Int. J. Renew. Energy Dev. 2025, 14 (6), 1213-1220

Contents list available at CBIORE journal website

International Journal of Renewable Energy Development

Journal homepage: https://ijred.cbiore.id

Research Article

Influence of missing wind measurements on wind turbine power production using various measure-correlate-predict methods and reanalysis datasets

Buyankhishig Amarzaya and Kyungnam Ko*

Faculty of Wind Energy Engineering, Graduate School, Jeju National University, 102 Jejudaehakro, Jeju, 63243, South Korea

Abstract. This study analyzes the impact of missing wind data points on the accuracy of Annual Energy Production (AEP) estimation in wind resource assessment (WRA). Evaluations were made under different scenarios using various Measure-Correlate-Predict (MCP) methods and reanalysis datasets. One year of wind measurements was collected from an inland met mast located in the Gashiri area of Jeju Island, South Korea. Three types of long-term reanalysis datasets- ERA-5, MERRA-2 and WRF (ERA-5)- were obtained, each exhibiting different levels of correlations with the met mast wind measurements. To simulate missing data points scenarios, a yearly percentage sampling method was applied to the one-year met mast wind data with sampling rates ranging from 10% to 90%. To ensure statistical reliability, random sampling was performed 12 times for each sampling rate. The MCP method was applied after pairing each sampled dataset with the reanalysis datasets. Long-term predictions were generated using four MCP approaches- two machine learning techniques (Random Forest and Gradient Boosting Regression) and two traditional methods (Regression and Matrix). AEP was calculted from these predictions and compared to the reference AEP derived from the complete dataset. Results show that accurate AEP estimation remained feasible even when using reanalysis datasets with low correlation to the measured data. Moreover, all four MCP methods demonstrated similar performance, with machine learning–based approaches producing results comparable to those of traditional methods. While conventional WRA practice recommends a data recovery rate above 90% for accurate AEP estimation, this study demonstrated reliable AEP estimates could be achieved with rates as low as 50%.

Keywords: Wind resource assessment, Measure-Correlate-Predict method, Annual energy production, Missing wind measurements, Reanalysis Datasets, Machine learning

Received: 9th July 2025; Revised: 8th Oct 2025 ; Accepted: 12th Oct 2025; Available online: 18th Oct 2025

1. Introduction

Renewable energy generation has been steadily replacing conventional thermal power production, leading to significant reductions in fossil fuel consumption and carbon dioxide emissions (Cai *et al.*, 2024; Li *et al.*, 2020; Loza *et al.*, 2024). Wind energy has become one of the most prominent renewable energy sources due to its substantial potential, minimal environmental impact, high economic viability, and ability to contribute to local economic development (Khosravi *et al.*, 2018). As reported by the Global Wind Energy Council, the total global installed wind power capacity exceeded 1,136 GW as of the end of 2024, with approximately 127 GW added that year (GWEC, 2025). Furthermore, the International Energy Agency (IEA) projects that annual wind power additions should quadruple by 2030 compared with 2020 levels to achieve the Net Zero Emissions goal by 2050 (IEA, 2021).

Accurate wind resource assessment (WRA) plays an essential role in planning and developing wind energy projects, as it determines the wind energy potential and expected energy yield at a specific site (Pelser *et al.*, 2024; Spiru & Simona, 2024). In other words, a reliable assessment of wind resources is crucial for estimating energy production, optimizing turbine placement and minimizing financial risk (Wang *et al.*, 2016; Zhou *et al.*, 2011). WRA generally follows a structured, multi-step process to ensure accurate characterization of wind conditions at a candidate project site (Brower, 2012). Typically, the process begins with site selection and the installation of a met mast to collect wind measurements and other meteorological parameters over a monitoring period of at least one year (Kang *et al.*, 2023; Weekes *et al.*, 2015; X. Zhang *et al.*, 2017). Following data collection, wind data validation procedures are conducted to filter low-quality data (Brower, 2012). Once data is validated, the short-term on-site measurements are statistically correlated with nearby long-term reference datasets using the Measure–Correlate–Predict (MCP) method to estimate long-term wind conditions (Ali *et al.*, 2018; Basse *et al.*, 2021; Carta *et al.*, 2013; Miguel *et al.*, 2019). Subsequently, wind flow modeling and terrain analysis are performed to evaluate spatial wind distribution and support turbine micro-siting (Song *et al.*, 2014). Finally, Annual Energy Productions (AEPs) are estimated at specific wind turbine positions, which are used to assess overall project economic feasibility (Bodini *et al.*, 2020).

Reanalysis datasets, among one of the long-term reference datasets, are gridded global atmospheric fields generated by assimilating historical weather observations into numerical weather prediction (NWP) models (Buontempo *et al.*, 2025). These datasets provide long-term meteorological data such as

* Corresponding author
Email: gnkor2@jejunu.ac.kr (K. Ko)

https://doi.org/10.61435/ijred.2025.61557
ISSN: 2252-4940/© 2025.The Author(s). Published by CBIORE

wind speed, wind direction, temperature and other parameters across multiple atmospheric height levels (Mavromatis, 2022). In wind energy applications, reanalysis datasets are a key component in the effective execution of the MCP method, as they directly influence the accuracy of long-term wind predictions (Davidson & Millstein, 2022; Gualtieri, 2022; Zekeik *et al*., 2024). In practice, the correlation between met mast data and reanalysis datasets is often regarded as a primary indicator of dataset suitability for long-term wind forecasting. Correlation coefficients below 0.6 are generally considered poor, 0.6 to 0.7 weak, 0.7 to 0.8 moderate, and above 0.8 good (M. H. Zhang, 2015). Therefore, the careful selection of reanalysis datasets is important for ensuring accurate AEP estimation over a wind farm's lifetime (Manwell *et al*., 2010).

However, missing or low-quality data points in wind measurement causes wind data loss leading to poor AEP estimation, which has been a common issue in WRA. It may occur for various reasons, including sensor malfunctions, data logger failures, tower shadow effects, power outages, communication disruptions and so on (Coville *et al*., 2011; Gottschall & Dörenkämper, 2021). Although conventional WRA typically relies on at least one year of wind measurements with over 90% wind data availability (Brower, 2012), it is typically required to estimate AEP using incomplete datasets due to significant time and financial investments.

