atmosphere

MDPI

Article

Spatialization Study of Monthly Global Solar Radiation in Sparse Observation Area Based on Environmental Similarity and Spatial Proximity

Mao-Fen Li [1], Peng-Tao Guo [1,*], A-Xing Zhu [2] and Xuan Yu [3]

[1] School of Geographical Sciences and Tourism/Jinsha River Basin Research Center, Zhaotong University, Zhaotong 657000, China; 20230012@ztu.edu.cn
[2] Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA; azhu@wisc.edu
[3] Institute of Scientific and Technical Information, Chinese Academy of Tropical Agriculture Sciences, Haikou 571101, China; xuanyu@catas.cn
* Correspondence: glpt2011@126.com; Tel.: +86-18217888998

Abstract

Global Solar Radiation (Rs) is essential for ecological and climatic modeling, yet its spatialization is often hampered by sparse observation networks. Conventional methods demand a well-distributed set of stations with global representativeness—a requirement rarely met in practice. To address this gap, we propose a spatialization method based on environmental similarity and spatial proximity (ES-SP), which integrates the Law of Geographic Similarity and Tobler's First Law of Geography. Using monthly Rs data from 11 stations in Tropical China (2015), we evaluated ES-SP against Ordinary Kriging (OK) and Local Polynomial Interpolation (LP) through leave-one-out cross-validation (LOOCV), with root mean square error (RMSE), relative RMSE, and mean absolute percentage error (MAPE) as accuracy metrics. Topographic and monthly meteorological covariates were selected dynamically via random forest (RF), and the performance differences among the three methods were tested statistically using the Wilcoxon signed-rank test. Results show that ES-SP outperforms both OK and LP in accuracy and stability, achieving the lowest error metrics in most months—e.g., RMSE as low as 37.23 MJ·m^{-2} in December and MAPE as low as 4.34% in August—along with a narrow interquartile range, indicating consistent performance across seasons. Spatially, ES-SP accurately reproduces the coastal–inland gradient during the rainy season (May) and the latitudinal gradient in the dry season (January), whereas OK yields overly smooth distributions that obscure local details, and LP exhibits extreme instability and unrealistic spatial discontinuities. The study demonstrates that the ES-SP method effectively overcomes the reliance on globally representative station samples, providing a robust technical pathway for generating continuous Rs datasets in data-sparse regions such as Tropical China. Further research should focus on extending the geographic scope and refining the covariate set to enhance generalizability.

Keywords: global solar radiation; spatial interpolation; environmental similarity; Tropical China

Check for updates

Academic Editor: Abd Al Karim Haj Ismail

Received: 14 January 2026
Revised: 8 February 2026
Accepted: 11 February 2026
Published: 12 February 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.

1. Introduction

Global Solar Radiation (Rs) is the ultimate energy source for all physical, chemical, and biological processes in the earth–atmosphere system. It serves as an indispensable fundamental dataset for the operation of crop models, ecological models, atmospheric

circulation models, and other models including models for assessing human exposure to UV radiation at different scales such as field, regional, and global levels [1–4]. Consequently, spatially continuous Rs datasets are a prerequisite for studying regional and global radiation changes, parameterizing environmental models, and understanding ecosystem dynamics [5].

However, Rs measurements from ground stations only represent conditions at specific points [6]. For vast, unmeasured areas, Rs values must be estimated indirectly, typically through spatial interpolation of data from discrete stations. The high cost and maintenance requirements of radiometric equipment have resulted in a sparse and uneven global distribution of Rs stations, particularly in developing countries [7,8]. In China, for instance, Rs stations constitute only about 13% of surface meteorological stations, with a denser concentration in the eastern regions compared to the west [9].

There have also been numerous studies on inferring the spatial distribution of Rs on horizontal surfaces using methods like land surface process models, climate models, and satellite remote sensing models. Existing Rs spatialization methods can be broadly categorized into three groups: (1) methods based on spatial auto-correlation (e.g., Kriging, IDW) [10–12]; (2) methods based on factor correlation using linear or non-linear models with meteorological parameters [13]; and (3) hybrid methods combining both spatial auto-correlation and factor correlation [14]. A common critical limitation across these methods is their reliance on a sufficiently large number of well-distributed stations to establish a reliable spatial variogram or a stable relationship with environmental factors—i.e., they require samples with good global representativeness [15,16]. This requirement is often unmet in practice.

To address this challenge, this study proposes a Rs estimation method based on environmental similarity and spatial proximity. Grounded in the Third Law of Geography [17] and Tobler's First Law of Geography [18], the method operates on the premise that greater environmental similarity and closer spatial distance between a target point and sample points lead to greater similarity in Rs values. Building on this premise, the product of environmental similarity and spatial proximity is used as a weighting factor, and the Rs at the target point is estimated via a weighted average of measurements from sample points. To validate the effectiveness of the method, it was applied to spatialize solar radiation in the tropical region of China for the year 2015. The main objectives of this study are: (1) to develop a novel Rs estimation method that integrates environmental similarity and spatial proximity (ES-SP); (2) to compare the predictive performance of this method with commonly used approaches, such as Ordinary Kriging (OK) and Local Polynomial Interpolation (LP), in order to evaluate its validity and advantages; and (3) to produce a high-quality spatially continuous Rs dataset for Tropical China in 2015 using limited station observations.

2. Materials and Methods

2.1. Study Area and Data

2.1.1. Study Area

The study area encompasses approximately 540,000 km^2 in Southern China (Figure 1), a region critical for national food and ecological security [19]. It features complex topography with a maximum relief of over 2000 m, making it an ideal testing ground for comparing spatialization methods.

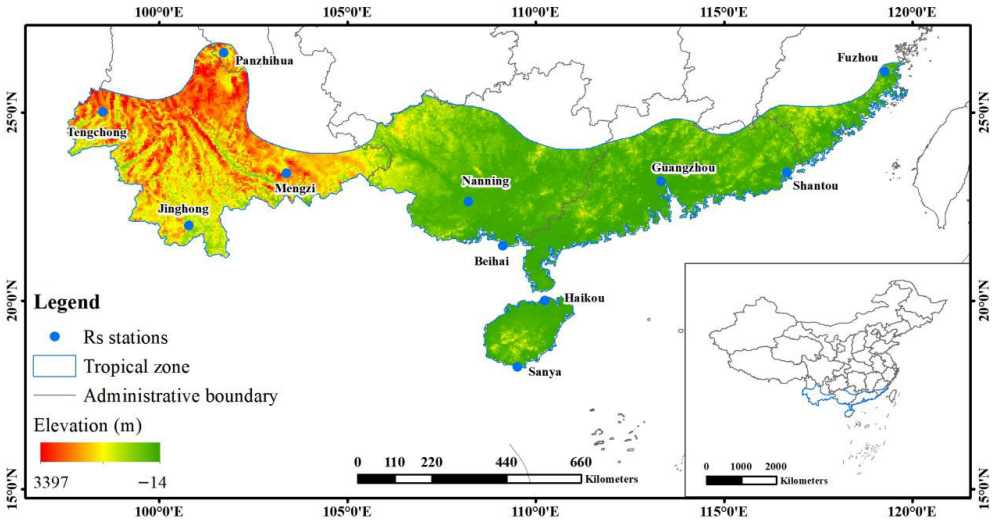


Figure 1. Study area of the present study.

