energies

MDPI

Article

An Advanced CBR-Based Photovoltaic Power Forecasting Framework Incorporating Kriging Meteorological Interpolation

Minjoon Kim [1] and Woncheol Lee [2,*]

[1] Department of Energy Policy and Technology Convergence, Soongsil University, Seoul 06978, Republic of Korea; minjoon99@soongsil.ac.kr
[2] School of Electronic Engineering, Soongsil University, Seoul 06978, Republic of Korea
* Correspondence: wlee@ssu.ac.kr; Tel.: +82-2-816-6606

Abstract

Accurate meteorological data at photovoltaic (PV) sites are essential for reliable PV forecasting, but direct observations are often unavailable because of the spatial mismatch between meteorological stations and PV sites. This study proposes a Case-Based Reasoning and Ordinary Kriging (CBR-OK) framework for site-specific meteorological interpolation and subsequent PV forecasting. Historical variogram characteristics and model-selection results were organized into a case database, and ANN-derived feature importance was used to weight features for KNN-based case retrieval. An SVM trained on KNN-derived labels was then used to classify the variogram model for new cases, after which the corresponding historical variogram parameters were retrieved for OK interpolation. CBR-OK achieved interpolation accuracy comparable to Direct-OK, although end-to-end evaluation showed additional computational overhead from the CBR inference process. Using the interpolated meteorological inputs, LSTM, Bi-LSTM, and GRU models were evaluated for 1-, 12-, and 24-h-ahead PV forecasting, with the most consistent benefits observed at the 1-h horizon. As an example of absolute forecasting performance, the CBR-OK-based LSTM achieved mean absolute errors (MAEs) of 25.3, 45.7, and 47.3 kW at the 1-, 12-, and 24-h horizons, respectively, across the six PV sites. These results demonstrate the applicability of CBR-OK as an adaptive variogram-configuration and meteorological-interpolation framework for short-term PV forecasting.

Keywords: case-based reasoning; variogram selection; photovoltaic power forecasting

1. Introduction

Recent advances in artificial intelligence (AI) and the widespread adoption of generative AI services have increased computational demand in data centers, raising concerns regarding the associated growth in electricity consumption [1]. In response to increasing electricity demand and the need to reduce greenhouse gas emissions, the utilization of renewable energy has expanded, with photovoltaic (PV) power becoming an important renewable energy source [2].

However, PV power generation is strongly affected by meteorological conditions, including solar radiation, air temperature, wind speed, and humidity, and exhibits substantial variability and uncertainty, particularly because of the spatiotemporal variability of solar radiation [2,3]. Accurate PV power forecasting is therefore important for reliable grid operation and efficient power supply management. Forecasting performance depends

Check for updates

Academic Editor: Santiago Silvestre

Received: 31 August 2026
Revised: 19 October 2026
Accepted: 23 October 2026
Published: 28 October 2026

Copyright: © 2026 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.

not only on the forecasting model but also on the quality of meteorological inputs, forecast horizon, and regional characteristics [2,3].

PV power forecasting requires meteorological inputs that adequately represent weather conditions at PV plant locations. However, meteorological stations are not available at all PV plants, and direct observations at target locations may therefore be unavailable. Although observations from nearby stations can be used, spatial separation between the station and PV plant may result in differences between the observed and actual meteorological conditions at the target location [4]. This issue is particularly relevant for variables with substantial spatial variability, such as solar radiation [4,5]. Spatial interpolation can therefore be used to estimate meteorological conditions at PV plant locations from surrounding observations.

Previous studies have demonstrated the usefulness of spatial information for PV-related interpolation and forecasting. Jamaly and Kleissl [5] applied spatial and spatiotemporal Kriging to irradiance interpolation and forecasting at unobserved locations. Agoua et al. [6] evaluated multiple spatiotemporal information sources for short-term PV power forecasting. Nam and Hur [7] combined Kriging-derived spatial information with a Naïve Bayes classifier for probabilistic PV power forecasting, while Lee et al. [4] applied spatially interpolated meteorological data to PV power prediction. These studies demonstrate the value of spatially distributed information for estimating conditions at unobserved locations and supporting PV power forecasting. However, the systematic reuse of variogram characteristics and model-evaluation results derived from historical meteorological data for selecting variogram models for newly acquired data has received comparatively limited attention.

Inverse Distance Weighting (IDW) and Kriging are widely used spatial interpolation methods. IDW assigns weights according to the distance between observation and prediction locations but does not explicitly model spatial autocorrelation [8]. In contrast, Kriging represents spatial autocorrelation using an experimental semivariogram and incorporates this structure into interpolation [5,9]. Its performance, however, can depend on the selected semivariogram model and parameterization [9]. Jian et al. [10] presented a weighted least-squares approach for fitting theoretical semivariogram models and estimating their parameters, while Ly et al. [11] demonstrated that interpolation performance can vary among variogram models.

For long-term multivariate meteorological datasets, experimental variogram estimation, candidate-model fitting, and model evaluation are repeatedly performed across variables and time points. When new meteorological data become available, direct variogram fitting provides a conventional means of determining the variogram configuration for spatial interpolation. However, historical datasets may already contain variogram characteristics, model parameters, and model-evaluation results obtained under similar spatial conditions. This motivates the reuse of historical variogram information as case knowledge to support variogram model determination for newly acquired meteorological data.

Case-Based Reasoning (CBR) retrieves previous cases similar to a new problem and reuses their solution information [12]. Case retrieval depends on case representation, similarity measures, and feature weighting [13–15]. Park and Han [13] investigated feature-weighted KNN-based CBR retrieval, Fei and Feng [14] incorporated feature importance and interactions into case retrieval, and Lin and Chen [15] considered parameter tuning, feature selection, and feature-weight assignment. Previous research has also demonstrated that neural-network-derived feature weights can guide CBR retrieval [16]. These characteristics make CBR suitable for reusing historical variogram information through similarity-based retrieval.

Accordingly, this study proposes CBR-OK, which combines CBR-based variogram model selection with Ordinary Kriging (OK) for meteorological spatial interpolation. Historical variogram parameters and model-evaluation results are organized as cases, and feature importance is incorporated into similarity-based retrieval to support variogram

model determination for newly acquired meteorological data. The interpolation performance of CBR-OK is evaluated against Nearest Neighbor (NN), Inverse Distance Weighting (IDW), Static Ordinary Kriging (Static-OK), and direct-fit Ordinary Kriging (Direct-OK) for air temperature, relative humidity, wind components, solar radiation, and sunshine hours. The resulting meteorological datasets are subsequently used with LSTM, Bi-LSTM, and GRU models to evaluate PV power forecasting performance at six PV plants over 1-, 12-, and 24-h forecast horizons.

The contribution of this study lies in the development and evaluation of an integrated CBR-OK framework that combines ANN-derived feature weighting, similarity-based case retrieval, SVM-based variogram model classification, historical variogram-parameter reuse, and Ordinary Kriging for adaptive meteorological spatial interpolation and subsequent PV power forecasting.

2. Materials and Methods

2.1. Overall Framework of the Proposed CBR-OK Method

The proposed CBR-OK method combines case-based reasoning (CBR)-based variogram model selection with Ordinary Kriging (OK) to estimate meteorological variables at PV plant locations. First, variogram modeling and model evaluation are performed using historical meteorological data, and a case database is constructed from the resulting variogram characteristics, parameters, and model-selection outcomes (Section 2.4). ANN-based permutation feature importance is then used to derive feature-specific weights for case retrieval (Section 2.5), and the similarity measure and number of nearest neighbors, *K*, are determined through weighted K-nearest-neighbor (KNN) retrieval (Section 2.6). The KNN-derived labels obtained under the selected retrieval configuration are subsequently used as target classes to train a Support Vector Machine (SVM) classifier (Section 2.7). Figure 1 summarizes the overall framework of the proposed CBR-OK method.

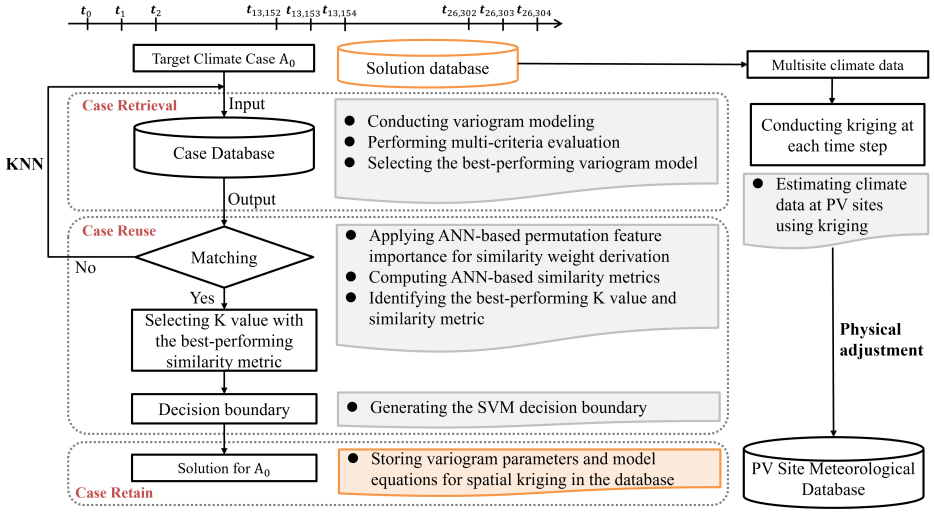


Figure 1. Framework of the proposed CBR-OK method for variogram model selection and meteorological spatial interpolation.

To further illustrate the inference procedure for a newly acquired case, Figure 2 presents a representative example of the complete retrieval and kriging process. The new meteorological case is first classified into a variogram-model class by the trained SVM, after which historical cases with the same model label are considered as retrieval candidates. The most similar historical case is then identified using the selected weighted similarity measure, and its variogram model and associated parameters are reused for Ordinary Kriging interpolation at the target PV site (Section 3.1). In the representative example

shown in Figure 2, the retrieved historical case is associated with an Exponential variogram model, and the corresponding nugget, partial sill, and range parameters are subsequently applied to generate the interpolated meteorological data.

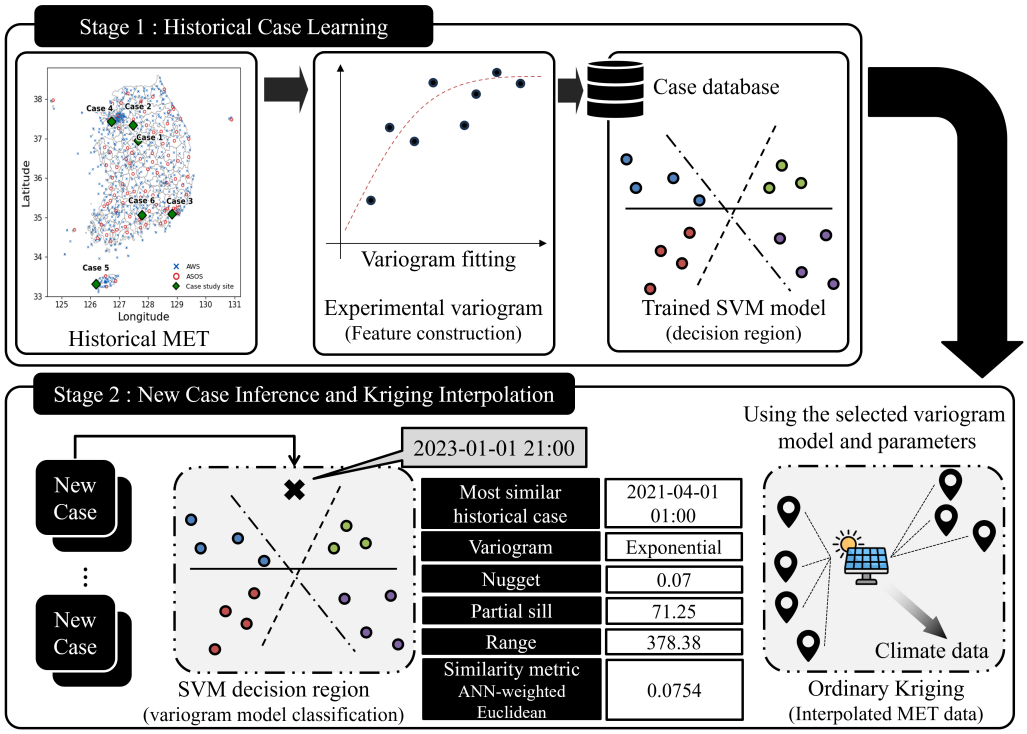


Figure 2. Representative example of the CBR-OK retrieval and interpolation process. The new case is classified using the trained SVM, and the most similar historical case is identified using the selected ANN-weighted similarity measure. The retrieved variogram model and parameters are then used for Ordinary Kriging interpolation at the target PV site.

2.2. Overall Research Framework for Meteorological Data Interpolation and PV Power Forecasting

This study establishes an integrated framework that combines CBR-OK-based meteorological data interpolation with PV power forecasting to address the spatial limitations of meteorological observations and evaluate the effects of site-specific meteorological inputs on forecasting performance. Figure 3 illustrates the overall research procedure, from meteorological data acquisition and spatial-dependence analysis to meteorological interpolation at PV plant locations and multi-horizon PV power forecasting.

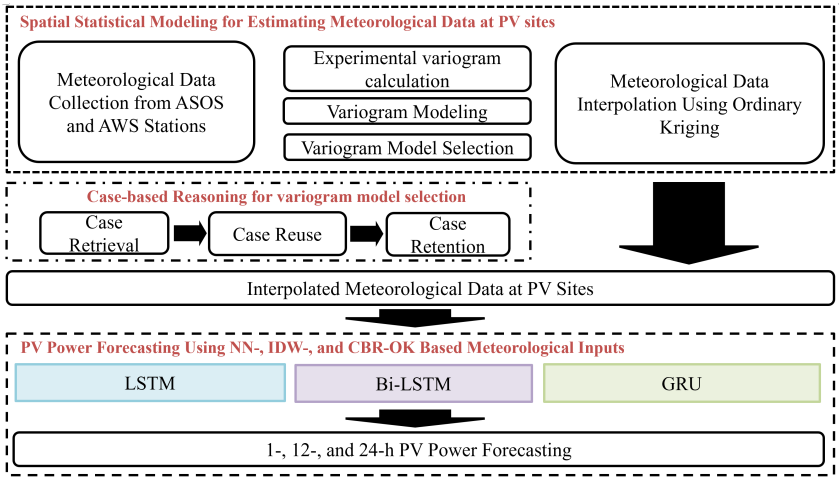


Figure 3. Overall framework of the proposed CBR-OK-based meteorological data interpolation and PV power forecasting method.

In the first stage, meteorological observations collected from the Automated Synoptic Observing System (ASOS) and Automatic Weather Station (AWS) networks are preprocessed, and the spatial dependence of each meteorological variable is characterized through variogram analysis. The CBR-OK procedure described in Section 2.1 is then used to determine the variogram model and associated parameters for newly acquired meteorological data. Detailed procedures for data preprocessing, variogram modeling, and CBR-based model selection are presented in Sections 2.3–2.7.

In the second stage, the selected variogram model and associated parameters are applied to Ordinary Kriging (OK) to interpolate meteorological variables at each PV plant location. The target variables include air temperature, relative humidity, wind speed, wind direction, solar radiation, and sunshine hours. Wind speed and wind direction are transformed into orthogonal wind components prior to interpolation and subsequently reconstructed at each PV plant location. The resulting site-specific meteorological time series are then used as inputs for PV power forecasting.

In the final stage, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU) models are used to evaluate PV power forecasting performance at 1-, 12-, and 24-h forecast horizons. To examine the downstream effects of meteorological interpolation, NN-, IDW-, and CBR-OK-based meteorological inputs are evaluated using identical forecasting architectures and training configurations within each case. This design enables the effects of the interpolation method on PV power forecasting performance to be assessed under consistent forecasting conditions. The forecasting procedure and evaluation are described in Section 4.

2.3. Meteorological Data Preprocessing

Air temperature, relative humidity, wind speed, wind direction, solar radiation, and sunshine hours were considered as meteorological variables in this study. Each variable was preprocessed according to its physical characteristics and data distribution, as summarized in Table 1. Air temperature and relative humidity were used directly without transformation. For wind data, orthogonal north–south and west–east components were derived from wind speed and direction to account for the circular nature of wind direction. Solar radiation and sunshine hours, which contain frequent structurally occurring zero values, were subjected to zero-case screening, and positive observations were subsequently used for variogram analysis.

Table 1. Preprocessing of meteorological variables.

Category	Variable	Preprocessing
Direct-use variables	Temperature Relative humidity	Original observations
Wind-derived variables	Wind_{NS} Wind_{WE}	Derived from wind speed and direction
Variables with frequent zeros	Solar radiation Sunshine hours	Zero-case screening for variogram analysis

2.3.1. Transformation of Wind Direction and Speed into Orthogonal Components

Wind direction is a circular variable for which 0° and 360° represent the same direction [17]. Therefore, treating wind direction directly as a conventional continuous variable may not adequately preserve directional continuity at the angular boundary. To account for this circularity while incorporating wind speed, the wind vector was decomposed into orthogonal north–south (NS) and west–east (WE) components [18]. With wind direction defined by an angle θ measured with respect to north, the components were calculated as follows:

$$\text{Wind}_{NS} = v\cos\left(\frac{\pi\theta}{180}\right), \tag{1}$$

$$\text{Wind}_{WE} = v\sin\left(\frac{\pi\theta}{180}\right), \tag{2}$$

where v denotes wind speed and θ denotes wind direction in degrees. The resulting Wind_{NS} and Wind_{WE} components were treated as separate meteorological variables for variogram modeling and spatial interpolation.

2.3.2. Zero-Case Processing for Solar Radiation and Sunshine Hours

Solar radiation and sunshine hours exhibit pronounced diurnal variability, with zero values generally associated with nighttime conditions. To avoid constructing spatial models from zero-dominated fields or an insufficient number of positive observations, a screening procedure was applied before empirical variogram construction. At each timestamp t, the proportion of zero-valued station observations was defined as

$$R_{0,t} = \frac{N_{0,t}}{N_t}, \tag{3}$$

where $N_{0,t}$ denotes the number of stations reporting zero-valued observations at time t and N_t denotes the number of stations with valid observations. A timestamp was excluded from variogram fitting when $R_{0,t} \geq 0.8$ or fewer than three positive observations were available. The zero-ratio threshold was intended to exclude spatial fields dominated by zero values, while the minimum-positive-observation criterion was used to screen timestamps with insufficient positive observations. Timestamps without any valid station observations were treated separately and were not included in the screening denominator.

To evaluate the sensitivity of the minimum-positive-observation criterion, thresholds ranging from two to five observations were examined. Table 2 presents the corresponding screening rates and their changes relative to the baseline requirement of three positive observations.

Table 2. Sensitivity of zero-case screening to the minimum number of positive observations.

Variable	Min. Positive Obs.	Screening Rate (%)	Change from Baseline (pp)
SH	2	2.148	−1.550
	3 (baseline)	**3.698**	**0.000**
	4	4.894	+1.196
	5	6.156	+2.458
SR	2	0.643	−0.249
	3 (baseline)	**0.892**	**0.000**
	4	1.121	+0.229
	5	1.245	+0.353

Note: SH and SR denote sunshine hours and solar radiation, respectively. The sensitivity analysis covers the complete 2025 hourly observation period. The screening rate was calculated as the number of screened timestamps divided by the total number of valid timestamps for which at least one non-missing station observation was available. At each timestamp, the zero ratio was defined as the number of zero-valued station observations divided by the total number of valid station observations. A timestamp was screened when the zero ratio was ≥ 0.80 or when the number of positive observations was below the specified minimum. The baseline required at least three positive observations, and changes are reported in percentage points (pp) relative to this baseline.

Under the baseline criterion, the screening rates were 3.698% for sunshine hours and 0.892% for solar radiation. Reducing the minimum requirement to two positive observations decreased the screening rates by 1.550 and 0.249 percentage points, respectively.

Conversely, increasing the requirement to four or five positive observations progressively increased the screening rates. The three-observation threshold was retained as the baseline screening criterion to limit the exclusion of timestamps with sparse positive observations.

The screening rates were calculated relative to timestamps containing at least one valid station observation rather than the complete annual time series. In 2025, 3942 solar radiation timestamps and 4244 sunshine hour timestamps contained no valid station observations, leaving 4818 and 4516 timestamps eligible for screening, respectively. Of these, 43 solar radiation timestamps and 167 sunshine hour timestamps were excluded under the baseline criterion. These exclusions correspond to the screening rates reported in Table 2.

The original station observations examined during the 2025 verification contained no explicitly recorded zero-valued measurements for either variable. Accordingly, the screening exclusions identified in this verification were attributable to insufficient positive observations rather than the zero-ratio threshold. The reported screening rates therefore quantify exclusions among timestamps with available station observations and should not be interpreted as the proportion of nighttime timestamps or timestamps without valid observations.

The absence of explicit zero values in the examined records does not, by itself, establish whether earlier data preprocessing affected the representation of nighttime observations. Table 3 summarises the number of timestamps at each screening and subsequent variogram-processing stage for solar radiation and sunshine hours in 2025.

Table 3. Screening and subsequent variogram-processing outcomes for solar radiation and sunshine hours in 2025.

Processing Stage	Solar Radiation	Sunshine Hours
Total timestamps	8760	8760
No valid observations	3942	4244
Timestamps eligible for screening	4818	4516
Excluded by screening	43	167
Screening rate (%)	0.892	3.698
Passed screening	4775	4349
Failed subsequent variogram processing	22	160
Valid variogram parameter records	4753	4189

Note: Screening rates were calculated relative to timestamps with at least one valid station observation. Subsequent variogram-processing failures occurred after screening and were not included in the screening rejection rates.

Passing the initial screening criterion did not necessarily result in successful variogram fitting and model selection. Among the timestamps that passed screening, 22 solar radiation timestamps and 160 sunshine hour timestamps did not yield valid variogram parameter records. For solar radiation, 18 timestamps had fewer than five valid observations during subsequent fitting, while four did not yield a valid model ranking because of insufficient spatial variability. The corresponding counts for sunshine hours were 111 and 49 timestamps, respectively. Consequently, 4753 solar radiation timestamps and 4189 sunshine hour timestamps yielded valid variogram parameter records in 2025. These subsequent processing failures were distinguished from the initial screening exclusions and were not included in the screening rates.

When kriging outputs were unavailable, the corresponding meteorological inputs in the final PV forecasting datasets were assigned zero values rather than being estimated through CBR-OK. This fallback was applied to timestamps without valid station observations, timestamps excluded by screening, and timestamps for which subsequent variogram processing was unsuccessful. In 2025, zero-value completion was applied to 4007 solar

radiation timestamps and 4571 sunshine hour timestamps at each of the six PV sites. Accordingly, the complete annual sequence of 8760 hourly timestamps was preserved, with no remaining missing values for either variable. This procedure maintained the temporal continuity of the forecasting inputs but did not reconstruct unobserved solar conditions at the affected timestamps.

The sensitivity of variogram fitting to the number of lag bins was additionally examined using configurations of 10, 15, and 20 lag bins. Table 4 reports the fitting performance and model-selection agreement relative to the 15-lag baseline.

Table 4. Sensitivity of variogram fitting to the number of lag bins for sunshine hours and solar radiation.

Variable	Lag Bins	Successful Fits	Mean Effective Lags	Mean RMSE	Mean MAE	Median RMSE	Agreement with 15 Lags (%)
SH	10	580	9.90	0.0712	0.0626	0.0711	98.62
	15	**580**	**14.71**	**0.0726**	**0.0627**	**0.0717**	**100.00**
	20	580	19.40	0.0732	0.0619	0.0741	98.28
SR	10	595	9.98	0.1642	0.1327	0.0865	88.74
	15	**595**	**14.94**	**0.1378**	**0.1077**	**0.0846**	**100.00**
	20	595	19.87	0.1327	0.0960	0.0874	88.40

Note: The 15-lag configuration was used as the baseline. Agreement indicates the percentage of cases for which the selected variogram model was identical to that obtained using 15 lag bins.

For sunshine hours, model-selection agreement with the 15-lag baseline was 98.62% and 98.28% for the 10- and 20-lag configurations, respectively, with only minor differences in fitting errors. Solar radiation exhibited greater sensitivity to the number of lag bins, with corresponding agreement rates of 88.74% and 88.40%. Nevertheless, the fitting errors remained of comparable magnitude across the evaluated configurations. Overall, these results indicate relatively stable variogram-fitting performance under the examined lag-bin configurations, although solar radiation was more sensitive to lag-bin selection than sunshine hours. The 15-lag configuration was therefore retained for the subsequent analysis.

2.3.3. Coordinate Transformation and Spatial Distance Calculation

A common projected coordinate system was used to calculate spatial distances for all meteorological variables. The geographic coordinates of the observation stations were transformed from WGS 84 (EPSG:4326) to KGD2002/Unified CS (EPSG:5179), and Euclidean distances between stations were calculated using the projected coordinates in meters. Variogram modeling was subsequently performed under the assumption of spatial isotropy, with semivariance modeled solely as a function of separation distance. Directional anisotropy was not explicitly modeled in the present study.

2.4. Variogram Modeling and Case Database Construction

2.4.1. Empirical Semivariogram Calculation

Empirical semivariograms were calculated at each time step using the preprocessed meteorological observations and were subsequently used for variogram model fitting and CBR case database construction. The empirical semivariance was calculated as

$$\hat{\gamma}(h) = \frac{1}{2|P(h)|} \sum_{(i,j)\in P(h)} \left[Z(s_i) - Z(s_j)\right]^2, \quad (4)$$

where $\hat{\gamma}(h)$ denotes the empirical semivariance at separation distance h, $P(h)$ denotes the set of station pairs assigned to the corresponding lag bin, and $|P(h)|$ denotes the number of station pairs in that bin. $Z(s_i)$ and $Z(s_j)$ denote the observations at locations s_i and s_j, respectively.

The maximum lag distance was set to 50% of the maximum inter-station distance. For air temperature, relative humidity, and the orthogonal wind components Wind_{NS} and Wind_{WE}, this distance range was divided into 20 lag bins. For solar radiation and sunshine hours, 15 lag bins were used together with the zero-case screening criteria described in Section 2.3.2, which also presents the corresponding lag-bin sensitivity analysis. For each lag bin, the empirical semivariogram was represented by the mean separation distance and the mean empirical semivariance of the station pairs assigned to that bin.

For air temperature and relative humidity, a lag bin was retained only when at least 30 station pairs were available, and variogram fitting was performed only when at least eight valid lag bins and 30 valid stations were available. For Wind_{NS} and Wind_{WE}, a lag bin was retained when it contained at least one station pair, and variogram fitting required at least three valid lag bins and five valid stations. For solar radiation and sunshine hours, the 15-bin configuration was applied after the zero-case screening procedure. The variable-specific fitting criteria and implementation settings are summarized in Appendix A.

2.4.2. Theoretical Variogram Model Fitting

Four theoretical variogram models—spherical, exponential, rational quadratic, and pentaspherical—were considered as candidate models for representing the spatial dependence characterized by the empirical semivariogram. Their functional forms and parameter definitions are presented in Table 5. For each candidate model, the nugget c_0, partial sill c, and distance parameter a were estimated, and the total sill $c_0 + c$ was stored in the case database.

Table 5. Theoretical variogram models used in this study.

Model	Variogram Function
Spherical	$\gamma(h) = \begin{cases} c_0 + c\left[\frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], & h \le a, \\ c_0 + c, & h > a \end{cases}$
Exponential	$\gamma(h) = c_0 + c\left[1 - \exp\left(-\frac{3h}{a}\right)\right]$
Rational quadratic	$\gamma(h) = c_0 + c\left[\frac{19(h/a)^2}{1 + 19(h/a)^2}\right]$
Pentaspherical	$\gamma(h) = \begin{cases} c_0 + c\left[\frac{15}{8}\left(\frac{h}{a}\right) - \frac{5}{4}\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5\right], & h \le a, \\ c_0 + c, & h > a \end{cases}$

Note: c_0, c, and a denote the nugget, partial sill, and practical-range parameter, respectively; $c_0 + c$ represents the total sill. For the exponential model, a corresponds to approximately 95% of the partial sill.

The parameters of each candidate model were estimated using weighted least squares (WLS) [19]. Specifically, a distance-based weight of $1/(h_k + \epsilon)$ was applied to assign greater weight to residuals at shorter separation distances. The parameter vector θ was estimated by minimizing

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{L} \frac{[\hat{\gamma}(h_k) - \gamma(h_k;\theta)]^2}{h_k + \epsilon}, \tag{5}$$

where L denotes the number of valid lag bins, h_k is the mean separation distance of the kth lag bin, θ is the parameter vector of the corresponding candidate variogram model, and ϵ is a small positive constant introduced to prevent division by zero. Non-negativity constraints were imposed on the nugget and partial sill, whereas the distance parameter was constrained to be positive. The L-BFGS-B algorithm was used to minimize the objective function subject to these parameter bounds [20]. Additional details on the variogram-fitting procedure and implementation settings are provided in Appendix A.

2.4.3. Variogram Model Evaluation and Case Database Construction

Each fitted candidate model was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) between the empirical semivariances and the corresponding fitted variogram values. Candidate models were ranked in ascending order of RMSE and MAE and in descending order of R^2. For each time step, the model with the minimum sum of the three metric ranks was selected:

$$m^* = \arg\min_{m \in M} [r_{\mathrm{RMSE}}(m) + r_{\mathrm{MAE}}(m) + r_{R^2}(m)], \quad (6)$$

where M denotes the set of four candidate variogram models and $r_{\mathrm{RMSE}}(m)$, $r_{\mathrm{MAE}}(m)$, and $r_{R^2}(m)$ denote the corresponding ranks of candidate model m. Ties within an individual metric were assigned average ranks. When multiple models produced the same final rank-sum score, the first model in the predefined candidate-model order was selected to ensure deterministic model labeling.

The three criteria in Equation (6) were assigned equal weight in the rank-sum criterion so that no individual variogram-fitting metric was given an a priori preference. Because RMSE, MAE, and R^2 evaluate agreement with the empirical semivariogram rather than spatial predictive performance directly, the robustness of the rank-based selection procedure was additionally assessed using spatial cross-validation. At selected time points distributed across the analyzed data, the model chosen by the rank-based criterion was compared with the model yielding the lowest spatial cross-validation error. Sensitivity to the number of lag bins and bootstrap uncertainty in the fitted variogram parameters were also examined. The corresponding analyses are provided in Appendix B.