In this context, machine learning techniques offer powerful tools for wind speed forecasting, energy production estimation and management of missing data points. Since these techniques effectively capture both linear and nonlinear relationships between variables, they are well-suited for handling wind data gaps and applying the MCP method (Salah *et al*., 2022). Numerous studies have been carried out for the application of machine learning techniques to the wind energy research field. Offshore wind farm power outputs were analyzed using several machine learning algorithms, including multiple linear regression (MLR), artificial neural networks (ANNs), decision trees (DTs) and support vector regression (SVR). The results indicated that MLR was the most effective algorithm for MCP execution, followed closely by ANN, whereas DT and SVR exhibited higher error levels (Mifsud *et al*., 2020). Also, a random forest (RF)-based model was the best choice to extrapolate near-surface wind speeds up to 200 meters, achieving higher accuracy, lower bias, and stronger correlation compared to conventional MCP-corrected ERA5 profiles (Rouholahnejad & Gottschall, 2025). Chen *et al*., (2022) developed an MCP model that combines a neural network with the frozen flow hypothesis to capture large-scale wind velocity fluctuations. The results exhibited high predictive accuracy, particularly when the spanwise offset between the target and reference points was small.

To estimate long-term wind conditions, Zhang *et al*., (2014) proposed a hybrid MCP method that utilized data from multiple reference stations. The result indicated that the optimal combination of MCP algorithms was dependent on the length of the concurrent short-term correlation period. In addition, Ayuso-Virgili *et al*., (2024) evaluated four MCP models (linear regression, variance ratio, ANN, and SVR) to enhance the estimation performance of wind and wave energy availability at a semi-exposed coastal site. The findings noted improvement in prediction accuracy for all four models when swell heading was performed as a sorting strategy during peak waves.

The Gradient Boosting Regression (GBR) model was also applied for wind power forecasting, leading to superior predictive accuracy (Singh *et al*., 2021). Furthermore, Jonietz Alvarez *et al*., (2024) introduced K-Nearest Neighbors (KNN), linear interpolation, and sector average deviation MCP methods for filling data gaps, resulting in the KNN-based MCP method performing best overall. These studies have demonstrated the capabilities of machine learning techniques in enhancing the MCP performance and supporting wind energy analysis.

The objective of this study is to clarify the effect of missing wind data points on the accuracy of AEP estimates with various correlation coefficients between site measurements and reference data as well as various MCP algorithms. Different reanalysis datasets with various correlation coefficients between the datasets and on-site measurements were utilized for the investigation. Four MCP methods were also employed: two traditional MCP methods (Regression and Matrix) and two machine learning approaches (Random Forest and Gradient Boosting Regression). The one-year measured wind data points were sampled at rates ranging from 10% to 90%, which were used for MCP execution. Finally, relative errors of the estimated AEP values were analyzed to find the influence of data loss and dataset correlation on AEP prediction accuracy. The novel contribution of this study lies in evaluating whether AEP estimation remains reliable under conditions of specific data point loss, even when correlations with reanalysis datasets are limited.

2. Site setup and methodology

2.1 Site setup

This study was carried out in the Gashiri region of Jeju Island, South Korea. Fig. 1 shows the location of Jeju Island and

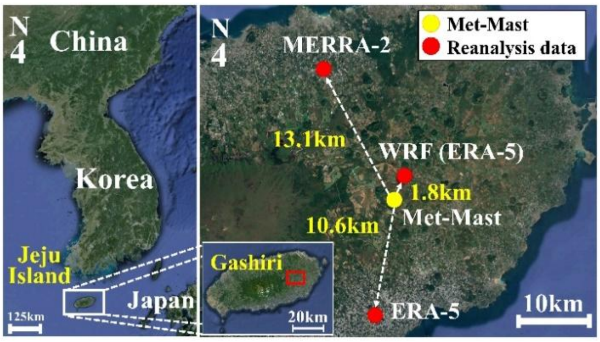

Fig. 1 Location of Jeju Island and measurement site

Table 1
Information on the met mast and the reanalysis data

Data types	Location (UTM-WGS84)	Mean wind speeds [m/s]	Data periods	Analyzed heights WS/WD [m]
Met-Mast	E: 289,641 N: 3,697,537	6.81	2012.05.01 –2013.04.30 1 year	70/65
ERA-5	E: 287,718 N: 3,687,098	5.12	1994.05.01 – 2013.04.30 19 years	100/100
MERRA-2	E: 283,461 N: 3,709,151	6.50	1993.05.01 – 2013.04.30 20 years	50/55
WRF (ERA-5)	E: 290,375 N: 3,699,108	7.75	1999.05.01 – 2013.04.30 14 years	75/75

Table 2
Specifications of the wind turbine

Item	Category	Specification
Wind turbine	Model	GE 3.2-103
	Rated power	3,200 kW
	Hub height	70 m
	Rotor diameter	103 m
	Cut-in wind speed	3 m/s
	Rated wind speed	14.8 m/s
	Cut-out wind speed	25 m/s