2.1.2. Solar Radiation and Environmental Data

Monthly Rs data for 2015 were obtained from 11 radiation stations within the study area, provided by the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA). The data underwent strict quality control according to WMO standards.

Two categories of environmental covariates were collected, topographic covariates and meteorological covariates, as shown in Table 1. Topographic covariates were derived from a 1000 m resolution DEM, including elevation (Ele), slope (Slo), aspect, hillshade (Hishade), and topographic wetness index (TWI). Meteorological covariates including monthly maximum and minimum temperature (Tmx, Tnx), precipitation (Pre), sunshine duration (Sun), air pressure (Ap), water vapor pressure (Wp), and total cloud cover (TC) were obtained from CMA, calculated extraterrestrial radiation (R) and area solar radiation (AreaSol) through ArcGIS 10.8, respectively. Covariates of sunshine duration (Sun), air pressure (Ap), and water vapor pressure (Wp) were interpolated from a denser non-Rs station network; the station network and interpolation method are available in reference [19]. Covariate of aspect was converted to sine and cosine components. All the covariates were normalized to heterogeneous units and magnitudes.

Table 1. Environmental covariates used in the study.

Environmental Variables		Abbreviation of Environmental Variables and Unit
Topographic covariates	Elevation	Ele, m
	Slope	Slo, °
	Sine component of aspect	Aspectsin, /
	Cosine component of aspect	Aspectcos, /
	Hillshade	Hishade, °
	Topographic wetness index	Twi, -
Weather covariates	Monthly Maximum Temperature	Tmx, °C
	Monthly Minimum Temperature	Tnx, °C
	Monthly Precipitation	Pre, mm
	Monthly sunshine duration	Sun, h
	Monthly air pressure	Ap, hPa
	Monthly water vapor pressure	Wp, hPa
	Monthly total cloud cover	TC, %
	Monthly duration of possible sunshine	N, h
	Extraterrestrial solar radiation	R, MJ/m^2
	Area solar radiation	AreaSol, WH/m^2

2.2. Methods

2.2.1. Basic Idea and Overall Design

According to radiative transfer theory, empirical solar radiation (Rs) models rely on the relationship between Rs and the environmental factors influencing it [20]. This relationship allows the spatial distribution of Rs to be inferred from that of the environmental drivers—greater similarity in environmental conditions generally corresponds to greater similarity in Rs values [17]. Moreover, as a geographical environmental variable, the spatial distribution of Rs follows the First Law of Geography (Tobler's Law) [18], which states that nearer locations tend to exhibit more similar attributes. From this perspective, we consider the spatial distribution of Rs to be jointly shaped by environmental similarity and spatial proximity: the more alike the environmental conditions and the shorter the spatial distance, the more similar the Rs values are expected to be. Based on this premise, we propose a new algorithm for estimating solar radiation that integrates both environmental similarity and spatial proximity.

The overall procedure of ES-SP for estimating Rs at unobserved locations is as follows (Figure 2): First, environmental variables are identified to represent the conditions at known sample points and the target point. By comparing the similarity of environmental conditions, the representativeness of sample points to the target point—that is, the environmental similarity of the target variable—is assessed. Subsequently, the spatial proximity between sample points and the target point is calculated. The environmental similarity and spatial proximity are then multiplied to produce an integrated similarity measure. Based on the similarity scores, sample points that exhibit high integrated similarity with the target point are selected. These selected sample points are finally used to estimate the target variable value at the unobserved location.

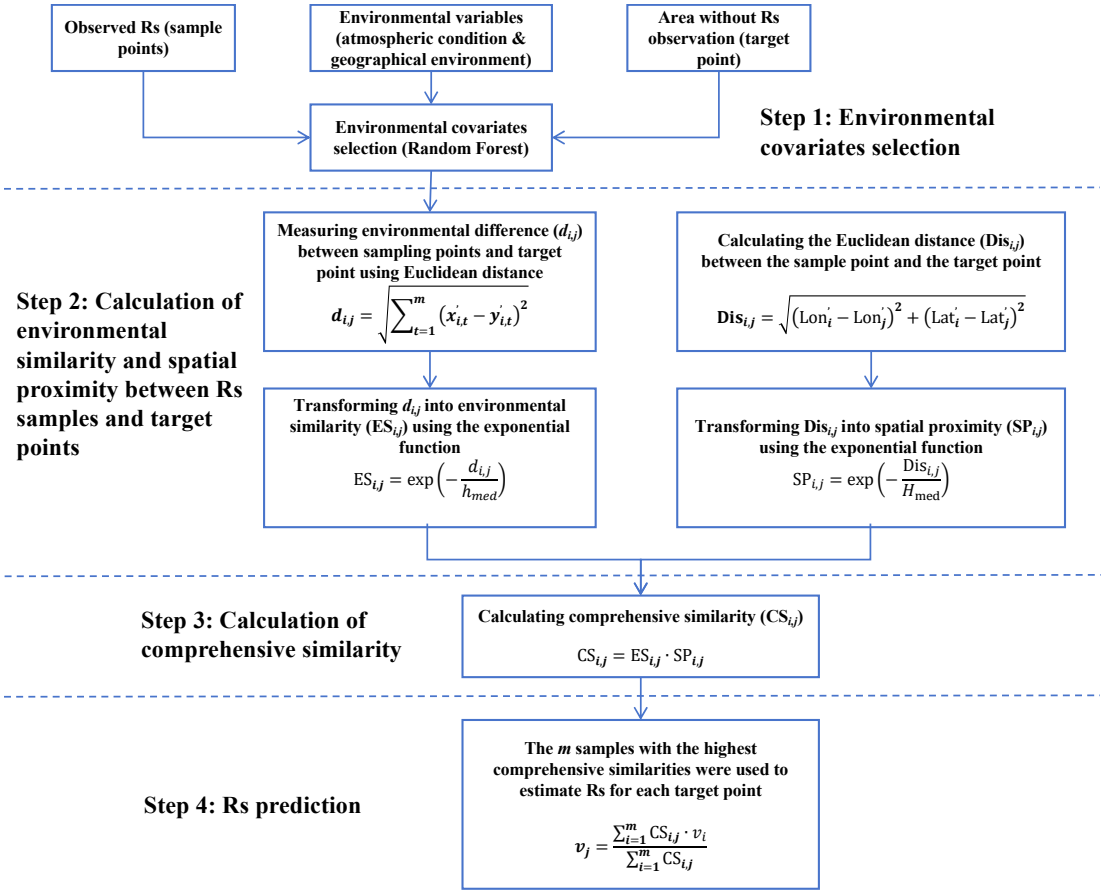


Figure 2. The flowchart of the environmental similarity and spatial proximity (ES-SP)-based approach.