The resulting case database therefore contained the variogram parameters and fitting-performance metrics of all four candidate models, together with the selected-model label for each time step, as summarized in Table 6. Zero cases for solar radiation and sunshine hours were handled according to the screening procedure described in Section 2.3.2.

Table 6. Structure of the CBR case database.

Category	Variable	Description
Variogram parameter	Nugget	Nugget effect for each candidate model
	Total sill	Sum of the nugget and partial sill for each candidate model
	Distance parameter	Model-specific distance parameter for each candidate model
Performance	RMSE	Root mean square error for each candidate model
	MAE	Mean absolute error for each candidate model
	R^2	Coefficient of determination for each candidate model
Model selection	Selected model	Model with the minimum rank-sum score

Note: Each of the six variables (nugget, total sill, distance parameter, RMSE, MAE, and R^2) was stored separately for each of the four candidate variogram models, resulting in 24 variogram-related features for each time step. The selected model was stored separately as the model-selection label.

2.5. ANN-Based Feature Importance and Similarity Weight Determination

2.5.1. ANN-Based Variogram Model Classification

An Artificial Neural Network (ANN) classifier was trained using the 24 variogram-related features constructed in Section 2.4.3 to predict the variogram-model class. These input features comprised the nugget, total sill, distance parameter, RMSE, MAE, and R^2 for each of the four candidate variogram models. The model selected using the rank-sum criterion described in Section 2.4.3 was used as the model-selection label for ANN classification. Permutation feature importance was subsequently applied to the trained

ANN to quantify the contribution of each input feature, and the resulting importance values were used to derive similarity weights for CBR case retrieval.

The case data from 2019–2021 were combined and randomly divided into 80% training and 20% validation subsets using stratified sampling to preserve the variogram-model class distribution. The random state was fixed at 42 for reproducibility. ANN hyperparameters, including the hidden-layer architecture, activation function, L2 regularization coefficient, and initial learning rate, were selected through grid search using five-fold cross-validation on the training subset. Macro-F1 was used as the hyperparameter-selection criterion to account for class imbalance.

Min–max scaling was applied to the ANN input features. The scaling parameters were estimated exclusively from the training subset and subsequently applied unchanged to the validation subset. Values outside the range observed in the training subset were transformed using the same fitted scaling parameters without clipping to the $[0, 1]$ interval. After hyperparameter selection, the selected ANN was evaluated on the independent 20% validation subset. Permutation feature importance was then calculated using this validation subset with 10 permutation repetitions and a fixed random state of 42. Negative importance values were set to zero, and the remaining importance values were normalized to derive the feature weights used for subsequent CBR case retrieval.

2.5.2. Permutation Feature Importance and Weight Normalization

Permutation feature importance was applied to the trained ANN to quantify the contribution of each input feature to variogram model classification. This model-agnostic approach measures feature importance based on the decrease in predictive performance after randomly permuting the values of a given feature [21,22]. The permutation importance I_s of feature s was calculated as

$$I_s = \frac{1}{R}\sum_{r=1}^{R}(A_0 - A_{s,r}), \tag{7}$$

where A_0 denotes the classification accuracy obtained using the original validation data, $A_{s,r}$ denotes the classification accuracy after the rth random permutation of feature s, and R denotes the number of permutation repetitions. Each feature was permuted 10 times.

Only positive permutation importance values were retained for weight derivation. Negative importance values were set to zero, and the resulting values were normalized as

$$w_s = \frac{\max(I_s, 0)}{\sum_{j=1}^{S}\max(I_j, 0)}, \qquad \sum_{s=1}^{S} w_s = 1, \tag{8}$$

where $S = 24$ denotes the total number of input features. If all permutation importance values were non-positive, equal weights were assigned to all input features. The resulting feature weights were subsequently incorporated into the similarity measures used for CBR case retrieval.

To further examine the temporal stability of the permutation-based feature importance, an additional evaluation was conducted using temporally separated monthly periods for air temperature, relative humidity, Wind_{NS}, and Wind_{WE}. Detailed results are provided in Appendix C.

2.5.3. Weighted Similarity Measures

Before computing the weighted similarity measures, the condition attributes of the historical and query cases were normalised using feature-wise Min–Max scaling. The

minimum and maximum values were determined exclusively from the historical case database and applied consistently to both historical and query features. The query case was therefore excluded from the estimation of scaling parameters.

Table 7 summarises the three weighted similarity measures used for CBR case retrieval: weighted Euclidean, weighted Manhattan, and weighted grey-relational similarity. These measures were calculated using the normalised features and the feature weights derived in Section 2.5.2. The grey-relational similarity measure was based on the general concept of grey relational analysis [23] and implemented using feature-wise minimum and maximum differences across the candidate case library. The resulting similarity values were subsequently used for KNN-based case retrieval in Section 2.6.

Table 7. Weighted similarity measures used in this study.

Similarity Measure	Formula
Weighted Euclidean similarity	$SIM_E(A,B) = \frac{1}{1+\sqrt{\sum_{s=1}^{S}(w_s\Delta_s)^2}}$
Weighted Manhattan similarity	$SIM_M(A,B) = \frac{1}{1+\sum_{s=1}^{S} w_s\Delta_s}$
Weighted grey-relational similarity	$SIM_G(A,B) = \sum_{s=1}^{S}(w_s\xi_s)^2$ $\xi_s = \frac{2\Delta_{s,\min}+\Delta_{s,\max}}{2\Delta_s+\Delta_{s,\max}+\epsilon}$

Note: A and B denote the target and historical cases, respectively; w_s is the normalized weight of the sth feature; $\Delta_s = |a_s - b_s|$; S is the total number of features; $\Delta_{s,\min}$ and $\Delta_{s,\max}$ denote the minimum and maximum absolute differences of feature s across the candidate case library, respectively; ξ_s denotes the feature-wise grey-relational coefficient; and $\epsilon = 10^{-12}$ is a small positive constant used to prevent division by zero.

2.6. Similarity-Based Case Retrieval

The 2022 dataset was used to evaluate the similarity-based case retrieval procedure. For each input case, similarity was calculated against the historical cases in the 2019–2021 case database using the feature weights and weighted similarity measures defined in Section 2.5.3. Historical cases were ranked in descending order of similarity, and the K most similar cases were retrieved using a K-nearest neighbor (KNN) approach [24]. The variogram-model label of the input case was initially determined by majority voting among the labels of the K retrieved cases.

When two or more variogram models received the same highest number of votes, the similarity values associated with the retrieved cases of each tied model were summed, and the model with the highest cumulative similarity was selected. Thus, cumulative similarity was used as a secondary criterion when the vote count alone could not determine the retrieval label.

Retrieval configurations were evaluated for $K = 2, \ldots, 10$ using weighted Euclidean, weighted Manhattan, and weighted grey-relational similarities. Each configuration was evaluated using Accuracy and Macro-F1 after applying the similarity-based tie-breaking rule described above. To additionally penalize configurations that frequently produced voting ties, Tie-as-Incorrect Accuracy and Tie-as-Incorrect Macro-F1 were calculated by treating any case with an initial voting tie as an incorrect classification, irrespective of the label subsequently assigned by the tie-breaking rule. Macro-F1 was included to account for class-wise performance under class imbalance [25]. The four evaluation metrics were combined with equal weights using the Overall Score (OS):

$$OS = \frac{A + F_{\text{macro}} + A_{\text{tie}} + F_{\text{tie}}}{4}, \tag{9}$$

where A and F_{macro} denote Accuracy and Macro-F1, respectively, and A_{tie} and F_{tie} denote Tie-as-Incorrect Accuracy and Tie-as-Incorrect Macro-F1, respectively. The combination of similarity measure and K yielding the highest OS was selected as the final retrieval configuration.

The variogram-model label produced by the selected retrieval configuration was defined as the CBR retrieval label. In contrast to the model-selection label obtained directly from the rank-sum evaluation in Section 2.4.3, the CBR retrieval label was inferred from similar historical cases through KNN-based retrieval. This label was subsequently used as the target class for SVM training in Section 2.7.

2.7. SVM-Based Variogram Model Classification and Parameter Retrieval

A support vector machine (SVM) constructs a decision boundary by maximizing the margin between classes. Kernel functions enable nonlinear classification by implicitly mapping input features into a higher-dimensional feature space [26,27]. The decision hyperplane of a linear SVM is defined as

$$\mathbf{w}^T\mathbf{x} + b = 0, \tag{10}$$

where $\mathbf{x}$ is the input feature vector, $\mathbf{w}$ is the normal vector to the decision hyperplane, and b is the bias term. For nonseparable data, the soft-margin SVM balances margin maximization and classification error and is formulated as [26]

$$\begin{aligned} &\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \quad y_i\left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \qquad i = 1, \ldots, N. \end{aligned} \tag{11}$$

Here, C controls the trade-off between margin maximization and classification error, and ξ_i is the slack variable for the ith case. The mapping function $\phi(\cdot)$ transforms the input features into a feature space, and the corresponding kernel function is defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ [27]. For multiclass classification of the four candidate variogram models, the one-against-one strategy implemented in LIBSVM was adopted [28].

The SVM classification dataset was constructed from the 2022 KNN-based case-retrieval results obtained using the selected similarity measure and number of neighbors, K, for each meteorological variable. The variogram-related features defined in Section 2.4.3 were used as input features. Cases for which the KNN-derived CBR retrieval label agreed with the model-selection label obtained from the rank-sum criterion in Section 2.4.3 were retained. The corresponding CBR retrieval labels were then used as the target classes for SVM training.

For SVM model development, the resulting dataset was randomly divided into 80% training and 20% validation subsets using stratified sampling to preserve the variogram-model class distribution. A fixed random state of 42 was used for reproducibility. The input features were standardized using `StandardScaler` within the SVM pipeline. The scaler was fitted exclusively to the training data, and the fitted transformation was subsequently applied to the validation data, thereby preventing information from the validation subset from being used to estimate the scaling parameters.

Linear, radial basis function (RBF), sigmoid, and polynomial kernels were evaluated. Kernel-specific hyperparameters were selected through grid search using five-fold cross-validation on the 80% training subset, with Macro-F1 as the selection criterion to account for class imbalance. The selected SVM configuration was subsequently evaluated on the independent 20% validation subset. The evaluated kernel functions are summarized in Table 8.

Table 8. Kernel functions evaluated for SVM-based variogram model classification.

Kernel	Kernel Function
Linear	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
RBF	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right)$
Sigmoid	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma \mathbf{x}_i^T \mathbf{x}_j + r\right)$
Polynomial	$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\gamma \mathbf{x}_i^T \mathbf{x}_j + r\right)^d$

Note: γ denotes the kernel coefficient, r denotes the independent term (`coef0`) , and d denotes the polynomial degree.

The selected SVM was applied to each new case to predict its variogram-model class. Historical cases assigned to the same variogram-model class as the SVM prediction were retained as retrieval candidates. The selected weighted similarity measure from Section 2.6 was then applied within this candidate set, and the most similar historical case was retrieved. The variogram parameters associated with the retrieved case were subsequently used for Ordinary Kriging (OK) interpolation at the PV plant locations, as described in Section 3.1.

2.7.1. Comparison with Alternative Multi-Class Classifiers

To assess the classification performance and computational implications of the proposed multi-stage procedure, the KNN-SVM configuration was compared with Random Forest (RF)- and XGBoost-based alternatives. For each classifier family, two configurations were considered: a retrieval-assisted configuration, in which KNN-generated retrieval labels were used as target classes for RF or XGBoost training, and a direct multi-class configuration, in which RF or XGBoost was trained directly using the same variogram-related input features. This comparison was conducted to examine whether the SVM-based classification stage provided an advantage over alternative tree-based classifiers.

Table 9 summarizes the classification accuracy and inference time for the six meteorological variables. KNN-SVM achieved the highest classification accuracy for temperature, relative humidity, Wind_{NS}, and Wind_{WE}. For temperature and relative humidity, KNN-SVM achieved accuracies of 98.55% and 95.46%, respectively, compared with 90.97% and 83.20% for direct RF and 95.34% and 90.25% for direct XGBoost. Similar patterns were observed for the two wind components, although KNN-SVM generally required longer inference times than the tree-based alternatives.

Table 9. Classification accuracy and inference time of the proposed KNN-SVM method and RF- and XGBoost-based alternatives for the six meteorological variables.

(a) Comparison with RF-Based Classifiers

Meteorological Variable	KNN-SVM Accuracy (%)	KNN-RF Accuracy (%)	RF Accuracy (%)	KNN-SVM Time (s)	KNN-RF Time (s)	RF Time (s)
Temperature	98.55	90.67	90.97	0.061	0.042	0.211
Relative humidity	95.46	80.94	83.20	0.118	0.148	0.252
Wind_{NS}	89.82	84.01	85.25	0.700	0.060	0.060
Wind_{WE}	91.04	85.33	86.77	0.861	0.101	0.106
Solar radiation	87.66	90.10	91.36	0.090	0.093	0.060
Sunshine hours	89.37	91.54	91.06	0.144	0.061	0.088

Table 9. *Cont.*

(b) Comparison with XGBoost-Based Classifiers						
Meteorological Variable	**KNN-SVM Accuracy (%)**	**KNN-XGBoost Accuracy (%)**	**XGBoost Accuracy (%)**	**KNN-SVM Time (s)**	**KNN-XGBoost Time (s)**	**XGBoost Time (s)**
Temperature	98.55	94.77	95.34	0.060	0.028	0.031
Relative humidity	95.46	87.58	90.25	0.122	0.043	0.063
Wind_{NS}	89.82	88.04	89.10	0.706	0.022	0.023
Wind_{WE}	91.04	89.82	90.01	0.159	0.037	0.017
Solar radiation	87.66	90.63	92.28	0.100	0.006	0.014
Sunshine hours	89.37	91.70	81.84	0.146	0.019	0.005

Note: Classification performance was evaluated on the pooled external 2023–2025 dataset using the `best_model` variogram classes as reference labels. KNN-SVM, KNN-RF, and KNN-XGBoost use retrieval-based training cases, whereas RF and XGBoost are direct multi-class classifiers trained using the same variogram-related input features. Time denotes the measured inference time.

For solar radiation and sunshine hours, some tree-based alternatives achieved higher classification accuracy than KNN-SVM. For solar radiation, direct RF and direct XGBoost achieved accuracies of 91.36% and 92.28%, respectively, compared with 87.66% for KNN-SVM. For sunshine hours, KNN-RF and KNN-XGBoost achieved 91.54% and 91.70%, respectively, compared with 89.37% for KNN-SVM. The tree-based alternatives also generally required shorter inference times. These results demonstrate a trade-off between classification accuracy and computational cost: KNN-SVM provided higher accuracy for temperature, relative humidity, and the two wind components, whereas RF- and XGBoost-based alternatives were more computationally efficient and achieved higher accuracy for some solar-variable configurations.

2.7.2. Validation of SVM Predictions Against Rank-Sum References

To further evaluate whether the SVM classifier merely reproduced the KNN-derived labels or also remained consistent with the original variogram-model references, the SVM predictions were directly compared with the rank-sum reference labels defined in Section 2.4.3. The evaluation was conducted separately for each meteorological variable using Accuracy and Macro-F1.

Table 10 compares the SVM predictions against both the KNN-derived labels and the rank-sum references. For temperature and relative humidity, the SVM showed high agreement with the rank-sum references, achieving an accuracy of 99.636% and a Macro-F1 score of 99.375% for both variables. For Wind_{NS} and Wind_{WE}, the corresponding accuracies were 90.525% and 90.411%, with Macro-F1 scores of 77.236% and 70.763%, respectively.

Table 10. Comparison of SVM variogram-model classification performance against the KNN-derived labels and rank-sum reference labels.

Meteorological Variable	**SVM vs. KNN Labels**		**SVM vs. Rank-Sum References**		**SVM Error vs. Rank-Sum (%)**
	Accuracy (%)	**Macro-F1 (%)**	**Accuracy (%)**	**Macro-F1 (%)**	
Temperature	99.636	99.375	99.636	99.375	0.364
Relative humidity	98.746	98.387	99.636	99.375	0.364
Wind_{NS}	95.434	87.843	90.525	77.236	9.475
Wind_{WE}	92.979	77.284	90.411	70.763	9.589
Solar radiation	87.660	62.873	82.766	58.663	17.234
Sunshine hours	94.641	60.340	92.549	58.072	7.451

Note: Results are based on the 2019–2021 historical cases. KNN labels denote the retrieval-derived variogram-model labels described in Section 2.6, whereas Rank-Sum labels denote the variogram-model reference labels determined using the rank-sum criterion in Section 2.4.3. SVM performance is reported against both label definitions, and SVM error denotes the percentage of SVM predictions that differ from the corresponding rank-sum reference labels.

The largest discrepancy from the rank-sum references was observed for solar radiation, for which the SVM achieved an accuracy of 82.766% and a Macro-F1 score of 58.663%. For sunshine hours, the corresponding values were 92.549% and 58.072%, respectively. The relatively lower Macro-F1 scores for the solar variables, particularly in comparison with their accuracy values, indicate greater variation in class-wise classification performance. Overall, agreement with the rank-sum references was variable-dependent, and high agreement with the KNN-derived labels did not necessarily imply equivalent agreement with the original rank-sum labels. This additional comparison therefore provides a reference-based assessment of the final SVM classification beyond its agreement with the KNN-derived labels.

3. Meteorological Data Interpolation and Performance Evaluation

In this study, meteorological observations were spatially interpolated to six PV plant locations using the proposed CBR-OK method and four baseline interpolation methods: Direct-Fit Ordinary Kriging (Direct-OK), Static Ordinary Kriging (Static-OK), Nearest Neighbor (NN), and Inverse Distance Weighting (IDW). In CBR-OK, the variogram model was determined using the SVM-based classification procedure described in Section 2.7, and the corresponding variogram parameters were retrieved from the most similar historical case. Direct-OK fitted a variogram directly to the observations at each time step, whereas Static-OK used a fixed variogram configuration. NN and IDW were included as non-kriging interpolation baselines.

The spatial interpolation performance was evaluated using a station-based spatial holdout procedure, following the general principle of evaluating spatial interpolation at stations excluded from the interpolation inputs [29]. At each evaluation timestamp, 25 observation stations were randomly withheld simultaneously, while the remaining stations were used as interpolation inputs. The withheld stations and their observations were excluded from both query-feature construction and interpolation and were used only as spatial validation targets. Predictions at the withheld station locations were compared with the corresponding observations using RMSE and MAE. This procedure therefore provides a station-based assessment of spatial interpolation performance while retaining the remaining observation network for spatial estimation. Missing or invalid observations were excluded, whereas physically valid zero-valued observations were retained.

3.1. Meteorological Data Interpolation Using Ordinary Kriging

In the proposed CBR-OK method, Ordinary Kriging (OK) was performed using the variogram-model class predicted by the SVM in Section 2.7 and the nugget, total sill, and distance parameter retrieved from the most similar historical case within the predicted class. OK estimates the value at an unobserved location as a weighted linear combination of surrounding observations, with the weights determined according to the spatial dependence represented by the variogram [30]. Kriging-based methods have also been applied to the spatial interpolation of meteorological variables using observations from surrounding stations [29].

The meteorological variable at target location s_0 was estimated from n observations as

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i), \tag{12}$$

where $Z(s_i)$ denotes the observed value at the ith station and λ_i is the corresponding kriging weight. To ensure unbiasedness of the OK estimator, the weights were constrained as [30]

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{13}$$

The kriging weights were determined from the theoretical variogram values between observation stations and between each observation station and the target location through the following system of equations [30]:

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix}, \tag{14}$$

where γ_{ij} denotes the theoretical variogram value between observation stations i and j, γ_{i0} denotes the theoretical variogram value between observation station i and target location s_0, and μ is the Lagrange multiplier associated with the unbiasedness constraint. At each time step, interpolation was performed only when at least three valid observation stations were available.

Air temperature, relative humidity, solar radiation, and sunshine hours were directly interpolated from the preprocessed observations. For wind speed and direction, the orthogonal west–east and north–south components, $\mathrm{Wind}_{\mathrm{WE}}$ and $\mathrm{Wind}_{\mathrm{NS}}$, defined in Section 2.3, were interpolated separately and subsequently reconstructed at the target location. Wind speed was reconstructed as

$$\widehat{WS} = \sqrt{\widehat{W}_{WE}^2 + \widehat{W}_{NS}^2}, \tag{15}$$

and wind direction was reconstructed as

$$\widehat{WD} = \left[\frac{180}{\pi} \operatorname{atan2}\left(\widehat{W}_{WE}, \widehat{W}_{NS} \right) + 360 \right] \bmod 360. \tag{16}$$

Here, $\widehat{W}_{WE}$ and $\widehat{W}_{NS}$ denote the interpolated west–east and north–south wind components, respectively, whereas $\widehat{WS}$ and $\widehat{WD}$ denote the reconstructed wind speed and wind direction at the target location. Equation (16) is consistent with the positive sine and cosine component definitions in Section 2.3.1, with wind direction represented clockwise from true north and normalized to the interval $[0, 360^\circ)$. Thus, a wind direction of 0° is reconstructed as 0° rather than 180°.

This procedure was repeated at each time step to generate site-specific time series of air temperature, relative humidity, wind speed, wind direction, solar radiation, and sunshine hours at the six PV plant locations.

To ensure the physical consistency of the interpolated meteorological variables, physical constraints were applied as a post-processing step. Negative solar radiation estimates were set to zero, sunshine-hour estimates were constrained to the range of 0–1 h, and relative humidity was constrained to 0–100%. Wind speed and wind direction were reconstructed from the interpolated west–east and north–south components according to Equations (15) and (16), respectively, ensuring non-negative wind-speed values and wind-direction values within $[0, 360^\circ)$.

3.2. Baseline Interpolation Methods

Four baseline interpolation approaches were considered: Direct-Fit Ordinary Kriging (Direct-OK), Static Ordinary Kriging (Static-OK), Inverse Distance Weighting (IDW), and Nearest Neighbor (NN). These baselines were used to distinguish the contribution of CBR-based variogram retrieval from the general effects of spatial interpolation and kriging.

Direct-OK fits a variogram directly to the observations available at each time step, providing a direct comparison with CBR-OK without historical case retrieval. Static-OK applies each candidate variogram model separately using a fixed configuration without

dynamic model selection, allowing the effect of adaptive variogram-model selection to be evaluated. IDW provides a conventional deterministic spatial interpolation baseline, whereas NN provides a simple local spatial interpolation baseline.

3.2.1. Direct-Fit Ordinary Kriging

Direct-Fit Ordinary Kriging (Direct-OK) was employed to evaluate the contribution of CBR-based variogram model and parameter retrieval. Unlike CBR-OK, which retrieves the variogram model and its parameters from historical cases, Direct-OK determines them directly from the meteorological observations available at each time step.

For each time step, an empirical semivariogram was first constructed from the available observations. The four candidate variogram models were then fitted using the same weighted least-squares procedure described in Section 2.4.2. The fitted models were evaluated using the same rank-based criterion defined in Section 2.4.3, and the selected variogram model and its directly fitted nugget, total sill, and distance parameter were subsequently applied to Ordinary Kriging.

Except for the procedure used to determine the variogram model and parameters, Direct-OK and CBR-OK used the same observations, spatial coordinates, target locations, and interpolation settings. This comparison therefore isolates the contribution of CBR-based variogram retrieval under an otherwise identical Ordinary Kriging framework.

3.2.2. Static Ordinary Kriging

Static Ordinary Kriging (Static-OK) was employed to evaluate the effect of dynamic variogram-model determination in CBR-OK relative to a fixed variogram model family. Four Static-OK configurations were considered using the Exponential, Spherical, Rational Quadratic, and Penta-spherical models. In each configuration, only the variogram model family was fixed, whereas the variogram parameters were independently estimated at each timestamp. Specifically, an empirical semivariogram was constructed from the contemporaneous training observations, and the nugget, partial sill, and distance parameter were fitted for the predefined model family. The resulting timestamp-specific parameters were then used for Ordinary Kriging.

Static-OK and CBR-OK used the same meteorological observations, spatial coordinates, target locations, and Ordinary Kriging procedure. Therefore, the comparison was designed to evaluate the effect of dynamic variogram-model determination while maintaining the remaining interpolation conditions consistently across the methods.

3.2.3. Inverse Distance Weighting

Inverse Distance Weighting (IDW) was employed as a deterministic spatial interpolation baseline for comparison with CBR-OK. IDW estimates the value at an unsampled location as a weighted average of surrounding observations, where the contribution of each observation decreases with increasing distance from the target location [31,32]. Unlike Ordinary Kriging, IDW does not require variogram model selection or parameter estimation, thereby providing a conventional non-geostatistical baseline for evaluating the interpolation performance of CBR-OK.

The IDW estimate at a target location s_0 is expressed as

$$\hat{Z}(s_0) = \frac{\sum_{i=1}^{N} d(s_0, s_i)^{-p} Z(s_i)}{\sum_{i=1}^{N} d(s_0, s_i)^{-p}}, \tag{17}$$

where $\widehat{Z}(s_0)$ is the estimated value at the target location, $Z(s_i)$ is the observed value at station i, $d(s_0, s_i)$ denotes the distance between the target location and station i, N is the number of observations used for interpolation, and p is the distance-decay power parameter. A larger value of p assigns greater influence to observations located closer to the target location [31,32]. In this study, p was fixed at 2 for all IDW interpolations. Distances were calculated using station and target coordinates projected to EPSG:5179.

For comparison, IDW and CBR-OK used the same meteorological observations, spatial coordinates, target locations, evaluation periods, and spatial holdout procedure. This configuration enabled the two methods to be evaluated under consistent interpolation and evaluation conditions.

3.2.4. Nearest Neighbor

Nearest Neighbor (NN) was employed as the simplest spatial baseline for comparison with CBR-OK. For each target PV site, the meteorological observation from the geographically closest available observation station was directly assigned to the target location without combining observations from multiple neighboring stations or modeling an explicit spatial covariance structure.

Let s_0 denote the target location and s_i the locations of the available meteorological stations. The NN estimate is expressed as

$$\widehat{Z}(s_0) = Z(s_{i^*}), \qquad i^* = \arg\min_i d(s_0, s_i), \tag{18}$$

where $d(s_0, s_i)$ denotes the spatial distance between target location s_0 and observation station s_i. NN therefore provides a simple spatial reference for evaluating whether the interpolation methods provide additional benefit beyond directly using the observation from the nearest available station.

For comparison, NN and CBR-OK used the same meteorological observations, spatial coordinates, target locations, evaluation periods, and spatial holdout procedure. This configuration enabled the two methods to be evaluated under consistent interpolation and evaluation conditions.

3.3. Interpolation Performance Evaluation

The spatial interpolation performance of CBR-OK and the four baseline approaches—Direct-OK, Static-OK, IDW, and NN—was evaluated using the station-based spatial holdout procedure described in Section 3. For Static-OK, the four fixed variogram-model configurations described in Section 3.2.2 were evaluated separately. At each evaluation timestamp, 25 observation stations were randomly withheld simultaneously, and their observations were used only as spatial validation targets. The remaining stations were used as interpolation inputs. Evaluation was performed only when at least three valid observation stations remained available for interpolation.

Interpolation performance was quantified using the root mean square error (RMSE) and mean absolute error (MAE) [33]. Owing to its squared-error term, RMSE is more sensitive to relatively large errors, whereas MAE represents the average magnitude of the absolute errors [33].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{Z}_i - Z_i\right)^2}, \tag{19}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{Z}_i - Z_i\right|, \tag{20}$$

where Z_i and $\widehat{Z}_i$ denote the observed and spatially interpolated values for the ith validation sample, respectively, and N is the total number of validation samples. Lower RMSE and MAE values indicate better interpolation performance.

4. PV Power Forecasting Models

4.1. PV Sites and Meteorological Input Data

PV power forecasting models were developed for the six PV sites shown in Figure 4 using meteorological inputs obtained using the NN, IDW, and CBR-OK methods. Six meteorological variables were considered: air temperature, relative humidity, wind speed, wind direction, solar radiation, and sunshine hours. For the NN inputs, the first four variables were obtained from the nearest Automatic Weather Station (AWS), whereas solar radiation and sunshine hours were obtained from the nearest Automated Synoptic Observing System (ASOS) station. The corresponding stations and distances from each PV site are summarized in Table 11.

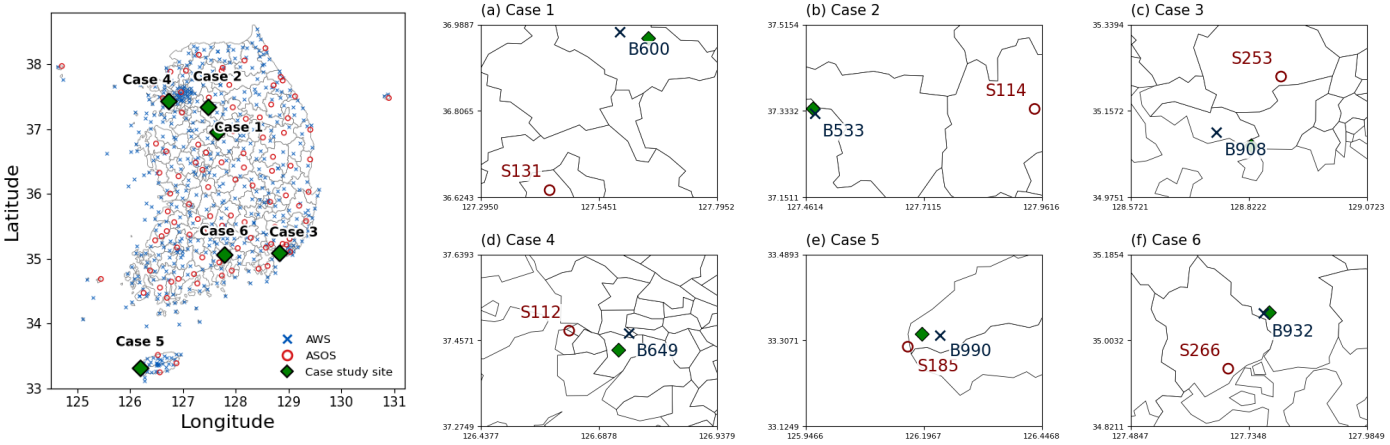


Figure 4. Spatial distribution of meteorological observation stations and detailed locations of the six PV plant case-study sites. The overview map shows the locations of the six PV plants together with AWS and ASOS stations, while the detailed maps show the PV site, solar observation station, and basic meteorological station for (**a**) Case 1, (**b**) Case 2, (**c**) Case 3, (**d**) Case 4, (**e**) Case 5, and (**f**) Case 6.