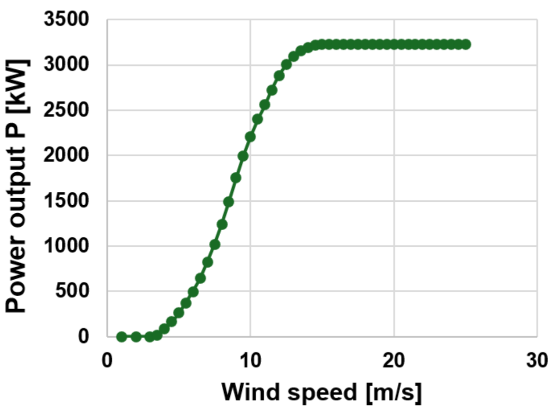

Fig. 2 Power curve of wind turbine

the 70-meter-high met mast position in Gashiri. A Thies First Class anemometer and wind vane were equipped to collect wind speeds and direction. To apply the MCP methods, long-term reanalysis datasets of ERA-5, MERRA-2 and EMD-WRF (ERA-5) were collected. The spatial distances from the met mast to the respective reanalysis dataset grid points were 10.6 km, 13.1 km, and 1.8 km, respectively. Table 1 presents the information regarding the met mast and the reanalysis data. The wind measurement campaign spanned one year (May 1, 2012-April 30, 2013), obtaining 10-minute average wind data. Furthermore, the ERA-5, MERRA-2 and WRF (ERA-5) reanalysis datasets were collected for the periods of 19, 20, and 14 years, respectively. The mean wind speeds were listed at the analyzed heights. Table 2 shows the specifications of the wind turbine used for the investigation. The AEP was calculated using a 3.2 MW GE-103 wind turbine. The power curve of the wind turbine is presented in Fig. 2. The cut-in, rated and cut-out wind speeds of the wind turbine are 3 m/s, 14.8 m/s and 25 m/s, respectively.

2.2 Methodology and workflow

Two machine learning algorithms- Random Forest (RF) and Gradient Boosting Regression (GBR)- were applied in this study. RF, introduced by Breiman (2001), constructs multiple decision trees during training and outputs their average prediction for regression problems (Breiman, 2001; Ishak, 2016). It utilizes bootstrap sampling and random feature selection to reduce overfitting and improve generalization. Since the RF algorithm is effective at capturing nonlinear relationships and handling high-dimensional datasets, it is well-suited for complex prediction tasks (Ho *et al*., 2023; Sathyaraj & Sankardoss, 2024). GBR also operates by constructing decision trees sequentially with each new tree trained to correct the residual errors of its predecessor (Boldini *et al*., 2023; Park *et al*., 2023). By minimizing a specified loss function through gradient descent, GBR enables the model to iteratively focus on the most challenging samples. The GBR algorithm is effective for modeling complex relationships and is widely acknowledged for its strong predictive performance in regression applications (Friedman, 2001).

In addition to the machine learning algorithms, two traditional MCP methods (Regression and Matrix) were also applied for comparison. These approaches have been widely used for MCP applications due to their simplicity, interpretability and practicality in long-term wind data prediction (Carta *et al*., 2013; M. H. Zhang, 2015).

AEP was calculated by the following equation:

$$\mathrm{AEP}\ [\mathrm{kWh}] = \sum_{i=1}^{8760}[P(V_i)] \tag{1}$$

where P (V_i) is the power output at the i^{th} wind speed and V_i is wind speed at i^{th} time stamp.

Then, the relative error of AEPs was calculated using the AEP reference value derived from one-year estimates based on 100% of the met-mast data points. The AEP relative error, ε_{AEP}, was expressed by the following equation:

$$\varepsilon_{AEP}\ [\%] = \frac{\mathrm{AEP_{ref}} - \mathrm{AEP_{calculated}}}{\mathrm{AEP_{ref}}} \times 100\% \tag{2}$$

where, AEP_{ref} is the reference AEP and $AEP_{calculated}$ is the AEP calculated from each MCP result.

Fig. 3 shows the overall workflow of the study. Wind data were collected from the met mast and three types of long-term hourly reanalysis datasets- ERA-5, MERRA-2 and WRF (ERA-5). A data quality check was then performed for the met mast data, resulting in a data recovery rate of 99%. The original 10-minute interval wind data from the met mast were aggregated to hourly averages to align with the temporal resolution of the reanalysis data. Subsequently, to generate missing data point scenarios, the wind data points from the met mast were sampled using a percentage data sampling method over a measurement year. Following this approach, the one-year wind data points were randomly sampled at 10% increments ranging from 10% to 90%.

To verify that every sampling dataset preserves temporal wind variation, seasonal and diurnal wind speed variations of each sampled dataset were compared with the full one-year measurement dataset. Fig. 4 shows an example of diurnal wind speed variation using the full and 50% sampled datasets. Fig. 5 illustrates an example of seasonal wind speed variation using the full and 50% sampled datasets. As shown in Figs. 4 and 5, the sampled datasets retained the patterns of the full dataset, confirming that temporal variation was preserved. The random sampling process was repeated twelve times for each percentage level to ensure consistency and reduce sampling bias. Then, all sampling datasets were paired with the three types of reanalysis data to apply the MCP method.