2.2.2. Selection of Environmental Variables

This study employs a dynamic and adaptive environmental variable screening strategy, which is tightly embedded within a leave-one-out cross-validation framework. Specifically, for each sample point designated as the test case, a random forest regression model [21,22] is first constructed using the remaining samples as the training set. The out-of-bag permutation importance metric is then applied to rank all environmental variables based on their predictive contributions. Subsequently, to identify the optimal variable subset, a second round of leave-one-out cross-validation is performed internally on the training set. Following the importance ranking, variable subsets of increasing size—from one variable up to a predefined maximum—are sequentially tested, and the average prediction error for each subset is computed through internal validation. By comparing these errors, the subset size—and the corresponding subset of top-ranked variables—that minimizes the internal validation error is selected. Using only this optimized subset, environmental similarity between the test point and training samples is calculated, followed by prediction of the target variable. This dual cross-validation approach ensures that variable selection is rigorously data-driven and adaptively identifies the most explanatory combination of environmental factors for each spatial location.

2.2.3. Calculation of Environmental Similarity Between Rs Samples and Locations Without Observations

First, the Euclidean distance is used to calculate the environmental difference $d_{i,j}$ between the i-th observed Rs point and the j-th Rs estimation point. Prior to calculating $d_{i,j}$, scaling by standard deviation is applied to eliminate differences in measurement units among environmental variables. Specifically, for each environmental variable, the standard deviation σ_t is computed from the known samples, and all sample values—including those from known and unobserved points—are scaled by the corresponding standard deviation:

$$x'_{i,t} = \frac{x_{i,t}}{\sigma_t} \tag{1}$$

$$y'_{j,t} = \frac{y_{j,t}}{\sigma_t} \tag{2}$$

where $x'_{i,t}$ and $y'_{i,t}$ denote the standardized values of the t-th environmental variable at the i-th observation point and the j-th estimation point, respectively; $x_{i,t}$ and $y_{i,t}$ are the corresponding raw values; and σ_t is the standard deviation of the t-th environmental variable calculated from the known samples.

Based on the standardized environmental variables, the environmental difference $d_{i,j}$ between the i-th observed Rs sample and the j-th Rs estimation point is computed using the Euclidean distance formula:

$$d_{i,j} = \sqrt{\sum_{t=1}^{m} \left(x'_{i,t} - y'_{j,t}\right)^2} \tag{3}$$

where m represents the total number of environmental variables.

Subsequently, the environmental difference $d_{i,j}$ is converted into an environmental similarity measure $ES_{i,j}$ using an exponential function. This similarity reflects how closely the environmental conditions at the i-th known Rs point resemble those at the j-th prediction location. The approach has been previously applied to assess environmental similarity between sampled and unsampled locations [23]. The environmental similarity $ES_{i,j}$ is given by:

$$ES_{i,j} = \exp\left(-\frac{d_{i,j}}{h_{\text{med}}}\right) \tag{4}$$

where, exp(·) is the exponential function, and h_{med} is the bandwidth parameter. To dynamically determine an optimal bandwidth for each prediction point, a median-based adaptive bandwidth selection method is employed. The procedure is as follows: (1) For each target point z_0, compute the Euclidean distances of environmental variables to all training samples. (2) Exclude any zero-distance values to avoid self-comparison or duplicate points. (3) Use the median of the remaining distances as the bandwidth h_{med} for that point.

2.2.4. Calculation of Spatial Proximity Between Rs Samples and Locations Without Observations

After calculating environmental similarity, it is necessary to further compute spatial proximity. The calculation of spatial proximity begins with determining the Euclidean distance Dis between the sample point and the target point. Prior to this calculation, the original latitude and longitude coordinates must be standardized using the same method applied to environmental variables. This standardization step is critical: since raw spatial distances can vary widely (from a few kilometers to hundreds of kilometers), directly applying them to the subsequent exponential function would result in extreme weighting—nearby points would be assigned excessively high weights, while distant points would be weighted too weakly, distorting the reasonable distribution of spatial weights. Standardization transforms distance values to a relatively uniform scale, leading to a more balanced and robust characterization of spatial relationships.

Thus, the Euclidean distance $\text{Dis}_{i,j}$ between sample point i and target point j is calculated as follows:

$$\text{Dis}_{i,j} = \sqrt{(\text{Lon}'_i - \text{Lon}'_j)^2 + (\text{Lat}'_i - \text{Lat}'_j)^2} \tag{5}$$

where Lon'_i and Lat'_i represent the standardized longitude and latitude coordinates of the i-th sample point, and Lon'_j and Lat'_j denote the standardized longitude and latitude coordinates of the j-th target point.

After obtaining the spatial distance, it is similarly converted into spatial proximity $\text{SP}_{i,j}$ using an exponential function:

$$\text{SP}_{i,j} = \exp\left(-\frac{\text{Dis}_{i,j}}{H_{\text{med}}}\right) \tag{6}$$

where, $\text{SP}_{i,j}$ denotes the spatial proximity between the i-th sample point and the j-th target point; and H_{med} is the bandwidth parameter which is determined using the same method as the bandwidth for environmental similarity in Equation (4).

2.2.5. Calculation of Comprehensive Similarity

Following the approach of Qin et al. [24], the comprehensive similarity $\text{CS}_{i,j}$ between the i-th sample point and the j-th target point is obtained by multiplying the environmental similarity $\text{ES}_{i,j}$ by the spatial proximity $\text{SP}_{i,j}$. The calculation is expressed as:

$$\text{CS}_{i,j} = \text{ES}_{i,j} \cdot \text{SP}_{i,j} \tag{7}$$

where $\text{CS}_{i,j}$ denotes the comprehensive similarity, $\text{ES}_{i,j}$ the environmental similarity, and $\text{SP}_{i,j}$ the spatial proximity between the two points.

2.2.6. Rs Spatialization

For the j-th target point, if the m samples with the highest comprehensive similarity (where m is less than the total number of samples n) are selected for estimation, the predicted solar radiation value v_j at the target point can be calculated as a weighted average based

on the measured solar radiation values v_i of these m samples and their comprehensive similarity $CS_{i,j}$ to the target point. The formula is as follows [23]:

$$v_j = \frac{\sum_{i=1}^{m} CS_{i,j} \cdot v_i}{\sum_{i=1}^{m} CS_{i,j}} \quad (8)$$

where, v_j represents the predicted solar radiation value at the j-th target point; v_i is the measured solar radiation value at the i-th sample point; and m is the number of similar samples used to estimate the target value. The optimal value of m is determined through cross-validation: values from 1 to 10 are tested iteratively, and the value of m that yields the smallest prediction error is selected as the optimal parameter.

In the present study, two other spatial interpolation methods were implemented and compared. The first is OK, a geostatistical method that relies on spatial autocorrelation modeled by a variogram [25], and the other is LP, a deterministic method that fits a polynomial surface within a moving window, suitable for capturing local trends [26]. Both OK and LP were implemented in the Geostatistical Analyst extension of ArcGIS 10.8, with all parameters tuned to minimize prediction error. Through parameter calibration, the polynomial order of the LP model was set to 1 for all months, the neighborhood type was specified as standard, and the kernel function varied by month. Exponential for January, March, August to December; Gaussian for months of February, April, May and July; and Epanechnikov for June. For the OK model across all months, the trend type was set to Constant, anisotropy was 0, and the variogram model was Stable (Exponential for September and Gaussian for November); the kernel function was Exponential (Constant for April and Gaussian for November).

