistical forecasting models include autoregressive moving average (ARMA) using time series [12–14], autoregressive integrated moving average (ARIMA) [15,16], seasonal autoregressive integrated moving average (SARIMA) [17,18], and generalised autoregressive conditional heteroscedasticity (GARCH) [19,20]. Researchers have also used multiple linear regression (MLR) models to forecast the output and performance [21,22].

Recently, models to reduce the error between the actual output and predicted value have been developed, such as the machine learning techniques of artificial neural networks (ANNs) [23], deep neural networks (DNNs) [24], and convolutional neural networks (CNNs) [25], as well as methods using support vector machine (SVM) [26].

Although there are various forecasting models for wind power generation, the forecasting power may greatly differ according to the forecasted time range, wind turbine capacity, etc. The output of a wind farm is forecasted in monthly or hourly units on a long-term basis, with the capacity ranging from tens to hundreds of MW. Tae-Hui Park et al. [27] used CNN and ANN models to forecast the hourly output of a 46 MW wind farm, where the ANN and CNN showed mean absolute errors (MAEs) of 1.504 MW and 0.982 MW, respectively. Bigdeli et al. [28] used a neural network model to forecast the hourly output of a 150 MW wind farm, which showed an MAE of approximately 7.6 MW.

Researchers have investigated the superior performance of short-term minute-based forecasting models according to the forecasting power for one turbine. Wang et al. [29] presented a 15-min-based wind power forecasting model based on ARMA, which yielded an error of approximately 9%. Duan et al. [30] used long short-term memory (LSTM) neural networks and deep belief networks based on particle swarm optimisation to forecast the output of a 1.1 MW wind turbine, and reported an MAE of 35 kW.

Researchers have recently forecasted the minute-based outputs of large-capacity wind farms by combining multiple models and achieved excellent forecasting power. Qin et al. [31] used a DNN model to measure the outputs of Wind Farm-1 (49.5 MW), Wind Farm-2 (48 MW), and Wind Farm-3 (48 MW) in 15 min units, with MAEs of 3.3 MW, 2.8 MW, and 3.1 MW, respectively. Liu et al. [32] proposed a multivariate phase space reconstruction model and MLR by using the wind speed, wind direction, and atmospheric pressure as the variables, and forecasted the power in 15 min units with an error of approximately 6.46%. The ARIMA model showed an error of approximately 12.7%, which suggested a considerable improvement in the forecasting performance. Additionally, Qin et al. [33] trained a hybrid model linked with variational mode decomposition by using the actual data of wind farms capable of output of up to 150 MW, and then forecasted the output of a wind farm operating from 0 to 80 MW in 5 min units. The proposed model showed an MAE of approximately 3 MW. Finally, Yu et al. [34] trained a complexity-trait-driven rolling decomposition-reconstruction-ensemble model with 5 min unit output data of a 2500 MW wind farm and used the model to forecast the output. The method had an MAE of approximately 27 MW, demonstrating excellent forecasting power.

Thus, researchers have continuously upgraded and modified the various forecasting models and greatly improved the forecasting performance for wind power generation. However, despite the advantages of linear statistical models such as ARIMA and MLR when compared with the machine learning models—simplicity and transparency, parameter interpretation, and reliable results even with small datasets—their forecasting power is

relatively poor owing to issues such as increased noise caused by unrefined data; hence, they have been neglected in recent studies.

This study sought to reduce the uncertainty of the data provided to the forecasting model by applying special operating conditions such as torque control in addition to the collection of actual wind turbine operation data. In addition, we aimed to demonstrate that the forecasting power of statistical models such as ARIMA and MLR, which have an error rate of 7–12% according to a literature review, can be improved by considering appropriate variables and applying data refinement, and that the models can be used for minute-based forecasting of wind power generation. The main contributions of this paper are the review of various model performance through a validation using actual operation data. The SVM model proved to be better than conventional statistical model in short-term forecasting accuracy. However, by comparing with SVM, a machine learning technique, we sought to demonstrate that the conventional statistical model such as MLR do not yield inferior performance.

2. Method

2.1. Data Source

To forecast the wind power generation, we used the operational data of the Yeonggwang Baeksu wind farm located in Yeonggwang, Jeollanam-do, South Korea (see Figure 1). The Yeonggwang Baeksu wind farm was built in 2015; it is operated by Korea East–West Power, and can produce 40 MW of output. To develop a forecasting model for wind power generation, we used data such as the average power (kW) produced by one 2.3 MW wind turbine and the wind speed. The data were collected at 10 min intervals from January 1 to December 31, 2020, through a public data portal (www.data.go.kr, accessed on 2 March 2022) [35]. In addition, 51,161 cases were calculated using the data from the Automatic Weather System provided by the Korea Meteorological Administration (www.weather.go.kr, accessed on 2 March 2022), such as temperature, pressure, and humidity in the region. Table 1 shows the summarised values of the collected data.

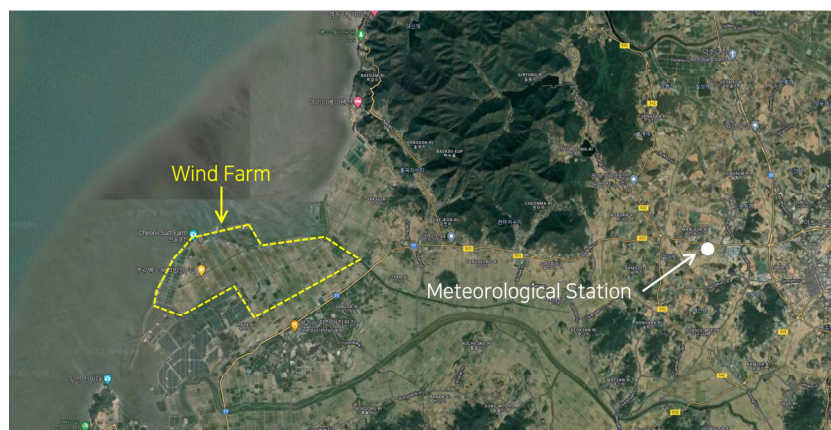


Figure 1. Location of Yeonggwang Baeksu wind farm.

Table 1. Statistical summary of weather information.