Table 11. Nearest AWS and ASOS stations used for NN meteorological inputs at the six PV sites.

ID	City	Basic Meteorological Elements Measured by AWS			Solar Variables Measured by ASOS		
		Closest AWS	Closest AWS ID	Distance to Closest AWS (km)	Closest ASOS	Closest ASOS ID	Distance to Closest ASOS (km)
Case 1	Chungbuk	Geumwang	600	5.52	Cheongju	131	40.09
Case 2	Gyeonggi	Baeksa	533	1.26	Wonju	114	41.65
Case 3	Busan	Jinhae	908	7.30	Gimhae	253	17.13
Case 4	Incheon	Bupyeong	649	4.38	Incheon	112	10.32
Case 5	Jeju	Nakcheon	990	3.59	Gosan	185	4.00
Case 6	Hadong	Hadong	932	1.09	Gwangyang	266	15.45

Note: AWS observations include air temperature, relative humidity, wind speed, and wind direction, whereas ASOS observations provide solar radiation and sunshine hours.

For the IDW and CBR-OK inputs, the same six meteorological variables were spatially interpolated at each PV site using the procedures described in Sections 3.2.3 and 3.1, respectively. For CBR-OK, variogram-model classification and parameter retrieval followed the procedure described in Section 2.7. The meteorological inputs obtained using the three methods were aligned with the observed PV power at the corresponding site to construct

the forecasting datasets. The observed PV power was used exclusively as the prediction target and was not included as an input feature of the forecasting models.

4.2. Forecasting Models and Experimental Design

To evaluate the effect of meteorological inputs generated by different spatial interpolation methods on PV power forecasting, identical preprocessing and model-training procedures were applied to the NN-, IDW-, and CBR-OK-based datasets. Three recurrent neural network architectures—Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU)—were employed because of their ability to model temporal dependencies in sequential data and their successful application to PV power forecasting [34].

Forecast horizons of 1, 12, and 24 h were evaluated for each model, enabling a consistent comparison among NN-, IDW-, and CBR-OK-based meteorological inputs under the same forecasting conditions.

4.3. Meteorological Data and Time-Series Input Preprocessing

4.3.1. Cyclical Feature Encoding

The meteorological datasets generated by NN, IDW, and CBR-OK were aligned to common timestamps and subjected to identical preprocessing. To preserve the circular nature of wind direction and avoid discontinuity at the 0°/360° boundary, wind direction was transformed into sine and cosine components [35]:

$$WD_{\sin} = \sin\left(\frac{\pi WD}{180}\right), \qquad WD_{\cos} = \cos\left(\frac{\pi WD}{180}\right), \tag{21}$$

where WD denotes wind direction in degrees. To represent diurnal and seasonal periodicities relevant to PV power generation and meteorological conditions, the hour of the day (H) and day of the year (DOY) were similarly encoded using sine and cosine functions [36]:

$$H_{\sin} = \sin\left(\frac{2\pi H}{24}\right), \qquad H_{\cos} = \cos\left(\frac{2\pi H}{24}\right), \tag{22}$$

$$DOY_{\sin} = \sin\left(\frac{2\pi (DOY-1)}{365.25}\right), \qquad DOY_{\cos} = \cos\left(\frac{2\pi (DOY-1)}{365.25}\right). \tag{23}$$

Here, H and DOY denote the hour of the day and day of the year, respectively. A period of 365.25 days was used to provide a consistent representation of the annual cycle across multiple years. The resulting forecasting input vector comprised 11 variables: air temperature, relative humidity, wind speed, $WD_{\sin}$, $WD_{\cos}$, solar radiation, sunshine hours, $H_{\sin}$, $H_{\cos}$, $DOY_{\sin}$, and $DOY_{\cos}$. Historical PV power was not included in the input vector and was used only as the prediction target.

For an h-hour-ahead forecast issued at time t with a look-back window of L hours, the forecasting model used the input sequence

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \ldots, \mathbf{x}_t], \tag{24}$$

to predict the PV power P_{t+h}, where

$$\mathbf{x}_t = [T_t, RH_t, WS_t, WD_{\sin,t}, WD_{\cos,t}, SR_t, SH_t, H_{\sin,t}, H_{\cos,t}, DOY_{\sin,t}, DOY_{\cos,t}]. \tag{25}$$

Accordingly, the 1-, 12-, and 24-h forecasting experiments used only information available at or before the forecast origin t. No future meteorological observations or meteorological forecasts beyond the forecast origin were supplied to the LSTM, Bi-LSTM, or GRU models.

4.3.2. Data Standardization and Time-Series Splitting

All input features were standardized using the mean and standard deviation calculated exclusively from the training set. The same standardization parameters were subsequently applied to the validation and test sets to prevent information leakage. The dataset was chronologically divided, without random shuffling, into training (70%), validation (15%), and test (15%) sets while preserving the temporal order of the observations.

For each forecast origin, a fixed-length sequence of historical input features was constructed, with PV power at 1, 12, or 24 h ahead defined as the prediction target depending on the forecast horizon. The input sequence length was determined through hyperparameter tuning for each forecasting model.

4.4. Long Short-Term Memory (LSTM)

4.4.1. LSTM Architecture

Long Short-Term Memory (LSTM) was introduced to address the difficulty of learning long-term dependencies in recurrent neural networks (RNNs) through gating mechanisms and a cell state [37]. In this study, LSTM was employed to predict PV power from historical sequences of the meteorological and temporal features described in Section 4.3. At time step t, the LSTM operations were defined as

$$f_t = \sigma\Big(W_f[h_{t-1}, x_t] + b_f\Big), \tag{26}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{27}$$

$$\widetilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \tag{28}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t, \tag{29}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{30}$$

$$h_t = o_t \odot \tanh(C_t). \tag{31}$$

Here, x_t denotes the input feature vector at time step t; h_t and C_t denote the hidden and cell states, respectively; and $\widetilde{C}_t$ denotes the candidate cell state. The variables f_t, i_t, and o_t represent the forget, input, and output gates, respectively. W and b denote trainable weight matrices and bias terms, while $\sigma(\cdot)$ and $\tanh(\cdot)$ denote the sigmoid and hyperbolic tangent activation functions, respectively.

The final hidden state of the input sequence was mapped to the predicted PV power through a fully connected layer. Forecasts were generated for horizons of 1, 12, and 24 h.

4.4.2. Model Training and Hyperparameter Optimization

The forecasting model architectures and training configurations were determined through a hyperparameter search over predefined candidate values based on validation performance. The tuned hyperparameters included the input sequence length (lookback), number of hidden units, number of recurrent layers, dropout rate, learning rate, and batch size. Hyperparameter optimisation was conducted separately for each PV site and forecasting architecture using the 1-h forecast horizon.

To ensure that the selected hyperparameter configuration was not specific to a single meteorological interpolation method, each candidate configuration was evaluated separately using the validation datasets constructed from NN-, IDW-, and CBR-OK-based meteorological inputs. A joint validation objective was used to assess performance across the three input methods. Specifically, the objective combined the mean normalised root mean squared error (NRMSE) with a penalty for variability in validation performance, as follows:

$$J(\boldsymbol{\theta}) = \frac{1}{3} \sum_{m \in \mathcal{M}} \mathrm{NRMSE}_m(\boldsymbol{\theta}) + \lambda \, \mathrm{SD}\big(\{\mathrm{NRMSE}_m(\boldsymbol{\theta})\}_{m \in \mathcal{M}}\big), \tag{32}$$

where $\boldsymbol{\theta}$ denotes a candidate hyperparameter configuration. The set $\mathcal{M}$ comprises the three meteorological input methods: NN, IDW, and CBR-OK. For each method m, the corresponding validation NRMSE under configuration $\boldsymbol{\theta}$ is denoted by $\mathrm{NRMSE}_m(\boldsymbol{\theta})$. The operator SD denotes the population standard deviation of the three validation NRMSE values, and the penalty coefficient was set to $\lambda = 0.25$.

For hyperparameter selection, the validation NRMSE was calculated by dividing the RMSE by the range of observed PV power in the training subset, defined as the difference between its maximum and minimum values:

$$\mathrm{NRMSE}_m(\boldsymbol{\theta}) = \frac{\mathrm{RMSE}_{m,\mathrm{val}}(\boldsymbol{\theta})}{P_{\mathrm{train,max}} - P_{\mathrm{train,min}}}, \tag{33}$$

where $\mathrm{RMSE}_{m,\mathrm{val}}(\boldsymbol{\theta})$ denotes the validation RMSE for meteorological input method m and $P_{\mathrm{train,max}}$ and $P_{\mathrm{train,min}}$ represent the maximum and minimum observed PV power in the training subset, respectively. The normalisation factor was calculated exclusively from the training data and applied consistently across the three meteorological input methods for each PV site. This training-range-based NRMSE was used specifically for hyperparameter selection, whereas the capacity-normalised metrics reported in Section 5.6 were used for the final forecasting performance evaluation.

For each PV site and forecasting architecture, the configuration minimising $J(\boldsymbol{\theta})$ was selected to balance low average validation error and limited performance variability across the three meteorological input methods. This joint selection yielded a single configuration for each site and architecture, rather than separate configurations for NN, IDW, and CBR-OK. The selected configuration was subsequently applied unchanged to the 1-, 12-, and 24-h forecasting experiments. The selected configurations for the six PV sites are summarised in Table 20.

The models were trained using the Adam optimizer [38] with a peak-weighted Huber loss. The standard Huber loss [39] is defined as

$$L_\delta(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta, \end{cases} \tag{34}$$

where e denotes the error between the predicted and observed PV power in the standardised target space and δ is the threshold separating the quadratic and linear regions of the loss function. Prior to model training, the target PV power was standardised using the mean and standard deviation calculated exclusively from the training data as

$$P_t^* = \frac{P_t - \mu_{\mathrm{train}}}{\sigma_{\mathrm{train}}}, \tag{35}$$

where P_t^* denotes the standardised PV power and μ_{train} and σ_{train} denote the mean and standard deviation of the training target, respectively. The Huber-loss threshold was set to $\delta = 1$ in the standardised target space; therefore, the threshold corresponds to one training-set standard deviation rather than a fixed error of 1 kW.

To place greater emphasis on periods of high PV generation, the Huber loss was weighted according to the observed PV power. Samples corresponding to the upper 20% of the training-target distribution were assigned a weight of 3, whereas the remaining samples were assigned a weight of 1. The resulting peak-weighted loss preserves the robustness of the Huber loss while placing greater emphasis on high-generation periods during model training. Early stopping was applied based on the validation loss.

For a fair comparison, the selected site- and architecture-specific configuration and training conditions were applied consistently across the NN-, IDW-, and CBR-OK-based datasets at each forecasting horizon. Final forecasting performance was evaluated on an independent test set that was not used for model training or hyperparameter selection.

4.5. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) combines two LSTM networks that process an input sequence in the forward and backward directions, enabling temporal dependencies in both directions within the sequence to be captured [40,41]. In this study, the backward LSTM processed only the historical input sequence available before each forecast origin in reverse order and did not use observations occurring after the forecast origin. The forward and backward hidden states at time step t were defined as

$$\overrightarrow{h}_t = \mathrm{LSTM}_f\left(x_t, \overrightarrow{h}_{t-1}\right), \qquad \overleftarrow{h}_t = \mathrm{LSTM}_b\left(x_t, \overleftarrow{h}_{t+1}\right), \tag{36}$$

respectively. The two hidden states were concatenated as

$$h_t^{\mathrm{Bi}} = \left[\overrightarrow{h}_t; \overleftarrow{h}_t\right], \tag{37}$$

where x_t denotes the input feature vector at time step t and $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ denote the forward and backward hidden states, respectively. The resulting bidirectional representation was mapped to the predicted PV power through a fully connected layer.

The meteorological and temporal input sequences described in Section 4.3 were used as inputs to the Bi-LSTM model. The same forecast horizons and hyperparameter-search procedure described for the LSTM model in Section 4.4.2 were applied. Within each case, identical Bi-LSTM architectures, hyperparameters, and training conditions were used for the NN-, IDW-, and CBR-OK-based input datasets, and final performance was evaluated on the independent test set.

4.6. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a recurrent neural network architecture that controls the retention and updating of sequential information through update and reset gates [42]. Unlike LSTM, GRU does not employ a separate cell state and directly updates the hidden state, providing a compact architecture for learning temporal dependencies [42]. At time step t, the update and reset gates were defined as

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \tag{38}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r). \tag{39}$$

The candidate hidden state was computed as

$$\widetilde{h}_t = \tanh[W_h x_t + U_h(r_t \odot h_{t-1}) + b_h], \tag{40}$$

and the hidden state was updated as

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h}_t. \tag{41}$$

Here, x_t and h_t denote the input feature vector and hidden state at time step t, respectively, while z_t and r_t denote the update and reset gates. W_* and U_* are trainable weight matrices; b_* denotes the corresponding bias vectors; $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangent activation functions, respectively; and $\odot$ denotes element-wise multiplication.

The meteorological and temporal input sequences described in Section 4.3 were used as inputs to the GRU model, and the final hidden state was mapped to the predicted PV power through a fully connected layer. GRU-based architectures have also been successfully applied to short-term PV power forecasting [43]. The same forecast horizons and hyperparameter-search procedure described in Section 4.4.2 were applied. Within each case, identical GRU architectures, hyperparameters, and training conditions were used for the NN-, IDW-, and CBR-OK-based input datasets, and final performance was evaluated on the independent test set.

5. Results and Discussion

5.1. ANN-Based Variogram Model Classification and Feature Importance

ANN-based variogram model classifiers were developed separately for each meteorological variable following the procedure described in Section 2.5.1. The 2019–2021 cases were combined and randomly divided into 80% training and 20% validation subsets using stratified sampling. ANN hyperparameters were selected using five-fold cross-validation on the training subset, with Macro-F1 as the selection criterion. The reference labels were the variogram-model classes selected using the rank-sum criterion described in Section 2.4.3. AWS observations were used for temperature, relative humidity, and wind components, whereas ASOS observations were used for solar radiation and sunshine hours. Table 12 summarizes the selected ANN hyperparameters and cross-validation performance, and Figure 5 presents the normalized permutation feature importance calculated using the independent 20% validation subset.

Table 12. Selected ANN hyperparameters and five-fold cross-validation performance using the 2019–2021 case database.

Meteorological Variable	Activation	α	Hidden Layer Sizes	Learning Rate	CV Macro-F1 (%)	Accuracy (%)
Temperature	tanh	0.0001	(64, 32, 16)	0.0001	91.31 ± 0.60	96.8
Relative Humidity	tanh	0.0010	(128, 64)	0.0010	91.94 ± 0.31	92.9
Wind_{NS}	tanh	0.0001	(128, 64)	0.0010	82.37 ± 0.95	93.3
Wind_{WE}	tanh	0.0010	(64, 32, 16)	0.0001	81.55 ± 1.19	92.3
Solar Radiation	tanh	0.0001	(64, 32, 16)	0.0001	74.72 ± 3.17	89.7
Sunshine Hours	tanh	0.0001	(64, 32)	0.0010	59.41 ± 1.11	91.5

Note: The 2019–2021 case data were randomly stratified into 80% training and 20% validation subsets. CV Macro-F1 values are reported as the mean ± standard deviation across the five cross-validation folds of the training subset. The Accuracy values were calculated using the independent 20% validation subset.

The highest mean CV Macro-F1 scores were obtained for relative humidity ($91.94 \pm 0.31\%$) and temperature ($91.31 \pm 0.60\%$), followed by Wind_{NS} ($82.37 \pm 0.95\%$) and Wind_{WE} ($81.55 \pm 1.19\%$). Solar radiation and sunshine hours yielded lower scores of $74.72 \pm 3.17\%$ and $59.41 \pm 1.11\%$, respectively, despite achieving accuracies of 89.7% and 91.5%. Solar radiation also showed greater variation across the cross-validation folds. The differences between Accuracy and Macro-F1 for solar radiation and sunshine hours indicate less balanced classification performance across the variogram classes.

The permutation feature importance results in Figure 5 show that the contributions of the variogram-related features differed across meteorological variables and candidate variogram models. R^2 was particularly influential for temperature and the two wind components. For the Exponential model, its normalized weights were 0.221, 0.139, and 0.164 for temperature, Wind_{NS}, and Wind_{WE}, respectively. For relative humidity, R^2 showed the highest importance, followed by MAE and RMSE, whereas MAE and RMSE were among the more influential for solar radiation. For sunshine hours, RMSE showed

the highest importance, followed by sill. Range and nugget generally showed relatively low importance across most meteorological variables and candidate variogram models.

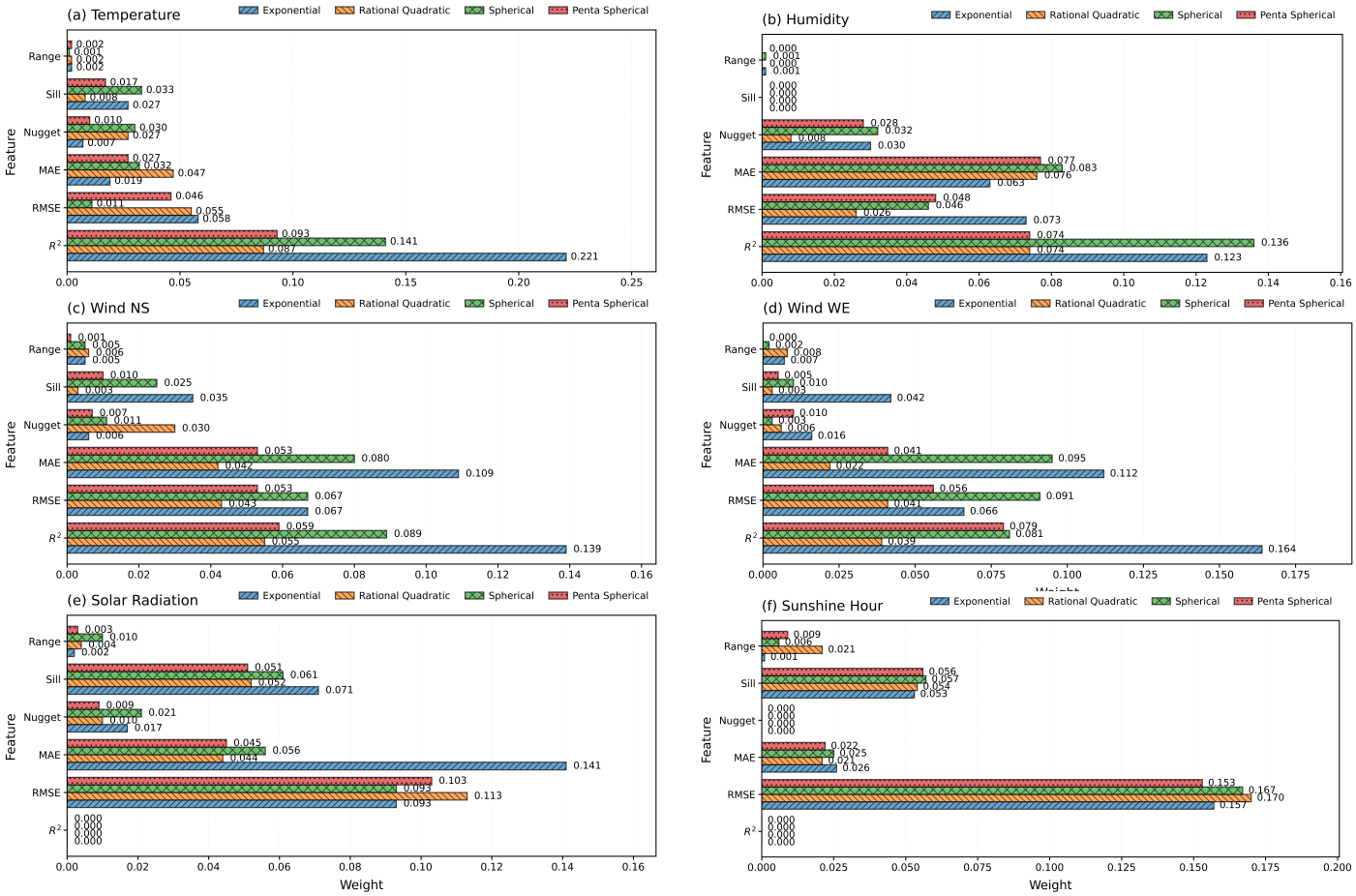


Figure 5. Normalized permutation feature importance calculated using the independent 20% validation subset of the 2019–2021 case database for the six meteorological variables: (**a**) temperature, (**b**) relative humidity, (**c**) Wind_{NS}, (**d**) Wind_{WE}, (**e**) solar radiation, and (**f**) sunshine hours.

Accordingly, the normalized permutation importance for each meteorological variable was used as a feature-specific weight in the subsequent similarity calculation rather than assigning uniform weights to all features. These weights were applied to KNN-based case retrieval, as described in Section 2.6.

Temporal robustness was further evaluated using monthly blocked evaluation during 2023–2025 without additional hyperparameter tuning. All six meteorological variables were included. For solar radiation and sunshine hours, valid daytime observations from 09:00–16:59 KST were used to prevent structurally occurring nighttime values from dominating the evaluation. The detailed validation procedure and results are provided in Appendix D.

5.2. Similarity-Based Case Retrieval and Retrieval Condition Selection

KNN-based case retrieval performance for the 2022 dataset was evaluated using the ANN-derived feature weights obtained in Section 2.5. The three similarity measures were compared across different values of K. Figure 6 presents temperature as a representative example. Accuracy and Macro-F1 showed relatively small variations across K, whereas the Tie-as-Incorrect metrics, particularly Macro-F1, exhibited substantially greater variation. This suggests that ties in multiclass majority voting can affect class-wise retrieval performance. $K = 1$ was included only as a reference, while $K = 2, \ldots, 10$, as defined in Section 2.6, were considered for retrieval-condition selection.

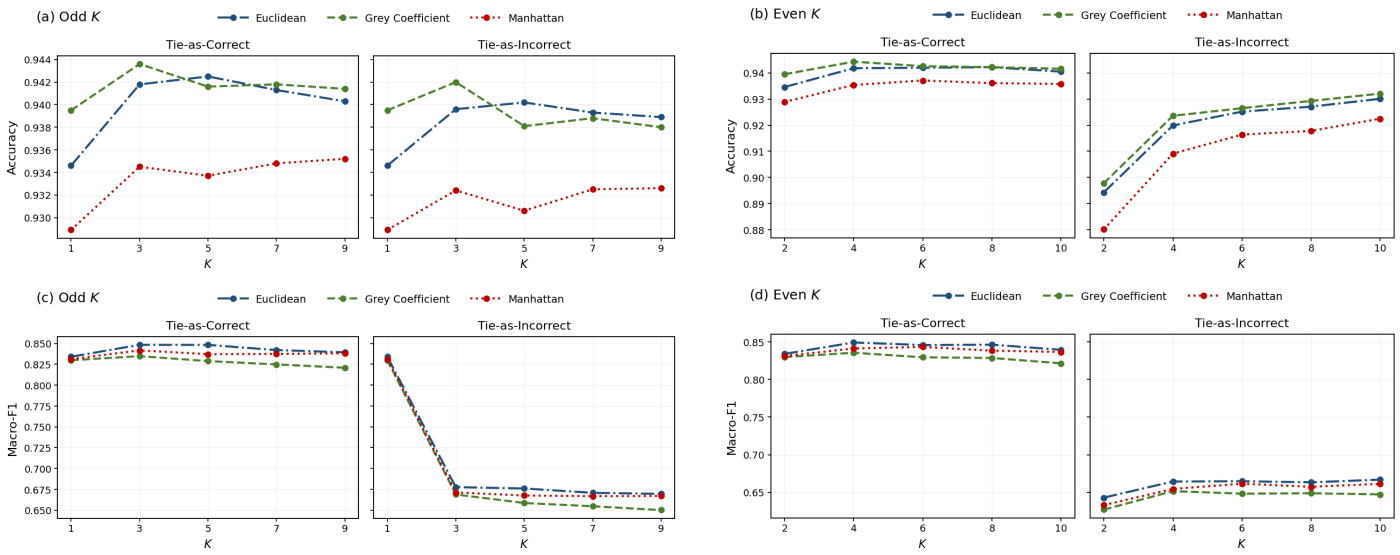


Figure 6. KNN-based case-retrieval performance for temperature according to similarity measure and number of nearest neighbors, *K*: (**a**) Accuracy for odd *K*; (**b**) Accuracy for even *K*; (**c**) Macro-F1 for odd *K*; and (**d**) Macro-F1 for even *K*.

For each meteorological variable, the retrieval condition was selected from combinations of $K = 2, \ldots, 10$ and the three similarity measures using the Overall Score defined in Section 2.6. As summarized in Table 13, Euclidean similarity was selected for temperature ($K = 3$) and Wind NS ($K = 9$), Grey Relational similarity for relative humidity ($K = 7$) and solar radiation ($K = 3$), and Manhattan similarity for Wind WE ($K = 5$) and sunshine hours ($K = 5$).

Table 13. Selected similarity measure, *K*, and KNN-based case-retrieval performance for each meteorological variable using the 2022 evaluation dataset.

Meteorological Variable	Similarity Measure	*K*	Accuracy (%)	Macro-F1 (%)	Tie-as-Incorrect Accuracy (%)	Tie-as-Incorrect Macro-F1 (%)
Temperature	Euclidean	3	94.18	84.83	93.96	67.77
Relative Humidity	Grey Relational	7	81.91	77.75	80.96	62.11
Wind NS	Euclidean	9	90.25	75.18	90.05	59.96
Wind WE	Manhattan	5	88.85	70.27	88.33	55.61
Solar Radiation	Grey Relational	3	84.56	68.86	83.92	53.95
Sunshine Hours	Manhattan	5	91.42	60.26	91.05	47.95

Note: Retrieval performance was evaluated using the 2022 dataset with the ANN-derived feature weights obtained from the preceding model-development stage. The similarity measure and *K* were selected from the evaluated combinations using the Overall Score defined in Section 2.6.

Under the selected conditions, temperature achieved the highest Accuracy (94.18%) and Macro-F1 (84.83%). The Tie-as-Incorrect Macro-F1 was lower than the conventional Macro-F1 for all meteorological variables, with particularly low values for solar radiation (53.95%) and sunshine hours (47.95%). These results indicate that tied retrieval outcomes had a greater effect on class-wise performance than on overall accuracy.

Although all selected values of *K* were odd, ties were not necessarily eliminated because multiclass majority voting can produce ties even for odd *K*. The variation in the selected similarity measures and *K* values across the meteorological variables indicates that the appropriate case-retrieval condition was variable-dependent. Accordingly, the variable-specific retrieval conditions summarized in Table 13 were used in the subsequent analysis.

The variogram model determined under each selected retrieval condition was defined as the KNN-derived label. Rather than representing the original rank-sum reference label,

the KNN-derived label was obtained through retrieval from the case database constructed in Section 2.4 using the procedure described in Section 2.6. These KNN-derived labels were subsequently used as target classes for the SVM described in Section 2.7, allowing the retrieval-based variogram model selection patterns to be learned by the SVM classifier.

5.3. SVM-Based Variogram Model Decision Boundaries and Classification Performance

The SVM framework described in Section 2.7 was used to learn the variogram-model selection patterns represented by the KNN-derived labels obtained in Section 2.6. The classification dataset was constructed from the 2022 KNN-based case-retrieval results for each meteorological variable, with the KNN-derived labels used as the target classes. The dataset was stratified into 80% training and 20% validation subsets. SVM hyperparameters were selected using five-fold cross-validation on the training subset with Macro-F1 as the selection criterion, and the selected configurations were evaluated on the independent 20% validation subset. The resulting configurations and classification performance are summarized in Table 14.

Table 14. Selected SVM configurations and classification performance for each meteorological variable.

Meteorological Variable	Selected Kernel	C	γ	Degree	r	CV Macro-F1 (%)	Validation Accuracy (%)	Validation Macro-F1 (%)
Temperature	Linear	100	–	–	–	98.4 ± 0.9	99.6	99.3
Humidity	Linear	100	–	–	–	97.7 ± 0.49	98.7	98.3
Wind NS	Polynomial	0.5	0.01	2	50	94.7 ± 2.7	97.4	91.4
Wind WE	Polynomial	0.5	0.01	2	50	89.4 ± 2.69	97.1	93.7
Solar Radiation	Linear	100	–	–	–	80.4 ± 7.58	94.1	93.7
Sunshine Hours	RBF	–	scale	–	–	72.43 ± 12.19	98.0	90.1

Note: Results are based on the KNN-retrieved 2022 cases, using the KNN-derived `pred_label` as the target class. The data were stratified into 80% training and 20% validation subsets, with five-fold cross-validation performed on the training subset for SVM configuration selection using Macro-F1. CV Macro-F1 is reported as the mean ± standard deviation across the five folds, whereas Validation Accuracy and Macro-F1 were calculated on the independent 20% validation subset. Feature standardization was performed within the SVM pipeline. Here, r corresponds to `coef0`, and "scale" denotes the automatically determined RBF kernel coefficient.

Linear kernels were selected for temperature, relative humidity, and solar radiation; polynomial kernels for Wind_{NS} and Wind_{WE}; and an RBF kernel for sunshine hours. The mean CV Macro-F1 ranged from $72.43 \pm 12.19\%$ to $98.4 \pm 0.9\%$. Validation accuracy ranged from 94.1% to 99.6%, while validation Macro-F1 ranged from 90.1% to 99.3% across the six meteorological variables.

Generalization was additionally evaluated using unseen 2023–2025 variogram cases. This external evaluation was distinct from the 2022 internal cross-validation and validation results reported in Table 14. For solar radiation and sunshine hours, the zero-case screening criteria described in Section 2.3.2 were applied, whereas all available valid cases were retained for the other variables. Performance was evaluated over progressively increasing sample sizes using 20 randomized repetitions. Representative results for temperature and solar radiation are shown in Figures 7 and 8, respectively.

For temperature, the linear kernel maintained the highest mean Accuracy and Macro-F1 across the evaluated sample sizes. For solar radiation, the linear kernel also showed the highest mean Macro-F1, while the class-wise F1 scores varied substantially among the variogram-model classes. In particular, the Rational Quadratic class showed lower F1 scores than the other classes in Figure 8. Thus, the external evaluation showed differences in classification performance across meteorological variables, kernels, and variogram-model classes.