2.2.7. Validation and Accuracy Assessment

The limited number of samples (only 11 global solar radiation observation stations in the study area) made it impossible to split the dataset into an independent training set and an external test set (a basic requirement for the gold standard of model validation). And the small sample size brought multiple non-negligible limitations to the model validation, accuracy assessment and subsequent parameter optimization processes. The leave-one-out cross-validation (LOOCV) used in this study and its results can only reflect the interpolation ability of the model for the existing sample points, but cannot fully verify the prediction ability of the model for the unobserved points in the entire study area, which is a key limitation for the generalization of the model results.

Three error indices, namely root mean squared error (RMSE), relative RMSE (rRMSE), and mean absolute percentage error (MAPE), were used to evaluate the in situ performances of three Rs spatialization models. The smaller the value of these indicators, the better the estimation accuracy. Specifically, RMSE reflects the overall deviation between predicted and measured values, MAPE quantifies the relative error level, and rRMSE comprehensively evaluates the stability and accuracy of spatialization methods by combining deviation and dispersion characteristics.

2.2.8. Statistical Test

To scientifically compare the accuracy differences among three spatialization methods, namely OK, LP and ES-SP, a hierarchical statistical analysis process of hypothesis-testing–nonparametric paired comparison was constructed. All tests were implemented by programming on the MATLAB R2023a platform. The Shapiro–Wilk test [27] was adopted to verify the data distribution characteristics, which is particularly suitable for small-sample analysis with $n \leq 50$.

The first-order autocorrelation coefficient of the sequence was calculated to judge data independence, avoiding the interference of sequence correlation on the test results. A judgment threshold of $|ac| < 0.3$ was set. This threshold is consistent with the common criteria in spatial data analysis to ensure the reliability of subsequent paired comparison results. Since the sample data did not satisfy the normality assumption, the Wilcoxon signed-rank test [28] was used for paired sample comparison. This method is a standard approach for analyzing differences between paired samples under the nonparametric test framework, which does not require the data to follow a specific distribution and is robust to outliers.

The small sample size may lead to a lower statistical power of the nonparametric test. This means that the test may fail to detect the potential small-to-medium magnitude of actual differences between the models. For pairwise comparisons among the three methods (OK vs. LP, OK vs. ES-SP, LP vs. ES-SP), the Bonferroni correction method [29] was used to adjust the significance level, and the corrected threshold was $a = 0.05/3 \approx 0.0167$. This correction effectively reduces the risk of type I errors (false positives) caused by multiple comparisons, ensuring the rigor of the statistical conclusions. Meanwhile, the effect size r was calculated to evaluate the practical significance of the differences. The classification criteria of the effect size are as follows: $|r| < 0.1$ indicates no effect, $0.1 \leq |r| < 0.3$ indicates a small effect, $0.3 \leq |r| < 0.5$ indicates a medium effect, and $|r| \geq 0.5$ indicates a large effect. The effect size supplements the statistical significance, reflecting the actual impact of the differences between methods in practical applications.

3. Results

3.1. Performance Comparison of In Situ Rs Estimation

Figure 3 compares the performance of three spatial methods, OK, LP, and ES-SP in predicting monthly Rs across 12 months. Overall, ES-SP demonstrated superior performance in most months, with a consistently lower error index compared to OK and LP, while LP exhibited the most significant monthly fluctuations in spatialization accuracy.

In terms of RMSE (Figure 3), ES-SP achieved the lowest values in all months except April, where its RMSE (42.48 MJ/m^2) was slightly higher than that of LP (41.59 MJ/m^2). Notably, ES-SP showed remarkable advantages in May and December, with RMSE values of 51.01 and 37.23 MJ/m^2 respectively, far lower than those of OK (89.18 and 80.25 MJ/m^2) and LP (93.51 and 95.66). In contrast, LP displayed extreme instability, with the highest RMSE (95.66 MJ/m^2) in December and the lowest (24.51 MJ/m^2) in August, indicating significant sensitivity to seasonal variations in Rs.

The rRMSE results (Figure 3) were consistent with RMSE, reflecting the relative deviation between predicted and measured Rs values. ES-SP maintained the lowest rRMSE across most months, particularly in May (9.12%) and December (13.07%), which were less than half of the corresponding values of OK (15.94% and 28.16%). OK and LP showed similar rRMSE patterns, with no significant differences in most months, especially in January and June, where both methods had nearly identical rRMSE values.

For MAPE (Figure 3), which quantifies the relative error level, ES-SP also outperformed the other two methods. The lowest MAPE value of ES (4.34%) was observed in August, while its highest value (12.56%) in March was still lower than the highest MAPE values of OK (24.87% in December) and LP (26.49% in December). LP exhibited relatively low MAPE in August (3.80%) but poor performance in other months, whereas OK showed moderate but stable MAPE variations throughout the year.

