Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.36773

Data-Driven Analysis of Renewable Energy Generation Using Meteorological Indicators

CHEN RONG QI

Hugh McRoberts Secondary School, Richmond, Canada
aaron20221026@gmail.com

Abstract. Renewable energy generation is affected by various reasons related to the weather, and the specific reasons and analysis methods for solar, wind and hydropower systems are different. This paper studies the influence of irradiance, temperature, cloud cover, humidity, wind speed, air density, precipitation, runoff and seasonal changes in hydrology on renewable electricity generation. First, it introduces the key generation and weather indicators; then, it presents data preprocessing, descriptive statistics, correlation analysis, regression, time-series modelling and machine-learning methods. Representative studies are compared according to data sources, time resolution, areas covered, model design and evaluation indicators. Based on the above evidence, solar irradiance is a main meteorological driver of photovoltaic power generation; wind speed affects turbine power nonlinearly; and hydropower generation is also subject to hydroclimatic conditions and operational decisions. The main deficiencies are missing or misaligned observations, non-meteorological curtailment, regional heterogeneity, inconsistent temporal scales, limited model interpretability, and weak cross-site generalisation. Future work will combine multi-source weather and equipment data, physical constraints, external validation and uncertainty quantification. Data-driven analysis can be employed to investigate the reasons for changes in renewable energy generation under different weather conditions and other technology or operating factors.

Keywords: renewable energy generation, meteorological indicators, solar power, wind power, hydropower

1. Introduction

Climate change and rising energy demand around the world have increased the demand for renewable energy to reduce greenhouse gas emissions and improve energy efficiency in an environmentally friendly way. Rapid reductions in fossil-fuel consumption and greenhouse-gas emissions are necessary to limit global warming. Therefore, renewable electricity is gradually being integrated into the construction of a new-energy system around the world. Therefore, in the future, solar, wind and hydropower will be used to expand the electricity system and reduce the proportion of fossil fuels. Renewable energy generation is not stable and will change from time to time due to various reasons. Solar power is based on solar irradiance, temperature, the angle of the sun, etc. Wind power generation is based on wind speed, air density and wind direction; hydropower

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.36773

generation changes with water flow and rainfall fluctuations over time [1]. Changes in space and time of the above meteorological factors lead to fluctuations in renewable energy generation. Recently, climate data has been combined with energy data at higher frequencies and at more granular levels to explore the connection between them systematically [2]. Therefore, data-driven analysis has been used to study the fluctuations in renewable energy generation for some time now. Recently, research on solar energy has identified solar irradiance as a type of climate factor that affects photovoltaic power generation, and how surrounding or module temperatures change this efficiency has also been studied [3]. Cloud Cover and atmospheric haze will reduce or alter the amount of radiation reaching the photovoltaic modules. Research using ground-based cloud images showed that cloud-cover and aerosol information can improve short-term photovoltaic-power forecasting under hazy conditions [4]. Temperature should therefore be considered together with irradiance. Although sunny and warm conditions may provide strong irradiance, module heating can reduce conversion efficiency. For wind power, wind speed is a major explanatory variable, but other atmospheric and site conditions should also be considered. Reviews classify wind-power forecasting methods as physical, statistical, hybrid, artificial intelligence (AI) and deep learning (DL) models [5]. Other atmospheric and site conditions are also important. Based on the PLUSWIND database, the model of wind speed will change with changes in weather, location and time. Based on the above results, air-density adjustment can be relevant for high-altitude wind farms and across seasons [6]. Alterations in the quantity of precipitation, streamflow and runoff will also affect hydropower generation. Climate and land-use changes will affect the supply of water and hydropower generation in different river basins and areas to varying degrees and in different directions [7]. Different Areas, time periods, data sources, plant characteristics and evaluation indicators will lead to different research results. This paper studies the relationship between weather factors and renewable energy generation using data. This paper identifies the main climate factors that affect solar, wind and hydropower generation, summarises general data analysis methods, and compares representative results from the three areas. Similarities and Differences in energy sensitivity to changes in weather are also shown. First, this paper presents the basic ideas of renewable energy and meteorological data; then, it reviews data processing and statistical methods, examines the results from solar, wind and hydropower studies, determines the current shortcomings, and finally proposes directions for future research. The above structure will be used to present how data science helps to understand the fluctuations in renewable energy.