Variables	Unit	Min.–Max.	Mean \pm SD
Wind power	kW	0.0–2357.9	536.8 \pm 729.6
Wind speed	m/s	0.0–36.9	7.2 \pm 4.6
Wind direction	$^{\circ}$	0.0–360.0	146.9 \pm 120.9
Temperature	$^{\circ}$ C	−9.9–34.5	13.3 \pm 9.3
Pressure	hPa	981.7–1032.7	1012.5 \pm 8.7
Humidity	%	16.9–99.9	75.7 \pm 17.6

As is commonly known, when converting wind into energy, according to Betz's law, the threshold of efficiency is approximately 59.36%, whereas the efficiency of the current commercialised wind turbines is 15–45% [36]. Accordingly, researchers have attempted to reduce the gap between the efficiency of commercial wind turbines and the theoretical efficiency limit [37]. As part of these efforts, along with the development of wind power forecasting models, data refinement procedures to improve the forecasting power are also critical [38]. When operating a wind turbine, even under good wind conditions, particular operating points may need to be considered, such as the output limits due to capacity problems in the power grid. Therefore, a procedure to refine the operational data that may be unrelated to weather is essential. This study analysed the regional wind direction and wind speed data, refined the power generation data based on the output limits and specific operating points, and then conducted an analysis.

2.2. Forecasting Model

2.2.1. ARIMA and ARIMAX

Autoregressive integrated moving average (ARIMA) model, a linear combination of autoregressive (AR) and moving average (MA) equations, is a representative model using time-series data. First, in the AR model of order p , the data observed at time t (y_t) consists of a linear combination of the past data ($y_{t-1}, y_{t-2} \dots$) and weights (ϕ_i) with error term (ε_t) as Equation (1). In the MA model of order q , y_t is represented as a linear combination of the present and past error terms (ε_{t-i}) with the weights (θ_i) as Equation (2).

$$\text{AR}(p) : y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

$$\text{MA}(q) : y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Using the backshift operator ($B, B^p y_t = y_{t-p}$), Equations (1) and (2) can be re-written as $(1 - \phi_1 B - \dots - \phi_p B^p) y_t = \varepsilon_t$ and $y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$. With the operator $\phi_p(B)$ defined as $1 - \phi_1 B - \dots - \phi_p B^p$, the AR(p) and MA(q) can be represented more concisely as $\phi_p(B) y_t = \varepsilon_t$ and $y_t = \theta_q(B) \varepsilon_t$. Adding the differencing operator, the ARIMA model can be expressed as

$$\text{ARIMA}(p, d, q) : \phi_p(B)(1 - B)^d y_t = \theta_q(B) \varepsilon_t \quad (3)$$

where p is the order of lag, d is the degree of differencing, and q is the order of the moving average. Exogenous variables, represented as X , are added as independent variables to the ARIMA model, so called ARIMAX. We assume that the independent and dependent variables are stationary time series, and the error follows the ARIMA model. If x_{it} is the i -th independent variable and β_{it} is the regression coefficient of the corresponding independent variable, then Equation (3) can be expanded as follows.

$$\text{ARIMAX}(p, d, q) : \phi_p(B)(1 - B)^d (y_t - \sum \beta_i x_{it}) = \theta_q(B) \varepsilon_t \quad (4)$$

2.2.2. SARIMA and SARIMAX

As wind has seasonal characteristics that vary with the month and quarter, the seasonal autoregressive integrated moving average (SARIMA) model with seasonality was also used to verify the performance. The wind power forecasting model simultaneously includes both seasonal and non-seasonal patterns. Therefore, a multiplicative ARIMA model is applied, and the (P, D, Q, s) model considering non-seasonal orders (p, d, q) and the seasonal order can be expressed as Equation (5) where s means the time span of seasonal pattern and $\Phi_P(B^s)$ and $\theta_Q(B^s)$ represent the seasonal AR and MA operator.

$$\text{SARIMA}(p, d, q)(P, D, Q, s) : \phi_p(B) \Phi_P(B^s) (1 - B)^d (1 - B^s)^D y_t = \theta_q(B) \theta_Q(B^s) \varepsilon_t \quad (5)$$

Exogenous variables are added as independent variables to the SARIMA model. As with SARIMAX, the i -th independent variable x_{it} and dependent variable y_{it} are assumed to be a stationary time series, and the error follows SARIMA, which can be expressed as

$$\text{SARIMAX}(p,d,q)(P,D,Q,s):$$

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D(y_t - \sum \beta_i x_{it}) = \theta_q(B)\theta_Q(B^s)\varepsilon_t \quad (6)$$

2.2.3. GARCH

GARCH (generalised auto regressive conditional heteroskedasticity), the most basic model of heteroskedasticity, considers the error term for stationary time-series variables. It is composed with respect to the term (ε^2_{t-1}) of the residual squared for the conditional mean and the predicted term (σ^2_{t-1}) for the conditional variance. The GARCH (P,Q) model is defined as

$$\sigma_t^2 = \omega + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 \dots + \alpha_P \sigma_{t-P}^2 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \varepsilon_{t-2}^2 \dots + \beta_Q \varepsilon_{t-Q}^2 \quad (7)$$

2.2.4. MLR

The next method is multiple linear regression (MLR) analysis in which several independent variables affect one dependent variable. Through previous studies, it was confirmed that the major variables such as wind speed, wind direction, and temperature influence the power generation. Hence, these existing variables and additional variables are calculated and expressed as Equation (8).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_t, \varepsilon_t \sim N(0, 1^2) \quad (8)$$

Variables such as wind speed (WS), wind direction (WD), atmospheric pressure (AP), air temperature (AT), and humidity (HM) are commonly used to forecast the wind power generation [27]. Because wind has seasonality, a season (SS) variable can be added, and because the wind directions during day and night are different near a coast, a time category (TC) variable can be used.

2.2.5. SVR

SVR (support vector regression) is a machine learning technique that finds the optimal hyperplane that contains as much data as possible within the distance between the vectors and then uses this to forecast the wind power generation. It can be expressed as shown in Equation (9), where x is the input vector, y is the output vector, ξ_i is the slack variable, b is the bias, C is the regularisation parameter, and ω is the coefficient vector orthogonal to the hyperplane.

$$\min. \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i$$

$$\text{s.t. } y_i(\omega^t x_i + b) \geq 1 - \xi_i \text{ for } i = 1, \dots, n$$

$$\xi_i \geq 0 \quad (9)$$

2.3. Model Development and Evaluation

To find the optimal form of each model and verify its performance, we divided the data into a training set and a test set. When building an optimal model, overfitting may occur, wherein the learned data fit the model well but the prediction rate for new data is poor. We attempted to solve this problem through cross-validation. However, because general K-fold cross-validation is not suitable for time-series data in which the training data must precede the verification data, we applied time-series cross-validation (rolling origin recalibration) [39] and selected the optimal model through 10-fold time-series cross-validation.