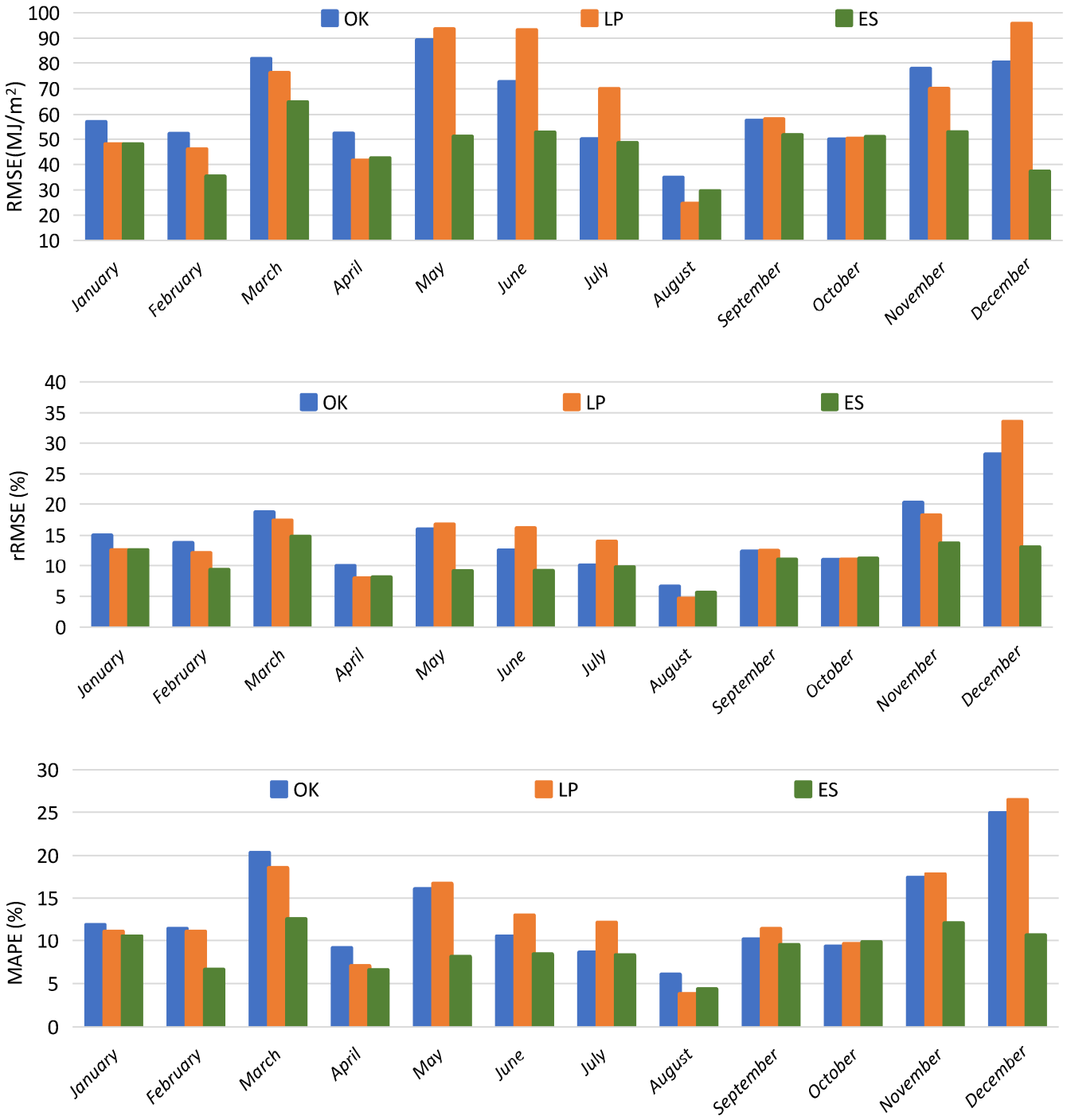


Figure 3. Comparative performance of OK, LP and ES-SP for Rs spatialization across months in 2015.

To statistically confirm performance differences among the three methods and characterize error distributions, a Wilcoxon signed-rank test with Bonferroni correction ($p = 0.05/3 \approx 0.0167$) was conducted (Table 2), paired with boxplot analysis of error metrics (Figure 4). Pairwise comparisons across RMSE, MAPE, and rRMSE were performed with effect size r to assess practical significance, while boxplots illustrated error distribution, median, and dispersion, jointly verifying accuracy and stability differences.

Table 2. Wilcoxon Signed-Rank Test Results ($p = 0.05/3$).

Comparison Group	Error Index	z-Statistic	*p*-Value	Effect Size *r*	Significance	Effect Size Magnitude
LP vs. OK	RMSE	0.039	0.9687	0.011	Not significant	No effect
	MAPE	−0.353	0.724	0.102	Not significant	Small effect
	rRMSE	0.039	0.969	0.011	Not significant	No effect
ES-SP vs. OK	RMSE	−2.942	0.003	0.849	Significant	Large effect
	MAPE	−2.863	0.004	0.827	Significant	Large effect
	RRMSE	−2.942	0.003	0.849	Significant	Large effect
ES-SP vs. LP	RMSE	−2.236	0.025	0.645	Not significant	Large effect
	MAPE	−2.706	0.007	0.781	Significant	Large effect
	RRMSE	−2.236	0.025	0.645	Not significant	Large effect

Note: Effect size magnitude: $|r| < 0.1$ (no effect), $0.1 \leq |r| < 0.3$ (small effect), $0.3 \leq |r| < 0.5$ (medium effect), $|r| \geq 0.5$ (large effect).

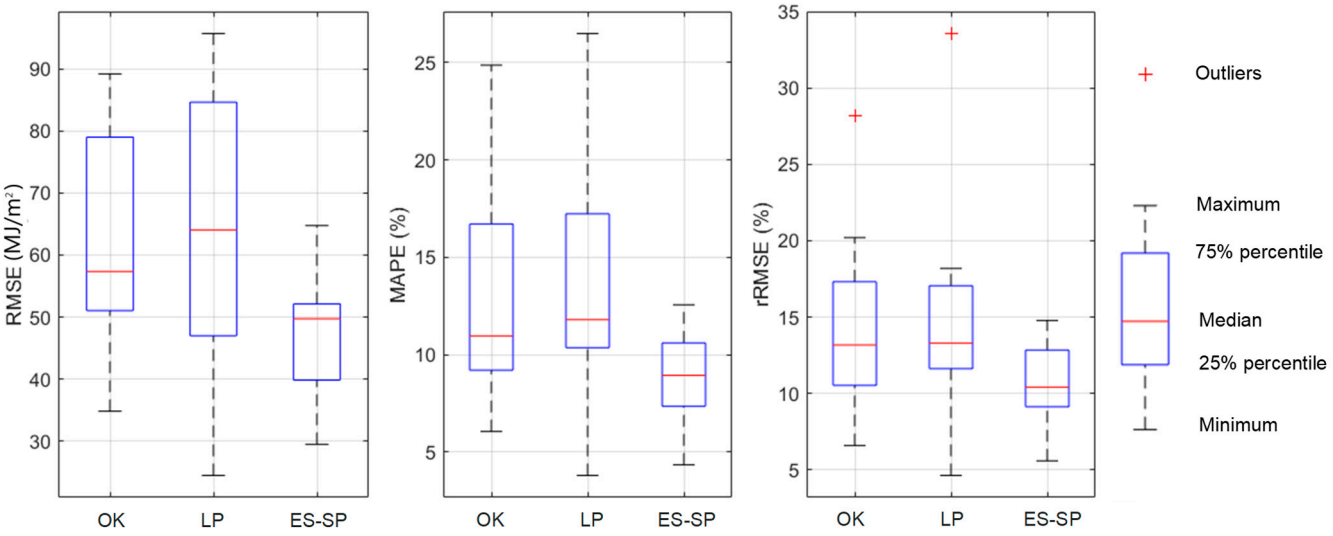


Figure 4. Boxplot of error indicators from 3 prediction methods: OK, LP and ES-SP.

Overall, consistent results were observed across RMSE, MAPE, and rRMSE. No significant differences were found between OK and LP ($p > 0.0167$), with no practical effect for RMSE/rRMSE and a small effect for MAPE, which was supported by their similar error distributions in Figure 4. In contrast, ES-SP differed significantly from OK ($p < 0.0167$), with large effect sizes for all metrics, and although not statistically significant with LP, it showed substantial practical differences (large effect sizes). Boxplot analysis confirmed ES-SP's superiority, characterized by the lowest median error and narrowest IQR (high stability), while LP exhibited the widest IQR (greatest variability), and OK showed moderate stability with higher median errors than ES-SP.

3.2. Performance Comparison of Rs Spatialization

Figures 5 and 6 present the spatial distribution patterns of solar radiation (Rs) estimated by the three methods for May (rainy season) and January (dry season), respectively. The overall trends are generally consistent across the methods.

During May (rainy season, Figure 5), all three methods show broadly similar spatial trends: lower values in coastal areas (e.g., Fuzhou, Shantou) and higher values in inland mountainous regions (e.g., Panzhihua, Tengchong). This pattern aligns with actual climatic conditions, where inland areas experience less cloud cover and coastal zones receive more rainfall. However, the OK method predicts more extreme values, with a lower minimum (302.43 MJ/m^2) and a higher maximum (755.88 MJ/m^2). Both OK and LP methods produce

relatively continuous and smooth spatial distributions. In contrast, the ES-SP method yields a less smooth surface with noticeable patchiness.