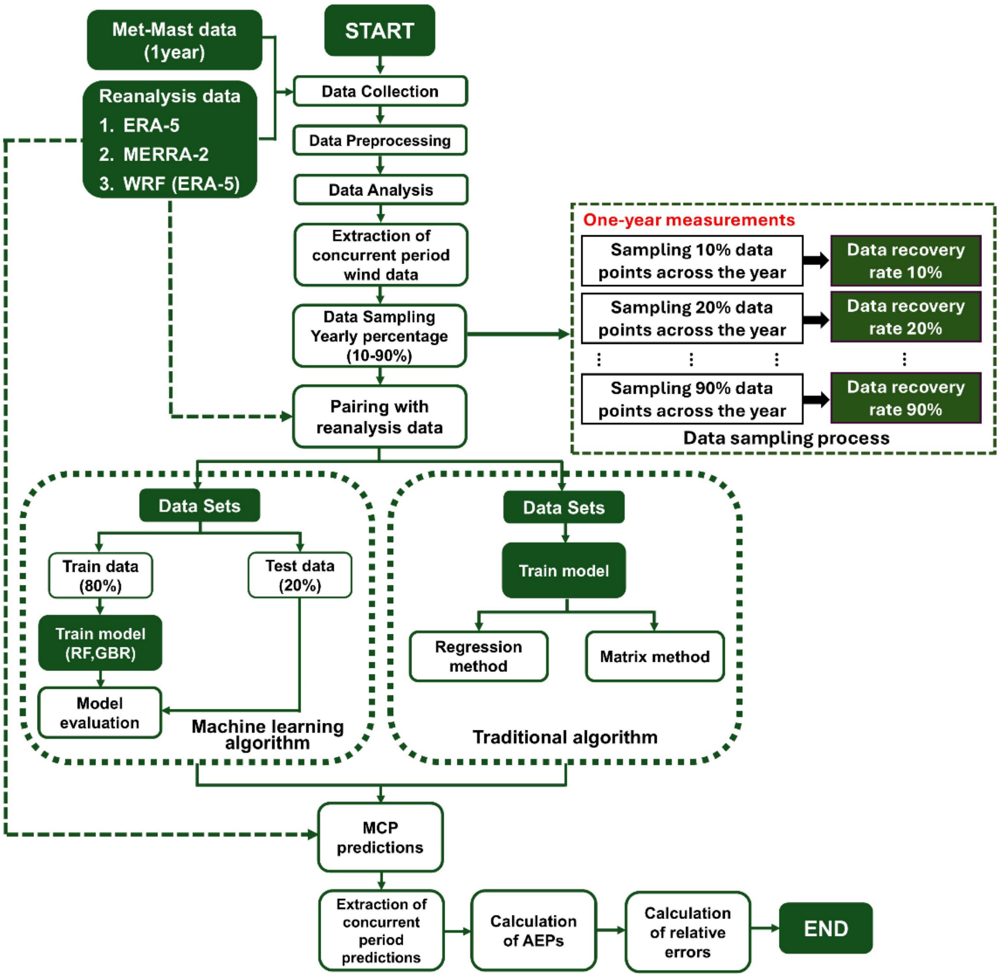

Fig. 3 Workflow for this study

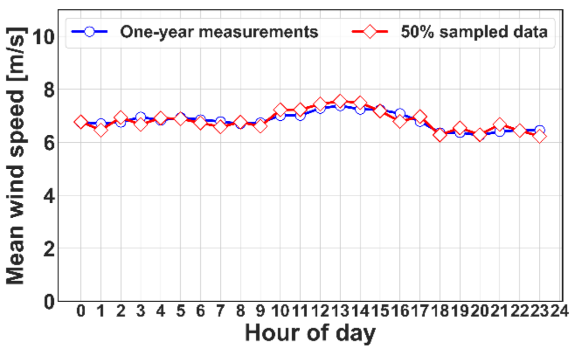

Fig. 4 Diurnal wind speed variation for the full and sampled datasets

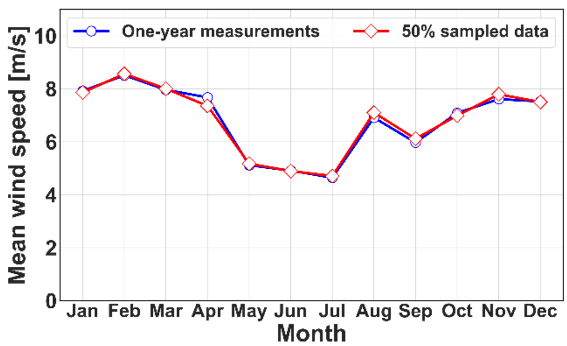

Fig. 5 Seasonal wind speed variation for the full and sampled

Each paired dataset was divided into a training set (80%) and a testing set (20%). The RF and GBR models were trained on the training dataset using optimized hyperparameters (RF: n_estimators=200, max_depth= none, random_state=42; GBR: n_estimators=200, learning_rate=0.1, max_depth=3, random_state=42), and their performance was subsequently evaluated on the testing dataset.

The resulting trained models were incorporated into the MCP framework. The traditional Regression and Matrix methods were also implemented for comparison. All the trained models and the traditional models were then applied to the three types of long-term reanalysis datasets, resulting in twelve types of MCP estimates. These were generated by combining four MCP methods with three different pairings of reanalysis and met mast datasets.

From these predicted MCP estimates, only the data points corresponding to the met mast measurement period were extracted for AEP calculations. For each one-year MCP estimates extracted, the AEP was calculated 108 times caused by repeating a random data selection process twelve times at each percentage interval from 10% to 90%. Finally, the AEP estimates were compared with the reference AEP value to derive the relative errors, from which conclusions were drawn.

3. Results and discussion

3.1 Linear regression analysis between met mast measurements and reanalysis data

Fig. 6 presents the results of the linear regression analysis between the wind speeds measured on the met mast and those obtained from each reanalysis dataset: (a) ERA-5, (b) MERRA-2 and (c) WRF (ERA-5). Reanalysis datasets with different levels of correlation to the met mast measurements were selected to find out the effect of correlation coefficients on AEP estimation. Among the three reanalysis datasets, WRF (ERA-5) exhibited the highest correlation, with a regression line slope of 0.93 and a correlation coefficient (R) of 0.84. MERRA-2 showed a moderate correlation with a slope of 0.60 and R=0.74, while ERA-5 demonstrated the lowest correlation with a slope of 0.47 and R=0.62.

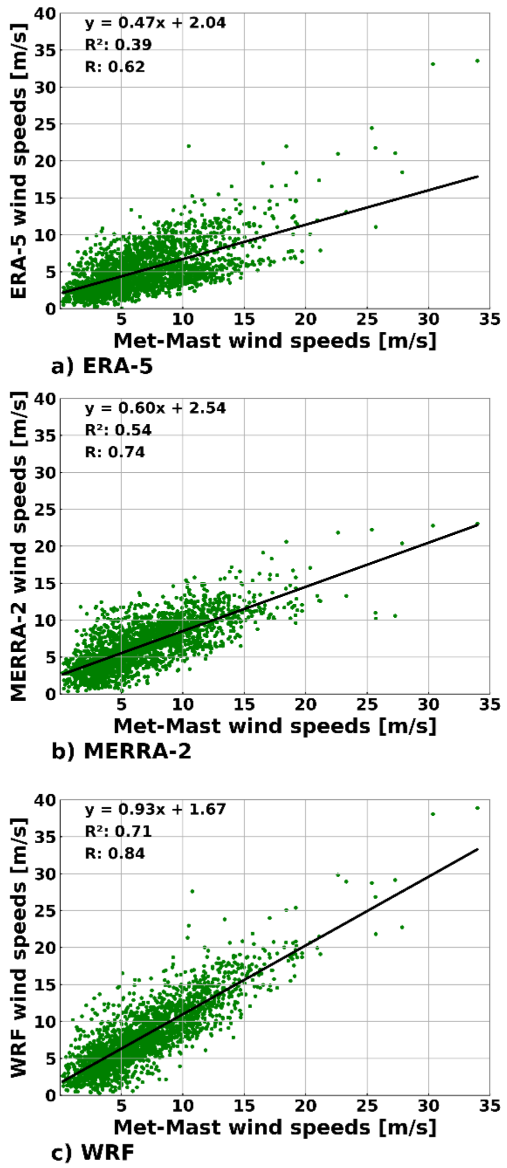

Fig. 6 Linear regression analysis between met mast wind speeds and (a) ERA-5, (b) MERRA-2 and (c) WRF (ERA-5) reanalysis datasets.