Based on a literature review [40,41], we used the mean absolute error (MAE), normalised mean absolute error (NMAE), root mean squared error (RMSE), and Akaike

information criterion (AIC) as the performance indicators to evaluate the performance of the estimated wind power forecasting models. The calculation formulas are as follows.

$$\text{MAE}(kW) : \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (10)$$

$$\text{NMAE}(\%) : \frac{1}{C N} \sum_{i=1}^N |\hat{y}_i - y_i| \times 100 \quad (11)$$

$$\text{RMSE} : \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (12)$$

$$\text{AIC} : -2\ln L + 2k \quad (13)$$

In the above equations, \hat{y}_i is the predicted value, y_i is the actual value, N is the number of observations and C applied to NMAE is the capacity of the forecasted wind farm. In addition, L is the log-likelihood and k is the estimated parameter of the model. Because the MAE greatly varies with the wind farm's capacity, we analysed the main results based on the NMAE, which indicates the error in comparison with the wind power capacity [42]. Figure 2 shows the overall analysis process.

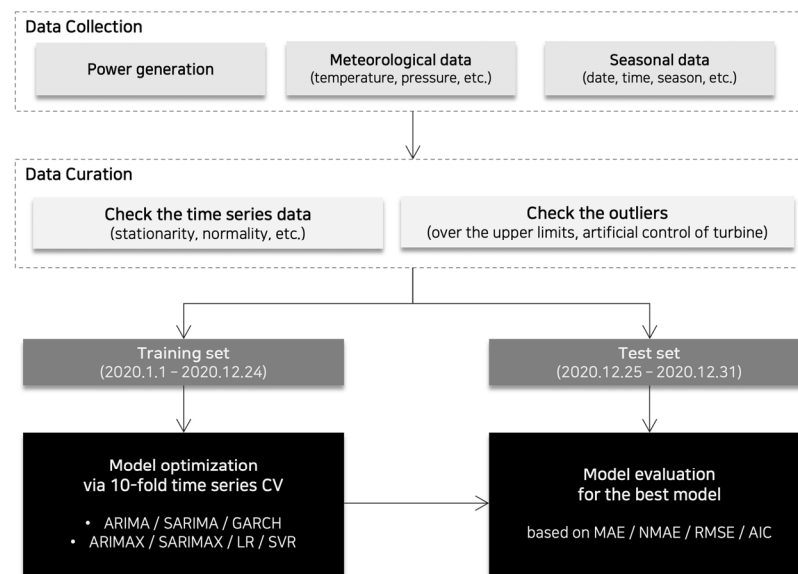


Figure 2. Process flow of data analysis.

2.4. Statistical Analyses

The data observed at 10 min intervals were considered to have a period of 144 (24 h \times 6 data points/h). The seasonal periods were analysed accordingly and the time-series model was fitted. For the time-series data and residuals of the model, we tested for normality based on the skewness and kurtosis using the Jarque–Bera test. When fitting the ARIMA, SARIMA, and GARCH models, the stationarity was tested based on the augmented Dickey–Fuller test and the Kwiatkowski, Phillips, Schmidt, and Shin test after being differencing at a time. When the number of differencing has been determined, the orders of p and q were chosen by minimizing the Akaike Information Criteria (AIC) within a predefined range of orders as shown in Table 2. Additionally, autocorrelation function (ACF) and partial autocorrelation function (PACF) were examined under the determined order of the final model. Finally, the partmanteau test was applied to check the residuals if it is a white noise or not. The correlation between power generation (the forecasting target) and the exogenous variable was calculated through Pearson's correlation coefficient.

When selecting the final model with exogenous variables, the grid search method was used based on AIC for all models except SVR. In the case of SVR, the hyperparameters that minimise the loss function were selected. For SVR, the libraries of Tensorflow 2.2.0 and Keras 2.3.1 in Python 3.7 were applied [43–45], and for the other models, the fpp2 and rugarch packages were applied in R 4.2.1 (The R Development Core Team, Vienna, Austria) for the analysis [46,47]. For the final model, we performed the Diebolde-Mariano (DM) test to confirm statistically significant differences between the models [48]. This test uses the prediction error of each model to check whether there is a significant difference in the loss function calculated with two models. Thus, it is used to compare the predictive power between two models.

Table 2. Range of parameters for grid search and the final model in model optimisation.

Model	Parameter	Candidate Range	Final Model
Single time-series model			
ARIMA	p, d, q	0–10	ARIMA(3,1,2)
SARIMA	p, d, q, P, D, Q	0–10 for p, d, q 0–12 for P, D, Q	SARIMA(4,1,2)(1,0,2) (144)
ARIMA-GARCH	p, d, q, P, Q	0–10 for p, d, q 0–5 for P, Q	ARIMA(3, 1, 2)-GARCH(1, 1)
With weather information			
MLR*	-	-	PG ~ PG(lag 1), WS, WD, AT, HM, TC, SS, AT × SS, HM × TC
ARIMAX	p, d, q	0–10	ARIMAX(1, 1, 1)
SARIMAX	p, d, q, P, D, Q	0–10 for p, d, q 0–12 for P, D, Q	SARIMA(3, 0, 1)(0,0,1) (144)
SVR	Kernel Cost Epsilon	linear, polynomial, radial bias 1–10 0.1 to 0.9 by 0.05	radial bias 2 0.15

MLR considered eight factors, namely, lag1 power generation (PG(1)), wind speed (WS), wind direction (WD), air temperature (AT), atmospheric pressure (AP), humidity (HM), time category (TC), and season (SS). ARIMAX, SARIMAX, ARIMA-GARCHX, and SVR considered wind speed, wind direction, air temperature, and humidity as covariates.

3. Results

3.1. Data Curation

The trends and seasonality of the data such as wind speed and average power (kW) produced by one 2.3 MW wind turbine in Yeonggwang for one year in 2020 were sequentially decomposed, and the results are shown in Figure 3a. Although the data did not exhibit a characteristic trend, the variation in the mean was very large, and a period was clearly observed. The average power generation per 10 min by season is shown in Figure 3b. Through this, we confirmed that the average level of wind power generation per 10 min varies with the season, and a very large deviation occurs according to time.