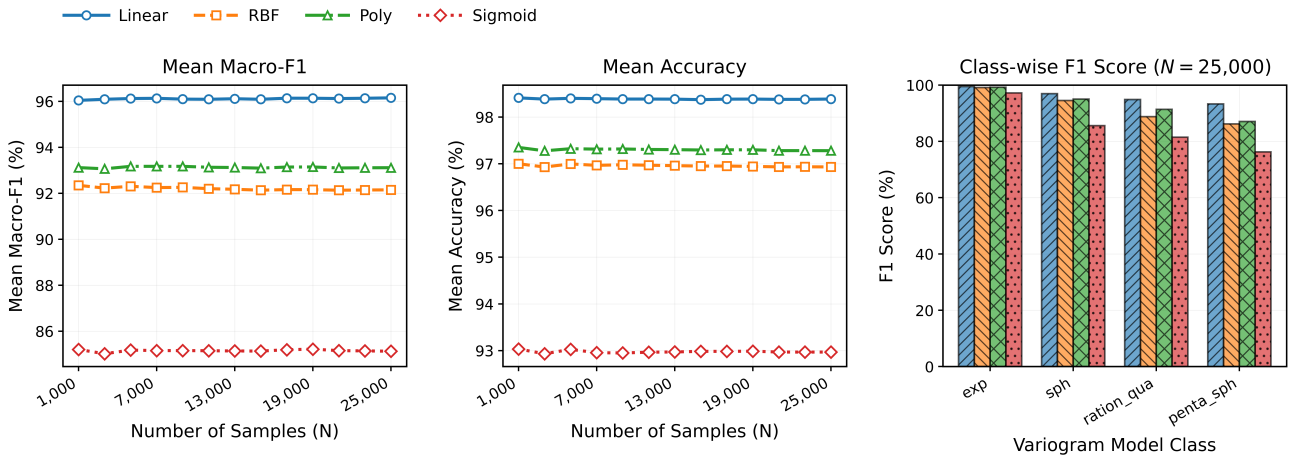


Figure 7. SVM classification performance for temperature across increasing sample sizes of unseen 2023–2025 variogram cases (class-wise F1 scores at $N = 25{,}000$).

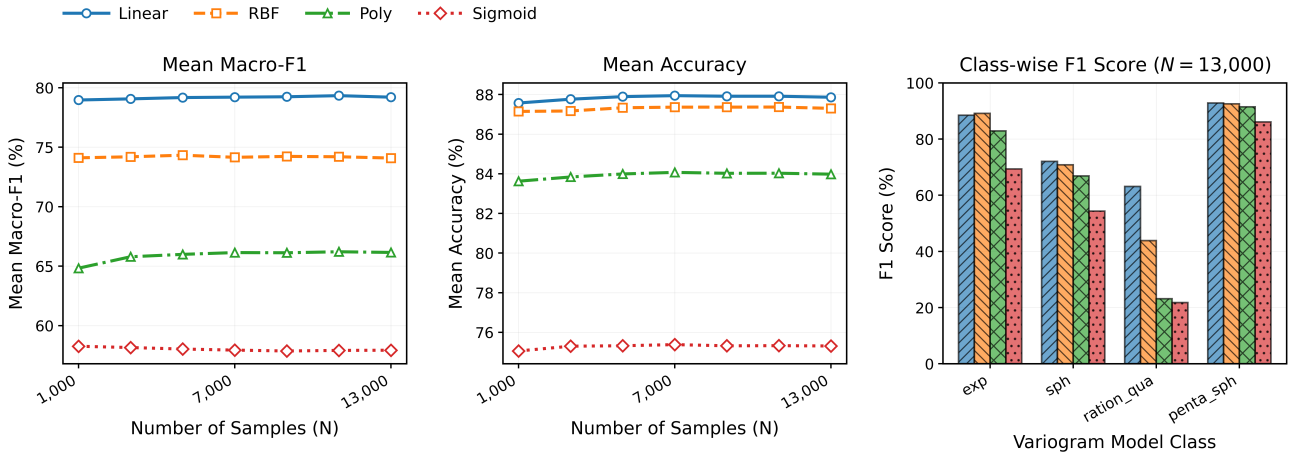


Figure 8. SVM classification performance for solar radiation across increasing sample sizes of unseen 2023–2025 variogram cases (class-wise F1 scores at $N = 13{,}000$).

5.4. Comparison of Meteorological Interpolation Performance

Figures 9–14 compare the interpolation performance of CBR-OK with Direct-OK, NN, IDW, and four Static-OK configurations during the 2023–2025 evaluation period. Monthly RMSE and MAE were calculated from the hourly interpolation errors and are presented chronologically from January 2023 to December 2025. In addition, the hourly win rate was used to quantify how frequently CBR-OK produced a lower interpolation error than each Static-OK configuration. For a given error metric E, the hourly win rate is defined as

$$W_m = \frac{1}{N} \sum_{t=1}^{N} \mathbb{I}(E_{\mathrm{CBR},t} < E_{m,t}) \times 100, \tag{42}$$

where W_m denotes the hourly win rate of CBR-OK against Static-OK model m, N is the total number of valid hourly comparisons, $E_{\mathrm{CBR},t}$ and $E_{m,t}$ denote the corresponding interpolation errors at hour t, and $\mathbb{I}(\cdot)$ is an indicator function equal to 1 when CBR-OK produces the lower error and 0 otherwise. This measure complements the monthly error profiles by quantifying the frequency with which CBR-OK produced a lower hourly error than each fixed variogram configuration.

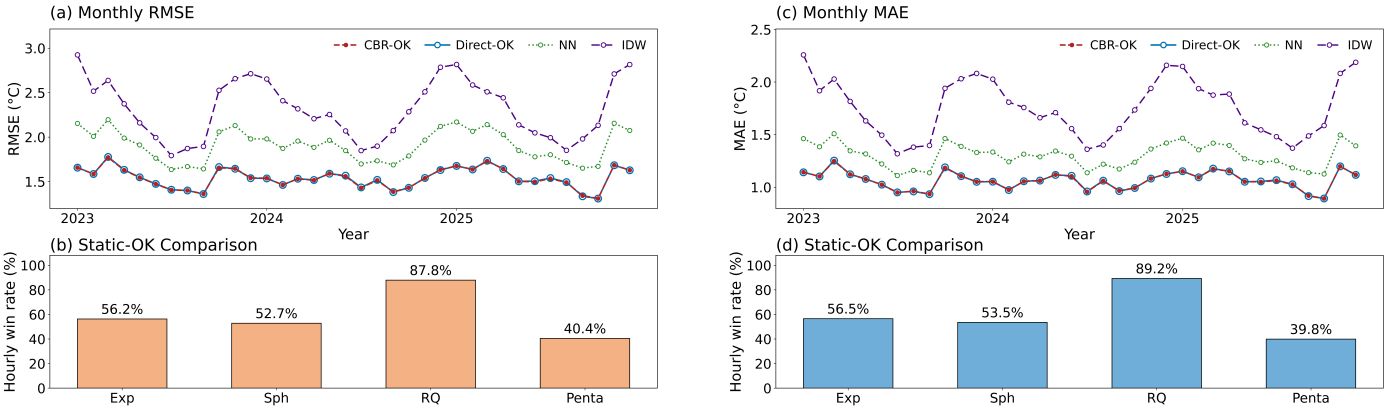


Figure 9. Monthly temperature interpolation performance during 2023–2025: (**a**) RMSE, (**b**) RMSE-based hourly win rates against Static-OK models, (**c**) MAE, and (**d**) MAE-based hourly win rates against Static-OK models.

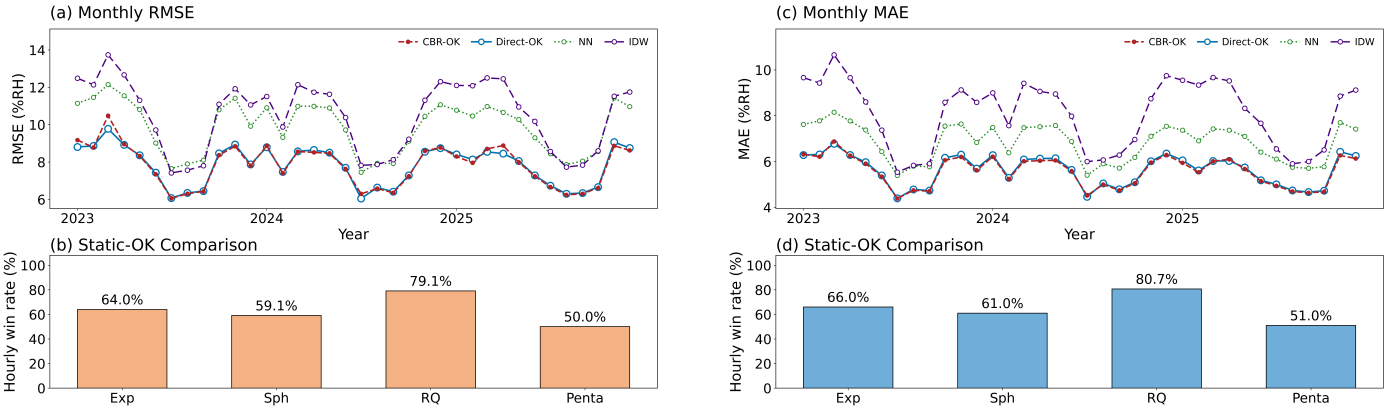


Figure 10. Monthly relative humidity interpolation performance during 2023–2025: (**a**) RMSE, (**b**) RMSE-based hourly win rates against Static-OK models, (**c**) MAE, and (**d**) MAE-based hourly win rates against Static-OK models.

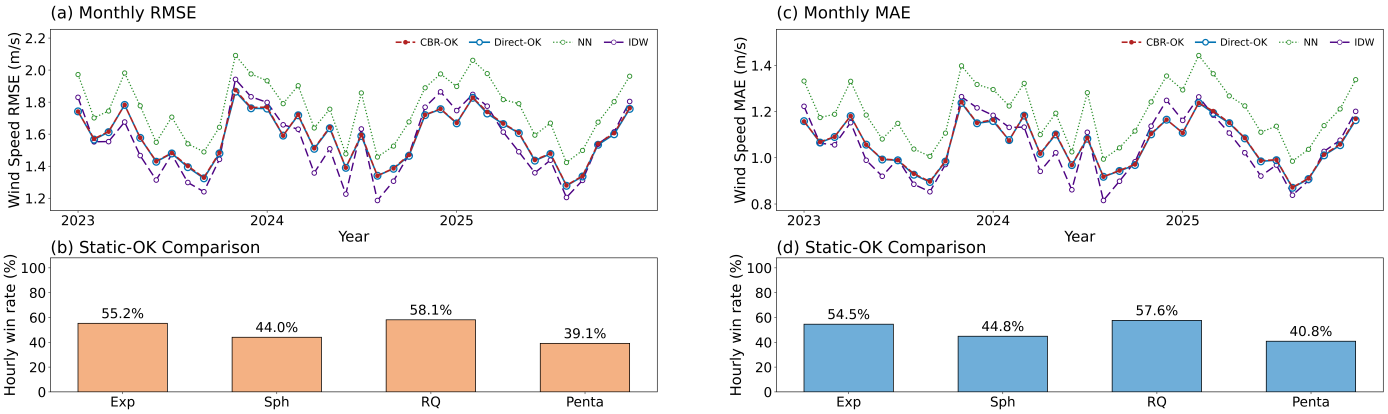


Figure 11. Monthly wind speed interpolation performance during 2023–2025: (**a**) RMSE, (**b**) RMSE-based hourly win rates against Static-OK models, (**c**) MAE, and (**d**) MAE-based hourly win rates against Static-OK models.

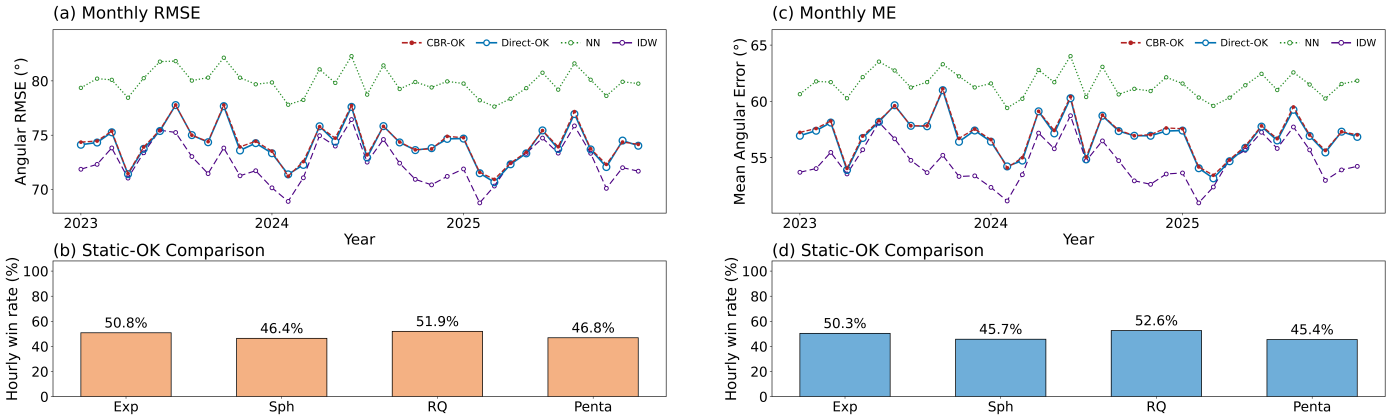


Figure 12. Monthly wind direction interpolation performance during 2023–2025: (**a**) angular RMSE, (**b**) angular RMSE-based hourly win rates against Static-OK models, (**c**) angular MAE, and (**d**) angular MAE-based hourly win rates against Static-OK models.

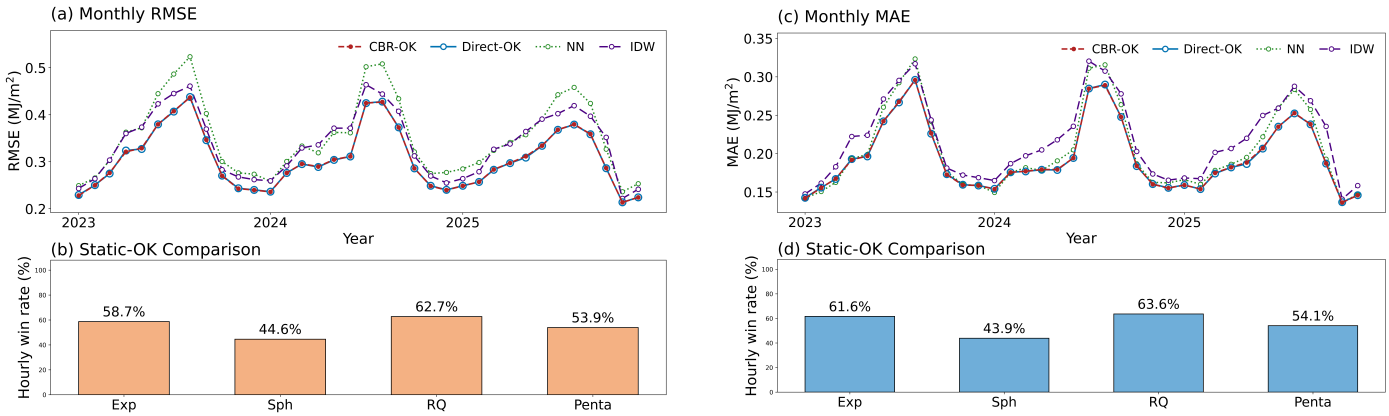


Figure 13. Monthly solar radiation interpolation performance during 2023–2025: (**a**) RMSE, (**b**) RMSE-based hourly win rates against Static-OK models, (**c**) MAE, and (**d**) MAE-based hourly win rates against Static-OK models.

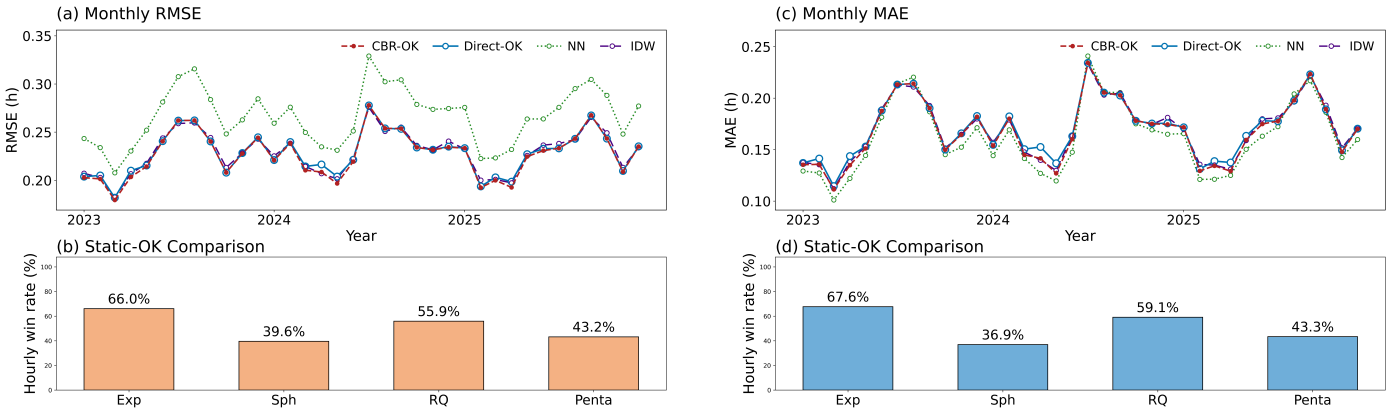


Figure 14. Monthly sunshine duration interpolation performance during 2023–2025: (**a**) RMSE, (**b**) RMSE-based hourly win rates against Static-OK models, (**c**) MAE, and and (**d**) MAE-based hourly win rates against Static-OK models.

For temperature (Figure 9), CBR-OK produced monthly RMSE values of approximately 1.3–1.8 °C and MAE values of approximately 0.9–1.3 °C. Its monthly error profile was lower than those of NN and IDW, while CBR-OK and Direct-OK showed nearly overlapping trajectories. The CBR-OK hourly win rates against Static-OK ranged from 40.4% to 87.8% for RMSE and from 39.8% to 89.2% for MAE. The highest win rates were observed against Static-

RQ, reaching 87.8% and 89.2% for RMSE and MAE, respectively, whereas the corresponding rates against Static-Penta were 40.4% and 39.8%.

For relative humidity (Figure 10), CBR-OK yielded monthly RMSE and MAE values of approximately 6.0–10.4 %RH and 4.4–6.8 %RH, respectively. CBR-OK and Direct-OK showed similar temporal patterns and generally lower errors than NN and IDW. The hourly win rates against Static-OK ranged from 50.0% to 79.1% for RMSE and from 51.0% to 80.7% for MAE. The highest rates were observed against Static-RQ.

For wind speed (Figure 11), CBR-OK yielded monthly RMSE values of approximately 1.3–1.9 m/s and MAE values of approximately 0.9–1.2 m/s. CBR-OK generally produced lower monthly errors than NN, whereas IDW produced lower errors than CBR-OK during several periods. The hourly win rates against Static-OK ranged from 39.1% to 58.1% for RMSE and from 40.8% to 57.6% for MAE.

Wind direction was evaluated separately using circular error measures because conventional linear differences do not account for the periodicity of angular observations. The angular difference at hour t was calculated as

$$e_{\theta,t} = \min\left(|\hat{\theta}_t - \theta_t|, 360^\circ - |\hat{\theta}_t - \theta_t|\right), \tag{43}$$

from which the angular RMSE and mean angular error were calculated as

$$\text{Angular RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_{\theta,t}^2}, \tag{44}$$

$$\text{Mean Angular Error} = \frac{1}{N}\sum_{t=1}^{N} e_{\theta,t}, \tag{45}$$

respectively.

As shown in Figure 12, CBR-OK produced monthly angular RMSE values of approximately 71–78° and mean angular errors of approximately 53–61°. NN showed higher angular errors than CBR-OK, whereas IDW produced lower errors during several periods. The CBR-OK hourly win rates against Static-OK ranged from 46.4% to 51.9% for angular RMSE and from 45.4% to 52.6% for mean angular error. Thus, the frequency with which CBR-OK produced lower errors than the individual Static-OK configurations was close to 50% for wind direction.

For solar radiation (Figure 13), CBR-OK yielded monthly RMSE values of approximately 0.21–0.44 MJ/m^2 and MAE values of approximately 0.14–0.30 MJ/m^2. CBR-OK and Direct-OK showed nearly overlapping monthly profiles and generally lower errors than NN and IDW. The hourly win rates of CBR-OK against Static-OK ranged from 44.6% to 62.7% for RMSE and from 43.9% to 63.6% for MAE. The win rates therefore varied across the four fixed variogram configurations.

For sunshine hours (Figure 14), CBR-OK produced monthly RMSE and MAE values of approximately 0.18–0.28 h and 0.11–0.23 h, respectively. CBR-OK generally produced lower RMSE than NN, while the differences among CBR-OK, Direct-OK, and IDW varied across the evaluation period. The hourly win rates against Static-OK ranged from 39.6% to 66.0% for RMSE and from 36.9% to 67.6% for MAE.

Overall, CBR-OK generally produced lower interpolation errors than NN and showed error profiles similar to Direct-OK, particularly for temperature, relative humidity, and solar radiation. Its performance relative to Static-OK varied across meteorological variables and fixed variogram configurations. The hourly win-rate ranges were wider for temperature and relative humidity, whereas the wind-direction win rates remained close to 50%. Therefore, the results do not indicate a uniform improvement of CBR-OK over

every fixed variogram configuration; rather, its relative performance depends on both the meteorological variable and the fixed variogram configuration used for comparison.

To complement the graphical comparisons in Figures 9–14, the annual and overall interpolation errors for all meteorological variables and interpolation methods are provided in Appendix E.

5.5. End-to-End Computational Cost and Scalability Analysis

The runtime analysis measured the complete end-to-end workflow from contemporaneous station observations to the final Ordinary Kriging predictions. For Direct-OK, the measured runtime includes empirical variogram construction, fitting and evaluation of the four candidate variogram models, model selection, and kriging. For CBR-OK, it includes empirical variogram construction, fitting of the four candidate models, construction of the 24-dimensional query vector, feature scaling, SVM classification, case retrieval, and kriging. Thus, candidate variogram fitting is included in the timing boundary of both methods.

All experiments were conducted using an Intel(R) Core(TM) i9-12900KF processor (16 cores and 24 logical processors), 32 GB of RAM, and Python 3.11.3. Each experiment was repeated 10 times under identical conditions. Component-wise runtimes are reported as mean values, while total end-to-end runtimes are reported as mean ± standard deviation. The total runtime is the sum of the corresponding computational components.

The computational overhead of CBR-OK relative to Direct-OK was calculated as

$$\text{Overhead}(\%) = \frac{T_{\text{CBR-OK}} - T_{\text{Direct-OK}}}{T_{\text{Direct-OK}}} \times 100, \tag{46}$$

where $T_{\text{Direct-OK}}$ and $T_{\text{CBR-OK}}$ denote the mean end-to-end computation times of Direct-OK and CBR-OK, respectively.

For temperature, Direct-OK required less computation time at 100, 200, and 400 stations, whereas the mean runtimes were comparable at 300 stations (0.0450 s for Direct-OK and 0.0449 s for CBR-OK). In the temporal analysis, CBR-OK showed an overhead of 53.5–58.2% relative to Direct-OK (Table 15).

Table 15. End-to-end computational scalability and runtime breakdown of Direct-OK and CBR-OK for temperature.

(a) Spatial Scalability and Component-Wise Runtime

Stations	Direct-OK			CBR-OK					
	Fit/Select (s)	Kriging (s)	Total (s)	Query (s)	Scaling (s)	SVM (s)	Retrieval (s)	Kriging (s)	Total (s)
100	0.0054	0.0041	0.0096 ± 0.0035	0.0043	0.0010	0.0013	0.0073	0.0037	0.0175 ± 0.0020
200	0.0067	0.0149	0.0216 ± 0.0101	0.0055	0.0009	0.0013	0.0069	0.0095	0.0241 ± 0.0058
300	0.0082	0.0367	0.0450 ± 0.0383	0.0068	0.0010	0.0014	0.0069	0.0287	0.0449 ± 0.0205
400	0.0095	0.0282	0.0377 ± 0.0252	0.0087	0.0010	0.0013	0.0064	0.0399	0.0573 ± 0.0360

(b) Temporal Scalability

Timestamps	Direct-OK Total (s)	CBR-OK Total (s)	CBR-OK Overhead (%)
100	2.753 ± 0.234	4.344 ± 0.217	57.8
200	5.451 ± 0.266	8.622 ± 0.317	58.2
400	11.239 ± 1.075	17.430 ± 1.059	55.1
600	16.961 ± 1.667	26.160 ± 1.467	54.2
800	22.744 ± 1.982	35.069 ± 2.003	54.2
1000	28.428 ± 2.529	43.632 ± 2.458	53.5

Note: Component-wise values are mean computation times. Total computation times are reported as mean ± standard deviation over 10 repeated runs. For Direct-OK, Fit/Select includes empirical variogram construction, fitting and evaluation of the four candidate variogram models, and variogram-model selection. For CBR-OK, Query includes empirical variogram construction, fitting of the four candidate variogram models, and construction of the 24-dimensional feature vector.

For wind factor, Direct-OK required less mean computation time at 100–300 stations, whereas CBR-OK showed a slightly lower mean runtime at 400 stations (0.0896 s versus 0.0963 s). In the temporal analysis, CBR-OK showed an overhead of 34.5–35.7% relative to Direct-OK (Table 16).

Table 16. End-to-end computational scalability and runtime breakdown of Direct-OK and CBR-OK for wind factor.

(a) Spatial Scalability and Component-Wise Runtime									
Stations	**Direct-OK**			**CBR-OK**					
	Fit/Select (s)	**Kriging (s)**	**Total (s)**	**Query (s)**	**Scaling (s)**	**SVM (s)**	**Retrieval (s)**	**Kriging (s)**	**Total (s)**
100	0.0112	0.0090	0.0203 ± 0.0062	0.0092	0.0018	0.0017	0.0043	0.0097	0.0266 ± 0.0037
200	0.0126	0.0170	0.0296 ± 0.0079	0.0105	0.0018	0.0017	0.0046	0.0328	0.0513 ± 0.0317
300	0.0155	0.0428	0.0583 ± 0.0289	0.0131	0.0018	0.0017	0.0050	0.0396	0.0612 ± 0.0204
400	0.0185	0.0778	0.0963 ± 0.0511	0.0166	0.0018	0.0017	0.0050	0.0646	0.0896 ± 0.0223

(b) Temporal Scalability			
Timestamps	**Direct-OK Total (s)**	**CBR-OK Total (s)**	**CBR-OK Overhead (%)**
100	4.449 ± 0.603	5.998 ± 0.715	34.8
200	8.696 ± 0.693	11.759 ± 0.853	35.2
400	17.368 ± 1.645	23.367 ± 1.608	34.5
600	25.811 ± 2.183	34.846 ± 1.905	35.0
800	34.048 ± 2.403	46.191 ± 2.150	35.7
1000	42.599 ± 3.325	57.775 ± 2.853	35.6

Note: Runtime components and statistical reporting follow the definitions provided in Table 15.

For relative humidity, Direct-OK showed lower mean runtimes at all evaluated spatial scales. In the temporal analysis, CBR-OK showed an overhead of 23.6–30.8% relative to Direct-OK (Table 17).

Table 17. End-to-end computational scalability and runtime breakdown of Direct-OK and CBR-OK for relative humidity.

(a) Spatial Scalability and Component-Wise Runtime									
Stations	**Direct-OK**			**CBR-OK**					
	Fit/Select (s)	**Kriging (s)**	**Total (s)**	**Query (s)**	**Scaling (s)**	**SVM (s)**	**Retrieval (s)**	**Kriging (s)**	**Total (s)**
100	0.0075	0.0041	0.0115 ± 0.0030	0.0064	0.0009	0.0012	0.0046	0.0091	0.0223 ± 0.0045
200	0.0086	0.0143	0.0230 ± 0.0202	0.0074	0.0009	0.0012	0.0049	0.0117	0.0261 ± 0.0123
300	0.0093	0.0144	0.0236 ± 0.0090	0.0081	0.0009	0.0012	0.0050	0.0216	0.0369 ± 0.0175
400	0.0111	0.0195	0.0305 ± 0.0034	0.0099	0.0009	0.0012	0.0036	0.0411	0.0567 ± 0.0179

(b) Temporal Scalability			
Timestamps	**Direct-OK Total (s)**	**CBR-OK Total (s)**	**CBR-OK Overhead (%)**
100	6.300 ± 7.265	7.786 ± 7.197	23.6
200	12.480 ± 10.642	15.654 ± 10.407	25.4
400	25.000 ± 16.387	31.297 ± 16.380	25.2
600	32.979 ± 17.087	42.637 ± 17.043	29.3
800	41.613 ± 17.046	54.434 ± 17.116	30.8
1000	53.370 ± 23.278	69.484 ± 23.379	30.2

Note: Runtime components and statistical reporting follow the definitions provided in Table 15.

Across all three variables, the cumulative runtime increased approximately linearly with the number of processed timestamps.

5.6. Evaluation of PV Power Forecasting Performance

For each of the six PV sites, the PV generation records were temporally aligned with the meteorological inputs obtained using nearest-neighbour interpolation (NN), inverse distance weighting (IDW), and the proposed CBR-OK method. The resulting datasets covered the period from 1 January 2023 to 31 December 2025, comprising 26,304 hourly observations per site.

In the forecasting implementation, records containing missing PV target values were excluded rather than imputed. Examination of the final model input datasets confirmed that no missing PV generation values, duplicate timestamps, or discontinuities in the hourly time index were present. Consequently, no records were removed because of missing PV targets, and no forecasting sequences were affected by missing PV observations or temporal gaps.

The datasets were partitioned chronologically into training, validation, and test subsets at an approximate ratio of 70:15:15, without random shuffling. The training period extended from 1 January 2023 at 00:00 to 6 February 2025 at 03:00, the validation period from 6 February 2025 at 04:00 to 20 July 2025 at 13:00, and the test period from 20 July 2025 at 14:00 to 31 December 2025 at 23:00. The corresponding observation counts were 18,412, 3946, and 3946, respectively, for each PV site, as summarised in Table 18.