3.2 Impact of reanalysis data selection on AEP estimation

Fig. 7 shows box plots of the relative errors and standard deviations in AEP estimations for different correlation coefficients of the three reanalysis datasets. For the four MCP methods, twelve AEP relative errors were calculated by repeating the random sampling process twelve times for each sampling percentage from 10% to 90%. For the 100% case, only a single AEP relative error was calculated by each MCP method, since the complete dataset was available and no random sampling was required. As a result, each box from 10% to 90% in Fig. 7 contained a total of 48 values (12 repetitions × 4 MCP methods), while the box for 100% had 4 values (only 4 MCP

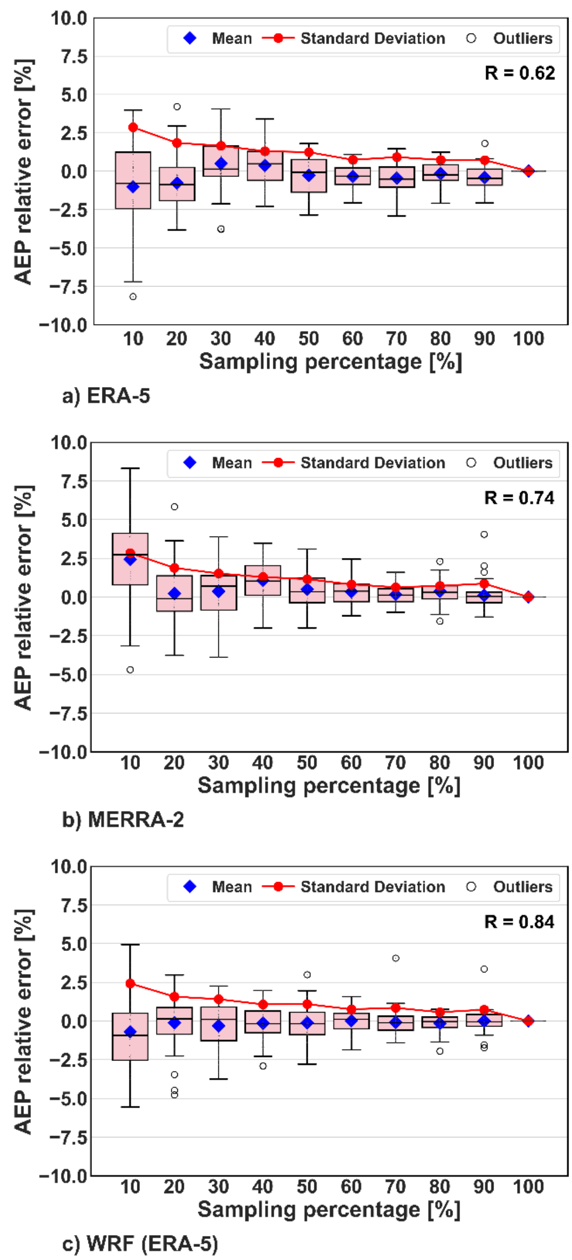

Fig. 7 Box plots of AEP relative errors and standard deviations for (a) ERA-5, (b) MERRA-2, and (c) WRF (ERA-5) reanalysis datasets across different data recovery rates.

methods). The standard deviation was derived from 48 values of each percentage.

Over the three box plots of Fig. 7, the mean values of the relative errors for the sampling percentages from 20% to 90% were close to those at the 100% (reference) case. The median values were also close to the mean values at all sampling percentages. The minimum and the maximum whisker values of each box decreased as the sampling rate increased. At the 10% sampling percentage, the minimum and the maximum whiskers were within ±8.5%, and they converged to ±2.5% at a 90% sampling percentage. The standard deviations of AEP relative errors for the three correlation coefficients were 2.85% at most at 10% sampling. Those decreased as the sampling percentage increased, finally reaching 0.86% at most at 90% sampling.

Although the reanalysis dataset with a correlation coefficient of 0.84 was considered a good choice for accurate AEP estimations, the dataset with a moderate correlation coefficient of 0.74 still produced comparable AEP results. Furthermore, even when the reanalysis dataset with the lowest correlation coefficient of 0.62 was used, it did not show a significant decline in AEP estimation performance. Though reanalysis datasets with strong correlations to met mast measurements are typically preferred for MCP execution, such high-quality datasets may not always be freely available for a given site. This finding show that even when only lower-correlation reanalysis datasets are available, they could nevertheless be used to generate long-term wind speed predictions comparable to those derived from higher-correlation datasets.

3.3 Impact of MCP methods selection on AEP estimation

Figs. 8 and 9 represent box plots of AEP relative errors and standard deviations for the four MCP methods. At each sampling percentage from 10% to 90%, each box includes a total of 36 AEP relative errors, which were from twelve random sampling repetitions with three types of reanalysis data. For the 100%, each box had 3 values from only reanalysis data.