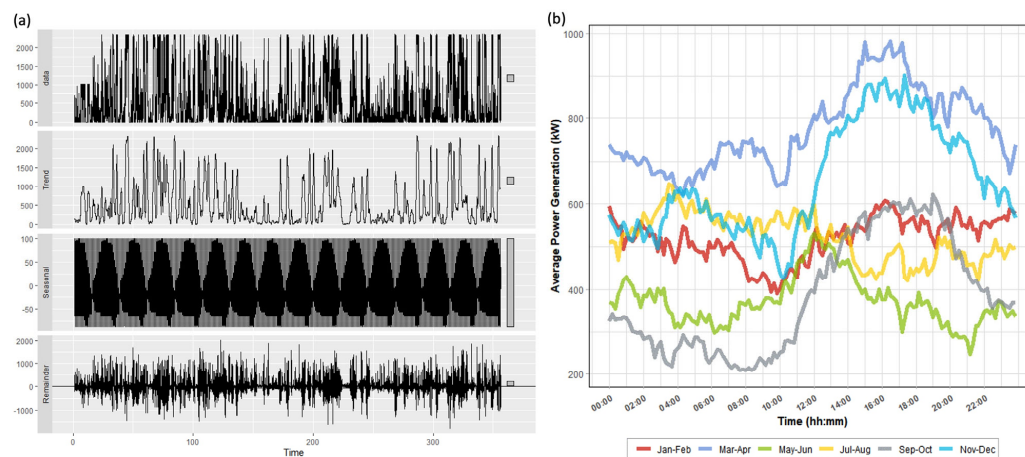


Figure 3. Time—series plot of wind power generation: (a) Time series plot of seasonal and trend decomposition using loess (STL decomposition), (b) Average 10 min power generation per season.

The change in output according to wind speed of the 2.3 MW wind turbine used in the analysis is shown in Figure 4. The maximum power of 2.3 MW can be produced under a wind speed of approximately 11 m/s, and the power produced varies with the main direction of the wind (orange line). Moreover, the operational data indicate that the output is limited, even at the wind speed that can produce the maximum power (yellow area). Accordingly, to improve the accuracy of the forecasting models, the operating conditions with artificially limited output should be excluded before developing the models.

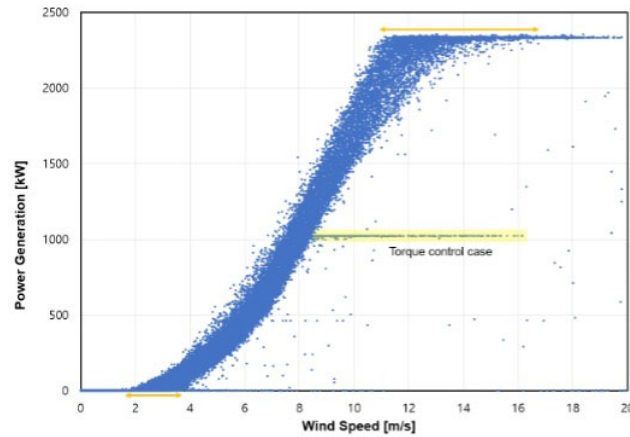


Figure 4. Actual wind power curve for 3 MW generation in Yeonggwang Baeksu wind farm.

Next, to express the relationship between wind direction and wind speed, the wind direction was divided into 16 directions and the wind speed into six levels at intervals of 3 m/s, as shown in Figure 5. According to the wind rose results, the main wind direction in the Yeonggwang-Baeksu wind farm is north (0°).

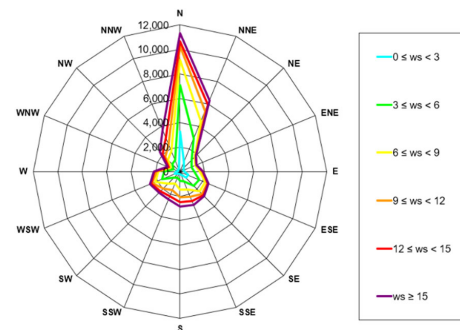


Figure 5. Actual wind direction at 2.3 MW plant in Yeonggwang Baeksu wind farm.

Figure 6 shows the energy distribution by direction. In the region where the turbine is installed, the highest energy distribution is in the north (0°) direction at 22.3%, followed by the distribution of 12.5% in the north-northeast (22.5°).

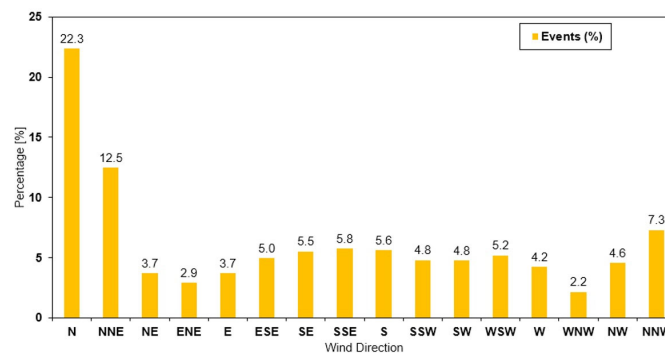


Figure 6. Distribution of wind direction.

Figure 7 shows the distribution of wind speed energy, which was 31.8% at 3–6 m/s and 22.1% at 6–9 m/s. Considering that the cut-in wind speed of a wind turbine is generally 3 m/s, approximately 82.4% of the wind speed is effectively utilised for energy conversion. However, this value decreases if the wind direction is considered.

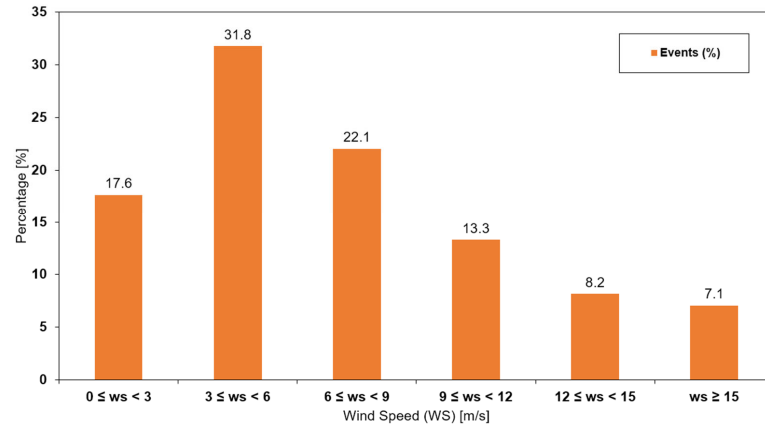


Figure 7. Distribution of wind speed classes.

