hydrology

MDPI

Article

Interpolation Strategy Selection for Areal Rainfall Estimation in an Extremely Sparse-Gauge Small Catchment: An Event-Scale Comparison Using Gauge and Radar References

Yongli Ma [1,2], Cheng Chen [1,2], Furong Xu [1,2], Haigang Li [3], Xiaojun Zhang [3], Yanzhi Liu [4], Qinghui Jiang [5] and Xiaobo Zhang [1,2,*]

[1] School of Infrastructure Engineering, Nanchang University, Nanchang 330031, China
[2] Jiangxi Provincial Key Laboratory of Hydraulic Geotechnical Engineering and Safety, Nanchang University, Nanchang 330031, China
[3] Jiangxi Academy of Emergency Management Science, Nanchang 330038, China
[4] Jiangxi Duchang Jinding Tungsten Molybdenum Mining Co., Ltd., Jiujiang 332611, China
[5] School of Civil Engineering, Wuhan University, Wuhan 430072, China
* Correspondence: zhangxb@ncu.edu.cn

Abstract

Accurate areal rainfall estimation is essential for hydrological modeling and flood forecasting, yet method selection remains uncertain in small catchments with extremely sparse gauge networks. This event-scale study compared arithmetic mean (AM), Thiessen polygon (TP), inverse distance weighting (IDW), precipitation–elevation linear regression (ELR), multiple linear regression (MLR), and a multi-layer perceptron (MLP) in a 0.719 km^2 catchment monitored by three gauges. Two complementary evaluations were conducted. Station-wise leave-one-out cross-validation (LOOCV) assessed prediction at an omitted gauge, whereas a radar-referenced comparison assessed catchment-average estimates obtained from the complete gauge network. MLR produced the lowest LOOCV error (RMSE = 0.433 mm; CC = 0.777). In the radar comparison, MLP and MLR produced nearly identical RMSE values of 0.668 and 0.669 mm, respectively, and are therefore interpreted as practically similar rather than meaningfully different. All method rankings are conditional on the selected 60 h event, the three-gauge arrangement, and uncertainty in the radar reference. The findings demonstrate that station-omission performance and full-network areal estimation address different operational questions and should be considered together when selecting an interpolation method for extremely sparse networks.

Keywords: areal rainfall; spatial interpolation; sparse rain-gauge network; leave-one-out cross-validation; small catchment

Check for updates

Academic Editor: Ali A. Assani

Received: 17 July 2026
Revised: 31 August 2026
Accepted: 8 September 2026
Published: 11 September 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

Areal rainfall is a fundamental hydrological input for runoff generation, flood forecasting, reservoir operation, soil-moisture estimation, and water-resource management. Because precipitation varies substantially in space while gauges provide point observations, transforming point rainfall into a catchment-average estimate remains a persistent source of uncertainty in hydrological analysis [1–3]. This uncertainty is particularly consequential in small catchments, where localized rainfall variability can dominate the runoff response [2,3].

A wide range of methods has been developed for areal rainfall estimation. Deterministic approaches, including arithmetic mean (AM), Thiessen polygon (TP), and inverse distance

weighting (IDW), remain common because they are transparent and computationally inexpensive [1,4–7]. TP represents gauges through area weights, whereas IDW assigns greater influence to nearby observations. Regression-based approaches incorporate geographic or topographic covariates, most commonly elevation and planar coordinates [8–10]. More recently, machine-learning methods have been applied to spatial interpolation because of their capacity to represent nonlinear relations and interactions among covariates [11–13]. Their performance, however, depends strongly on sample density, predictor relevance, and the spatial configuration of observations [11,12].

Most comparative studies have been conducted with monitoring networks containing substantially more than three gauges [9,14–17]. Such networks provide enough independent spatial locations to estimate model parameters and evaluate spatial dependence. Many engineering catchments—including small reservoirs, tailings facilities, mountain valleys, and remote headwaters—are instead monitored by only a few gauges because of economic, access, and maintenance constraints. Under these conditions, geometric weighting may become highly sensitive to station placement, topographic regression may be poorly identified, and flexible learning algorithms may overfit the limited spatial information. Evidence on method behavior under this degree of data scarcity remains limited [6,14].

A further issue is the choice of validation framework. Rainfall-interpolation studies commonly rely on internal station-wise cross-validation [9,15,16]. Although leave-one-out cross-validation (LOOCV) measures the ability to predict at an omitted gauge, it also reduces the already small number of independent station locations available for model construction. The resulting ranking may therefore differ from operational performance when all gauges are used. Independent comparisons against radar-derived areal precipitation or denser temporary networks provide a complementary perspective, but radar estimates have their own measurement and retrieval uncertainties [3,18,19]. The interaction between validation design, network density, and model complexity has not been sufficiently examined for extremely sparse small catchments [20–23].