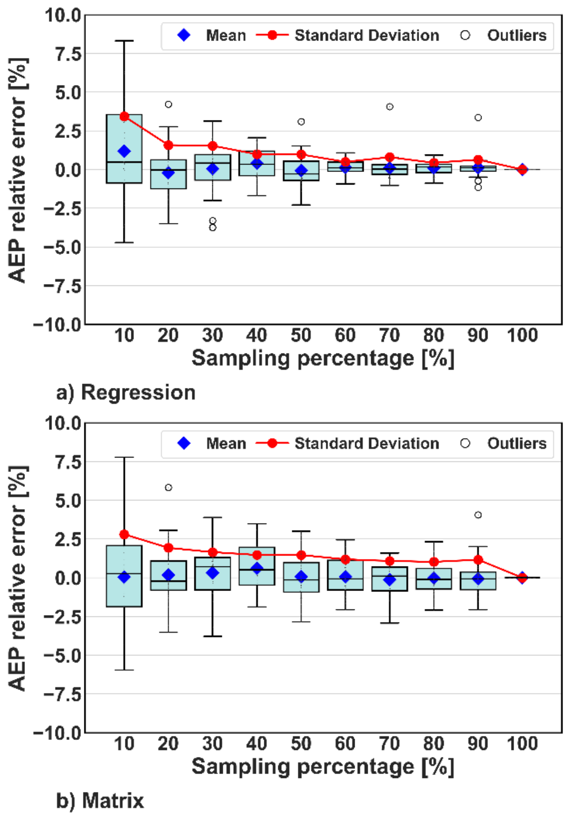

Fig. 8 Box plots of AEP relative errors and standard deviations for traditional MCP methods: (a) Regression and (b) Matrix

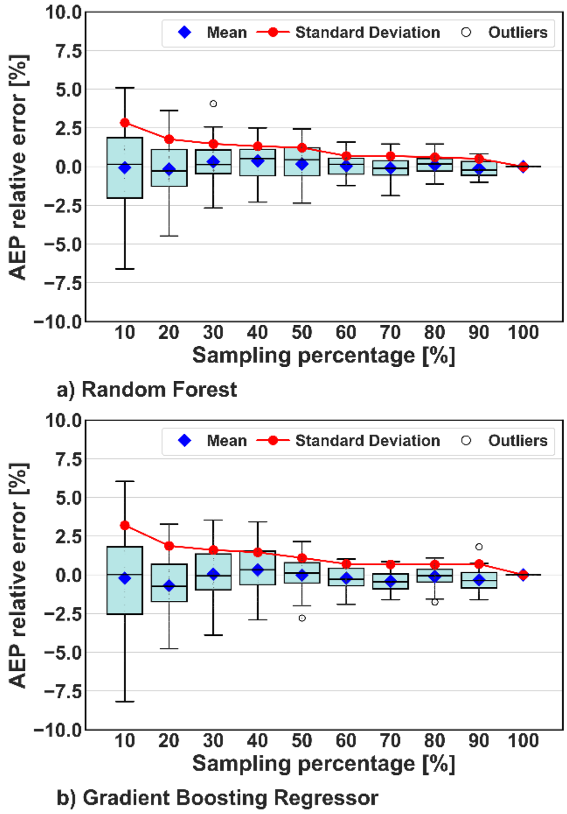

Fig. 9 Box plot of AEP relative errors and standard deviations for machine learning-based MCP methods: (a) RF and (b) GBR

Fig. 8 depicts the results for two traditional MCP methods: (a) Regression and (b) Matrix. For the two methods, the trends and the values of the mean, the median, the maximum and the minimum whiskers were similar to those of Fig. 7. At the 10% sampling level, the standard deviations of the AEP relative errors for the Regression and the Matrix were 3.4% and 2.8%, respectively. Those gradually decreased as the sampling percentage increased, converging to 0.64% and 1.16% at the 90% sampling level, respectively.

Fig. 9 displays box plots of AEP relative errors and standard deviations for two machine learning–based MCP methods: (a) RF and (b) GBR. Consistent with the traditional MCP methods, the machine learning–based methods exhibited comparable AEP estimation performances, with no significant trend differences observed in the mean and the median relative errors over all the sampling percentages. The RF exhibited an initial standard deviation of about 2.85% at the 10% sampling level, and then steadily declined toward 0.51% at the 90% level. For the GBR, the standard deviation of the AEP relative errors started at 3.2% at the 10% level, which gradually decreased as data recovering rate increased and finally reached to 0.69% at the 90% level. These findings suggest that both machine learning–based and traditional MCP methods demonstrated comparable levels of accuracy and variability in AEP estimation across all sampling percentages considered in this study.

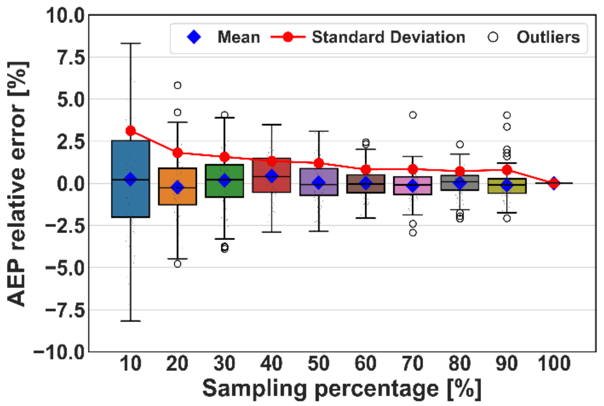

Fig. 10 Box plot of AEP relative errors and standard deviations for different sampling percentages

3.4 Impact of missing wind measurements on AEP estimation

Fig. 10 displays box plots of AEP relative errors and standard deviations across different sampling percentages, combining the results from all previous box plots. At each sampling percentage from 10% to 90%, each box and whisker was created using a total of 144 AEP relative errors, which were calculated from twelve random samplings for the combination of three reanalysis datasets and four MCP methods. The relative error and the standard deviation were all zero at the 100% case, which was a reference.

At lower sampling percentages, particularly at 10%, the mean AEP relative error was approximately +0.24%, with a standard deviation of 3.14%. The errors ranged from −8.17% to +8.31%, showing considerable variability. As the sampling percentage increased to 20% and 30%, the mean errors were closer to zero and the variability also decreased with standard deviations of 1.83% and 1.58%, respectively. At 40%, the mean error slightly rose to +0.43% with a standard deviation of 1.33%, though variability continued to decline. From 50% to 90%, the mean errors stabilized near zero with the standard deviation decreasing from 1.21% to 0.81%, respectively.