To enhance the forecasting accuracy for power generation, we analysed the correlation between power generation and the added exogenous variables of wind direction, wind speed, temperature, atmospheric pressure, and humidity, the results of which are shown in Figure 8. For all variables except wind speed, a strong correlation coefficient of at least 0.8 or correlation of at least 0.6 with power generation was not detected; hence, model selection was performed with all variables included in the analysis [49]. However, multicollinearity may occur when atmospheric pressure and temperature are considered simultaneously; hence, these two variables were separately input to the models and the results were compared to select the optimal variable.

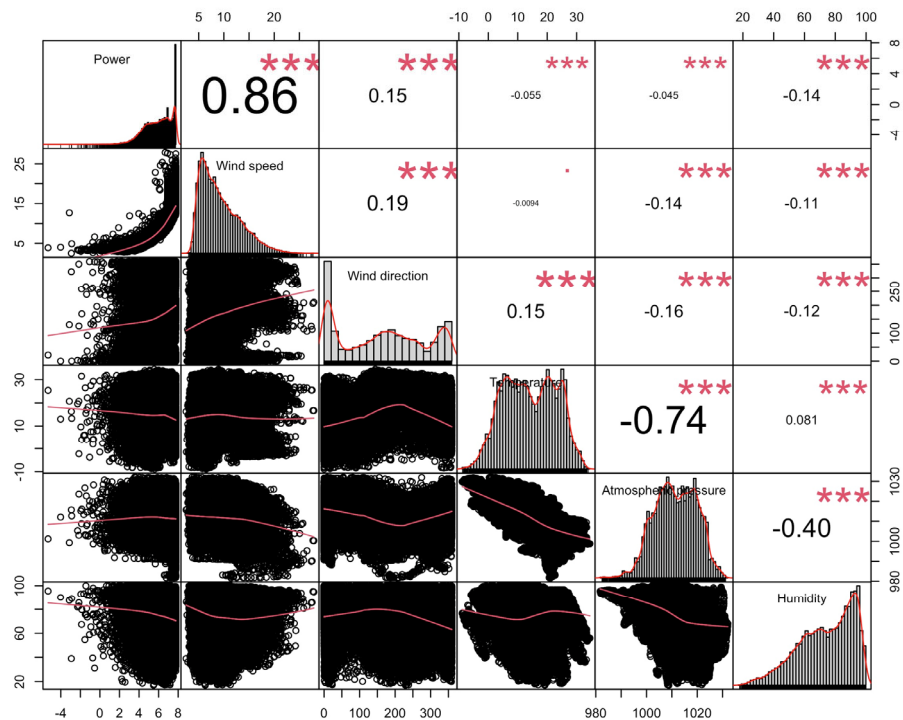


Figure 8. Correlation matrix plot of exogenous variables. (***) $p < 0.001$; ** $p < 0.01$; * $p < 0.05$).

3.2. Model Optimisation

The models were built based on data measured from 00:00 on 1 January to 23:50 on 24 December 2020. The candidate values used when selecting the optimal values of each model parameter through the grid search method and the final models are shown in Table 2. From the single time-series models, ARIMA (3,1,2), SARIMA (4,1,2) (1,0,2) (144), and ARIMA (3,1,2)-GARCH (1,1) were selected as the optimal models. Among the regression models that use weather information, the parameters lag1 power generation, wind speed, wind direction, temperature, humidity, time group, seasonal group, interaction between temperature and seasonal group, and interaction between humidity and time group were selected as the final variables for the multiple linear regression model. For ARIMAX and SARIMAX, model seasonality and autocorrelation were considered, whereas lag1 power generation, time group, and seasonal group were excluded; finally, ARIMAX(1,1,1) and SARIMAX(3,0,1)(0,0,1) (144) were selected with consideration of wind speed, wind direction, temperature, and humidity. For SVR, the final model was calculated under the hyperparameters of Cost 2 and Epsilon 0.15 based on the kernel function of the radial bias function.

3.3. Performance Comparison

To examine the forecasting power of each model, we calculated the performance evaluation indicators using a test dataset containing data from 00:00 on 25 December to 23:50 on 31 December 2020, that is, the data of last seven days. Table 3 shows the results. Smaller values of MAE, NMAE, RMSE, and AIC indicate higher performance. However, the number of orders to estimate differs for each model, and it is difficult to directly compare the AIC of a linear regression model or SVR with that of ARIMA. Hence, when analysed based on the NMAE, which can account for the maximum capacity of wind power, the SVR model has a low error rate of 6.31% when compared with the other models, indicating the best forecasting power. The lowest NMAE of the single time-series models ARIMA, SARIMA, and ARIMA-GARCH was inferior to those of the other models. However, ARIMAX and SARIMAX, in which weather variables are applied as exogenous variables to the single time-series model, showed improved NMAE at 14.52% and 9.61%, respectively. A probable reason for the performance of the linear models being similar to that of the traditional time-series models was the inclusion of lag1 power generation, which reflects autocorrelation. Among the single time-series models, ARIMA-GARCH had a relatively high error rate. In this regard, because wind power generation is influenced to a greater extent by weather conditions than the existing power generation level or variations in the parameters, the GARCH model, which is specialised for modelling conditional self-variability within a single time-series component, yielded poor performance.

Table 3. Comparison of performance of each model excluding artificial control case.

Model	MAE (kW)	NMAE (%)	RMSE (kW)	AIC
Single time-series model				
ARIMA	684.24	29.75	2847.11	150,593.8
SARIMA	521.46	22.67	2704.04	149,480.4
ARIMA-GARCH	1447.70	62.94	17,429.79	292,465.1
With weather information				
MLR	150.54	6.55	33.62	143,138.9
ARIMAX	350.28	15.23	93.13	147,805.1
SARIMAX	300.40	13.06	84.71	147,436.1
SVR	145.08	6.31	30.72	-

To demonstrate the effect of excluding artificial control case, six individual models were used for comparison in Experiment. The best results can be obtained from SVR model, then the NMAE values of six models are improved 2.14%, 7.22%, -0.27% , 1.93%, 3.32%, 4.36% and 0.31%, respectively from including artificial control case (Appendix A, Table A1).