This study evaluates areal rainfall estimation in a 0.719 km^2 catchment equipped with three automatic rain gauges. Six approaches—AM, TP, IDW, precipitation–elevation linear regression (ELR), multiple linear regression (MLR), and a lightweight multi-layer perceptron (MLP)—are compared through two complementary procedures. Station-wise LOOCV evaluates prediction at an omitted gauge when only two stations remain available, whereas the radar-referenced procedure evaluates catchment-average estimates produced using the complete three-gauge network. Because the procedures differ in station availability, prediction target, and reference dataset, they are not treated as a controlled experiment that isolates validation strategy alone. The objective is to examine how these methods behave under the two operational settings within the selected event, rather than to identify a universally optimal interpolator.

The study makes three bounded contributions. First, it compares deterministic, regression-based, and nonlinear approaches under an extreme three-gauge spatial design. Second, it distinguishes gauge-omission prediction from radar-referenced full-network areal estimation and makes the different information provided by the two procedures explicit. Third, it reports the original performance metrics together with a descriptive four-metric rank synthesis while avoiding statistical or physical interpretation of negligible numerical differences. The contribution is therefore an event-scale comparison and a transparent evaluation structure for sparse engineering catchments, rather than a universal ranking of the six methods.

2. Materials and Methods

2.1. Study Area

The study was conducted in a small engineered catchment in Jiangxi Province, southeastern China, with a drainage area of approximately 0.719 km^2 (Figure 1). The catchment is characterized by low-relief hilly terrain, vegetated slopes, exposed engineering surfaces, access roads, and water bodies visible in the UAV orthophoto. The region has a humid subtropical monsoon climate, with a pronounced flood-season rainfall period during July–August. The limited relief and compact catchment extent provide the spatial setting for interpreting the weak predictive contribution of elevation during the selected event.