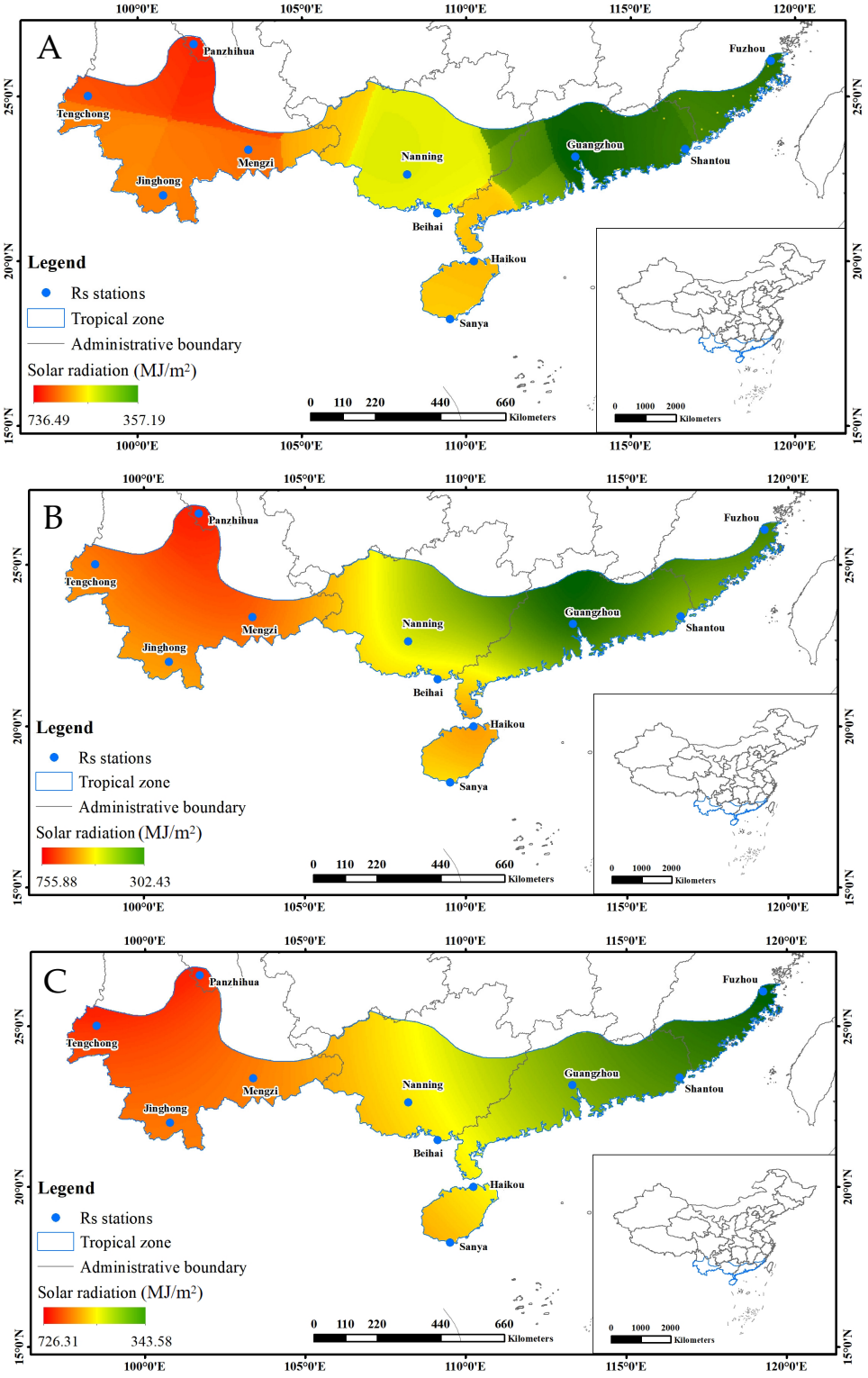


Figure 5. The spatial distribution of Rs in May in Tropical China predicted by three methods: environmental similarity and spatial proximity (**A**), Ordinary Kriging (**B**), Local Polynomial (**C**).

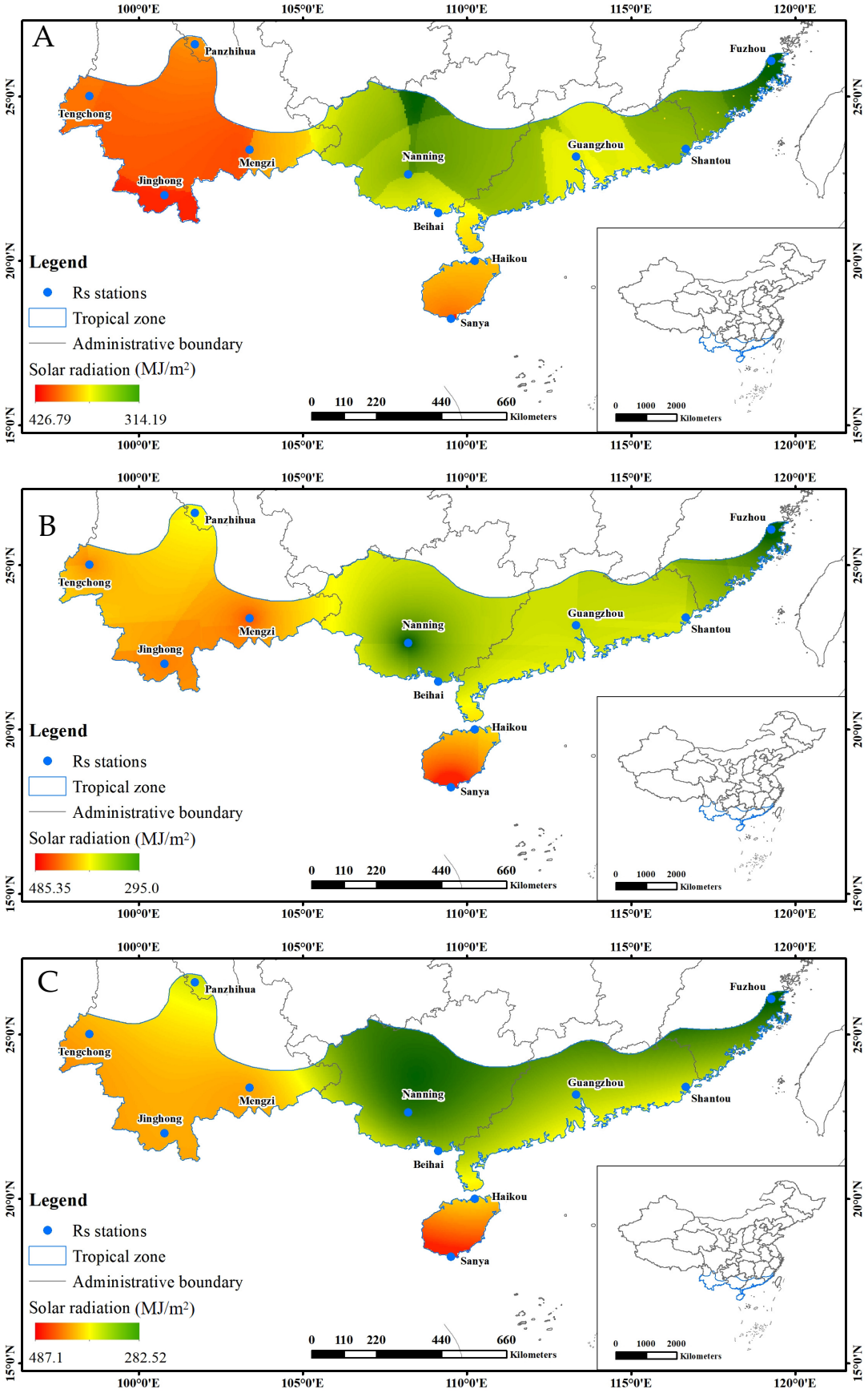


Figure 6. The spatial distribution of Rs in January in Tropical China predicted by three methods: environmental similarity and spatial proximity (**A**), Ordinary Kriging (**B**), Local Polynomial (**C**).