We conducted the DM test to investigate whether there was a significant difference in the forecasting power of the remaining six models based on SVR, which showed the best forecasting performance. Except for MLR, the others showed significantly lower forecasting power than SVR ($p < 0.001$ for all), whereas MLR showed no significant difference in forecasting power at the 0.05 level ($p = 0.186$ for MLR vs. SVR).

Figure 9 shows a time-series plot of the predicted values of the six models in the test set excluding ARIMA-GARCH model which showed the lower performance. The prediction errors of the single time-series models were relatively large when a sudden fluctuation occurred during the seven days. The other models with covariates also showed a relatively large error since a constraint for exceeding the output of 2.3 MW cannot be placed.

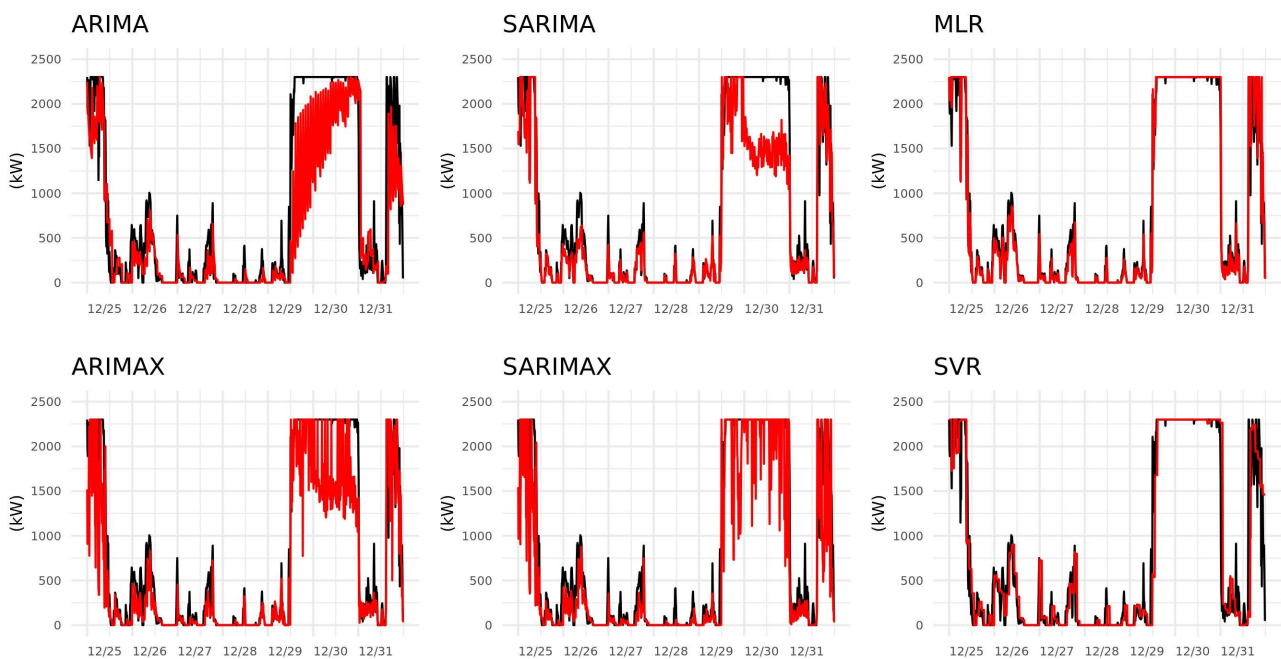


Figure 9. Actual (black) and predicted (red) value of power generation on 25 December to 31 December.

4. Discussion

As the share of power generated from renewable energy increases worldwide, various power grid-related issues have been noted, which can cause massive economic losses due to reasons such as blackouts. Hence, developing reliable models to forecast the power obtainable from renewable energy can contribute to not only stabilising the grid but also estimating the renewable energy capacity in a specific region.

In this study, the ARIMA, SARIMA, and GARCH models, which are time-series models traditionally used to forecast wind power generation, were applied along with the SVR model, a machine learning technique widely used in forecasting models lately; then, the accuracies of the 10 min wind power forecasting models were compared. Traditional time-series models are generally known to have very poor forecasting power. However, after removing white noise and operational data that artificially control the turbine output, their forecasting power was found to be greatly improved through the time-series rolling origin recalibration method. Moreover, we confirmed that sufficient forecasting power can be achieved if the models are trained after excluding the operational data that may produce less energy than the maximum possible value, such as torque control and blade pitch angle control.

The proposed models were trained using weather information and operational data from 1 January to 24 December 2015, and then used to forecast the power generation in 10 min intervals during the last week of the data collection period, that is, from 25 to 31 December 2015. According to the results, the SVR model, a machine learning technique,

outperformed the time-series models. However, building the SVR model becomes more complex as the number of support vectors increases when compared with the training dataset, thereby reducing its advantages when compared with the classical time-series forecasting techniques. Although the total size of the training set in this study was not large, by continuously accumulating historical data, a suitable time window size can be fixed to effectively realise these advantages. Nevertheless, a limitation is that the model object containing the data itself must be passed to perform forecasting. In contrast, for classical time-series models or multiple linear regression models, the regression coefficients can be intuitively interpreted, and the models can be easily reproduced without data or a separate object. For short-term forecasting models of 10 min or less, because it is necessary to anticipate the changes in power generation due to input variable fluctuations and modify the operating conditions accordingly, it is vital to predict and interpret the nature of effects of the input variables on the power generation. Hence, regardless of the forecasting power, it is important to explore additional input variables that can improve the intuitive models, such as classical statistical models of linear regression models. Additionally, proper data curation is as important as the to increase the accuracy of forecasting model.

This study has a few limitations. First, the data for only one year and from only one site were used. When using a training set with data of one year or more to build a model, seasonal variations throughout the year must be considered. Second, the models employing machine learning in this study were limited to SVR. Future research will compare the performances of deep learning models such as LSTM and DNN, which are known to be excellent for short-term forecasting. Third, additional exogenous variables that can affect the operating conditions aside from weather variables must be considered. Fourth, in the field, predicted values must be used for the weather variables, which are applied as the input variables. Hence, by collecting the predicted values reported by the Korea Meteorological Administration, the variation in the prediction error when predicting the input variables must also be considered in the model.