2. Basic concepts of renewable energy generation and meteorological data

Electricity is supplied by solar, wind and hydro power in daily life. A solar photovoltaic system converts the incoming solar radiation into electricity by means of semiconducting materials. The Power generated by a photovoltaic system depends on the intensity of sunlight and other factors that affect module and ambient temperature, etc. Wind turbines use the wind to rotate the rotor blades and generate electricity through this motion. The output is determined by the wind speed at the turbine's hub height and other features of the turbine. A turbine is employed to generate electricity by means of rotation in a water flow. Hydropower generation is determined by water supply, runoff and hydraulic head [8]. The main output indicators are power, energy production, capacity factor and conversion efficiency. Power describes the rate of electricity produced at a moment, measured usually in kilowatts or megawatts. Energy generation is the total amount of electricity produced in a certain time, usually in kilowatt-hours or megawatt-hours. Capacity factor compares practical generation with the maximum possible generation at rated capacity over the same period. The proportion of generated electricity to the initial energy is the conversion efficiency. Meteorological

DOI: 10.54254/2753-8818/2026.36773

factors are environmental variables used to explain alterations in output indicators. Solar irradiance measures incoming solar energy per unit area. The Amount of radiation received by photovoltaic panels is affected by air temperature, cloud cover and humidity, and thus their power generation varies accordingly. Wind speed is how fast the air is moving; wind direction is the direction in which the wind blows; and atmospheric pressure is related to the density of air. Precipitation is the amount of water that falls in a certain time. Runoff is water that flows over the ground or in streams. The above indicators can be observed and measured at the levels of hour, day, month and season. Hourly data capture rapid changes in weather and generation cycles; daily data reduce short-term fluctuations; monthly or seasonal data show large-scale climate patterns. Temporal aggregation may obscure a brief period of unusual or high-energy output [8]. Renewable-energy data require preparation before statistical analysis. Missing values and outliers should be handled appropriately; at the same time, irrelevant or redundant features may be removed to improve data quality and reduce computational costs. Exploratory Data Analysis can also find any odd cases, hidden patterns and correlations in the data. The above operations enhance the consistency and suitability of the dataset for subsequent analysis and modelling [9].

3. Data-driven methods for analyzing meteorological effects

Data-driven analysis generally starts with descriptive statistics to present the main features of the renewable-energy data. The mean is the average power generation over a certain time, and the variance shows the spread of the data. The minimum, maximum and range help to find the scale of fluctuation and periods of unusually high or low generation. The above measures can be used to compare different seasons, places and energy technologies, but they cannot explain why these changes in observation occur. Pearson's correlation coefficient is used to find out how closely related two sets of data are linearly, for example, solar radiation and energy generation. Spearman's rank correlation coefficient shows the extent to which two variables are monotonically related. Both coefficients are in the range [-1, 1], and a strong association does not imply causation. The Pearson correlation coefficient can be used to select weather variables before applying regression and further analysis [9]. Regression methods have constructed a model relating the weather to power generation. A single-variable linear regression model is used to express power generation as a function of a single independent variable, such as solar radiation, wind speed or water flow. At the same time, some linear regression models have included the variables of temperature, air density and humidity. The coefficients show how much the power generation changes when one variable increases by one unit and the others are held constant. Regression results are relatively simple to interpret and analyze, but the linear form is not suitable for modelling threshold effects and other non-linear relationships. Time-series analysis considers the order of the generation data in time to model trends, seasonality and irregular fluctuations. Machine Learning Models can learn more complex associations. A Decision Tree Uses Meteorological Thresholds for Power Generation Prediction. A single decision tree may be inaccurate; therefore, a random forest is employed to combine the predictions of multiple decision trees for better stability and accuracy. These methods can use historical power generation or multiple meteorological indicators as input and produce the expected power generation as output. Feature importance measures can show which factors are more effective in the prediction, but they do not offer the same coefficient-based interpretation as linear regression. Statistical models are therefore generally more interpretable, and machine learning methods can model more complex and non-linear patterns [10, 11]. Several indices of the model's performance are used simultaneously. The coefficient of determination (R^2) shows how much of the variation in the dependent variable is explained by the regression model; a larger value generally indicates a