Table 18. Chronological data partitioning for the six PV sites.

PV Site	Training	Validation	Test	Total
Cases 1–6	18,412	3946	3946	26,304

Note: Values indicate hourly observation counts per PV site. The same chronological partitioning was applied to all six sites.

Training input–target sequences were constructed from consecutive hourly observations within the training subset using the model-specific lookback length and forecasting horizon. For N_{train} training observations, a lookback length L, and a forecasting horizon h, the number of training sequences was calculated as $N_{\text{train}} - L - h + 1$. The corresponding sequence counts for the 1-, 12-, and 24-h forecasting horizons are summarised in Table 19. Differences in sequence counts arose from the selected lookback lengths and forecasting horizons rather than from missing PV observations or temporal gaps.

Table 19. Model-specific training sequence counts across forecasting horizons.

Case	Model	Lookback (h)	1 h	12 h	24 h
Case 1	LSTM/Bi-LSTM	48	18,364	18,353	18,341
	GRU	24	18,388	18,377	18,365
Case 2	LSTM/Bi-LSTM	48	18,364	18,353	18,341
	GRU	24	18,388	18,377	18,365
Case 3	LSTM	168	18,244	18,233	18,221
	Bi-LSTM/GRU	48	18,364	18,353	18,341
Case 4	LSTM	48	18,364	18,353	18,341
	Bi-LSTM/GRU	24	18,388	18,377	18,365
Case 5	LSTM/GRU	72	18,340	18,329	18,317
	Bi-LSTM	24	18,388	18,377	18,365
Case 6	LSTM	168	18,244	18,233	18,221
	Bi-LSTM	72	18,340	18,329	18,317
	GRU	24	18,388	18,377	18,365

Note: Values represent the number of training input–target sequences generated from 18,412 hourly training observations per site.

LSTM, Bi-LSTM, and GRU models were used to evaluate the effects of NN-, IDW-, and CBR-OK-based meteorological inputs on PV forecasting performance. For each PV

site and forecasting architecture, identical model-selection and training protocols were applied across the three meteorological input datasets. Hyperparameter optimization was performed at the 1-h forecast horizon, and the selected configuration was subsequently fixed and applied to the 1-, 12-, and 24-h forecasting experiments. The selected site- and model-specific configurations are summarized in Table 20.

Table 20. Selected model configurations of the LSTM, Bi-LSTM, and GRU models for the six PV sites.

Model	Case	Lookback (h)	Hidden Units 1	Hidden Units 2	Dropout	Learning Rate	Batch Size
LSTM	Case 1	48	128	16	0.2	0.0003	128
	Case 2	48	128	16	0.1	0.0003	128
	Case 3	168	96	–	0.3	0.0010	64
	Case 4	48	96	32	0.1	0.0003	32
	Case 5	72	96	64	0.4	0.0030	32
	Case 6	168	32	–	0.1	0.0010	32
Bi-LSTM	Case 1	48	128	16	0.1	0.0003	128
	Case 2	48	128	32	0.1	0.0003	128
	Case 3	48	128	32	0.2	0.0010	64
	Case 4	24	96	32	0.1	0.0003	32
	Case 5	24	96	16	0.3	0.0030	32
	Case 6	72	64	–	0.2	0.0003	32
GRU	Case 1	24	128	16	0.2	0.0003	64
	Case 2	24	96	32	0.1	0.0003	32
	Case 3	48	128	32	0.2	0.0003	32
	Case 4	24	96	32	0.1	0.0003	32
	Case 5	72	64	32	0.3	0.0030	64
	Case 6	24	96	32	0.1	0.0003	32

Note: A dash (–) indicates that the second recurrent layer was not used. Hyperparameter optimization was performed at the 1-h forecast horizon using the same model-selection procedure for the NN-, IDW-, and CBR-OK-based meteorological input datasets. The selected configuration for each site and forecasting architecture was subsequently fixed and applied to the 1-, 12-, and 24-h forecasting experiments.

The PV forecasting experiments were implemented in Python 3.13.15 using PyTorch 2.11.0, NumPy 2.1.3, Pandas 2.2.3, scikit-learn 1.6.1, and Optuna 5.0.0. Model training was performed in a GPU-enabled Google Colab environment using an NVIDIA RTX PRO 6000 Blackwell Server Edition GPU with CUDA 13.0. The random seed was fixed at 42 for reproducibility. Hyperparameter optimization was conducted using the Optuna Tree-structured Parzen Estimator (TPE) sampler with 20 trials, with a maximum of 60 epochs per trial and an early-stopping patience of 10 epochs. The final models were trained for up to 120 epochs using the selected configurations.

The hyperparameter search space included lookback windows of 24, 48, 72, and 168 h; first recurrent-layer sizes of 32, 64, 96, and 128 units; optional second recurrent-layer sizes of 16, 32, and 64 units; dropout rates from 0.1 to 0.4; learning rates of 1×10^{-4}, 3×10^{-4}, 1×10^{-3}, and 3×10^{-3}; and batch sizes of 32, 64, and 128.

Information on the specific PV module technology and mounting configuration of the six case-study plants was not available in the public source data. Therefore, plant technology was not used as an input variable or as a criterion for model configuration in the present study.

Although no missing PV target values or temporal gaps were identified in the final forecasting datasets, communication failures and PV plant outages were not explicitly identified or distinguished from other operational conditions in the available PV records. Consequently, their individual effects on forecasting performance could not be separately quantified, and no additional correction specific to communication failures or plant outages was applied. This limitation should be considered when interpreting the forecasting results.

Representative 1-h-ahead hourly PV output profiles are presented in Figure 15. Across Cases 1–6, the predictions generally reproduced the observed diurnal generation patterns, while differences among the NN-, IDW-, and CBR-OK-based inputs were more apparent during rapid changes and high-generation periods. The three input methods showed broadly similar temporal patterns, although deviations from the observed PV power output remained during some rapid transitions and peak generation periods.

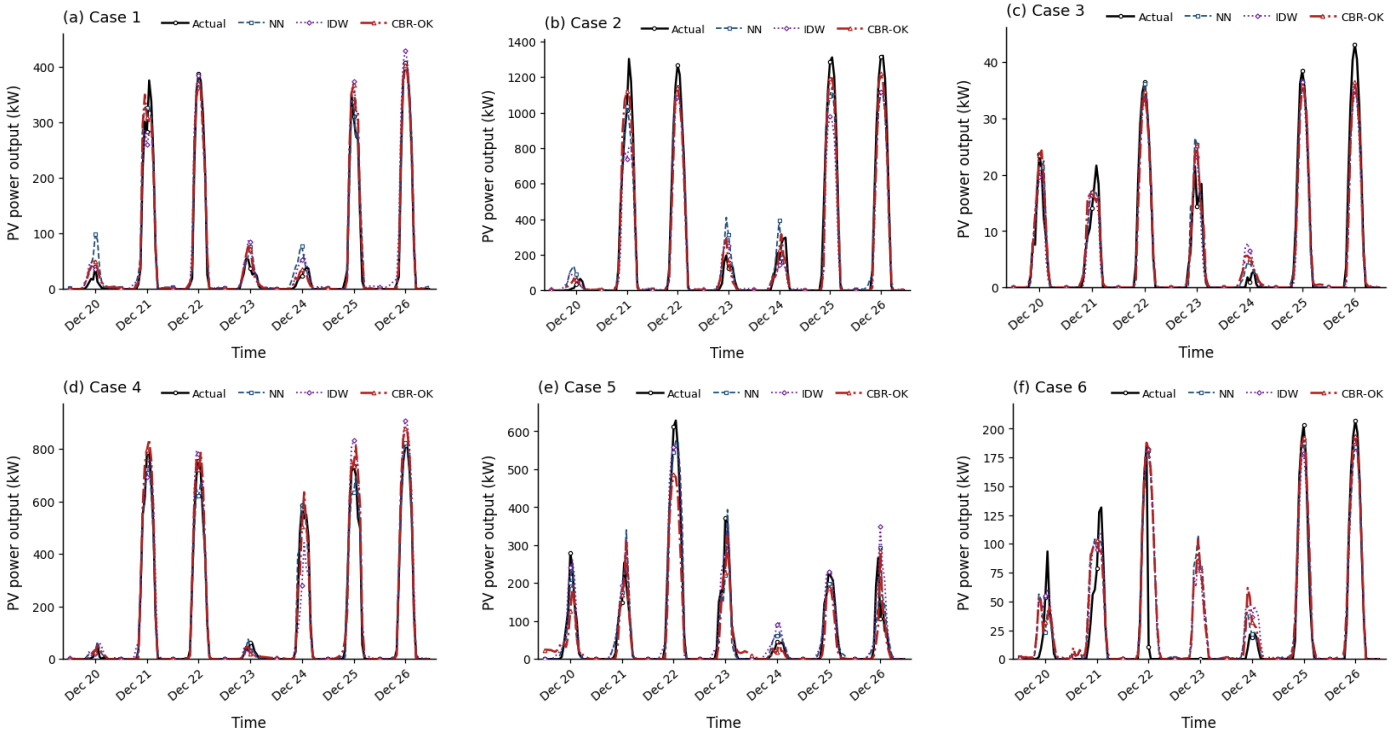


Figure 15. Examples of 1-h-ahead PV power output comparing actual values with predictions based on NN-, IDW-, and CBR-OK-based meteorological inputs: (**a**) Case 1; (**b**) Case 2; (**c**) Case 3; (**d**) Case 4; (**e**) Case 5; and (**f**) Case 6.

To account for differences in plant scale, Table 21 summarizes the site-specific all-hours forecasting performance obtained using CBR-OK-based meteorological inputs. RMSE and MAE are reported in kW, whereas NRMSE and NMAE are normalized by the installed PV capacity. Forecast skill (FS) was evaluated relative to a persistence benchmark. Detailed results for CBR-OK, NN, and IDW at identical test timestamps, together with the installed capacity and persistence errors for each site, model, and forecast horizon, are provided in Appendix F. The appendix reports these metrics separately for the all-hours and daylight evaluations.

Table 21. Site-specific all-hours PV forecasting performance using CBR-OK-based meteorological inputs across the six PV sites.

Case	Model	1 h		12 h		24 h	
		RMSE/MAE (kW)	NRMSE/NMAE/FS (%)	RMSE/MAE (kW)	NRMSE/NMAE/FS (%)	RMSE/MAE (kW)	NRMSE/NMAE/FS (%)
Case 1	LSTM	31.86/15.71	5.75/2.84/38.58	62.10/30.58	11.21/5.52/68.68	68.28/34.06	12.33/6.15/13.76
	Bi-LSTM	32.19/15.32	5.81/2.77/37.94	62.66/29.76	11.31/5.37/68.39	69.00/33.94	12.45/6.13/12.85
	GRU	31.83/15.04	5.75/2.71/38.63	64.42/32.20	11.63/5.81/67.51	67.06/33.19	12.10/5.99/15.30
Case 2	LSTM	97.46/47.80	5.31/2.61/39.81	193.96/95.54	10.58/5.21/69.77	215.74/108.89	11.76/5.94/16.19
	Bi-LSTM	101.95/50.35	5.56/2.75/37.04	194.48/96.44	10.60/5.26/69.69	212.22/107.24	11.57/5.85/17.55
	GRU	96.95/49.77	5.29/2.71/40.12	189.91/102.84	10.36/5.61/70.40	204.63/112.41	11.16/6.13/20.50

Table 21. *Cont.*

Case	Model	1 h		12 h		24 h	
		RMSE/MAE (kW)	NRMSE/NMAE/FS (%)	RMSE/MAE (kW)	NRMSE/NMAE/FS (%)	RMSE/MAE (kW)	NRMSE/NMAE/FS (%)
Case 3	LSTM	4.61/2.32	3.97/2.00/38.18	8.22/4.26	7.09/3.67/73.84	8.02/4.10	6.91/3.53/13.58
	Bi-LSTM	5.11/2.54	4.41/2.19/31.43	8.27/4.30	7.13/3.71/73.67	7.89/4.03	6.80/3.47/14.99
	GRU	4.59/2.50	3.96/2.16/38.37	7.86/4.06	6.78/3.50/74.97	7.95/3.91	6.85/3.37/14.29
Case 4	LSTM	57.97/26.92	5.81/2.70/35.11	104.98/50.98	10.52/5.11/69.95	114.04/57.07	11.43/5.72/23.73
	Bi-LSTM	60.00/27.51	6.01/2.76/32.84	101.48/50.12	10.17/5.02/70.96	112.26/56.10	11.25/5.62/24.92
	GRU	56.58/25.59	5.67/2.56/36.67	103.25/51.77	10.35/5.19/70.45	116.64/57.60	11.69/5.77/21.98
Case 5	LSTM	86.81/50.35	9.76/5.66/−14.46	139.63/77.26	15.71/8.69/54.00	119.66/64.98	13.46/7.31/−6.03
	Bi-LSTM	109.33/68.53	12.30/7.71/−44.15	147.32/84.97	16.57/9.56/51.46	102.47/57.48	11.53/6.47/9.20
	GRU	89.53/53.51	10.07/6.02/−18.05	143.88/103.39	16.18/11.63/52.59	140.08/77.60	15.76/8.73/−24.13
Case 6	LSTM	18.11/8.44	4.89/2.28/30.68	29.46/15.01	7.96/4.06/70.73	27.85/14.17	7.53/3.83/18.66
	Bi-LSTM	18.45/8.78	4.99/2.37/29.39	28.81/14.82	7.79/4.01/71.37	28.29/14.00	7.65/3.78/17.37
	GRU	18.77/9.20	5.07/2.49/28.16	29.83/16.55	8.06/4.47/70.35	28.61/14.17	7.73/3.83/16.43

Note: RMSE and MAE are expressed in kW, whereas NRMSE, NMAE, and FS are expressed as percentages. This table summarizes the all-hours CBR-OK forecasting results. Detailed CBR-OK, NN, and IDW results, including plant capacity, persistence errors, and daylight-only evaluations, are provided in Appendix F.

For an h-hour-ahead forecast issued at time t, the persistence prediction was defined as

$$\hat{P}_{t+h}^{\text{pers}} = P_t, \tag{47}$$

where P_t denotes the observed PV generation at the forecast origin. The RMSE-based forecast skill was calculated as

$$FS_{\text{RMSE}} = \left(1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{pers}}}\right) \times 100. \tag{48}$$

A positive FS indicates that the forecasting model produced a lower RMSE than persistence, whereas a negative FS indicates that its RMSE exceeded the corresponding persistence RMSE.

Absolute forecasting errors differed substantially among the six sites, consistent with their different installed capacities. After normalization, the results also varied by site and forecast horizon. Cases 1–4 and Case 6 generally showed lower NRMSE values at the 1-h horizon than at the 12- and 24-h horizons, whereas Case 5 showed larger normalized errors than the other sites across several configurations.

In particular, Case 5 showed negative forecast skill for all three CBR-OK-based models at the 1-h horizon, with FS values of −14.46%, −44.15%, and −18.05% for LSTM, Bi-LSTM, and GRU, respectively. These negative values indicate that the corresponding model RMSEs exceeded the persistence RMSE on the same test timestamps; they do not indicate negative PV generation predictions. The complete NN-, IDW-, and CBR-OK comparisons, including the daylight-only results, are provided in Appendix F.

To assess the statistical significance of forecasting differences, the Diebold–Mariano (DM) test was applied to paired forecast errors between CBR-OK and the NN and IDW baselines at each PV site. Squared-error and absolute-error loss functions were evaluated separately at a significance level of $p < 0.05$. Table 22 summarizes the results across the six sites, while detailed site-specific DM statistics, p-values, and uncertainty estimates are provided in Appendix G.

Table 22. Summary of Diebold–Mariano tests comparing CBR-OK-based forecasting with NN- and IDW-based forecasting across the six PV sites.

Model	Horizon	Baseline	Squared-Error Loss CBR/NS/Baseline	Absolute-Error Loss CBR/NS/Baseline
LSTM	1 h	NN	5/0/1	5/0/1
		IDW	4/2/0	5/1/0
	12 h	NN	3/2/1	2/2/2
		IDW	0/5/1	0/4/2
	24 h	NN	3/2/1	4/2/0
		IDW	1/2/3	1/2/3
Bi-LSTM	1 h	NN	4/1/1	4/1/1
		IDW	3/2/1	3/1/2
	12 h	NN	3/2/1	3/2/1
		IDW	0/5/1	1/4/1
	24 h	NN	4/1/1	5/1/0
		IDW	0/3/3	1/2/3
GRU	1 h	NN	6/0/0	6/0/0
		IDW	3/2/1	3/2/1
	12 h	NN	2/3/1	3/0/3
		IDW	1/3/2	1/1/4
	24 h	NN	3/2/1	4/1/1
		IDW	1/3/2	2/2/2

Note: Values indicate the number of PV sites classified as CBR-OK significantly better/not significant (NS)/baseline significantly better at $p < 0.05$. Each comparison includes six PV sites. Detailed site-specific DM statistics, p-values, and uncertainty estimates are reported in Appendix G.

The DM-test results showed the most consistent advantage of CBR-OK at the 1-h horizon. Against NN, CBR-OK produced significantly lower squared- and absolute-error losses at five of the six sites for LSTM and at all six sites for GRU. Comparisons with IDW were also generally favorable at the 1-h horizon, although less consistently than those with NN.

At the 12- and 24-h horizons, the DM-test results were more heterogeneous. CBR-OK remained significantly better than NN for several model–site combinations, whereas comparisons with IDW more frequently showed non-significant differences or significantly lower losses for IDW. Overall, the forecasting benefit of CBR-OK was most evident at the shorter horizon and became increasingly dependent on the forecasting architecture, PV site, and reference interpolation method at longer horizons.

5.7. Ablation Analysis of Meteorological Interpolation Methods

To evaluate the contribution of CBR-OK-based meteorological interpolation to downstream PV power forecasting, an ablation analysis was conducted across the six PV sites. CBR-OK-based meteorological inputs were compared with NN- and IDW-based inputs using LSTM, Bi-LSTM, and GRU models at forecast horizons of 1, 12, and 24 h. Five meteorological variable groups were considered: temperature, relative humidity, wind variables, solar radiation, and sunshine hours. For each experiment, one variable group in the NN- or IDW-based input dataset was replaced with the corresponding CBR-OK-based input, while all remaining input variables and forecasting conditions were kept unchanged.

The forecasting effect of replacing a baseline input with the corresponding CBR-OK-based input was quantified using the error reduction rate:

$$ER^{(q)}_{s,b,m,h,v} = \frac{E^{(q)}_{s,b,m,h,v} - E^{(q)}_{s,\text{CBR-OK},m,h,v}}{E^{(q)}_{s,b,m,h,v}} \times 100, \tag{49}$$

where s, b, m, h, v, and q denote the PV site, baseline interpolation method (NN or IDW), forecasting model, forecast horizon, meteorological variable group, and forecasting metric

(RMSE or MAE), respectively. Positive values indicate that replacing the corresponding baseline meteorological input with the CBR-OK-based input reduced the forecasting error, whereas negative values indicate an increase in forecasting error.

The mean error reduction across the six PV sites was calculated as

$$\overline{ER}^{(q)}_{b,m,h,v} = \frac{1}{6}\sum_{s=1}^{6} ER^{(q)}_{s,b,m,h,v}. \tag{50}$$

The number of PV sites showing a positive error reduction was also reported in each heatmap cell.

Figure 16 presents the aggregated ablation results. Relative to NN, solar radiation showed the clearest improvement at the 1 h horizon, with mean RMSE reductions of 5.6%, 4.5%, and 9.8% for LSTM, Bi-LSTM, and GRU, respectively. The corresponding MAE reductions were 7.5%, 5.2%, and 12.8%. Wind variables also showed improvements under selected 12 h configurations, including RMSE reductions of 5.0% and 5.7% for LSTM and Bi-LSTM, respectively, and an MAE reduction of 7.8% for Bi-LSTM.

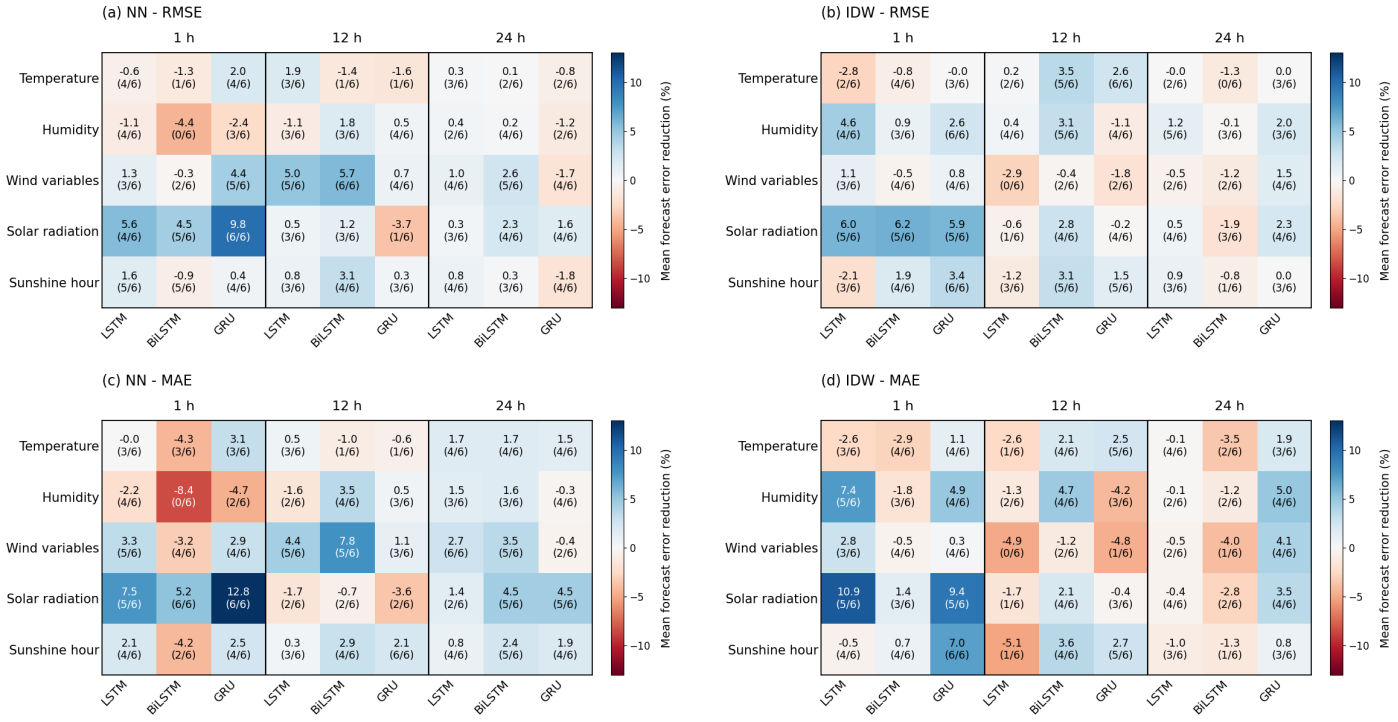


Figure 16. Aggregated ablation results across the six PV sites showing the mean forecast error reduction obtained by replacing NN- or IDW-based meteorological inputs with the corresponding CBR-OK-based inputs: (**a**) NN–RMSE, (**b**) IDW–RMSE, (**c**) NN–MAE, and (**d**) IDW–MAE. Values in parentheses indicate the number of sites with positive error reduction.

Relative to IDW, solar radiation again showed improvements at the 1 h horizon, with RMSE reductions of 6.0%, 6.2%, and 5.9% for LSTM, Bi-LSTM, and GRU, respectively. The corresponding MAE reductions were 10.9%, 1.4%, and 9.4%. At the 12 and 24 h horizons, however, the results became more dependent on the meteorological variable and forecasting architecture, with both positive and negative error reductions observed.

Overall, the ablation results indicate that the forecasting benefit of CBR-OK is variable- and configuration-dependent. The most consistent improvements were observed for solar radiation at the 1 h horizon, whereas wind-related variables showed benefits under selected 12 h configurations. These findings indicate that the effectiveness of meteorological interpolation should be assessed not only in terms of interpolation accuracy but also in terms of its downstream effect on PV power forecasting performance.

5.8. Temporal Attribution Analysis of PV Forecasting Models

To further examine the contribution of recent temporal information to PV power forecasting, a perturbation-based attribution analysis was conducted for the LSTM, Bi-LSTM, and GRU models using the NN-, IDW-, and CBR-OK-based meteorological input datasets at forecast horizons of 1, 12, and 24 h. The most recent 24 h of each input sequence were divided into four non-overlapping lag groups: 1 h, 2–6 h, 7–12 h, and 13–24 h.

For each lag group, the observations were not deleted from the input sequence. Instead, all standardized input features within the corresponding lag interval were replaced with zero, which corresponds to the training-set mean after standardization. Each source/model/forecast-horizon configuration was trained once using the original input data, and the trained model was then evaluated separately after perturbing each lag group. Therefore, no additional model retraining was performed for the individual perturbations. For configurations with lookback windows longer than 24 h, including the maximum 168-h lookback reported in Table 20, lags earlier than 24 h were retained unchanged. Accordingly, the present analysis evaluates the contribution of the most recent 24 h and does not quantify the contribution of the entire lookback window.

The percentage increase in RMSE was calculated as

$$\Delta\mathrm{RMSE}(\%) = \frac{\mathrm{RMSE}_{\mathrm{perturbed}} - \mathrm{RMSE}_{\mathrm{baseline}}}{\mathrm{RMSE}_{\mathrm{baseline}}} \times 100, \tag{51}$$

where $\mathrm{RMSE}_{\mathrm{baseline}}$ is obtained from the original unperturbed input sequence and $\mathrm{RMSE}_{\mathrm{perturbed}}$ is obtained after replacing the standardized features within the corresponding lag group with zero. A positive value indicates an increase in forecasting error after perturbation, whereas a negative value indicates that the perturbed configuration produced a lower RMSE than the corresponding unperturbed baseline. Results were summarized across the six PV sites using the mean percentage RMSE increase and the number of sites showing a positive increase.

Regarding the physical constraints on PV forecasts, negative predictions were clipped to zero to enforce non-negativity, whereas no upper clipping based on the installed PV capacity was applied. Examination of the reported predictions for all six case-study sites confirmed that the maximum predicted PV generation at each site remained below its corresponding installed capacity; thus, no upper-bound violations were observed.

Figure 17 shows that the effect of perturbing recent temporal information varied with forecast horizon and lag range. At the 1-h horizon, perturbation of the 1-h and 2–6-h lag groups generally produced larger RMSE increases than perturbation of the 7–12-h and 13–24-h groups across the three forecasting architectures. At the 12-h horizon, the effects were distributed across a broader range of recent lags, whereas at the 24-h horizon, the RMSE changes were generally smaller than the largest changes observed at the 1-h horizon. The 2–6-h lag group nevertheless produced positive RMSE increases in multiple configurations at both longer forecast horizons.

Overall, the contribution of recent temporal information depended on both forecast horizon and lag range, with stronger effects generally observed for more recent lags in short-term forecasting. Similar patterns were observed across the NN-, IDW-, and CBR-OK-based inputs. Because only the most recent 24 h were perturbed while earlier lags were retained unchanged, these results should be interpreted as an attribution analysis of the recent 24-h history rather than of the complete lookback window.

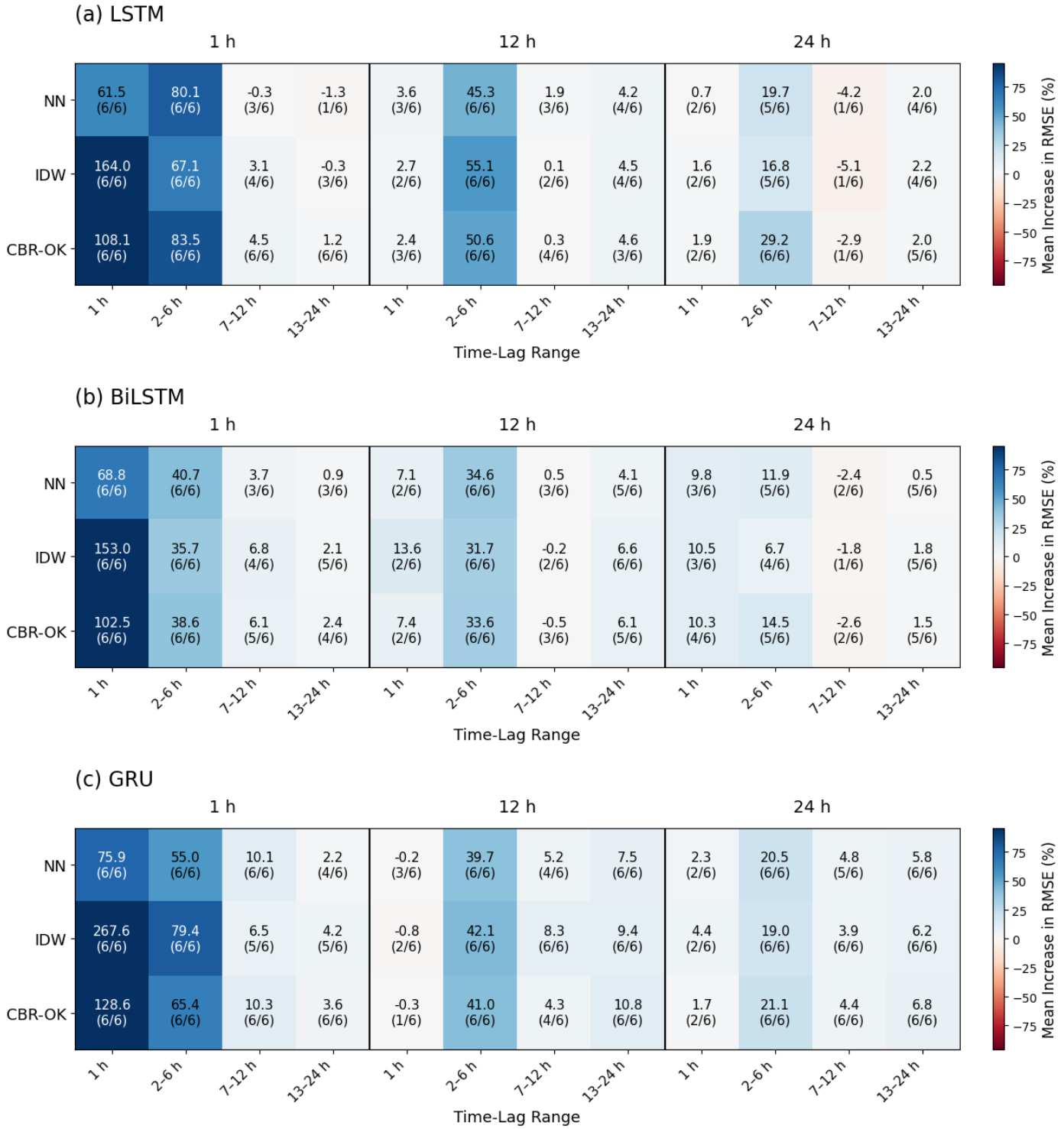


Figure 17. Temporal attribution analysis across the six PV sites for (**a**) LSTM, (**b**) Bi-LSTM, and (**c**) GRU using NN-, IDW-, and CBR-OK-based meteorological inputs at the 1-, 12-, and 24-h forecast horizons. Each cell represents the mean percentage increase in RMSE after replacing all standardized input features within the corresponding lag group with zero. Values in parentheses indicate the number of PV sites showing a positive RMSE increase out of the six evaluated sites.