5. Conclusions

This study demonstrated that in addition to being easy to interpret and reproduce, classical time-series models can achieve reasonable short-term forecasting performance when compared with machine learning by using exogenous variables and data refinement based on data from actual wind power generation and weather information. In future research, we will add exogenous variables and compare various machine learning and deep learning models to quantitatively evaluate the degree of performance improvement of classical statistical models.

Author Contributions: Conceptualization, J.-Y.R. and B.L.; methodology, B.L.; software, B.L.; validation, B.L. and S.H.; formal analysis, J.-Y.R.; investigation, J.-Y.R., H.P. and D.K.; resources, S.H.; data curation, J.-Y.R., C.L. and S.P.; writing—original draft preparation, J.-Y.R.; writing—review and editing, J.-Y.R. and B.L.; visualization, B.L.; supervision, B.L.; project administration, S.P.; funding acquisition, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Korean Energy Technology Evaluation and Planning, grant number 20213030020280.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Comparison of performance of each model including artificial control case.

Model	MAE (kW)	NMAE (%)	RMSE (kW)	AIC
Single time-series model				
ARIMA-i	733.58	31.89	2904.62	150,593.8
SARIMA-i	687.61	29.89	2873.40	149,480.4
ARIMA-GARCH-i	1442.76	62.70	1740.219	292,465.1
With weather information				
MLR-i	194.97	8.48	42.439	143,138.9
ARIMAX-i	426.56	18.55	104.064	147,805.1
SARIMAX-i	400.60	17.42	91.977	147,436.1
SVR-i	152.25	6.62	33.900	Not Available

i means the model 'including artificial control case'.

References

- Energy Outlook. 2020. Available online: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2020.pdf> (accessed on 26 September 2022).
- Cozzi, L.; Gould, T.; Bouckart, S.; Crow, D.; Kim, T.; Mcglade, C.; Olejarnik, P.; Wanner, B.; Wetzel, D. *World Energy Outlook 2020*; IEA: Paris, France, 2020; pp. 1–461.
- Arraño-Vargas, F.; Shen, Z.; Jiang, S.; Fletcher, J.; Konstantinou, G. Challenges and Mitigation Measures in Power Systems with High Share of Renewables—The Australian Experience. *Energies* **2022**, *15*, 429. [\[CrossRef\]](#)
- International Renewable Energy Agency Abu Dhabi (IRENA). *Global Renewables Outlook: Energy Transformation 2050*; IRENA: Masdar City, Abu Dhabi, 2020.
- Wei, J.; Zhang, Y.; Wang, J.; Wu, L.; Zhao, P.; Jiang, Z. Decentralized Demand Management Based on Alternating Direction Method of Multipliers Algorithm for Industrial Park with CHP Units and Thermal Storage. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 120–130. [\[CrossRef\]](#)
- Xiao, L.; Wang, J.; Dong, Y.; Wu, J. Combined forecasting models for wind energy forecasting: A case study in China. *Renew. Sustain. Energy Rev.* **2015**, *44*, 271–288. [\[CrossRef\]](#)
- Wang, J.; An, Y.; Li, Z.; Lu, H. A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting. *Energy* **2022**, *251*, 123960. [\[CrossRef\]](#)
- Frías-Paredes, L.; Mallor, F.; Gastón-Romeo, M.; León, T. Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors. *Energy Convers. Manag.* **2017**, *142*, 533–546. [\[CrossRef\]](#)
- Jung, J.; Broadwater, R.P. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* **2014**, *31*, 762–777. [\[CrossRef\]](#)
- Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A critical review of wind power forecasting methods—Past, present and future. *Energies* **2020**, *13*, 3764. [\[CrossRef\]](#)
- Bazionis, I.K.; Georgilakis, P.S. Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research. *Electricity* **2021**, *2*, 13–47. [\[CrossRef\]](#)
- Rajagopalan, S.; Santoso, S. Wind power forecasting and error analysis using the autoregressive moving average modeling. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 26–30 July 2009; pp. 1–6.
- Gomes, P.; Castro, R. Wind speed and wind power forecasting using statistical models: Autoregressive moving average (ARMA) and artificial neural networks (ANN). *Int. J. Sustain. Energy Dev.* **2012**, *1*. [\[CrossRef\]](#)
- Cao, Y.; Liu, Y.; Zhang, D.; Wang, W.; Chen, Z. Wind power ultra-short-term forecasting method combined with pattern-matching and ARMA-model. In Proceedings of the 2013 IEEE Grenoble Conference, Grenoble, France, 16–20 June 2013; pp. 1–4.
- Barbosa de Alencar, D.; de Mattos Affonso, C.; Limão de Oliveira, R.C.; Moya Rodriguez, J.L.; Leite, J.C.; Reston Filho, J.C. Different models for forecasting wind power generation: Case study. *Energies* **2017**, *10*, 1976. [\[CrossRef\]](#)
- Eldali, F.A.; Hansen, T.M.; Suryanarayanan, S.; Chong, E.K. Employing ARIMA models to improve wind power forecasts: A case study in ERCOT. In Proceedings of the 2016 North American Power Symposium (NAPS), Denver, CO, USA, 18–20 September 2016; pp. 1–6.
- Haddad, M.; Nicod, J.; Mainassara, Y.B.; Rabehasaina, L.; Al Masry, Z.; Péra, M. Wind and solar forecasting for renewable energy system using sarima-based model. In Proceedings of the International Conference on Time Series and Forecasting, Gran Canaria, Spain, 25–27 September 2019.
- Tena García, J.L.; Cadenas Calderón, E.; González Ávalos, G.; Rangel Heras, E.; Mbikayi Tshikala, A. Forecast of daily output energy of wind turbine using sARIMA and nonlinear autoregressive models. *Adv. Mech. Eng.* **2019**, *11*, 1687814018813464. [\[CrossRef\]](#)
- Chen, H.; Zhang, J.; Tao, Y.; Tan, F. Asymmetric GARCH type models for asymmetric volatility characteristics analysis and wind power forecasting. *Prot. Control. Mod. Power Syst.* **2019**, *4*, 29. [\[CrossRef\]](#)
- Chen, H.; Li, F.; Wang, Y. Wind power forecasting based on outlier smooth transition autoregressive GARCH model. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 532–539. [\[CrossRef\]](#)