These results demonstrated that AEP estimation became progressively more stable and reliable with an increase of data recovery rate. While larger errors and significant variability are observed at lower data sampling levels, particularly at 10%, reliable AEP predictions could be obtained with sampling percentages exceeding 50%, which were within AEP relative errors of ±3.1% at most. Although a data recovery rate above 90% is typically required for reliable WRA, field technicians often experience recovery rates lower than this threshold. It was found that reliable AEP predictions could be obtained with a data recovery rate of 50% or higher, when considering the corresponding relative errors in AEP. This finding may provide practical guidance for wind data analysts working under conditions of reduced data recovery.

4. Conclusions

This study evaluated the effect of missing wind data points on the accuracy of AEP estimates by applying four MCP methods- two traditional (Regression and Matrix) and two machine learning-based (RF and GBR) models- along with three reanalysis datasets- ERA-5, MERRA-2, and WRF (ERA-5)-exhibiting different correlation levels to the met mast measurements. A yearly percentage sampling approach was applied to one-year met mast wind data to simulate data points insufficiency, and AEP was estimated for each combination of MCP methods and reanalysis datasets. The estimated AEP values were then compared with reference AEP to determine relative errors.

The results first show that while the reanalysis dataset with the highest correlation coefficient (0.84) was considered to provide accurate AEP estimates, the datasets with moderate (0.74) and low (0.62) correlations produced similar AEP estimations in this study. Second, the performance of traditional and machine learning–based MCP methods was comparable in terms of accuracy and variability in AEP estimates. Third, AEP estimation became more stable and reliable as data recovery rates increased. Although larger errors and variability were shown at 10% data points sampling, reliable AEP predictions were achieved with relative errors within ±3.1% when the data recovery rate exceeded 50%. This highlighted that although a data recovery rate above 90% is widely recommended for accurate AEP estimation in WRA practice, accurate AEP predictions remained feasible with data recovery rates of 50% or higher when accounting for additional uncertainty.

Acknowledgments

This work was supported by the research grant of Jeju National University in 2024.

Author Contributions: BA.: Conceptualization, Methodology, Coding, Software, Formal analysis, Investigation, Writing - original draft, Visualization. KK.; Writing – review & editing, Supervision, Project administration. All authors have read and agreed to the published version of the manuscript.

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

References

Ali, S., Lee, S. M., & Jang, C. M. (2018). Forecasting the long-term wind data via measure-correlate-predict (MCP) methods. *Energies*, 11 (1541); https://doi.org/10.3390/en11061541

Ayuso-Virgili, G., Christakos, K., Lande-Sudall, D., & Lümmen, N. (2024). Measure-correlate-predict methods to improve the assessment of wind and wave energy availability at a semi-exposed coastal area. *Energy*, 309 (132904); https://doi.org/10.1016/j.energy.2024.132904

Basse, A., Callies, D., Grötzner, A., & Pauscher, L. (2021). Seasonal effects in the long-Term correction of short-Term wind measurements using reanalysis data. *Wind Energy Science*, 6(6), 1473–1490; https://doi.org/10.5194/wes-6-1473-2021

Bodini, N., Optis, M., & Optis, M. (2020). Operational-based annual energy production uncertainty: Are its components actually uncorrelated? *Wind Energy Science*, 5(4), 1435–1448; https://doi.org/10.5194/wes-5-1435-2020

Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L., & Sieber, S. A. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics*, 15(73); https://doi.org/10.1186/s13321-023-00743-7

Breiman, L. (2001). Random Forests, *Machine Learning; Kluwer Academic Publishers,* 45 5-32.

Buontempo, C., Cagnazzo, C., & Dee, D. (2025). Applications of reanalyses in climate services. *Meteorological Applications*, 32(1); https://doi.org/10.1002/met.70026

Cai, Y., Loi, E. H. N., Tung, C. L., Wang, H., & Zheng, J. (2024). Clean energy development and low-carbon transition in China's power industry. *Frontiers in Environmental Science*, 12 (1525047); https://doi.org/10.3389/fenvs.2024.1525047

Carta, J. A., Velázquez, S., & Cabrera, P. (2013). A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site. *Renewable and Sustainable Energy Reviews,* 27, 362–400; https://doi.org/10.1016/j.rser.2013.07.004

Chen, D., Zhou, Z., & Yang, X. (2022). A measure-correlate-predict model based on neural networks and frozen flow hypothesis for wind resource assessment. *Physics of Fluids*, 34 (045107); https://doi.org/10.1063/5.0086354

Coville, A., Siddiqui, A., & Vogstad, K. O. (2011). The effect of missing data on wind resource estimation. *Energy*, *36* (7), 4505–4517; https://doi.org/10.1016/j.energy.2011.03.067

Davidson, M. R., & Millstein, D. (2022). Limitations of reanalysis data for wind power applications. *Wind Energy*, 25(9), 1646–1653; https://doi.org/10.1002/we.2759

Denis Mifsud, M., Sant, T., & Nicholas Farrugia, R. (2020). Analysing uncertainties in offshore wind farm power output using measure-correlate-predict methodologies. *Wind Energy Science*, 5(2), 601–621; https://doi.org/10.5194/wes-5-601-2020

Energy Agency: Net Zero by 2050. (2021). *IEA 2021*. Available at: https://www.iea.org/reports/net-zero-by-2050

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics,* 29(5), 1189-1232.