6. Conclusions

This study proposed CBR-OK, a meteorological spatial interpolation framework that integrates case-based variogram model inference with Ordinary Kriging (OK), and evaluated its applicability to meteorological interpolation at unobserved locations and sub-

sequent PV power forecasting. The interpolation results showed that CBR-OK generally maintained accuracy comparable to Direct-OK, although its relative performance varied across meteorological variables and comparison methods. The end-to-end computational analysis showed that the current CBR-OK implementation did not provide a consistent runtime advantage over Direct-OK because construction of the query feature vector still requires fitting the candidate variogram models. These results indicate that the contribution of CBR-OK lies primarily in adaptive case-based variogram model selection rather than in reducing end-to-end computation time.

The downstream PV forecasting results further demonstrated that the effect of CBR-OK-based meteorological inputs depended on the forecasting model, PV site, and forecast horizon. The statistical comparisons showed the most consistent forecasting benefit at the 1 h horizon, particularly relative to NN, whereas the results at the 12 and 24 h horizons were more heterogeneous. The ablation analysis further showed that solar radiation produced the most consistent improvements at the 1 h horizon, while wind-related variables contributed to error reductions under selected 12 h configurations. In addition, the temporal attribution analysis indicated that, within the most recent 24 h of the input history, the 1 h and 2–6 h lag ranges generally had stronger effects on short-term forecasting than the more distant lag ranges. At longer forecast horizons, the effects of individual lag ranges became more dependent on the forecasting configuration.

Several limitations provide directions for future research. The current CBR-OK implementation introduces additional computational overhead because query-feature construction requires fitting the candidate variogram models, followed by feature scaling, SVM classification, case retrieval, and kriging. Developing more lightweight query representations that reduce the cost of candidate-model fitting could therefore improve computational efficiency. As the CBR case database expands, retrieving similar cases from the entire database may further increase computational cost. Cases with similar variogram characteristics could therefore be organized using unsupervised clustering methods such as self-organizing maps (SOMs), allowing retrieval to be restricted to relevant groups [44].

The kriging framework adopted in this study primarily represents spatial dependence and does not explicitly model temporal dependence jointly with spatial correlation. Future work could extend CBR-OK to spatio-temporal kriging using covariance structures that jointly represent spatial and temporal dependence [45]. More broadly, temporal dynamics could be incorporated explicitly into the interpolation framework to examine how evolving meteorological conditions affect case retrieval and spatial estimation over time. In addition, the spatial search strategy could be refined by restricting interpolation to stations within a variogram-derived search neighborhood, thereby reducing the influence and computational burden of distant stations with limited spatial relevance. Graph-based learning methods may provide a further extension by representing relationships among observation stations beyond conventional distance- and variogram-based spatial dependence [46].

The PV forecasting evaluation was conducted using six PV sites. Although the results demonstrate the applicability of the proposed framework across the evaluated sites, future studies should include a larger and more diverse set of PV plants to assess the robustness and generalizability of the forecasting results. Furthermore, the current case database was constructed from a predefined historical period. An online CBR framework in which newly evaluated cases are progressively retained in the case database could therefore be investigated, enabling the case knowledge base to adapt to new meteorological conditions over time.

Because the forecasting benefit of CBR-OK was most consistent at the 1 h horizon and became more configuration-dependent at the 12 and 24 h horizons, its applicability to longer forecast horizons should be further investigated using architectures designed for long-term time-series forecasting, such as PatchTST [47]. Finally, the present analysis

was limited to meteorological observations and PV sites in South Korea. Future studies should therefore evaluate CBR-OK under different geographical and climatic conditions to examine the transferability of the case-retrieval and variogram-selection framework beyond the study region. These extensions could improve the computational scalability, adaptive capability, and spatio-temporal representation of CBR-OK while broadening its applicability to PV power forecasting.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; software, M.K.; validation, M.K.; formal analysis, M.K.; investigation, M.K.; data curation, M.K.; visualization, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K. and W.L.; supervision, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Climate, Energy & Environment (MCEE) of the Republic of Korea (No. RS-2024-00398166).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The meteorological data used in this study are publicly available from the Korea Meteorological Administration (KMA) Data Open METEO Portal, and the PV power generation data are publicly available through the Korean Public Data Portal. Detailed data sources, dataset identifiers, PV site information, data sources for geographical coordinates, and the procedure used to identify the PV site locations are provided in Appendix H.

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

Appendix A. Variogram Fitting and Kriging Implementation Details

Additional implementation details are provided to clarify the variogram-fitting and kriging procedures used in this study. Because the fitting criteria and optimization settings differed across meteorological variable groups, the corresponding configurations are reported separately. For all variables, the maximum lag distance was restricted to 50% of the maximum inter-station distance.

For temperature and relative humidity, the empirical semivariogram was constructed using 20 lag bins. A lag bin was retained when at least 30 station pairs were available, and model fitting required at least eight valid lag bins and 30 valid stations. Variogram parameters were estimated using weighted least squares with the L-BFGS-B optimization algorithm. The optimization tolerances were set to $f_{\text{tol}} = 1 \times 10^{-9}$ and $g_{\text{tol}} = 1 \times 10^{-6}$, with a maximum of 500 iterations.

For Wind_{NS} and Wind_{WE}, the empirical semivariogram was also constructed using 20 lag bins. A lag bin was retained when at least one station pair was available, and model fitting required at least three valid lag bins and five valid stations. Variogram parameters were estimated using weighted least squares and optimized using L-BFGS-B with a maximum of 200 iterations. Consistent with the spatial-distance formulation described in Section 2.3.3, variogram fitting was performed under an isotropic assumption.

For solar radiation and sunshine hours, 15 lag bins were used after applying the zero-case screening procedure described in Section 2.3.2. Sensitivity to the number of lag bins was additionally evaluated using 10-, 15-, and 20-bin configurations, as reported in Section 2.3.2.

For all variables, the distance parameter was expressed using a common practical-range interpretation where applicable. In the kriging implementation, the semivariogram diagonal was explicitly defined as $\gamma(0) = 0$, while the nugget effect was represented as

a discontinuity at positive separation distances ($h > 0$). The principal implementation settings are summarized in Table A1.

Table A1. Variogram-fitting and kriging implementation settings used in this study.

Variable Group	Item	Setting
Temperature/ Relative Humidity	Number of lag bins	20
	Maximum lag distance	0.5× maximum inter-station distance
	Minimum pairs per retained lag bin	30
	Minimum valid lag bins	8
	Minimum valid stations	30
	ϵ	1×10^{-12}
	Optimizer	L-BFGS-B
	f_{tol}	1×10^{-9}
	g_{tol}	1×10^{-6}
	Maximum iterations	500
Wind_{NS}/ Wind_{WE}	Number of lag bins	20
	Minimum pairs per retained lag bin	1
	Minimum valid lag bins	3
	Minimum valid stations	5
	Optimizer	L-BFGS-B
	Maximum iterations	200
Solar Radiation/ Sunshine Hours	Number of lag bins	15
	Zero-case processing	Section 2.3.2
All Variables	Distance-parameter interpretation	Common practical-range definition
	Kriging diagonal	$\gamma(0) = 0$
	Nugget treatment	Discontinuity for $h > 0$

Note: Temperature, relative humidity, Wind_{NS}, and Wind_{WE} were modeled using 20 lag bins, whereas solar radiation and sunshine hours were modeled using 15 lag bins following zero-case screening. The sensitivity of variogram fitting for the solar variables to alternative lag-bin configurations is reported in Section 2.3.2.

Appendix B. Robustness Analysis of Variogram Model Selection

To examine the robustness of the rank-based variogram model-selection criterion defined in Equation (6), additional analyses were conducted using spatial cross-validation, alternative lag-bin configurations, and bootstrap uncertainty estimation. These analyses focused on air temperature, relative humidity, Wind_{NS}, and Wind_{WE}, for which 20 lag bins were used in the main analysis, as summarized in Table A1. Solar radiation and sunshine hours, for which 15 lag bins were used following zero-case screening, were not included in this appendix because their sensitivity to alternative lag-bin configurations was evaluated separately in Section 2.3.2.

The purpose of this additional analysis was not to redefine the model-selection rule in Equation (6) but to examine whether the rank-based fitting criterion produced spatial interpolation performance comparable to that obtained when the candidate variogram models were evaluated directly using spatial cross-validation.

Appendix B.1. Spatial Cross-Validation Performance

Table A2 summarizes the spatial cross-validation performance of the four candidate variogram models. For each variable, the table reports the mean and median RMSE, mean MAE, and mean R^2 obtained from the spatially withheld predictions.

Table A2. Spatial cross-validation performance of the candidate variogram models for temperature, relative humidity, Wind_{NS}, and Wind_{WE}.

Variable	Variogram	Mean RMSE	Median RMSE	Mean MAE	Mean R^2	Valid
Temperature	Spherical	2.379	2.308	1.749	0.380	2400
	Penta-spherical	2.379	2.356	1.763	0.372	2400
	Exponential	2.413	2.349	1.782	0.362	2400
	Rational quadratic	2.843	2.778	2.098	0.126	2400
Relative humidity	Exponential	11.032	9.817	8.464	0.183	2400
	Spherical	11.039	9.773	8.449	0.182	2400
	Penta-spherical	11.046	9.784	8.456	0.180	2400
	Rational quadratic	11.105	9.956	8.508	0.175	2400
Wind_{NS}	Exponential	1.390	1.350	0.986	−0.158	1800
	Spherical	1.400	1.345	0.998	−0.180	1800
	Penta-spherical	1.406	1.382	1.001	−0.181	1800
	Rational quadratic	3.499	1.450	2.580	−0.309	1800
Wind_{WE}	Exponential	1.344	1.266	0.956	−0.130	1800
	Spherical	1.353	1.292	0.967	−0.155	1800
	Penta-spherical	1.361	1.299	0.974	−0.170	1800
	Rational quadratic	1.384	1.306	0.994	−0.207	1800

Note: RMSE and MAE are expressed in the native units of each meteorological variable. Temperature errors are expressed in °C, relative-humidity errors in %RH, and Wind_{NS} and Wind_{WE} errors in m/s. "Valid" denotes the total number of successful spatial cross-validation predictions.

The spatial cross-validation results indicate that the relative performance of the candidate variogram models depended on the meteorological variable. For temperature, the Spherical and Penta-spherical models produced identical mean RMSE values, whereas the Exponential model yielded the lowest mean RMSE for relative humidity, Wind_{NS}, and Wind_{WE}. The differences among the leading models were small for relative humidity and the two wind components. In contrast, the Rational Quadratic model showed substantially larger errors for Wind_{NS} and yielded the lowest mean R^2 among the candidate models. Negative mean R^2 values were observed for both wind components, indicating that R^2 should be interpreted cautiously as a standalone spatial cross-validation metric. Accordingly, RMSE and MAE were considered as the primary spatial cross-validation metrics when assessing the robustness of the fitting-based model-selection criterion.

Appendix B.2. Comparison Between Fit-Rank and Spatial-CV-Based Selection

To directly evaluate whether the model selected by the rank-based fitting criterion also corresponded to the model yielding the lowest spatial cross-validation RMSE, the fit-rank selection was compared with the model yielding the lowest spatial-CV RMSE at 12 representative timestamps for each variable. Table A3 summarizes the agreement rate and the difference in spatial-CV RMSE between the two selections.

Table A3. Comparison between fit-rank-based and spatial-CV-based variogram selection at representative timestamps.

Variable	Same Model	Agreement (%)	Mean RMSE Difference (%)	Maximum RMSE Difference (%)
Temperature	5/12	41.7	1.525	9.331
Relative humidity	0/12	0.0	1.503	5.679
Wind_{NS}	4/12	33.3	0.819	2.657
Wind_{WE}	5/12	41.7	0.813	2.592

Note: The RMSE difference represents the relative increase in spatial cross-validation RMSE of the model selected by the fitting-based rank criterion compared with the model yielding the minimum spatial-CV RMSE at the same timestamp.

The model identity selected by the fitting-based criterion did not always coincide with that selected by spatial cross-validation. In particular, no identical selections were obtained for relative humidity at the 12 representative timestamps. However, disagreement in model identity was associated with only small differences in spatial prediction accuracy. The mean RMSE differences were approximately 1.53% for temperature, 1.50% for relative humidity, 0.82% for Wind_{NS}, and 0.81% for Wind_{WE}. These results indicate that fitting-based and prediction-based criteria may select different variogram forms while producing similar kriging errors at the evaluated timestamps.

Appendix B.3. Comparison of the Final SVM-Selected Configuration with the Spatial-CV-Best Configuration

To determine whether the final SVM-based model selection was consistent with the configuration identified directly from spatial cross-validation, the SVM-selected variogram model was compared with the spatial-CV-best model at the same 12 representative timestamps used in the preceding robustness analyses. The spatial-CV-best model denotes the candidate variogram model yielding the lowest spatial cross-validation RMSE at each timestamp.

Because the SVM-selected model could differ among spatial-validation folds, the modal SVM-selected model across the folds was used only to summarize model-selection agreement. Spatial prediction performance, however, was calculated using the fold-specific SVM-selected models. Therefore, the model-selection agreement and spatial-error comparison represent two related but distinct aspects of the evaluation. The comparison was performed using the spatial-CV RMSE obtained from the SVM-CBR-OK procedure and the RMSE of the spatial-CV-best configuration. The relative RMSE difference was calculated as

$$\Delta\text{RMSE}(\%) = \frac{\text{RMSE}_{\text{SVM-CBR}} - \text{RMSE}_{\text{CV-best}}}{\text{RMSE}_{\text{CV-best}}} \times 100 \quad \text{(A1)}$$

Table A4 summarizes the comparison for air temperature, relative humidity, Wind_{NS}, and Wind_{WE}.

Table A4. Comparison of the final SVM-selected variogram configuration with the spatial-CV-best configuration at the 12 representative timestamps.

Variable	Same Model	Agreement (%)	Mean RMSE Difference (%)	RMSE Difference Range (%)
Temperature	4/12	33.3	2.46	−2.17 to 8.59
Relative humidity	1/12	8.3	0.39	−3.99 to 2.20
Wind_{NS}	4/12	33.3	7.85	−11.83 to 17.58
Wind_{WE}	5/12	41.7	10.92	−10.61 to 32.57

Note: Same Model denotes the number of the 12 representative timestamps at which the modal SVM-selected variogram model across the spatial-validation folds was identical to the spatial-CV-best model. Agreement is the corresponding percentage. Mean RMSE Difference and RMSE Difference Range were calculated from the timestamp-specific relative RMSE differences defined in Equation (A1). Negative values indicate timestamps at which the SVM-CBR-OK spatial-CV RMSE was lower than the RMSE of the spatial-CV-best configuration. Spatial prediction errors for SVM-CBR-OK were calculated using the fold-specific SVM-selected models rather than the modal model reported for agreement.

The modal SVM-selected model was identical to the spatial-CV-best model at 4 of 12 timestamps for temperature, 1 of 12 for relative humidity, 4 of 12 for Wind_{NS}, and 5 of 12 for Wind_{WE}. Despite the relatively low model-selection agreement, the mean relative RMSE differences were 2.46% for temperature and 0.39% for relative humidity, indicating limited differences in spatial prediction error for these variables. Larger mean differences were observed for Wind_{NS} and Wind_{WE}, at 7.85% and 10.92%, respectively, with the largest timestamp-specific difference occurring for Wind_{WE} (32.57%). These results indicate

that agreement in variogram-model identity does not necessarily correspond directly to agreement in spatial prediction error and that the discrepancy between the SVM-selected and spatial-CV-best configurations was more pronounced for the wind components than for temperature and relative humidity.

Appendix B.4. Sensitivity to the Number of Lag Bins

For air temperature, relative humidity, Wind_{NS}, and Wind_{WE}, the sensitivity of variogram-model selection to the number of lag bins was examined using 10, 15, 20, and 25 bins. The 20-bin configuration was used as the reference, and agreement in the selected variogram model was evaluated at the same 12 representative timestamps for each variable. The resulting agreement rates for the alternative lag-bin configurations relative to the 20-bin reference are summarized in Table A5.

Table A5. Sensitivity of variogram-model selection to the number of lag bins relative to the 20-bin reference configuration.

Variable	10 vs. 20 bins (%)	15 vs. 20 bins (%)	25 vs. 20 bins (%)
Temperature	91.7	83.3	41.7
Relative humidity	83.3	83.3	83.3
Wind_{NS}	41.7	58.3	83.3
Wind_{WE}	91.7	91.7	100.0

Note: Values indicate the percentage of the 12 representative timestamps for which the variogram model selected using the alternative lag-bin setting was identical to that obtained using the 20-bin reference configuration.

The effect of the lag-bin configuration was variable-dependent. Temperature showed high agreement under the 10- and 15-bin configurations but lower agreement when the number of bins was increased to 25. In contrast, Wind_{WE} showed high agreement across all alternative settings, including complete agreement for 25 bins. Wind_{NS} showed lower agreement for 10 and 15 bins but increased to 83.3% for 25 bins, whereas relative humidity remained at 83.3% for all three alternative configurations. Overall, these results do not indicate a consistent monotonic relationship between the number of lag bins and model-selection stability across the meteorological variables.

Appendix B.5. Bootstrap Uncertainty of Fitted Variogram Parameters

Bootstrap uncertainty was evaluated for the fitted variogram parameters at 12 representative timestamps in 2024. The timestamps were selected at approximately monthly intervals across the year and at different hours of the day to provide temporal coverage across seasons and times of day, as listed in Table A6.

Table A6. Point estimates and empirical 95% bootstrap intervals of the fitted variogram parameters at the 12 representative timestamps in 2024.

Variable	Timestamp	Nugget	Total Sill	Distance Parameter
Temperature	2024-01-01 00:00	2.933 [0.000, 3.387]	13.474 [6.568, 14.234]	1032.287 [176.639, 1032.287]
	2024-02-03 06:00	0.844 [0.000, 1.493]	24.286 [10.802, 26.506]	1032.286 [176.640, 1032.286]
	2024-03-07 12:00	0.871 [0.124, 2.617]	7.779 [7.569, 8.887]	176.641 [176.641, 176.641]
	2024-04-09 19:00	0.701 [0.000, 2.195]	14.101 [13.575, 15.111]	171.377 [157.555, 208.149]
	2024-05-13 01:00	4.334 [3.444, 4.909]	8.131 [7.623, 8.726]	398.438 [176.640, 595.599]
	2024-06-15 08:00	1.858 [0.023, 2.085]	7.832 [3.983, 8.172]	1032.279 [176.640, 1032.279]
	2024-07-18 14:00	0.765 [0.007, 0.883]	17.693 [4.805, 14.313]	1032.307 [176.639, 817.515]
	2024-08-20 21:00	2.392 [2.061, 2.805]	4.032 [3.930, 4.196]	176.641 [176.641, 176.641]
	2024-09-23 03:00	0.732 [0.000, 1.051]	16.509 [6.895, 18.404]	1032.314 [176.638, 1032.314]
	2024-10-26 10:00	2.282 [0.923, 3.249]	7.952 [7.374, 9.004]	287.052 [176.632, 475.224]
	2024-11-28 16:00	1.274 [0.049, 2.532]	45.690 [11.912, 46.721]	1032.239 [176.652, 1032.239]
	2024-12-31 23:00	4.648 [3.732, 5.378]	23.834 [17.367, 24.466]	1032.269 [469.856, 1032.269]

Table A6. *Cont.*

Variable	Timestamp	Nugget	Total Sill	Distance Parameter
Relative humidity	2024-01-01 00:00	22.214 [13.688, 36.570]	387.854 [168.960, 402.878]	1032.298 [325.615, 1032.298]
	2024-02-03 06:00	65.117 [60.516, 67.913]	148.349 [128.589, 203.103]	552.455 [378.686, 1018.260]
	2024-03-07 12:00	77.135 [62.412, 91.667]	268.712 [250.995, 288.634]	536.035 [399.069, 660.494]
	2024-04-09 19:00	31.038 [18.786, 51.212]	1182.127 [520.617, 1204.836]	1032.282 [386.044, 1032.282]
	2024-05-13 01:00	120.533 [107.360, 136.008]	225.821 [217.213, 244.856]	363.665 [244.691, 579.094]
	2024-06-15 08:00	37.195 [32.398, 39.746]	170.235 [150.227, 174.415]	1032.270 [836.688, 1032.270]
	2024-07-18 14:00	18.819 [13.122, 24.631]	59.792 [57.043, 69.679]	253.112 [201.085, 378.251]
	2024-08-20 21:00	38.497 [22.892, 50.602]	86.066 [84.067, 88.789]	142.246 [95.066, 191.723]
	2024-09-23 03:00	49.354 [40.376, 60.126]	110.228 [108.205, 112.435]	112.148 [95.096, 142.507]
	2024-10-26 10:00	0.000 [0.000, 23.032]	119.787 [116.707, 122.307]	75.496 [66.920, 99.355]
	2024-11-28 16:00	37.069 [0.000, 75.032]	273.512 [0.000, 290.332]	138.665 [112.425, 199.346]
	2024-12-31 23:00	99.709 [74.819, 128.972]	411.208 [397.765, 436.635]	238.940 [212.856, 278.092]
$Wind_{NS}$	2024-01-01 00:00	0.000 [0.000, 6.171]	10.791 [7.148, 13.114]	400.313 [400.313, 400.313]
	2024-02-03 06:00	1.027 [0.833, 2.361]	2.190 [1.904, 2.690]	368.715 [368.715, 368.715]
	2024-03-07 12:00	1.326 [0.000, 6.652]	12.608 [9.447, 16.158]	659.064 [659.064, 659.064]
	2024-04-09 19:00	2.108 [0.469, 2.511]	3.856 [3.153, 4.156]	484.671 [484.671, 484.671]
	2024-05-13 01:00	0.443 [0.000, 5.254]	9.493 [6.450, 11.938]	415.936 [415.936, 415.936]
	2024-06-15 08:00	0.555 [0.542, 0.737]	0.897 [0.875, 1.068]	417.728 [417.728, 417.728]
	2024-07-18 14:00	2.975 [0.238, 4.266]	7.237 [6.262, 7.945]	415.936 [415.936, 415.936]
	2024-08-20 21:00	0.722 [0.000, 4.260]	13.340 [10.720, 15.602]	549.946 [549.946, 549.946]
	2024-09-23 03:00	1.708 [1.614, 2.398]	2.967 [2.812, 3.269]	484.618 [484.618, 484.618]
	2024-10-26 10:00	1.301 [1.269, 1.872]	2.341 [2.257, 2.647]	368.607 [368.607, 368.607]
	2024-11-28 16:00	3.971 [2.283, 5.478]	8.510 [7.592, 9.909]	484.457 [484.457, 484.457]
	2024-12-31 23:00	0.476 [0.000, 4.604]	7.376 [5.304, 9.808]	414.625 [414.625, 414.625]
$Wind_{WE}$	2024-01-01 00:00	1.129 [1.184, 2.063]	2.318 [2.073, 2.940]	578.218 [578.218, 578.218]
	2024-02-03 06:00	0.952 [0.184, 4.316]	5.006 [3.582, 6.635]	368.715 [368.715, 368.715]
	2024-03-07 12:00	2.123 [0.976, 3.594]	7.003 [5.673, 8.111]	368.711 [368.711, 368.711]
	2024-04-09 19:00	2.867 [2.019, 5.671]	7.044 [5.576, 7.957]	484.671 [484.671, 484.671]
	2024-05-13 01:00	1.159 [0.180, 2.870]	4.411 [3.422, 5.593]	577.670 [577.670, 577.670]
	2024-06-15 08:00	0.662 [0.703, 1.656]	1.429 [1.176, 2.098]	577.615 [577.615, 577.615]
	2024-07-18 14:00	3.156 [2.109, 3.994]	6.444 [5.927, 6.741]	368.681 [368.681, 368.681]
	2024-08-20 21:00	2.198 [1.767, 3.825]	5.591 [4.800, 6.487]	484.618 [484.618, 484.618]
	2024-09-23 03:00	1.398 [0.038, 3.841]	9.334 [7.503, 10.685]	400.313 [400.313, 400.313]
	2024-10-26 10:00	1.361 [0.110, 4.185]	7.506 [5.182, 8.967]	472.370 [472.370, 472.370]
	2024-11-28 16:00	3.938 [1.484, 9.229]	18.773 [14.825, 21.397]	484.457 [484.457, 484.457]
	2024-12-31 23:00	1.406 [0.739, 4.405]	5.328 [3.845, 6.644]	342.538 [342.538, 342.538]

Note: Values are reported as the fitted point estimate [2.5th percentile, 97.5th percentile]. Parameter uncertainty was obtained from 100 requested bootstrap resamples at each timestamp using the valid-lag-bin residual as the resampling unit. All bootstrap runs were successful for temperature, $Wind_{NS}$, and $Wind_{WE}$. For relative humidity, 100 runs were successful at 11 timestamps and 99 runs at 23 September 2024 03:00. The distance parameter is reported in kilometres for temperature and relative humidity and in metres for $Wind_{NS}$ and $Wind_{WE}$, consistent with the corresponding analysis outputs.

At each timestamp, 100 bootstrap resamples were requested for the variogram model selected by the fitting-based criterion. The resampling unit was the valid-lag-bin residual. For each bootstrap resample, the variogram parameters were re-estimated, and the resulting bootstrap distributions were used to calculate the 2.5th and 97.5th percentiles, which define the empirical 95% bootstrap interval. The evaluated parameters were the nugget, total sill, and distance parameter. The fitted point estimates and corresponding bootstrap interval bounds are reported in Table A6.

Across the four variables, the bootstrap interval bounds differed among parameters and timestamps. For temperature, the nugget, total-sill, and distance-parameter intervals varied across the 12 timestamps, with several distance-parameter intervals reaching the reported upper fitted values. For relative humidity, interval widths also differed across timestamps, including wider bounds for some nugget, total-sill, and distance-parameter

estimates. For Wind_{NS} and Wind_{WE}, the lower and upper bootstrap bounds for the distance parameter were frequently identical to the corresponding point estimate, whereas the nugget and total-sill bounds varied across timestamps. These results characterize the uncertainty associated with the fitted variogram parameters beyond that conveyed by the point estimates alone.

Appendix C. Temporal Stability Analysis of Permutation Feature Importance

To examine the temporal stability of the ANN-based feature-weighting procedure, permutation feature importance was additionally evaluated over 36 temporally separated monthly periods for air temperature, relative humidity, Wind_{NS}, and Wind_{WE}. For each input feature, the mean and standard deviation of permutation importance were calculated across the evaluation periods. Table A7 presents the ten highest-ranked features for each variable .

Table A7. Temporal stability of the ten highest-ranked permutation-importance features across 36 monthly evaluation periods.

Variable	Rank	Feature	Mean PI	SD PI
Temperature	1	exp_r2	0.4229	0.0818
	2	sph_r2	0.2846	0.0853
	3	exp_nugget	0.2003	0.1113
	4	ration_qua_r2	0.1721	0.0972
	5	ration_qua_nugget	0.1663	0.0969
	6	exp_rmse	0.1457	0.0886
	7	penta_sph_r2	0.1342	0.0773
	8	ration_qua_mae	0.1138	0.0874
	9	sph_nugget	0.1066	0.0918
	10	sph_mae	0.0869	0.1039
Relative Humidity	1	exp_r2	0.4110	0.0377
	2	sph_r2	0.2486	0.0268
	3	penta_sph_r2	0.1601	0.0308
	4	ration_qua_r2	0.1232	0.0172
	5	exp_nugget	0.0634	0.0415
	6	exp_mae	0.0600	0.0321
	7	sph_mae	0.0599	0.0295
	8	penta_sph_mae	0.0589	0.0229
	9	exp_rmse	0.0553	0.0218
	10	ration_qua_mae	0.0521	0.0236
Wind_{NS}	1	exp_r2	0.2982	0.0557
	2	sph_r2	0.2523	0.0432
	3	ration_qua_r2	0.2014	0.0524
	4	exp_mae	0.1553	0.0736
	5	exp_rmse	0.1489	0.0737
	6	ration_qua_rmse	0.1388	0.0635
	7	sph_rmse	0.1356	0.0652
	8	ration_qua_mae	0.1314	0.0591
	9	penta_sph_rmse	0.1302	0.0587
	10	sph_mae	0.1249	0.0661
Wind_{WE}	1	exp_r2	0.2677	0.0516
	2	sph_r2	0.2330	0.0490
	3	ration_qua_r2	0.1836	0.0465
	4	exp_mae	0.1377	0.0418
	5	exp_rmse	0.1361	0.0423
	6	ration_qua_rmse	0.1204	0.0442
	7	sph_rmse	0.1148	0.0353
	8	ration_qua_mae	0.1132	0.0435
	9	penta_sph_rmse	0.1117	0.0346
	10	penta_sph_r2	0.1107	0.0373

Note: PI denotes permutation importance. Mean PI and SD PI are calculated across the 36 monthly evaluation periods. The table reports the ten highest-ranked features for each meteorological variable from the full set of 24 input features.

The dominant features exhibited broadly consistent patterns across the temporal evaluation periods. In particular, the R^2-based features of the exponential and spherical variogram models occupied the highest ranks across all four variables. RMSE- and MAE-based features also showed relatively high contributions, particularly for the wind components. Although the magnitude of permutation importance varied across the evaluation periods, the recurrent prominence of the leading features indicates temporal stability in the dominant feature-importance patterns.