In January (dry season, Figure 6), the overall trends remain consistent among the three methods. The LP method generates the smoothest spatial distribution and the lowest minimum value (282.52 MJ/m^2). The OK method produces moderate Rs values but with less smoothness compared to LP. The ES-SP method results in the least smooth and most patchy distribution among the three.

Overall, the ES-SP method exhibits discontinuous and patchy spatial patterns in both seasons. This is likely due to the spatial proximity weighting introduced by ES-SP under sparse station density. The distances between Rs stations are much larger than the actual variation scale of Rs, making it difficult to capture its natural gradual transition. Moreover, spatial weighting amplifies the "zone of influence" around each station, causing predicted values within each zone to converge toward the station measurement. At zone boundaries, abrupt shifts between dominating stations lead to sharp transitions in predicted values, resulting in a discretized spatial pattern.

4. Discussion

This study develops a solar radiation (Rs) spatialization method that integrates environmental similarity (ES) and spatial proximity (SP), referred to as ES-SP. Its performance was evaluated against OK and LP methods in Tropical China, a region characterized by sparse station coverage and complex topography. The results demonstrate that ES-SP outperforms both conventional methods in terms of accuracy, stability, and spatial representativeness, effectively overcoming a key limitation of existing approaches: their heavy reliance on the global representativeness of station samples.

The superior performance of ES-SP stems from its unique methodological design. By employing Euclidean distance and an exponential function to quantify environmental similarity and spatial proximity respectively, the method reduces dependence on sample size—a constraint that strongly affects OK and LP. More importantly, ES-SP simultaneously incorporates the influence of environmental covariates (e.g., precipitation, cloud cover, elevation) and spatial neighborhood relationships. This dual mechanism distinguishes it from OK, which relies solely on spatial autocorrelation, and from LP, which focuses on local polynomial fitting. As a result, ES-SP captures both large-scale environmental controls and fine-scale spatial variations, leading to lower prediction errors and more realistic spatial patterns. For instance, ES-SP accurately reproduced Rs gradients during both the rainy (May) and dry (January) seasons, reflecting its ability to effectively integrate meteorological covariates (e.g., total cloud cover, precipitation) and topographic factors (e.g., elevation)—capabilities that OK and LP lack. These findings align with radiative transfer theory, which emphasizes the combined influence of environmental and spatial factors on Rs [20].

Our results are consistent with previous studies highlighting the limitations of traditional spatialization methods in data-sparse regions. For example, Kumari et al. [30] observed that OK tends to produce overly smoothed distributions in mountainous areas due to insufficient consideration of environmental covariates, matching our observation that OK poorly represents local details. Similarly, LP showed the largest interquartile range (IQR) and the highest monthly RMSE fluctuations in our study, indicating relatively lower stability. In contrast, ES-SP mitigates these issues through adaptive bandwidth selection and dynamic covariate screening, making it more suitable for regions with sparse station networks. This conclusion is supported by Qin et al. [24], who demonstrated that similarity-based methods outperform traditional interpolation when spatializing environmental variables with limited samples.

Despite its advantages, several limitations of this study should be noted. First, the small sample size (only 11 Rs stations) may restrict the generalizability of the results. Although LOOCV was applied, a larger station network would improve the robustness

of parameterization (e.g., bandwidth selection and covariate optimization). Second, the study area is confined to Tropical China; future work should test ES-SP in other climatic zones (e.g., arid or temperate regions) to verify its broader applicability. Third, the set of environmental covariates used here remains limited. Incorporating remote sensing data (e.g., MODIS cloud products, NDVI) could further enhance performance by improving the characterization of environmental similarity. Fourth, the computational efficiency of ES-SP requires optimization, as its adaptive bandwidth selection and covariate screening steps are more time-consuming than those of OK and LP.

Based on these findings, future research directions may include: (1) expanding the sample size and study area to improve generalization; (2) integrating remote sensing and ground-based data to optimize the environmental covariate system; (3) enhancing the computational efficiency of ES-SP through algorithmic improvements; (4) combining ES-SP with machine learning techniques (e.g., random forest, neural networks) to further increase prediction accuracy; and (5) extending ES-SP to the spatialization of other environmental variables, such as air temperature and precipitation, to explore its wider application potential.

5. Conclusions

Based on a systematic comparison of the ES-SP, OK, and LP methods for the spatialization of Rs in Tropical China—through analysis of monthly data from 11 stations in 2015, LOOCV, statistical tests, and spatial pattern evaluation—the following conclusions are drawn. The ES-SP method outperformed both OK and LP in terms of prediction accuracy and stability, achieving the lowest error metrics in most months, e.g., a minimum RMSE of 37.23 $MJ \cdot m^{-2}$ and a minimum MAPE as low as 4.34%, along with the narrowest interquartile range (IQR) in error distribution, demonstrating stable performance. Statistical tests confirmed significant differences between ES-SP and OK across all error metrics ($p < 0.0167$) with large effect sizes. For ES-SP versus LP, while MAPE showed a statistically significant difference ($p < 0.0167$) with large effect size, RMSE and rRMSE did not reach statistical significance ($p > 0.0167$) despite also exhibiting large effect sizes, indicating strong practical differences, albeit with limited statistical power under the small sample size. In terms of spatial representation, ES-SP accurately captured the coastal–inland gradient during the rainy season and the latitudinal gradient in the dry season, whereas OK produced overly smooth surfaces that obscured local details, and LP yielded unrealistic extreme values and spatial discontinuities. By integrating environmental similarity and spatial proximity, ES-SP effectively mitigates the reliance of conventional methods on a globally representative station network, making it particularly suitable for regions with sparse observations and complex topography. This approach provides a reliable technical pathway for generating continuous solar radiation surfaces in Tropical China and other similar regions, supporting ecological, agricultural, and climate-related research. Future work may further extend its applicability and potential by incorporating remote sensing data, expanding sample sizes, and improving computational efficiency.