DOI: 10.54254/2753-8818/2026.36773

better fit. Mean absolute error (MAE) is the average of the absolute errors between predictions and observations. Root mean squared error (RMSE) also measures error in the original unit but squares errors before averaging, so it penalizes large errors more strongly. Mean Absolute Percentage Error (MAPE) is the average of the absolute errors expressed as a percentage, and it can be used to compare models with different scales of actual values. Therefore, in selecting a standard for the model, all the above factors need to be considered simultaneously.

4. Recent studies of renewable energy generation using meteorological indicators

Recently, various ways have been used in research on renewable energy, such as simple physics equations, complex multivariable regression, time series and machine learning. However, new and complicated models are not always better. They are used according to the availability of data and the requirements for interpretability and accuracy. The following studies of solar, wind and hydropower show how data sources, regions, temporal resolutions, climate variables, evaluation methods and other auxiliary indicators affect the research results. Al-Dahidi and others have studied the electricity generation of a 264 kWp photovoltaic (PV) system at Applied Science University in Amman, Jordan, for solar energy. The four independent variables of the model are wind speed, relative humidity, ambient temperature and solar radiation. Five machine learning models were selected for comparison: multiple linear regression, decision tree regression, random forest regression, support vector regression and multi-layer perceptron. The models were optimised using the Chimpanzee Optimisation Algorithm. Among the above methods, the Multilayer Perceptron had the best reported results: RMSE = 0.503, MAE = 0.397 and $R^2 = 0.99$ [11]. However, as the study was carried out in a warm and dry place, the results may not be applicable to humid, cloudy, high-latitude or snowy areas. Multiple linear regression is still appropriate if the aim is to understand the direction and approximate size of the individual effects of weather, rather than to maximise estimation accuracy. Zhou and others have built a Spatial Dynamic Wind Power Forecasting dataset for wind power. This dataset is from a wind farm in Longyuan, China, and contains data for 134 wind turbines over the past two years (January 2020 - December 2021). The farm's supervisory control and data acquisition (SCADA) system samples every ten minutes. The variables are active power, wind speed, wind direction, outside temperature, inside temperature, nacelle direction and blade pitch angle. Weather information from the ERA5 reanalysis by the European Centre for Medium-Range Weather Forecasts, such as relative humidity, wind speed and wind direction, was also included [12]. These data sources are more informative because they include weather conditions, equipment status and spatial relationships, etc., and do not use only wind speed. According to the ablation study, adding weather information can reduce both MAE and RMSE for the tested forecasting model significantly. The study identified and addressed missing observations, equipment shutdowns, maintenance periods, and cases where a high-wind-speed event occurred but no power generation was recorded. Therefore, the above observations should not be considered typical weather-power relationships as non-meteorological operating factors influenced the recorded output. The dataset provides good support for machine learning comparison, but all the turbines belong to the same farm and use the same turbine type; thus, direct generalisation to other terrains, technologies and operating systems is limited. Hydropower is affected by the weather indirectly; thus, generation is influenced by precipitation, runoff, reservoir storage and operation decisions. Hao and others studied 36 hydropower stations and 13 reservoirs in the Dalälven River Basin in Sweden using monthly runoff data from 1961 to 2011. First, the years were divided into dry and wet years. Seven hydroclimatic scenarios were selected, and then stochastic runoff forecasting was added to a cascade hydropower management model. Monte Carlo simulations were performed on all scenarios for 90 and 180 days.