Appendix D. Additional Validation Analyses

To examine temporal dependence in the classification procedure, an additional time-aware validation was performed using the external 2023–2025 dataset. The `best_model` variogram classes were used as reference labels. For each year, the evaluation data were temporally ordered and divided into 12 monthly blocks, and the selected model configuration was retained without additional hyperparameter tuning. Accuracy and Macro-F1 were calculated for each monthly block and are reported in Table A8 as the annual mean ± standard deviation across the 12 blocks.

Table A8. Time-aware classification performance for the meteorological variables during 2023–2025.

Meteorological Variable	Year	Accuracy (%)	Macro-F1 (%)	Months
Temperature	2023	94.59 ± 3.78	87.05 ± 7.08	12
	2024	94.83 ± 3.17	84.75 ± 6.54	12
	2025	94.74 ± 3.53	80.86 ± 10.48	12
Relative Humidity	2023	87.70 ± 2.38	84.37 ± 4.45	12
	2024	86.60 ± 3.69	82.95 ± 4.38	12
	2025	88.21 ± 3.23	84.87 ± 3.03	12
Wind_{NS}	2023	90.56 ± 2.84	79.18 ± 4.23	12
	2024	90.26 ± 3.19	75.41 ± 6.75	12
	2025	89.17 ± 5.10	78.31 ± 4.10	12
Wind_{WE}	2023	89.80 ± 1.44	74.64 ± 4.06	12
	2024	90.56 ± 2.30	73.21 ± 5.26	12
	2025	91.20 ± 1.76	74.72 ± 4.19	12
Solar Radiation	2023	89.86 ± 2.43	74.80 ± 14.41	12
	2024	89.88 ± 3.76	73.26 ± 13.51	12
	2025	92.30 ± 2.38	74.23 ± 16.04	12
Sunshine Hours	2023	90.14 ± 2.68	54.48 ± 10.18	12
	2024	92.30 ± 3.15	56.60 ± 11.80	12
	2025	90.72 ± 3.91	52.69 ± 14.41	12

Note: Results are based on the 2023–2025 external dataset, using the `best_model` variogram classes as reference labels. Accuracy and Macro-F1 are reported as the mean ± standard deviation across the 12 monthly temporal evaluation blocks of each year. The selected model configuration was retained without additional hyperparameter tuning on these evaluation blocks. For solar radiation and sunshine hours, evaluation was restricted to valid daytime cases from 09:00–16:59 KST, with valid zero-valued observations retained.

For solar radiation and sunshine hours, the evaluation was restricted to valid daytime observations from 09:00–16:59 KST to avoid the influence of structurally occurring nighttime values. Valid zero-valued observations within this daytime window were retained. The resulting daytime evaluation included 2920, 2928, and 2920 cases in 2023, 2024, and 2025, respectively.

As shown in Table A8, classification performance varied across meteorological variables and years. Temperature showed the highest and most stable accuracy (94.59–94.83%), whereas sunshine hours showed relatively high accuracy (90.14–92.30%) but substantially lower Macro-F1 (52.69–56.60%), indicating a difference between overall accuracy and class-balanced performance.

These results are not directly comparable with those in Tables 8, 9 and 13 because the evaluations differ in terms of the datasets, reference-label definitions, and/or validation schemes used.

Appendix E. Annual and Overall Numerical Results of Meteorological Interpolation

This appendix provides the numerical interpolation results corresponding to the temporal comparisons presented in Figures 9–14. To provide a compact quantitative summary without reproducing all monthly values, interpolation errors are reported separately for 2023, 2024, and 2025, together with the overall performance across the complete 2023–2025 evaluation period. The comparison includes CBR-OK, Direct-OK, NN, IDW, and four Static-OK configurations based on the Exponential (Exp), Spherical (Sph), Rational Quadratic (RQ), and Penta-Spherical (Penta) variogram models.

Table A9. Annual and overall interpolation errors for temperature from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	RMSE	1.5603	1.5618	1.9377	2.3668	1.5709	1.5642	1.8866	1.5464
	MAE	1.0751	1.0758	1.3187	1.7742	1.0839	1.0791	1.2769	1.0652
2024	RMSE	1.5123	1.5133	1.8796	2.2922	1.5220	1.5152	1.8289	1.4981
	MAE	1.0461	1.0464	1.2807	1.7228	1.0534	1.0486	1.2552	1.0353
2025	RMSE	1.5605	1.5616	1.9334	2.3577	1.5720	1.5637	1.8611	1.5459
	MAE	1.0736	1.0743	1.3103	1.7654	1.0833	1.0767	1.2763	1.0630
Overall	RMSE	1.5445	1.5457	1.9170	2.3391	1.5551	1.5479	1.8590	1.5303
	MAE	1.0649	1.0655	1.3032	1.7541	1.0736	1.0681	1.2695	1.0545

Note: RMSE and MAE are reported in °C. Annual metrics were calculated using all test observations within each calendar year, whereas the overall metrics were calculated using the complete 2023–2025 evaluation period.

Table A10. Annual and overall interpolation errors for relative humidity from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	RMSE	7.5944	7.6106	10.2628	10.9267	8.1600	8.1005	9.2424	8.0315
	MAE	5.7145	5.7537	7.0041	8.2384	5.8250	5.7693	6.4056	5.7062
2024	RMSE	7.3805	7.4189	9.8185	10.4562	7.9150	7.8444	8.9359	7.7694
	MAE	5.5587	5.6111	6.7656	7.9826	5.6901	5.6286	6.2570	5.5600
2025	RMSE	7.2830	7.3323	9.8845	10.6585	7.8324	7.7734	9.0502	7.6988
	MAE	5.4775	5.5359	6.7501	8.0737	5.5990	5.5474	6.2286	5.4813
Overall	RMSE	7.4193	7.4539	9.9902	10.6819	7.9703	7.9073	9.0769	7.8345
	MAE	5.5835	5.6335	6.8398	8.0981	5.7047	5.6485	6.2971	5.5825

Note: RMSE and MAE are reported in %RH. Annual metrics were calculated using all test observations within each calendar year, whereas the overall metrics were calculated using the complete 2023–2025 evaluation period.

Table A11. Annual and overall interpolation errors for wind speed from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	RMSE	1.5964	1.5928	1.7749	1.5669	1.5992	1.5901	1.6514	1.5858
	MAE	1.0619	1.0603	1.1923	1.0489	1.0652	1.0568	1.0767	1.0539
2024	RMSE	1.5822	1.5787	1.7496	1.5501	1.5851	1.5779	1.6407	1.5724
	MAE	1.0593	1.0573	1.1833	1.0391	1.0626	1.0543	1.0762	1.0508
2025	RMSE	1.5867	1.5825	1.7722	1.5728	1.5884	1.5808	1.6214	1.5774
	MAE	1.0646	1.0619	1.2113	1.0550	1.0667	1.0596	1.0785	1.0571
Overall	RMSE	1.5884	1.5847	1.7656	1.5633	1.5909	1.5829	1.6379	1.5785
	MAE	1.0619	1.0598	1.1956	1.0477	1.0648	1.0569	1.0771	1.0539

Note: RMSE and MAE are reported in m/s. Annual metrics were calculated using all test observations within each calendar year, whereas the overall metrics were calculated using the complete 2023–2025 evaluation period.

Table A12. Annual and overall angular interpolation errors for wind direction from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	Angular RMSE	74.8927	74.7724	80.3768	72.8873	74.9430	74.4773	74.8945	74.5349
	Mean Angular Error	57.7617	57.6303	61.8796	54.8102	57.7990	57.3578	57.8702	57.3956
2024	Angular RMSE	74.2832	74.1836	79.8246	72.3315	74.2807	73.9009	74.1615	74.0372
	Mean Angular Error	57.1304	57.0125	61.5047	54.4851	57.1330	56.7106	57.1114	56.8193
2025	Angular RMSE	73.7156	73.6014	79.4520	72.3561	73.7698	73.4058	73.6137	73.4489
	Mean Angular Error	56.4344	56.2891	61.2095	54.6135	56.4462	56.0879	56.4094	56.1021
Overall	Angular RMSE	74.2987	74.1873	79.8853	72.5252	74.3327	73.9293	74.2250	74.0084
	Mean Angular Error	57.1088	56.9773	61.5313	54.6361	57.1261	56.7187	57.1303	56.7724

Note: Wind-direction errors are reported in degrees using circular error metrics. Overall values represent the complete 2023–2025 evaluation period.

Table A13. Annual and overall interpolation errors for solar radiation from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	RMSE	0.3263	0.3263	0.3766	0.3556	0.3280	0.3265	0.3305	0.3279
	MAE	0.2034	0.2034	0.2115	0.2222	0.2057	0.2032	0.2072	0.2045
2024	RMSE	0.3220	0.3220	0.3720	0.3571	0.3255	0.3227	0.3271	0.3235
	MAE	0.2023	0.2024	0.2118	0.2266	0.2058	0.2024	0.2058	0.2033
2025	RMSE	0.3076	0.3071	0.3594	0.3463	0.3097	0.3069	0.3121	0.3076
	MAE	0.1923	0.1919	0.2034	0.2186	0.1949	0.1916	0.1950	0.1923
Overall	RMSE	0.3186	0.3185	0.3693	0.3530	0.3211	0.3187	0.3233	0.3197
	MAE	0.1993	0.1992	0.2089	0.2225	0.2021	0.1990	0.2026	0.2000

Note: RMSE and MAE are reported in MJ/m^2. Annual metrics were calculated using all test observations within each calendar year, whereas the overall metrics were calculated using the complete 2023–2025 evaluation period.

Table A14. Annual and overall interpolation errors for sunshine hours from 2023 to 2025.

Period	Metric	CBR-OK	Direct-OK	NN	IDW	Exp	Sph	RQ	Penta
2023	RMSE	0.2250	0.2265	0.2643	0.2267	0.2274	0.2244	0.2258	0.2248
	MAE	0.1641	0.1665	0.1580	0.1648	0.1663	0.1633	0.1651	0.1639
2024	RMSE	0.2323	0.2345	0.2732	0.2337	0.2349	0.2316	0.2333	0.2322
	MAE	0.1728	0.1763	0.1675	0.1741	0.1754	0.1720	0.1740	0.1727
2025	RMSE	0.2259	0.2274	0.2650	0.2283	0.2286	0.2251	0.2269	0.2257
	MAE	0.1667	0.1690	0.1605	0.1683	0.1693	0.1658	0.1678	0.1665
Overall	RMSE	0.2277	0.2295	0.2675	0.2296	0.2303	0.2271	0.2287	0.2276
	MAE	0.1679	0.1706	0.1620	0.1691	0.1703	0.1670	0.1690	0.1677

Note: RMSE and MAE are reported in h. Annual metrics were calculated using all test observations within each calendar year, whereas the overall metrics were calculated using the complete 2023–2025 evaluation period.

The annual results show that the relative performance of the interpolation methods remained generally consistent across the three evaluation years. For temperature, CBR-OK closely matched Direct-OK throughout 2023–2025 and produced substantially lower errors than NN and IDW. Over the complete evaluation period, CBR-OK yielded an RMSE of 1.5445 °C and an MAE of 1.0649 °C, compared with 1.5457 °C and 1.0655 °C for Direct-OK, respectively. A similar pattern was observed for relative humidity, for which CBR-OK achieved overall RMSE and MAE values of 7.4193 %RH and 5.5835 %RH, respectively, closely matching Direct-OK while producing substantially lower errors than NN and IDW.

The results for the wind variables were more dependent on the interpolation method. For wind speed, CBR-OK remained close to Direct-OK and the Static-OK configurations, whereas IDW produced the lowest overall RMSE and MAE. A similar pattern was observed for wind direction: IDW yielded the lowest overall angular errors, while CBR-OK, Direct-OK, and the Static-OK configurations showed relatively similar performance. These

results indicate that adaptive variogram retrieval does not necessarily yield the minimum interpolation error for every meteorological variable.

For solar radiation, CBR-OK and Direct-OK produced nearly identical performance over the complete evaluation period, with overall RMSE values of 0.3186 and 0.3185 MJ/m^2, respectively. Both methods also produced lower overall RMSE and MAE than NN and IDW. For sunshine hours, the overall RMSE values of CBR-OK, Direct-OK, IDW, and the Static-OK configurations were relatively similar, whereas NN produced the lowest overall MAE despite its higher RMSE.

Taken together, these annual and overall results are consistent with the temporal patterns shown in Figures 9–14. The relative benefit of CBR-OK varies across meteorological variables and comparison methods rather than representing a uniform improvement in interpolation accuracy. Nevertheless, CBR-OK generally maintained interpolation performance comparable to Direct-OK across the 2023–2025 evaluation period.

Appendix F. Detailed Site-Specific PV Forecasting Results

Site-specific forecasting results are presented in Table A15 for the all-hours evaluation and Table A16 for the daylight-only evaluation, complementing the all-hours CBR-OK summary reported in Table 21. Both evaluations include persistence as a reference and compare the forecasting performance obtained using CBR-OK-, NN-, and IDW-based meteorological inputs in terms of RMSE, MAE, and forecast skill (FS). For each PV site, forecasting architecture, and forecast horizon, identical evaluation timestamps were used across the meteorological input methods and persistence. Negative FS values indicate that the forecasting RMSE exceeded the corresponding persistence RMSE.

Table A15. Detailed site-specific PV forecasting performance under the all-hours evaluation.

Case	Capacity	Model	h	Persistence RMSE/MAE	CBR-OK RMSE/MAE/FS	NN RMSE/MAE/FS	IDW RMSE/MAE/FS
Case 1	554	LSTM	1	51.87/27.67	31.86/15.71/38.58	35.54/16.92/31.49	34.94/16.83/32.64
			12	198.26/146.23	62.10/30.58/68.68	64.33/31.35/67.55	60.38/29.77/69.54
			24	79.17/35.83	68.28/34.06/13.76	69.70/34.83/11.96	65.93/32.67/16.72
		Bi-LSTM	1	51.87/27.67	32.19/15.32/37.94	35.19/16.86/32.17	35.59/16.91/31.38
			12	198.26/146.23	62.66/29.76/68.39	64.68/31.22/67.38	59.03/29.09/70.23
			24	79.17/35.83	69.00/33.94/12.85	66.00/33.76/16.64	65.17/32.34/17.68
		GRU	1	51.87/27.67	31.83/15.04/38.63	37.38/17.99/27.93	34.18/16.27/34.10
			12	198.26/146.23	64.42/32.20/67.51	65.03/33.96/67.20	60.42/29.24/69.53
			24	79.17/35.83	67.06/33.19/15.30	70.28/34.52/11.23	64.77/31.73/18.20
Case 2	1834	LSTM	1	161.91/89.56	97.46/47.80/39.81	125.94/60.78/22.22	114.77/57.10/29.12
			12	641.63/481.16	193.96/95.54/69.77	211.29/104.86/67.07	190.33/94.66/70.34
			24	257.40/119.54	215.74/108.89/16.19	228.84/114.92/11.10	217.50/108.42/15.50
		Bi-LSTM	1	161.91/89.56	101.95/50.35/37.04	125.66/61.75/22.39	127.63/69.30/21.17
			12	641.63/481.16	194.48/96.44/69.69	213.30/105.75/66.76	191.56/94.58/70.14
			24	257.40/119.54	212.22/107.24/17.55	230.31/115.69/10.53	212.81/107.00/17.32
		GRU	1	161.91/89.56	96.95/49.77/40.12	126.65/63.98/21.78	115.32/63.78/28.78
			12	641.63/481.16	189.91/102.84/70.40	206.20/108.97/67.86	189.65/95.43/70.44
			24	257.40/119.54	204.63/112.41/20.50	220.25/123.99/14.43	204.24/107.40/20.65

Table A15. *Cont.*

Case	Capacity	Model	h	Persistence RMSE/MAE	CBR-OK RMSE/MAE/FS	NN RMSE/MAE/FS	IDW RMSE/MAE/FS
Case 3	116	LSTM	1	7.45/4.17	4.61/2.32/38.18	5.20/2.86/30.26	4.52/2.25/39.38
			12	31.42/23.78	8.22/4.26/73.84	7.70/3.92/75.49	7.87/3.58/74.95
			24	9.28/3.95	8.02/4.10/13.58	7.93/4.04/14.51	7.77/3.90/16.31
		Bi-LSTM	1	7.45/4.17	5.11/2.54/31.43	5.41/2.63/27.35	4.97/2.43/33.28
			12	31.42/23.78	8.27/4.30/73.67	8.19/4.34/73.92	8.38/4.45/73.32
			24	9.28/3.95	7.89/4.03/14.99	8.25/4.23/11.05	7.68/3.84/17.26
		GRU	1	7.45/4.17	4.59/2.50/38.37	5.37/2.85/27.95	4.75/2.56/36.29
			12	31.42/23.78	7.86/4.06/74.97	7.55/3.86/75.98	8.10/4.31/74.22
			24	9.28/3.95	7.95/3.91/14.29	8.03/4.03/13.50	7.82/4.05/15.68
Case 4	998	LSTM	1	89.34/48.39	57.97/26.92/35.11	59.69/27.94/33.19	63.46/30.07/28.96
			12	349.39/260.13	104.98/50.98/69.95	113.83/56.57/67.42	103.19/50.06/70.47
			24	149.51/69.33	114.04/57.07/23.73	115.50/59.28/22.75	113.19/56.82/24.29
		Bi-LSTM	1	89.34/48.39	60.00/27.51/32.84	60.03/28.26/32.80	65.25/30.79/26.96
			12	349.39/260.13	101.48/50.12/70.96	111.47/57.13/68.10	104.30/51.33/70.15
			24	149.51/69.33	112.26/56.10/24.92	115.83/58.50/22.53	112.75/55.88/24.59
		GRU	1	89.34/48.39	56.58/25.59/36.67	59.06/26.62/33.89	62.78/29.11/29.72
			12	349.39/260.13	103.25/51.77/70.45	111.12/56.88/68.20	108.02/51.97/69.08
			24	149.51/69.33	116.64/57.60/21.98	120.59/63.33/19.34	115.64/59.91/22.66
Case 5	889	LSTM	1	75.84/42.25	86.81/50.35/−14.46	74.41/38.44/1.89	94.41/61.09/−24.49
			12	303.51/229.73	139.63/77.26/54.00	136.19/72.26/55.13	131.99/67.95/56.51
			24	112.85/51.89	119.66/64.98/−6.03	106.12/62.36/5.96	105.91/60.88/6.15
		Bi-LSTM	1	75.84/42.25	109.33/68.53/−44.15	79.40/39.91/−4.69	83.78/42.38/−10.47
			12	303.51/229.73	147.32/84.97/51.46	138.23/78.24/54.46	145.45/79.34/52.08
			24	112.85/51.89	102.47/57.48/9.20	111.93/67.95/0.81	98.63/52.68/12.60
		GRU	1	75.84/42.25	89.53/53.51/−18.05	92.58/60.83/−22.07	83.91/50.82/−10.64
			12	303.51/229.73	143.88/103.39/52.59	124.85/68.64/58.86	123.69/70.60/59.25
			24	112.85/51.89	140.08/77.60/−24.13	117.68/68.45/−4.28	135.22/77.56/−19.83
Case 6	370	LSTM	1	26.12/14.20	18.11/8.44/30.68	19.15/8.83/26.71	18.14/9.04/30.57
			12	100.62/76.15	29.46/15.01/70.73	29.67/15.31/70.51	28.83/14.74/71.35
			24	34.24/15.41	27.85/14.17/18.66	29.12/14.49/14.94	28.57/14.62/16.54
		Bi-LSTM	1	26.12/14.20	18.45/8.78/29.39	19.25/9.23/26.33	18.38/8.78/29.63
			12	100.62/76.15	28.81/14.82/71.37	29.01/15.00/71.17	29.27/14.89/70.91
			24	34.24/15.41	28.29/14.00/17.37	28.73/14.76/16.08	28.47/14.45/16.83
		GRU	1	26.12/14.20	18.77/9.20/28.16	19.44/9.89/25.58	18.68/9.35/28.49
			12	100.62/76.15	29.83/16.55/70.35	29.87/15.34/70.32	29.09/15.05/71.09
			24	34.24/15.41	28.61/14.17/16.43	29.92/15.62/12.59	30.22/16.59/11.73

Note: Capacity, RMSE, and MAE are expressed in kW, whereas FS is expressed in %. Persistence results are reported as RMSE/MAE, and CBR-OK, NN, and IDW results are reported as RMSE/MAE/FS. All methods were evaluated using identical all-hours test timestamps.

Table A16. Detailed site-specific PV forecasting performance under the daylight-only evaluation.

Case	Capacity	Model	h	Persistence RMSE/MAE	CBR-OK RMSE/MAE/FS	NN RMSE/MAE/FS	IDW RMSE/MAE/FS
Case 1	554	LSTM	1	73.64/55.34	45.18/30.34/38.64	50.43/32.85/31.52	49.57/32.56/32.68
			12	199.59/149.54	88.16/60.15/55.83	91.35/62.40/54.23	85.72/58.55/57.05
			24	112.54/72.33	97.03/67.39/13.79	99.05/68.91/11.99	93.69/64.60/16.75
		Bi-LSTM	1	73.64/55.34	45.66/29.79/38.00	49.89/32.64/32.26	50.48/32.62/31.44
			12	199.59/149.54	88.99/59.46/55.41	91.84/62.10/53.98	83.81/57.69/58.01
			24	112.54/72.33	98.06/67.73/12.87	93.80/67.52/16.66	92.62/64.60/17.70
		GRU	1	73.64/55.34	45.18/29.60/38.65	53.09/36.03/27.91	48.50/31.81/34.14
			12	199.59/149.54	91.46/64.03/54.17	92.32/67.69/53.75	85.79/58.45/57.02
			24	112.54/72.33	95.30/66.32/15.32	99.88/68.98/11.25	92.05/63.48/18.21

Table A16. *Cont.*

Case	Capacity	Model	h	Persistence RMSE/MAE	CBR-OK RMSE/MAE/FS	NN RMSE/MAE/FS	IDW RMSE/MAE/FS
Case 2	1834	LSTM	1	229.19/177.91	137.89/92.32/39.84	178.31/119.47/22.20	162.44/110.99/29.12
			12	644.60/490.20	274.67/189.34/57.39	299.21/208.21/53.58	269.54/187.53/58.18
			24	364.76/239.96	305.67/216.23/16.20	324.24/228.48/11.11	308.20/215.86/15.51
		Bi-LSTM	1	229.19/177.91	144.25/98.06/37.06	177.80/120.14/22.42	180.32/130.76/21.32
			12	644.60/490.20	275.40/191.19/57.28	302.06/209.32/53.14	271.28/188.23/57.91
			24	364.76/239.96	300.70/213.45/17.56	326.33/230.34/10.54	301.55/212.77/17.33
		GRU	1	229.19/177.91	136.97/93.86/40.24	179.13/122.79/21.84	162.11/113.80/29.27
			12	644.60/490.20	268.54/194.63/58.34	291.75/209.38/54.74	268.52/187.43/58.34
			24	364.76/239.96	289.50/211.97/20.63	311.18/228.85/14.69	289.21/208.40/20.71
Case 3	116	LSTM	1	10.98/8.89	6.81/4.81/37.98	7.59/5.47/30.89	6.68/4.66/39.19
			12	32.95/26.26	12.17/8.95/63.08	11.37/8.08/65.49	11.64/7.67/64.66
			24	13.76/8.66	11.88/8.77/13.66	11.72/8.50/14.83	11.51/8.46/16.37
		Bi-LSTM	1	10.98/8.89	7.57/5.47/31.04	8.02/5.75/27.00	7.37/5.34/32.86
			12	32.95/26.26	12.26/9.38/62.80	12.11/9.31/63.25	12.41/9.68/62.33
			24	13.76/8.66	11.69/8.77/15.04	12.24/9.25/11.09	11.38/8.38/17.34
		GRU	1	10.98/8.89	6.75/4.79/38.57	7.90/5.62/28.08	6.99/4.99/36.34
			12	32.95/26.26	11.64/8.76/64.67	11.16/8.12/66.13	11.99/9.24/63.61
			24	13.76/8.66	11.78/8.41/14.39	11.90/8.70/13.56	11.60/8.76/15.74
Case 4	998	LSTM	1	129.11/100.25	73.86/52.47/42.79	76.05/54.13/41.10	83.89/58.79/35.03
			12	357.96/274.64	148.10/102.70/58.63	163.55/115.30/54.31	144.34/100.86/59.68
			24	213.12/142.30	162.45/116.90/23.78	164.87/121.25/22.64	160.50/115.62/24.69
		Bi-LSTM	1	129.11/100.25	76.45/53.99/40.79	77.36/54.74/40.08	86.01/60.81/33.38
			12	357.96/274.64	143.41/101.08/59.94	158.41/115.23/55.75	145.82/103.37/59.26
			24	213.12/142.30	159.27/113.21/25.27	164.75/119.30/22.70	160.56/113.02/24.66
		GRU	1	129.11/100.25	71.45/50.37/44.66	74.23/52.13/42.51	82.24/56.90/36.30
			12	357.96/274.64	145.66/102.33/59.31	158.39/111.85/55.75	151.81/104.83/57.59
			24	213.12/142.30	165.45/115.81/22.37	171.71/123.36/19.43	163.75/117.88/23.17
Case 5	889	LSTM	1	109.70/87.35	113.47/85.52/−3.44	87.76/63.74/20.00	118.68/96.14/−8.19
			12	310.94/243.39	199.50/147.14/35.84	196.38/146.55/36.84	186.69/135.32/39.96
			24	161.91/107.33	168.64/125.66/−4.16	144.54/111.00/10.72	145.57/110.18/10.09
		Bi-LSTM	1	109.70/87.35	147.06/108.99/−34.06	89.73/63.43/18.20	104.54/73.13/4.70
			12	310.94/243.39	211.40/158.41/32.01	197.93/147.48/36.35	209.60/153.66/32.59
			24	161.91/107.33	137.23/104.73/15.24	148.56/118.50/8.24	129.51/97.18/20.01
		GRU	1	109.70/87.35	108.92/88.21/0.71	111.64/88.37/−1.77	102.70/82.79/6.38
			12	310.94/243.39	187.49/144.71/39.70	177.15/134.68/43.03	175.68/128.18/43.50
			24	161.91/107.33	200.27/150.15/−23.69	164.74/125.91/−1.75	190.84/144.96/−17.87
Case 6	370	LSTM	1	37.67/29.34	22.41/15.31/40.51	23.90/15.99/36.56	22.48/15.64/40.32
			12	102.69/79.63	39.73/28.10/61.31	40.15/28.72/60.90	38.85/27.34/62.17
			24	48.28/31.04	37.84/26.97/21.61	39.59/27.45/17.99	39.01/27.76/19.19
		Bi-LSTM	1	37.67/29.34	23.03/15.84/38.87	24.19/16.54/35.79	22.77/15.86/39.54
			12	102.69/79.63	38.97/27.54/62.05	39.12/27.89/61.90	39.37/27.76/61.66
			24	48.28/31.04	38.36/26.86/20.54	39.02/27.72/19.17	38.88/27.54/19.46
		GRU	1	37.67/29.34	23.55/16.55/37.48	24.51/17.10/34.95	23.36/16.57/37.99
			12	102.69/79.63	40.26/28.89/60.79	40.35/28.08/60.71	39.00/27.31/62.02
			24	48.28/31.04	38.88/27.08/19.47	40.78/28.99/15.54	41.20/29.55/14.66

Note: Capacity, RMSE, and MAE are expressed in kW, whereas FS is expressed in %. Persistence results are reported as RMSE/MAE, and CBR-OK, NN, and IDW results are reported as RMSE/MAE/FS. All methods were evaluated using identical daylight-only test timestamps.

The daylight-only evaluation was conducted by retaining observations for which the measured PV power generation exceeded 0.01 kW at the forecast target timestamp ($t + h$). In contrast, the all-hours evaluation included all available test observations. As summarised in Table A17, the all-hours evaluation comprised 3945, 3934, and 3922 observations per PV site

for the 1-, 12-, and 24-h forecasting horizons, respectively. The proportion of observations retained for the daylight-only evaluation ranged from 45.39% to 49.85% across the six PV sites and three forecast horizons. Accordingly, the all-hours and daylight-only evaluations were based on identical underlying predictions but different evaluation subsets.

Table A17. Numbers of all-hours and daylight-only evaluation observations across the six PV sites and three forecast horizons.

PV Site	Horizon	All-Hours (*N*)	Daylight-Only (*N*)	Excluded (*N*)	Daylight (%)
Case 1	1 h	3945	1955	1990	49.56
	12 h	3934	1950	1984	49.57
	24 h	3922	1941	1981	49.49
Case 2	1 h	3945	1966	1979	49.84
	12 h	3934	1961	1973	49.85
	24 h	3922	1953	1969	49.80
Case 3	1 h	3945	1793	2152	45.45
	12 h	3934	1788	2146	45.45
	24 h	3922	1780	2142	45.39
Case 4	1 h	3945	1886	2059	47.81
	12 h	3934	1881	2053	47.81
	24 h	3922	1873	2049	47.76
Case 5	1 h	3945	1879	2066	47.63
	12 h	3934	1874	2060	47.64
	24 h	3922	1866	2056	47.58
Case 6	1 h	3945	1955	1990	49.56
	12 h	3934	1950	1984	49.57
	24 h	3922	1941	1981	49.49

Note: Daylight-only observations were defined by observed PV power generation exceeding 0.01 kW at the forecast target time ($t + h$). Excluded (N) denotes observations not satisfying this criterion, and daylight (%) is their retained proportion relative to all-hours observations.

Forecasting performance varied across PV sites, forecasting architectures, forecast horizons, meteorological input methods, and evaluation periods. For Case 5, the CBR-OK forecast skill at the 1-h horizon was −14.46%, −44.15%, and −18.05% for LSTM, Bi-LSTM, and GRU, respectively, under the all-hours evaluation. Under the daylight-only evaluation, the corresponding values were −3.44%, −34.06%, and 0.71%, respectively. These differences indicate that the assessment of forecasting performance is sensitive to the selection of evaluation timestamps, even when the underlying forecast predictions remain unchanged.