21. Amral, N.; Ozveren, C.; King, D. Short term load forecasting using multiple linear regression. In Proceedings of the 2007 42nd International Universities Power Engineering Conference, Brighton, UK, 4–6 September 2007; pp. 1192–1198.
22. Ryu, J.; Cha, J.; Lee, B. Evaluation of Weather Information in Forecasting Daily Peak Load of Electricity Demand. *J. Korean Inst. Illum. Electr. Install. Eng* **2018**, *32*, 73–81.
23. Chen, Q.; Folly, K.A. Short-Term Wind Power Forecasting Using Mixed Input Feature-Based Cascade-connected Artificial Neural Networks. *Front. Energy Res.* **2021**, *9*, 634639. [[CrossRef](#)]
24. Wu, W.; Chen, K.; Qiao, Y.; Lu, Z. Probabilistic short-term wind power forecasting based on deep neural networks. In Proceedings of the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, China, 16–20 October 2016; pp. 1–8.
25. Mujeeb, S.; Javaid, N.; Gul, H.; Daood, N.; Shabbir, S.; Arif, A. Wind power forecasting based on efficient deep convolution neural networks. In Proceedings of the International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Antwerp, Belgium, 7–9 November 2019; pp. 47–56.
26. Zhang, H.; Chen, L.; Qu, Y.; Zhao, G.; Guo, Z. Support vector regression based on grid-search method for short-term wind power forecasting. *J. Appl. Math.* **2014**, *2014*, 835791. [[CrossRef](#)]
27. Park, T.-H.; Jang, D.-S.; Bae, G.-M.; Kim, K.-M.; Ahn, J.-H. Selection of Input variables and comparison of Artificial Neural Networks and one-dimensional Convolutional Neural Networks for Prediction of Wind Power Generation in Yeongheung Wind Power Plant. *J. Korean Soc. Environ. Eng.* **2021**, *43*, 219–229. [[CrossRef](#)]
28. Bigdeli, N.; Afshar, K.; Gazafroudi, A.S.; Ramandi, M.Y. A comparative study of optimal hybrid methods for wind power prediction in wind farm of Alberta, Canada. *Renew. Sustain. Energy Rev.* **2013**, *27*, 20–29. [[CrossRef](#)]
29. Wang, J.; Zhou, Q.; Zhang, X. Wind power forecasting based on time series ARMA model. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *199*, 022015. [[CrossRef](#)]
30. Duan, J.; Wang, P.; Ma, W.; Fang, S.; Hou, Z. A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107452. [[CrossRef](#)]
31. Qin, J.; Yang, J.; Chen, Y.; Ye, Q.; Li, H. Two-stage short-term wind power forecasting algorithm using different feature-learning models. *Fundam. Res.* **2021**, *1*, 472–481. [[CrossRef](#)]
32. Liu, R.; Peng, M.; Xiao, X. Ultra-short-term wind power prediction based on multivariate phase space reconstruction and multivariate linear regression. *Energies* **2018**, *11*, 2763. [[CrossRef](#)]
33. Qin, G.; Yan, Q.; Zhu, J.; Xu, C.; Kammen, D.M. Day-ahead wind power forecasting based on wind load data using hybrid optimization algorithm. *Sustainability* **2021**, *13*, 1164. [[CrossRef](#)]
34. Yu, L.; Ma, Y.; Ma, Y.; Zhang, G. A complexity-trait-driven rolling decomposition-reconstruction-ensemble model for short-term wind power forecasting. *Sustain. Energy Technol. Assess.* **2022**, *49*, 101794. [[CrossRef](#)]
35. Korea East-West Power Co., Ltd. Younggwang Baeksu Wind Power Complex Unit 1, 10-Minute Average Power Generation. 2022. Available online: <https://www.data.go.kr/data/15091978/fileData.do> (accessed on 13 October 2021).
36. Burton, T.; Jenkins, N.; Sharpe, D.; Bossanyi, E. *Wind Energy Handbook*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
37. Thönnißen, F.; Marnett, M.; Roidl, B.; Schröder, W. A numerical analysis to evaluate Betz’s Law for vertical axis wind turbines. *J. Phys. Conf. Ser.* **2016**, *753*, 022056. [[CrossRef](#)]
38. Tang, S.; Yuan, S.; Zhu, Y. Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery. *IEEE Access* **2020**, *8*, 149487–149496. [[CrossRef](#)]
39. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
40. Huang, C.-J.; Kuo, P.-H. A short-term wind speed forecasting model by using artificial neural networks with stochastic optimization for renewable energy systems. *Energies* **2018**, *11*, 2777. [[CrossRef](#)]
41. Al-Dahidi, S.; Ayadi, O.; Adee, J.; Alrbai, M.; Qawasmeh, B.R. Extreme learning machines for solar photovoltaic power predictions. *Energies* **2018**, *11*, 2725. [[CrossRef](#)]
42. Khazaei, S.; Ehsan, M.; Soleymani, S.; Mohammadnezhad-Shourkaei, H. A high-accuracy hybrid method for short-term wind power forecasting. *Energy* **2022**, *238*, 122020. [[CrossRef](#)]
43. McGrath, M. Python in Easy Steps: Covers Python 3.7. In *Easy Steps*; In Easy Steps Limited: Southham, UK, 2018.
44. Quang-Hung, N.; Doan, H.; Thoai, N. Performance evaluation of distributed training in Tensorflow 2. In Proceedings of the 2020 International Conference on Advanced Computing and Applications (ACOMP), Quy Nhon, Vietnam, 25–27 November 2020; pp. 155–159.
45. Ketkar, N. Introduction to Keras. In *Deep Learning with Python*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 97–111.
46. Hyndman, R.J.; Athanasopoulos, G.; Gally, S.; gridExtra, M.; Hyndman, R.; Hyndman, M.R. Package ‘fpp2’. 2020. Available online: <https://cran.r-project.org/web/packages/fpp2/index.html> (accessed on 9 September 2022).
47. Ghalanos, A.; Ghalanos, M.A.; Rcpp, L. Package ‘rugarch’; R Team Cooperation: Vienna, Austria, 2018. Available online: <https://cran.r-project.org/web/packages/rugarch/index.html> (accessed on 26 October 2022).
48. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [[CrossRef](#)]
49. Rahman, M.N.; Esmailpour, A.; Zhao, J. Machine learning with big data an efficient electricity generation forecasting system. *Big Data Res.* **2016**, *5*, 9–15. [[CrossRef](#)]