DOI: 10.54254/2753-8818/2026.36773

Based on the total electricity generated, production efficiency and a normalised Nash-Sutcliffe efficiency index, the results were evaluated and compared [13]. Based on the above study, the forecast of the actual wet-or-dry conditions was relatively accurate and effective. Unlike the solar and wind studies, this research had a long observation period and a coarser monthly time scale. Thus, it was more suitable for the study of seasonal and interannual changes in hydrology than short-term power fluctuations.

5. Limitations and prospects

Although data-driven methods have improved the analysis of the climate impact on renewable energy generation, some deficiencies in the current research limit their reliability and applicability. Data quality issues include missing values and measurement errors in the meteorological station, analysis dataset and power plant monitoring system. Power records may be affected by maintenance, shutdown and grid curtailment. In the SDWPF wind dataset, some periods had sufficient wind to generate electricity but no power output due to the effect of non-meteorological factors on turbine operation [12]. If the scenarios are assumed to be normal weather-power relationships without considering operational reasons, an incorrect conclusion will be drawn. Thus, the researchers need to present how to identify, handle or approximate irregular values. A second limitation is the difficulty of aligning weather and power-generation data. Weather observations are available at stations or grid cells, and generation data for individual panels, turbines, power plants, or whole areas may be available. The period of the SCADA data may be 10 minutes, hours or days. Aggregation of variables can simplify the analysis, but it may also hide short-term or extreme events, such as heavy rainfall or equipment failure [8]. Equipment Attributes and Site Conditions Can Affect the Results. Photovoltaic generation is affected by panel type, tilt, orientation, age, shading and maintenance. Turbine Design, hub height, terrain, air density, wake effect and control system all affect wind generation. Reservoir storage, hydraulic head, turbine efficiency and operator decisions all affect hydropower. As these uncontrolled factors differ in various places, a model trained in one area may not be suitable for others. Thus, cross-region validation is required. Multivariable regional models may improve off-site accuracy; however, more tests are still required. Method interpretation is also flawed. Pearson and Spearman correlations can only show a relationship, not a cause. The variables may appear to be correlated because they occur at the same time, or an unobserved factor may affect both simultaneously. Linear regression models are more interpretable, but they are sensitive to outliers and may not be suitable for non-linear relationships. Machine learning models are more accurate predictors in some cases, but they are generally less transparent. In the future, various data sources will be combined, such as weather stations, satellite observations, reanalysis products, forecasts, SCADA records and equipment metadata. Future studies should compare different regions, consider severe weather, and report prediction errors and uncertainties. Future studies should integrate physical knowledge of renewable-energy systems with machine learning to improve accuracy, transferability and interpretability. Future studies should improve data alignment, add external validation, perform uncertainty analysis and provide more intuitive explanations to make renewable-energy studies more practical for engineering operation and long-term planning.

6. Conclusion

This paper has analysed the relationship between meteorological indicators and renewable energy generation using data. Solar irradiance, temperature, cloud cover and humidity are major factors associated with photovoltaic power generation. Wind Speed, Wind Direction, Air Density and

DOI: 10.54254/2753-8818/2026.36773

Terrain Conditions affect Wind Power Generation. Changes in precipitation and runoff affect the hydrological cycle and reservoir conditions, thus changing hydropower generation. In general, the first kind of analysis is descriptive statistics; other types include correlation analysis, linear regression, time-series methods, machine learning, etc. These ways have different degrees of interpretability and prediction power, so none of them are ideal for both. Recently, nonlinear machine-learning models have been used to improve the accuracy of estimation for photovoltaic and wind power generation, time-series methods have captured temporal dependencies, and long-term hydrological analysis is more suitable for studying seasonal and interannual variations in hydropower. Other reasons for the uncertainty in the results include missing values, measurement errors, different operating conditions, regional differences and inconsistent time resolutions. Correlation does not imply causation; therefore, a model with good performance in one area may not be appropriate for others. In the future, several climate zones will be selected as the research subjects, and extreme weather events and uncertainty in interpretable and predictive models will be investigated. In short, data science offers some tools to explore the fluctuations of renewable energy generation and helps build a stable power system. A reliable evaluation should also handle missing values, curtailment and shutdown periods, and time alignment of weather and generation data transparently. Several error indicators should be reported together with site attributes and external validation to determine whether forecast improvements are genuine or specific to a particular dataset, area or operating condition.