Appendix G. Detailed Diebold–Mariano Test Results

This appendix provides the detailed site-specific Diebold–Mariano (DM) test results corresponding to the summary presented in Section 5.6. The tests compare PV power forecasts obtained using CBR-OK-based meteorological inputs with those obtained using nearest-neighbour (NN) and inverse-distance weighting (IDW) meteorological inputs. Squared-error (SE) and absolute-error (AE) losses are evaluated separately for LSTM, Bi-LSTM, and GRU at forecasting horizons of 1, 12, and 24 h.

For each comparison, the paired loss differential at time t is defined as

$$d_t = L(e_{\text{baseline},t}) - L(e_{\text{CBR}-\text{OK},t}), \tag{A2}$$

where $e_{\text{baseline},t}$ and $e_{\text{CBR}-\text{OK},t}$ denote the forecast errors of the baseline and CBR-OK methods, respectively. The loss function $L(e)$ is defined as

$$L(e) = \begin{cases} e^2, & \text{for squared-error loss (SE),} \\ |e|, & \text{for absolute-error loss (AE).} \end{cases} \tag{A3}$$

The mean loss differential is calculated as

$$\bar{d} = \frac{1}{T} \sum_{t=1}^{T} d_t, \tag{A4}$$

where T denotes the number of paired forecast errors. A positive value of $\bar{d}$ indicates that CBR-OK achieves a lower average forecasting loss than the corresponding baseline.

The DM statistic is computed as

$$DM = \frac{\bar{d}}{\sqrt{\widehat{\Omega}/T}}, \tag{A5}$$

where $\widehat{\Omega}$ denotes the estimated long-run variance of the loss differential. To account for serial dependence in multi-step forecast errors, the Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) estimator with Bartlett weights is employed:

$$\widehat{\Omega} = \widehat{\gamma}_0 + 2 \sum_{k=1}^{q} \left(1 - \frac{k}{q+1}\right) \widehat{\gamma}_k, \tag{A6}$$

where $\widehat{\gamma}_k$ denotes the sample autocovariance of the loss differential at lag k. The maximum lag is set to $q = h - 1$, corresponding to 0, 11, and 23 for the 1-, 12-, and 24-h forecasting horizons, respectively.

The HAC-based standard error of the mean loss differential is given by

$$\mathrm{SE}_{\mathrm{HAC}}(\bar{d}) = \sqrt{\frac{\widehat{\Omega}}{T}}. \tag{A7}$$

Accordingly, the 95% confidence interval for the mean loss differential is calculated as

$$CI_{95\%} = \left[\bar{d} - 1.96\,\mathrm{SE}_{\mathrm{HAC}}(\bar{d}),\ \bar{d} + 1.96\,\mathrm{SE}_{\mathrm{HAC}}(\bar{d})\right]. \tag{A8}$$

Table A18. Detailed site-specific Diebold–Mariano test results for squared-error loss.

Case	Model	h	CBR-OK vs. NN		CBR-OK vs. IDW	
			DM (p)	**$\bar{d}$ [95% CI]**	**DM (p)**	**$\bar{d}$ [95% CI]**
Case 1	LSTM	1	6.63 (<0.001)	247.03 [174.05, 320.02]	5.28 (<0.001)	205.62 [129.32, 281.92]
		12	2.74 (0.006)	284.34 [81.13, 487.56]	−1.94 (0.053)	−207.16 [−416.83, 2.50]
		24	2.18 (0.029)	192.88 [19.61, 366.16]	−3.12 (0.002)	−309.21 [−503.15, −115.26]
	Bi-LSTM	1	5.89 (<0.001)	201.32 [134.38, 268.26]	5.84 (<0.001)	230.53 [153.19, 307.88]
		12	2.59 (0.010)	253.62 [61.83, 445.41]	−3.59 (<0.001)	−437.67 [−676.60, −198.75]
		24	−3.15 (0.002)	−399.36 [−647.56, −151.16]	−4.71 (<0.001)	−514.79 [−729.00, −300.58]
	GRU	1	9.37 (<0.001)	381.94 [302.09, 461.80]	4.07 (<0.001)	154.89 [80.27, 229.51]
		12	0.50 (0.617)	75.13 [−219.15, 369.42]	−4.15 (<0.001)	−492.13 [−724.39, −259.87]
		24	4.20 (<0.001)	437.32 [233.06, 641.58]	−3.07 (0.002)	−296.36 [−485.69, −107.02]

Table A18. *Cont.*

Case	Model	*h*	CBR-OK vs. NN		CBR-OK vs. IDW	
			DM (*p*)	**$\bar{d}$ [95% CI]**	**DM (*p*)**	**$\bar{d}$ [95% CI]**
Case 2	LSTM	1	9.96 (<0.001)	6403.17 [5142.63, 7663.70]	8.02 (<0.001)	3690.95 [2789.08, 4592.82]
		12	2.92 (0.003)	6960.18 [2290.60, 11,629.76]	−0.84 (0.403)	−1452.49 [−4857.83, 1952.85]
		24	3.68 (<0.001)	5774.55 [2698.68, 8850.43]	0.82 (0.413)	775.79 [−1082.17, 2633.76]
	Bi-LSTM	1	9.91 (<0.001)	5427.96 [4354.49, 6501.43]	11.07 (<0.001)	5891.88 [4849.13, 6934.62]
		12	3.67 (<0.001)	7601.76 [3541.36, 11,662.17]	−1.00 (0.317)	−1188.64 [−3516.51, 1139.24]
		24	4.59 (<0.001)	7942.41 [4549.53, 11,335.29]	0.28 (0.781)	230.88 [−1393.47, 1855.23]
	GRU	1	11.06 (<0.001)	6668.04 [5486.06, 7850.01]	9.39 (<0.001)	3897.22 [3083.51, 4710.94]
		12	3.00 (0.003)	6349.17 [2203.63, 10,494.71]	−0.07 (0.944)	−108.73 [−3161.81, 2944.34]
		24	3.72 (<0.001)	6574.01 [3106.55, 10,041.48]	−0.10 (0.918)	−123.42 [−2468.90, 2222.07]
Case 3	LSTM	1	7.58 (<0.001)	5.80 [4.30, 7.30]	−1.02 (0.307)	−0.83 [−2.42, 0.76]
		12	−2.06 (0.039)	−8.21 [−16.01, −0.40]	−1.25 (0.211)	−5.51 [−14.14, 3.13]
		24	−0.66 (0.512)	−1.38 [−5.51, 2.75]	−4.44 (<0.001)	−4.02 [−5.80, −2.25]
	Bi-LSTM	1	4.02 (<0.001)	3.20 [1.64, 4.76]	−1.74 (0.082)	−1.41 [−3.00, 0.18]
		12	−0.52 (0.607)	−1.29 [−6.19, 3.61]	1.34 (0.179)	1.84 [−0.85, 4.54]
		24	3.13 (0.002)	5.84 [2.19, 9.50]	−3.20 (0.001)	−3.33 [−5.37, −1.29]
	GRU	1	8.70 (<0.001)	7.74 [6.00, 9.49]	1.77 (0.078)	1.44 [−0.16, 3.03]
		12	−1.57 (0.116)	−4.78 [−10.74, 1.18]	2.30 (0.021)	4.00 [0.59, 7.41]
		24	0.80 (0.424)	1.12 [−1.62, 3.85]	−0.80 (0.421)	−1.85 [−6.34, 2.65]
Case 4	LSTM	1	2.53 (0.011)	198.31 [44.59, 352.03]	4.89 (<0.001)	668.28 [400.18, 936.39]
		12	2.18 (0.029)	1905.61 [190.41, 3620.82]	−1.05 (0.293)	−373.75 [−1069.85, 322.35]
		24	0.68 (0.498)	319.41 [−604.34, 1243.16]	−0.64 (0.519)	−191.84 [−775.10, 391.42]
	Bi-LSTM	1	0.00 (0.997)	0.33 [−177.38, 178.04]	4.65 (<0.001)	662.61 [383.09, 942.13]
		12	2.76 (0.006)	2060.86 [599.26, 3522.47]	1.21 (0.227)	597.54 [−371.92, 1567.00]
		24	1.99 (0.046)	795.03 [12.44, 1577.62]	0.36 (0.722)	116.65 [−525.49, 758.80]
	GRU	1	4.03 (<0.001)	284.69 [146.31, 423.07]	5.62 (<0.001)	741.76 [483.12, 1000.39]
		12	2.44 (0.015)	1646.84 [326.04, 2967.64]	1.87 (0.062)	1018.73 [−49.83, 2087.29]
		24	1.67 (0.096)	895.60 [−158.58, 1949.78]	−0.66 (0.510)	−265.25 [−1053.91, 523.42]
Case 5	LSTM	1	−12.29 (<0.001)	−2005.17 [−2325.07, −1685.27]	7.32 (<0.001)	1378.39 [1009.47, 1747.31]
		12	−1.58 (0.115)	−928.19 [−2082.84, 226.46]	−3.79 (<0.001)	−2090.53 [−3172.74, −1008.33]
		24	−4.69 (<0.001)	−3126.64 [−4433.22, −1820.07]	−5.77 (<0.001)	−3156.21 [−4229.15, −2083.26]
	Bi-LSTM	1	−13.19 (<0.001)	−5671.10 [−6513.55, −4828.64]	−14.23 (<0.001)	−4944.86 [−5625.98, −4263.74]
		12	−6.46 (<0.001)	−2582.35 [−3366.17, −1798.53]	−1.30 (0.192)	−526.45 [−1317.39, 264.48]
		24	6.28 (<0.001)	2022.12 [1390.78, 2653.45]	−2.51 (0.012)	−775.43 [−1379.92, −170.93]
	GRU	1	3.16 (0.002)	555.99 [210.87, 901.12]	−8.49 (<0.001)	−974.56 [−1199.66, −749.47]
		12	−9.68 (<0.001)	−5151.63 [−6194.68, −4108.58]	−9.54 (<0.001)	−5415.86 [−6528.60, −4303.13]
		24	−10.55 (<0.001)	−5778.30 [−6851.35, −4705.25]	−2.94 (0.003)	−1289.73 [−2150.20, −429.26]

Table A18. *Cont.*

Case	Model	*h*	CBR-OK vs. NN		CBR-OK vs. IDW	
			DM (*p*)	**$\bar{d}$ [95% CI]**	**DM (*p*)**	**$\bar{d}$ [95% CI]**
Case 6	LSTM	1	5.35 (<0.001)	38.73 [24.55, 52.90]	0.11 (0.911)	0.89 [−14.83, 16.61]
		12	0.70 (0.484)	14.13 [−25.41, 53.66]	−1.32 (0.186)	−37.54 [−93.16, 18.07]
		24	3.86 (<0.001)	73.94 [36.45, 111.44]	2.56 (0.010)	36.72 [8.61, 64.83]
	Bi-LSTM	1	4.61 (<0.001)	30.12 [17.31, 42.93]	−0.31 (0.755)	−2.48 [−18.03, 13.08]
		12	0.69 (0.488)	14.87 [−27.14, 56.89]	1.03 (0.301)	26.84 [−24.00, 77.68]
		24	1.73 (0.083)	28.35 [−3.75, 60.44]	0.69 (0.490)	10.41 [−19.16, 39.98]
	GRU	1	3.52 (<0.001)	25.76 [11.44, 40.09]	−0.41 (0.680)	−3.19 [−18.31, 11.93]
		12	0.27 (0.788)	5.64 [−35.40, 46.67]	−1.72 (0.086)	−43.74 [−93.66, 6.18]
		24	4.20 (<0.001)	73.54 [39.24, 107.84]	5.92 (<0.001)	90.26 [60.36, 120.15]

Note: $\bar{d}$ represents the mean squared-error loss difference between the baseline and CBR-OK (baseline minus CBR-OK). Positive values favour CBR-OK, whereas negative values favour the baseline. The 95% confidence intervals are based on Newey–West HAC standard errors. Values are rounded for presentation.

Table A19. Detailed site-specific Diebold–Mariano test results for absolute-error loss.

Case	Model	*h*	CBR-OK vs. NN		CBR-OK vs. IDW	
			DM (*p*)	**$\bar{d}$ [95% CI]**	**DM (*p*)**	**$\bar{d}$ [95% CI]**
Case 1	LSTM	1	4.76 (<0.001)	1.21 [0.71, 1.70]	4.08 (<0.001)	1.12 [0.58, 1.66]
		12	1.89 (0.059)	0.78 [−0.03, 1.58]	−1.82 (0.069)	−0.79 [−1.65, 0.06]
		24	2.04 (0.041)	0.75 [0.03, 1.47]	−3.53 (<0.001)	−1.38 [−2.15, −0.62]
	Bi-LSTM	1	6.39 (<0.001)	1.54 [1.07, 2.01]	5.75 (<0.001)	1.58 [1.04, 2.13]
		12	3.49 (<0.001)	1.45 [0.64, 2.26]	−1.35 (0.176)	−0.66 [−1.61, 0.30]
		24	−0.31 (0.755)	−0.17 [−1.26, 0.92]	−3.57 (<0.001)	−1.65 [−2.55, −0.74]
	GRU	1	10.20 (<0.001)	2.93 [2.37, 3.50]	4.48 (<0.001)	1.22 [0.69, 1.76]
		12	2.55 (0.011)	1.74 [0.40, 3.08]	−5.82 (<0.001)	−2.92 [−3.90, −1.94]
		24	3.06 (0.002)	1.30 [0.47, 2.14]	−3.44 (<0.001)	−1.45 [−2.28, −0.63]
Case 2	LSTM	1	11.11 (<0.001)	13.04 [10.74, 15.34]	9.42 (<0.001)	9.33 [7.39, 11.27]
		12	3.16 (0.002)	9.25 [3.51, 14.99]	−0.46 (0.646)	−0.98 [−5.19, 3.22]
		24	3.09 (0.002)	6.03 [2.20, 9.86]	−0.41 (0.681)	−0.53 [−3.06, 2.00]
	Bi-LSTM	1	11.28 (<0.001)	11.44 [9.45, 13.43]	17.30 (<0.001)	18.94 [16.79, 21.09]
		12	3.54 (<0.001)	9.21 [4.11, 14.31]	−1.22 (0.223)	−2.00 [−5.20, 1.21]
		24	3.90 (<0.001)	8.44 [4.20, 12.68]	−0.23 (0.821)	−0.27 [−2.58, 2.05]
	GRU	1	12.54 (<0.001)	14.24 [12.02, 16.47]	15.14 (<0.001)	14.01 [12.20, 15.83]
		12	2.24 (0.025)	6.01 [0.75, 11.27]	−3.53 (<0.001)	−7.39 [-11.49, −3.28]
		24	4.75 (<0.001)	11.64 [6.84, 16.44]	−3.19 (0.001)	−4.94 [−7.98, −1.90]
Case 3	LSTM	1	15.30 (<0.001)	0.54 [0.47, 0.61]	−1.83 (0.067)	−0.06 [−0.13, 0.00]
		12	−2.97 (0.003)	−0.34 [−0.56, −0.12]	−5.17 (<0.001)	−0.67 [−0.92, −0.42]
		24	−0.76 (0.446)	−0.05 [−0.19, 0.08]	−5.93 (<0.001)	−0.20 [−0.27, −0.13]
	Bi-LSTM	1	2.79 (0.005)	0.09 [0.03, 0.16]	−3.08 (0.002)	−0.10 [−0.17, −0.04]
		12	0.48 (0.635)	0.04 [−0.12, 0.20]	3.00 (0.003)	0.15 [0.05, 0.25]
		24	3.31 (<0.001)	0.20 [0.08, 0.32]	−4.92 (<0.001)	−0.19 [−0.26, −0.11]
	GRU	1	9.03 (<0.001)	0.36 [0.28, 0.43]	1.84 (0.066)	0.07 [−0.00, 0.14]
		12	−2.17 (0.030)	−0.20 [−0.38, −0.02]	4.37 (<0.001)	0.26 [0.14, 0.38]
		24	1.89 (0.059)	0.12 [−0.00, 0.24]	1.76 (0.078)	0.15 [−0.02, 0.31]

Table A19. *Cont.*

Case	Model	h	CBR-OK vs. NN		CBR-OK vs. IDW	
			DM (p)	$\bar{d}$ [95% CI]	DM (p)	$\bar{d}$ [95% CI]
Case 4	LSTM	1	2.76 (0.006)	1.02 [0.29, 1.74]	6.22 (<0.001)	3.15 [2.16, 4.15]
		12	3.28 (0.001)	5.51 [2.22, 8.80]	−1.16 (0.247)	−0.92 [−2.48, 0.64]
		24	2.05 (0.040)	2.18 [0.10, 4.25]	−0.37 (0.708)	−0.24 [−1.50, 1.02]
	Bi-LSTM	1	1.80 (0.071)	0.74 [−0.06, 1.55]	6.29 (<0.001)	3.29 [2.26, 4.31]
		12	4.16 (<0.001)	6.86 [3.62, 10.09]	1.08 (0.279)	1.24 [−1.00, 3.49]
		24	2.51 (0.012)	2.35 [0.52, 4.18]	−0.25 (0.799)	−0.19 [−1.65, 1.27]
	GRU	1	3.44 (<0.001)	1.03 [0.44, 1.61]	7.08 (<0.001)	3.53 [2.55, 4.50]
		12	3.36 (<0.001)	5.01 [2.09, 7.93]	0.18 (0.855)	0.23 [−2.24, 2.71]
		24	4.79 (<0.001)	5.64 [3.33, 7.95]	2.38 (0.017)	2.23 [0.39, 4.06]
Case 5	LSTM	1	−22.15 (<0.001)	−11.92 [−12.97, −10.86]	12.80 (<0.001)	10.75 [9.10, 12.39]
		12	−4.24 (<0.001)	−4.96 [−7.25, −2.67]	−8.29 (<0.001)	−9.33 [−11.54, −7.13]
		24	−1.88 (0.060)	−2.74 [−5.59, 0.11]	−3.38 (<0.001)	−4.19 [−6.62, −1.76]
	Bi-LSTM	1	−25.06 (<0.001)	−28.65 [−30.89, −26.41]	−28.58 (<0.001)	−26.16 [−27.95, −24.37]
		12	−7.65 (<0.001)	−6.71 [−8.44, −4.99]	−7.32 (<0.001)	−5.60 [−7.10, −4.10]
		24	8.18 (<0.001)	10.49 [7.98, 13.00]	−4.96 (<0.001)	−4.78 [−6.67, −2.89]
	GRU	1	9.29 (<0.001)	7.33 [5.78, 8.87]	−4.95 (<0.001)	−2.68 [−3.74, −1.62]
		12	−14.71 (<0.001)	−34.83 [−39.47, −30.19]	−14.41 (<0.001)	−32.84 [−37.31, −28.37]
		24	−7.97 (<0.001)	−9.15 [−11.40, −6.90]	0.02 (0.980)	0.03 [−2.06, 2.11]
Case 6	LSTM	1	3.97 (<0.001)	0.39 [0.20, 0.58]	5.13 (<0.001)	0.60 [0.37, 0.83]
		12	1.60 (0.110)	0.30 [−0.07, 0.67]	−1.09 (0.275)	−0.28 [−0.77, 0.22]
		24	2.07 (0.038)	0.34 [0.02, 0.66]	2.83 (0.005)	0.42 [0.13, 0.72]
	Bi-LSTM	1	5.29 (<0.001)	0.45 [0.28, 0.62]	−0.03 (0.979)	−0.00 [−0.22, 0.21]
		12	1.11 (0.265)	0.19 [−0.15, 0.54]	0.33 (0.740)	0.07 [−0.35, 0.49]
		24	4.96 (<0.001)	0.79 [0.48, 1.11]	3.00 (0.003)	0.45 [0.16, 0.74]
	GRU	1	7.18 (<0.001)	0.69 [0.50, 0.88]	1.32 (0.186)	0.15 [−0.07, 0.37]
		12	−5.96 (<0.001)	−1.18 [−1.57, −0.79]	−6.17 (<0.001)	−1.50 [−1.97, −1.02]
		24	8.11 (<0.001)	1.44 [1.09, 1.79]	13.60 (<0.001)	2.38 [2.04, 2.73]

Note: $\bar{d}$ represents the mean absolute-error loss difference between the baseline and CBR-OK (baseline minus CBR-OK). Positive values favour CBR-OK, whereas negative values favour the baseline. The 95% confidence intervals are based on Newey–West HAC standard errors. Values are rounded for presentation.

Two-sided p-values are reported, with statistical significance assessed at $p < 0.05$. Table A18 presents the detailed site-specific results for squared-error loss, whereas Table A19 presents the corresponding results for absolute-error loss. For both loss functions, the tables report the DM statistic, p-value, mean loss differential, and 95% confidence interval for each PV site, forecasting model, forecast horizon, and baseline comparison.

Appendix H. Data Sources

All meteorological and photovoltaic (PV) power generation data used in this study are publicly available. The meteorological observations were obtained from the Korea Meteorological Administration (KMA) Data Open METEO Portal. The Automated Synoptic Observing System (ASOS) provided air temperature, relative humidity, wind speed, wind direction, solar radiation, and sunshine duration, while the Automatic Weather Station (AWS) provided air temperature, relative humidity, wind speed, and wind direction.

The PV power generation data for the six case-study sites were obtained from the Korean Public Data Portal and provided by Korea Southern Power Co., Ltd. (KOSPO). The geographical coordinates of the PV sites were identified by matching the plant information

in the PV generation datasets with the publicly available KOSPO PV plant location dataset, which provides plant names, unit identifiers, latitude, and longitude. The data sources and corresponding access URLs are summarized in Table A20.

Table A20. Public data sources used in this study.

Data Type	Dataset	Provider	Source
ASOS	Synoptic Weather Observation (ASOS)	KMA	https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36&tabNo=1
AWS	Automatic Weather Station (AWS)	KMA	https://data.kma.go.kr/data/grnd/selectAwsRltmList.do?pgmNo=56
PV site location	PV Plant Location Information	KOSPO	https://www.data.go.kr/data/15147894/fileData.do
Case 1	Gamuri PV Power Generation	KOSPO	https://www.data.go.kr/data/15109994/fileData.do
Case 2	Icheon-si Baeksa-myeon A PV Power Generation	KOSPO	https://www.data.go.kr/data/15109999/fileData.do
Case 3	Busan New Port PV Power Generation	KOSPO	https://www.data.go.kr/data/15043388/fileData.do
Case 4	Incheon Susan Water Purification Plant PV Power Generation	KOSPO	https://www.data.go.kr/data/15043402/fileData.do
Case 5	Yongsu-ri PV Power Generation	KOSPO	https://www.data.go.kr/data/15095404/fileData.do
Case 6	Hadong Public Stadium PV Power Generation	KOSPO	https://www.data.go.kr/data/15043404/fileData.do

Note: All website links were accessed on 20 August 2026.

References

1. de Vries, A. The Growing Energy Footprint of Artificial Intelligence. *Joule* **2023**, *7*, 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004.
2. Tian, J.; Ooka, R.; Lee, D. Multi-Scale Solar Radiation and Photovoltaic Power Forecasting with Machine Learning Algorithms in Urban Environment: A State-of-the-Art Review. *J. Clean. Prod.* **2023**, *426*, 139040. https://doi.org/10.1016/j.jclepro.2023.139040.
3. Yagli, G.M.; Yang, D.; Srinivasan, D. Automatic Hourly Solar Forecasting Using Machine Learning Models. *Renew. Sustain. Energy Rev.* **2019**, *105*, 487–498. https://doi.org/10.1016/j.rser.2019.02.006.
4. Lee, Y.; Choi, D.; Jung, Y.; Ko, M. Application of Technology to Develop a Framework for Predicting Power Output of a PV System Based on a Spatial Interpolation Technique: A Case Study in South Korea. *Energies* **2022**, *15*, 8755. https://doi.org/10.3390/en15228755.
5. Jamaly, M.; Kleissl, J. Spatiotemporal Interpolation and Forecast of Irradiance Data Using Kriging. *Sol. Energy* **2017**, *158*, 407–423. https://doi.org/10.1016/j.solener.2017.09.057.
6. Agoua, X.G.; Girard, R.; Kariniotakis, G. Photovoltaic Power Forecasting: Assessment of the Impact of Multiple Sources of Spatio-Temporal Data on Forecast Accuracy. *Energies* **2021**, *14*, 1432. https://doi.org/10.3390/en14051432.
7. Nam, S.; Hur, J. Probabilistic Forecasting Model of Solar Power Outputs Based on the Naïve Bayes Classifier and Kriging Models. *Energies* **2018**, *11*, 2982. https://doi.org/10.3390/en11112982.
8. Lu, G.Y.; Wong, D.W. An Adaptive Inverse-Distance Weighting Spatial Interpolation Technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010.
9. Arétouyap, Z.; Njandjock Nouck, P.; Nouayou, R.; Kemgang, F.E.G.; Toko, A.D.P.; Asfahani, J. Lessening the Adverse Effect of the Semivariogram Model Selection on an Interpolative Survey Using Kriging Technique. *SpringerPlus* **2016**, *5*, 549. https://doi.org/10.1186/s40064-016-2142-4.
10. Jian, X.; Olea, R.A.; Yu, Y.-S. Semivariogram Modeling by Weighted Least Squares. *Comput. Geosci.* **1996**, *22*, 387–397. https://doi.org/10.1016/0098-3004(95)00095-X.
11. Ly, S.; Charles, C.; Degré, A. Geostatistical Interpolation of Daily Rainfall at Catchment Scale: The Use of Several Variogram Models in the Ourthe and Ambleve Catchments, Belgium. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2259–2274. https://doi.org/10.5194/hess-15-2259-2011.
12. Aamodt, A.; Plaza, E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.* **1994**, *7*, 39–59. https://doi.org/10.3233/AIC-1994-7104.
13. Park, C.-S.; Han, I. A Case-Based Reasoning with the Feature Weights Derived by Analytic Hierarchy Process for Bankruptcy Prediction. *Expert Syst. Appl.* **2002**, *23*, 255–264. https://doi.org/10.1016/S0957-4174(02)00045-3.
14. Fei, L.; Feng, Y. A Novel Retrieval Strategy for Case-Based Reasoning Based on Attitudinal Choquet Integral. *Eng. Appl. Artif. Intell.* **2020**, *94*, 103791. https://doi.org/10.1016/j.engappai.2020.103791.

15. Lin, S.-W.; Chen, S.-C. Parameter Tuning, Feature Selection and Weight Assignment of Features for Case-Based Reasoning by Artificial Immune System. *Appl. Soft Comput.* **2011**, *11*, 5042–5052. https://doi.org/10.1016/j.asoc.2011.05.054.
16. Im, K.H.; Park, S.C. Case-Based Reasoning and Neural Network Based Expert System for Personalization. *Expert Syst. Appl.* **2007**, *32*, 77–85. https://doi.org/10.1016/j.eswa.2005.11.020.
17. Fisher, N.I. *Statistical Analysis of Circular Data*; Cambridge University Press: Cambridge, UK, 1993. https://doi.org/10.1017/CBO9780511564345.
18. Soukissian, T.; Apostolou, V. The Effects of Offshore Wind Interpolation Methods on Wind Power Density and Energy Assessment. *Energies* **2026**, *19*, 971. https://doi.org/10.3390/en19040971.
19. Cressie, N. Fitting Variogram Models by Weighted Least Squares. *J. Int. Assoc. Math. Geol.* **1985**, *17*, 563–586. https://doi.org/10.1007/BF01032109.
20. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. https://doi.org/10.1137/0916069.
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.
22. Fisher, A.; Rudin, C.; Dominici, F. All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
23. Mahmoudi, A.; Javed, S.A.; Liu, S.; Deng, X. Distinguishing coefficient driven sensitivity analysis of GRA model for intelligent decisions: Application in project management. *Technol. Econ. Dev. Econ.* **2020**, *26*, 621–641. https://doi.org/10.3846/tede.2020.11890.
24. Cover, T.M.; Hart, P.E. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. https://doi.org/10.1109/TIT.1967.1053964.
25. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
26. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. https://doi.org/10.1007/BF00994018.
27. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. https://doi.org/10.1023/A:1009715923555.
28. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. https://doi.org/10.1145/1961189.1961199.
29. Youn, Y.; Kafatos, M.; Kim, S.H.; Lee, Y. Observation-Based Reconstruction of High-Resolution Daily Temperature Field Using Lapse-Rate-Constrained Kriging in Complex Terrain: A Nationwide Dataset for South Korea. *Atmosphere* **2026**, *17*, 148. https://doi.org/10.3390/atmos17020148.
30. Olea, R.A. Ordinary Kriging. In *Geostatistics for Engineers and Earth Scientists*; Springer: Boston, MA, USA, 1999; pp. 39–65. https://doi.org/10.1007/978-1-4615-5001-3_4.
31. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*; ACM: New York, NY, USA, 1968; pp. 517–524. https://doi.org/10.1145/800186.810616.
32. Wong, D.W.S. Interpolation: Inverse-Distance Weighting. In *International Encyclopedia of Geography: People, the Earth, Environment and Technology*; Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A., Eds.; John Wiley & Sons, Ltd.: Chichester, UK, 2017. https://doi.org/10.1002/9781118786352.wbieg0066.
33. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.
34. Mellit, A.; Massi Pavan, A.; Lughi, V. Deep learning neural networks for short-term photovoltaic power forecasting. *Renew. Energy* **2021**, *172*, 276–288. https://doi.org/10.1016/j.renene.2021.02.166.
35. Pospíchal, J.; Kubovčík, M.; Dirgová Luptáková, I. Solar irradiance forecasting with transformer model. *Appl. Sci.* **2022**, *12*, 8852. https://doi.org/10.3390/app12178852.
36. Meisenbacher, S.; Turowski, M.; Phipps, K.; Rätz, M.; Müller, D.; Hagenmeyer, V.; Mikut, R. Review of automated time series forecasting pipelines. *WIREs Data Min. Knowl. Discov.* **2022**, *12*, e1475. https://doi.org/10.1002/widm.1475.
37. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
39. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. https://doi.org/10.1214/aoms/1177703732.
40. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. https://doi.org/10.1109/78.650093.
41. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

42. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. https://doi.org/10.3115/v1/D14-1179.
43. Wang, Y.; Liao, W.; Chang, Y. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies* **2018**, *11*, 2163. https://doi.org/10.3390/en11082163.
44. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. https://doi.org/10.1109/5.58325.
45. Gneiting, T. Nonseparable, stationary covariance functions for space-time data. *J. Am. Stat. Assoc.* **2002**, *97*, 590–600. https://doi.org/10.1198/016214502760047113.
46. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
47. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.