Gottschall, J., & Dörenkämper, M. (2021). Understanding and mitigating the impact of data gaps on offshore wind resource estimates. *Wind Energy Science*, 6(2), 505–520; https://doi.org/10.5194/wes-6-505-2021

Gualtieri, G. (2022). Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review. *Renewable and Sustainable Energy Reviews,* 167 (112741); https://doi.org/10.1016/j.rser.2022.112741

GWEC. (2025). *Global Wind Report*. Available at; www.gwec.net

Ho, C. Y., Cheng, K. S., & Ang, C. H. (2023). Utilizing the Random Forest Method for Short-Term Wind Speed Forecasting in the Coastal Area of Central Taiwan. *Energies*, 16 (1843); https://doi.org/10.3390/en16031374

Ishak, A. Ben. (2016). Variable selection using support vector regression and random forests: A comparative study. *Intelligent Data Analysis*, 20(1), 83–104; https://doi.org/10.3233/IDA-150795

Jonietz Alvarez, M. G., Watson, W., & Gottschall, J. (2024). Understanding the impact of data gaps on long-term offshore wind resource estimates. *Wind Energy Science*, 9(11), 2217–2233; https://doi.org/10.5194/wes-9-2217-2024

Kang, S., Hadadi, S., Yu, S. H., Lee, S. Il, Seo, D. W., Oh, J., & Lee, J. H. (2023). Wind resource assessment and potential development of wind farms along the entire coast of South Korea using public data from the Korea meteorological administration. *Journal of Cleaner Production*, 430 (139378); https://doi.org/10.1016/j.jclepro.2023.139378

Khosravi, A., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018). Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustainable Energy Technologies and Assessments*, 25, 146–160; https://doi.org/10.1016/j.seta.2018.01.001

Li, J., Li, S., & Wu, F. (2020). Research on carbon emission reduction benefit of wind power project based on life cycle assessment theory. *Renewable Energy*, 155, 456–468; https://doi.org/10.1016/j.renene.2020.03.133

Loza, B., Minchala, L. I., Ochoa-Correa, D., & Martinez, S. (2024). Grid-Friendly Integration of Wind Energy: A Review of Power Forecasting and Frequency Control Techniques. *Sustainability,* 16 (9535); https://doi.org/10.3390/su16219535

Manwell, J. F., McGowan, J. G., & Rogers, A. L. (2010). *Wind Energy Explained: Theory, Design and Application*.

Mavromatis, T. (2022). Evaluation of Reanalysis Data in Meteorological and Climatological Applications: Spatial and Temporal Considerations. *Water,* 14 (2769); https://doi.org/10.3390/w14172769

Michael Brower. (2012). *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*. ISBN: 978-1-118-02232-0. First Edition. Michael Brower et al. © 2012 John Wiley & Sons, Inc

Miguel, J. V. P., Fadigas, E. A., & Sauer, I. L. (2019). The influence of the wind measurement campaign duration on a measure-correlate-predict (MCP)-based wind resource assessment. *Energies*, 12 (3606); https://doi.org/10.3390/en12193606

Park, S., Jung, S., Lee, J., & Hur, J. (2023). A Short-Term Forecasting of Wind Power Outputs Based on Gradient Boosting Regression Tree Algorithms. *Energies*, 16 (1132); https://doi.org/10.3390/en16031132

Pelser, T., Weinand, J. M., Kuckertz, P., McKenna, R., Linssen, J., & Stolten, D. (2024). Reviewing accuracy & reproducibility of large-scale wind resource assessments. *Advances in Applied Energy,* 13 (100158); https://doi.org/10.1016/j.adapen.2023.100158

Rouholahnejad, F., & Gottschall, J. (2025). Characterization of local wind profiles: a random forest approach for enhanced wind profile extrapolation. *Wind Energy Science*, 10 (1), 143–159; https://doi.org/10.5194/wes-10-143-2025

Salah, S., Alsamamra, H. R., & Shoqeir, J. H. (2022). Exploring Wind Speed for Energy Considerations in Eastern Jerusalem-Palestine Using Machine-Learning Algorithms. *Energies*, 15 (2602); https://doi.org/10.3390/en15072602

Sathyaraj, J., & Sankardoss, V. (2024). A novel analysis of random forest regression model for wind speed forecasting. *Cogent Engineering*, 11 (1) (2434617); https://doi.org/10.1080/23311916.2024.2434617

Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies*, 14 (5196). https://doi.org/10.3390/en14165196

Song, M. X., Chen, K., He, Z. Y., & Zhang, X. (2014). Optimization of wind farm micro-siting for complex terrain using greedy algorithm. *Energy*, *67,* 454–459; https://doi.org/10.1016/j.energy.2014.01.082

Spiru, P., & Simona, P. L. (2024). Wind energy resource assessment and wind turbine selection analysis for sustainable energy production. *Scientific Reports*, 14 (10708); https://doi.org/10.1038/s41598-024-61350-6

Wang, J., Hu, J., & Ma, K. (2016). Wind speed probability distribution estimation and wind energy assessment. *Renewable and Sustainable Energy Reviews,* 60, 881–899; https://doi.org/10.1016/j.rser.2016.01.057

Weekes, S. M., Tomlin, A. S., Vosper, S. B., Skea, A. K., Gallani, M. L., & Standen, J. J. (2015). Long-term wind resource assessment for small and medium-scale turbines using operational forecast data and measure-correlate-predict. *Renewable Energy*, 81, 760–769; https://doi.org/10.1016/j.renene.2015.03.066

Zekeik, Y., OrtizBevia, M. J., Alvarez-Garcia, F. J., Haddi, A., El Mourabit, Y., & Alrubaye, A. (2024). Improved wind power assessments by bias adjusted reanalysed data with applications near Morocco's coast. *Scientific Reports*, 14 (1), 27711; https://doi.org/10.1038/s41598-024-77765-0

Zhang, J., Chowdhury, S., Messac, A., & Hodge, B. M. (2014). A hybrid measure-correlate-predict method for long-term wind condition assessment. *Energy Conversion and Management*, 87, 697–710; https://doi.org/10.1016/j.enconman.2014.07.057

Zhang, M.H. Wind Resource Assessment and Micro-Siting; Wiley: Hoboken, NJ, United States, 2015; https://doi.org/10.1002/9781118900116

Zhang, X., Liu, Y., Bai, K., Song, P., Liu, J., Cui, Y., & Zhou, S. (2017). Study on optimal site selection of a met mast in interconnected wind farm. *The Journal of Engineering*, 2017 (13), 1468–1472; https://doi.org/10.1049/joe.2017.0575

Zhou, Y., Wu, W. X., & Liu, G. X. (2011). Assessment of onshore wind energy resource and wind-generated electricity potential in Jiangsu, China. *Energy Procedia*, 5, 418–422; https://doi.org/10.1016/j.egypro.2011.03.072