References

[1] Liu, L., He, G., Wu, M., Liu, G., Zhang, H., Chen, Y., Shen, J., & Li, S. (2023). Climate change impacts on planned supply–demand match in global wind and solar energy systems. *Nature Energy, 8*, 870–880. https://doi.org/10.1038/s41560-023-01304-w

[2] Formayer, H., Nadeem, I., Leidinger, D., Maier, P., Schöniger, F., Suna, D., Resch, G., Totschnig, G., & Lehner, F. (2023). SECURES-Met: A European meteorological data set suitable for electricity modelling applications. *Scientific Data, 10*, Article 590. https://doi.org/10.1038/s41597-023-02494-4

[3] Keddouda, A., Ihaddadene, R., Boukhari, A., Atia, A., Arıcı, M., Lebbihiat, N., & Ihaddadene, N. (2023). Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Conversion and Management, 288*, Article 117186. https://doi.org/10.1016/j.enconman.2023.117186

[4] Lu, Z., Chen, W., Yan, Q., Li, X., & Nie, B. (2023). Photovoltaic power forecasting approach based on ground-based cloud images in hazy weather. *Sustainability, 15*(23), Article 16233. https://doi.org/10.3390/su152316233

[5] Tsai, W.-C., Hong, C.-M., Tu, C.-S., Lin, W.-M., & Chen, C.-H. (2023). A review of modern wind power generation forecasting technologies. *Sustainability, 15*(14), Article 10757. https://doi.org/10.3390/su151410757

[6] Millstein, D., Jeong, S., Ancell, A., & Wiser, R. (2023). A database of hourly wind speed and modeled generation for US wind plants based on three meteorological models. *Scientific Data, 10*, Article 883. https://doi.org/10.1038/s41597-023-02804-w

[7] Ahialey, E. K., Kabo-Bah, A. T., & Gyamfi, S. (2023). Impacts of LULC and climate changes on hydropower generation and development: A systematic review. *Heliyon, 9*(11), e21247. https://doi.org/10.1016/j.heliyon.2023.e21247

[8] Antonini, E. G. A., Di Bella, A., Savelli, I., Drouet, L., & Tavoni, M. (2024). Weather- and climate-driven power supply and demand time series for power and energy system analyses. *Scientific Data, 11*, Article 1324. https://doi.org/10.1038/s41597-024-04129-8

[9] Falope, T. O., Lao, L., & Hanak, D. (2024). A three-step weather data approach in solar energy prediction using machine learning. *Renewable Energy Focus, 50*, Article 100615. https://doi.org/10.1016/j.ref.2024.100615

[10] Dou, Y., Tan, S., & Xie, D. (2023). Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: A mini review. *Frontiers in Energy Research, 11*, Article 1218603. https://doi.org/10.3389/fenrg.2023.1218603

[11] Al-Dahidi, S., Alrbai, M., Alahmer, H., Rinchi, B., & Alahmer, A. (2024). Enhancing solar photovoltaic energy production prediction using diverse machine learning models tuned with the chimp optimization algorithm.

DOI: 10.54254/2753-8818/2026.36773

Scientific Reports, 14, Article 18583. https://doi.org/10.1038/s41598-024-69544-8

[12] Zhou, J., Lu, X., Xiao, Y., Tang, J., Su, J., Li, Y., Liu, J., Lyu, J., Ma, Y., & Dou, D. (2024). SDWPF: A dataset for spatial dynamic wind power forecasting over a large turbine array. *Scientific Data, 11*, Article 649. https://doi.org/10.1038/s41597-024-03427-5

[13] Hao, S., Wörman, A., & Brandimarte, L. (2024). The impact of hydroclimate-driven periodic runoff on hydropower production and management. *Scientific Reports, 14*, Article 25967. https://doi.org/10.1038/s41598-024-76461-3