Author Contributions: Conceptualization, A.-X.Z. and P.-T.G.; methodology, P.-T.G. and M.-F.L.; formal analysis, M.-F.L. and X.Y.; writing, M.-F.L.; visualization, M.-F.L. and X.Y.; data curation, P.-T.G. and M.-F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the PhD Research Startup Grant and the Yunnan Provincial Department of Science and Technology (Grant No. 202401BA070001-123).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

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

References

1. Qin, J.; Chen, Z.; Yang, K.; Liang, S.; Tang, W. Estimation of monthly-mean daily global solarradiation based on MODIS and TRMM products. *Appl. Energy* **2011**, *88*, 2480–2489. [CrossRef]
2. Li, M.F.; Guo, P.T.; Dai, S.; Luo, H.; Liu, E. Empirical estimation of daily global solar radiation with contrasting seasons of rain and drought characterize over tropical China. *J. Clean. Prod.* **2020**, *266*, 121915. [CrossRef]
3. Salvadori, G.; Leccese, F.; Lista, D.; Burattini, C.; Bisegna, F. Use of smartphone apps to monitor human exposure to solar radiation: Comparison between predicted and measured UV index values. *Environ. Res.* **2020**, *183*, 10942. [CrossRef] [PubMed]
4. Visser, W.I.; Amien, A.; Moola, H.; Naidoo, K. Visible Light Protection Strategies for Diverse Populations. *Dermatol. Ther.* **2026**, *102*, 67–78. [CrossRef]
5. Luković, J.B.; Bajat, B.J.; Kilibarda, M.S.; Filpović, D.J. High Resolution Grid of Potential Incoming Solar radiation for Serbia. *Therm. Sci.* **2015**, *19*, 427–435. [CrossRef]
6. Abed, A.; Khanjer, E.F.; Abdullah, S.A. Evolution and set up the maps for solar radiation of Iraq using Data observation and Angstrom model during monthly July2017. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing Ltd.: Bristol, UK, 2020; Volume 757, p. 012038. [CrossRef]
7. Wu, L.; Huang, G.; Fan, J.; Zhang, F.; Wang, X.; Zeng, W. Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Convers. Manag.* **2019**, *183*, 280–295. [CrossRef]
8. Zhou, Y.; Wang, D.; Liu, Y.; Liu, J. Diffuse solar radiation models for different climate zones in China: Model evaluation and general model development. *Energy Convers. Manag.* **2019**, *185*, 518–536. [CrossRef]
9. Feng, Y.; Hao, W.; Li, H.; Cui, N.; Gong, D.; Gao, L. Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renew. Sustain. Energy Rev.* **2020**, *118*, 109393. [CrossRef]
10. Bessafi, M.; Oree, V.; Khoodaruth, A.; Chabriat, J. Impact of Decomposition and Kriging Models on The Solar Irradiance Downscaling Accuracy in Regions with Complex Topography. *Renew. Energy* **2020**, *162*, 1992–2003. [CrossRef]
11. Bayray, M.; Gebreyohannes, Y.; Gebrehiwot, H.; Teklemichael, S.; Mustefa, A.; Haileslassie, A.; Gebray, P.; Kebedom, A.; Filli, F. Temporal and Spatial Solar Resource Variation By Analysis of Measured Irradiance in Geba Catchment, North Ethiopia. *Sustain. Energy Technol. Assess.* **2021**, *44*, 101110. [CrossRef]
12. Song, Z.; Cao, S.; Yang, H. Assessment of solar radiation resource and photovoltaic power potential across China based on optimized interpretable machine learning model and GIS-based approaches. *Appl. Energy* **2023**, *339*, 121005. [CrossRef]
13. Chen, J.-L.; Li, G.-S. Parameterization and Mapping of Solar Radiation in Data Sparse Regions. *Asia-Pac. J. Atmos. Sci.* **2012**, *48*, 423–431. [CrossRef]
14. Meher, C.; Yves, G.; Jompob, W. Solar radiation mapping using sunshine duration-based models and interpolation techniques: Application to Tunisia. *Energy Convers. Manag.* **2015**, *101*, 203–215. [CrossRef]
15. Bhowmik, A.K.; Costa, A.C. Representativeness impacts on accuracy and precision of climate spatial interpolation in data-scarce regions. *Meteorol. Appl.* **2015**, *22*, 368–377. [CrossRef]
16. Remus, P.; Cristian, P.; Georgeta, B. Spatial assessment of solar energy potential at global scale. A geographical approach. *J. Clean. Prod.* **2019**, *209*, 692–721. [CrossRef]
17. Zhu, A.X.; Lu, G.; Liu, J.; Qin, C.Z.; Zhou, C. Spatial prediction based on Third Law of Geography. *Ann. GIS* **2018**, *24*, 225–240. [CrossRef]
18. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [CrossRef]
19. Yu, X.; Yi, X.; Li, M.-F.; Dai, S.; Li, H.; Luo, H.; Zheng, Q.; Hu, Y. Calibration of the Ångström–Prescott Model for Accurately Estimating Solar Radiation Spatial Distribution in Areas with Few Global Solar Radiation Stations: A Case Study of the China Tropical Zone. *Atmosphere* **2023**, *14*, 1825. [CrossRef]
20. Krishnan Kumar, K.R.; Inda, C.S. How solar radiation forecasting impacts the utilization of solar energy: A critical review. *J. Clean. Prod.* **2023**, *338*, 135860. [CrossRef]
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
22. Liaw, A.; Wiener, M. Package "randomForest"—Breiman and cutler's random forests for classification and regression. *R News* **2022**, *2*, 18–22.
23. Zhu, A.X.; Liu, J.; Du, F.; Zhang, S.J.; Qin, C.Z.; Burt, J.; Behrens, T.; Scholten, T. Predictive soil mapping with limited sample data. *Eur. J. Soil Sci.* **2015**, *66*, 535–547. [CrossRef]

24. Qin, C.Z.; An, Y.M.; Liang, P.; Zhu, A.X.; Yang, L. Soil property mapping by combining spatial distance information into the Soil Land Inference Model (SoLIM). *Pedosphere* **2021**, *31*, 638–644. [CrossRef]
25. Margaret, A.O.; Richard, W. *Basic Steps in Geostatistics: The Variogram and Kriging*; Springer International Publishing: Cham, Switzerland, 2015.
26. Gouri, S.B.; Pravat, K.S.; Ramkrishna, M. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *J. Saudi Soc. Agric. Sci.* **2018**, *17*, 114–126. [CrossRef]
27. Avram, C.; Mărușteri, M. Normality Assessment, Few Paradigms and Use Cases. *Rev. Română Med. Lab.* **2022**, *30*, 251–259. [CrossRef]
28. Conover, W.J. On Methods of Handling Ties in the Wilcoxon Signed-Rank Test. *J. Am. Stat. Assoc.* **1973**, *68*, 985–988. [CrossRef]
29. Armstrong, R.A. When to use the Bonferroni correction. *Ophthalmic Physiol. Opt.* **2014**, *34*, 502–508. [CrossRef]
30. Kumari, M.; Basistha, A.; Bakimchandra, O.; Singh, C.K. Comparison of Spatial Interpolation Methods for Mapping Rainfall in Indian Himalayas of Uttarakhand Region. In *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment*; Raju, N., Ed.; Springer International Publishing: Cham, Switzerland, 2016; pp. 159–168.

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.