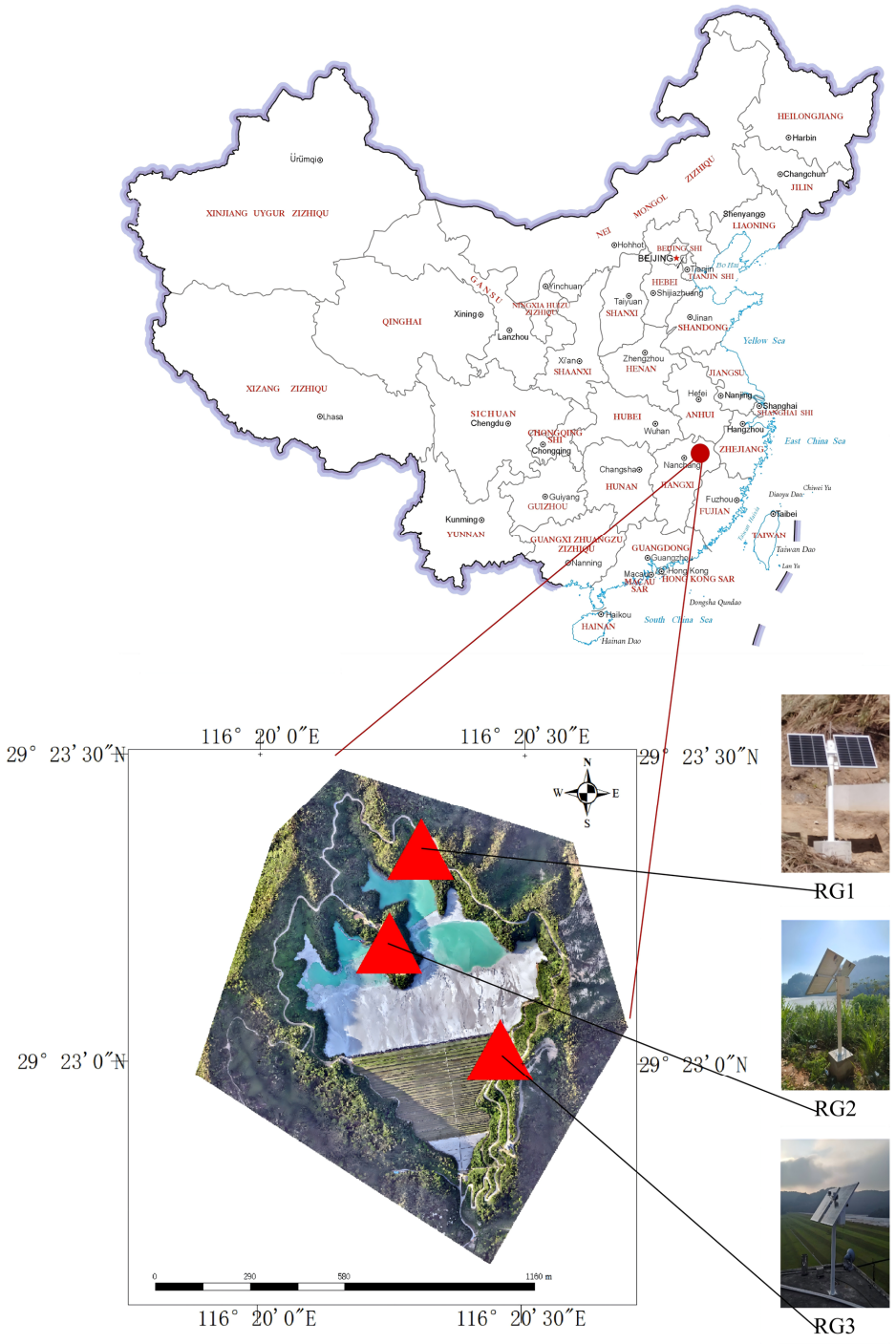

Figure 1. Study-area setting and gauge-network geometry: location of the engineered catchment in Jiangxi Province and a detailed orthophoto showing the gauge network (rain gauges RG1–RG3).

Three automatic rain gauges (RG1, RG2, and RG3) form an irregular north-to-south monitoring network. Their positions reflect terrain, access, and engineering constraints rather than a regular sampling design. Figure 1 presents the national-to-site context and a detailed UAV orthophoto with geographic coordinates, a north arrow, a scale bar, the catchment setting, and the relative positions of the three gauges. Because TP and IDW weights depend directly on station geometry, the reported comparison is specific to this arrangement.

2.2. Data Sources

Rainfall observations were collected from the three automatic rain gauges deployed within the catchment. The analysis used a continuous 60 h rainfall period in August 2025, comprising 240 synchronized, non-overlapping 15 min intervals at each gauge and 720 gauge-interval records in total. The period was selected according to simultaneous data availability, completeness, and temporal consistency at the three gauges. Figure 2 presents the synchronized RG1, RG2, and RG3 rainfall series. The dataset also included observation timestamps, station coordinates, and elevations for the spatial interpolation procedures.

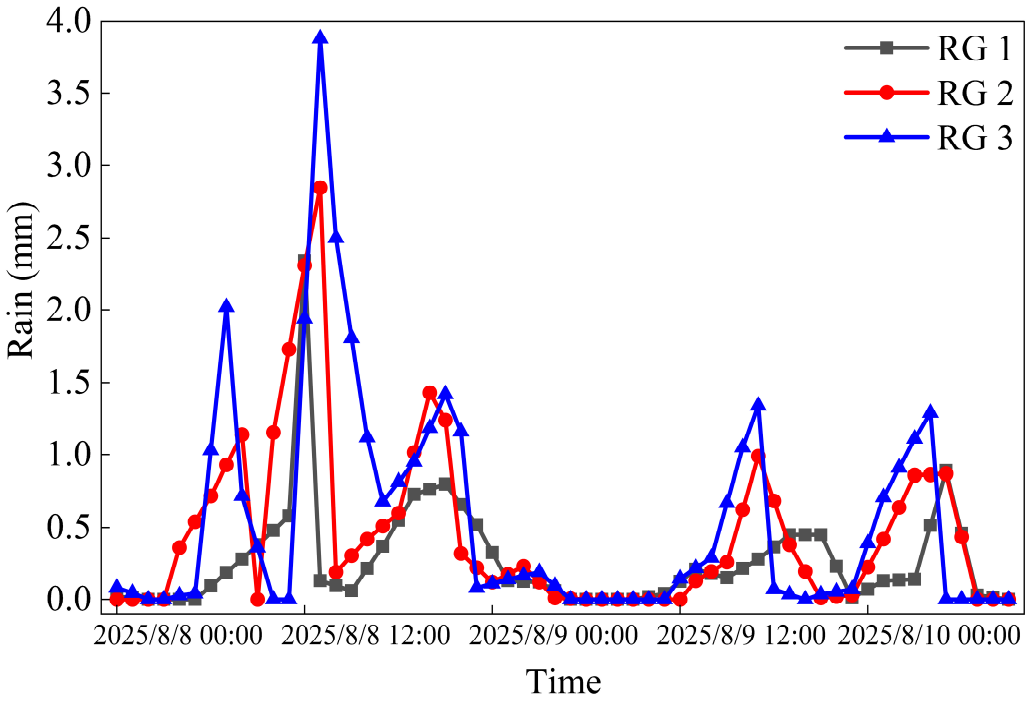

Figure 2. Fifteen-minute rainfall accumulations recorded by RG1, RG2, and RG3 during the selected continuous 60 h period in August 2025.

High-resolution topographic information was obtained from a project-specific unmanned-aerial-vehicle-derived digital orthophoto map (DOM) and digital elevation model (DEM). Planar station coordinates were used for distance calculations in TP and IDW, and elevation was used in ELR and MLR. Because of the limited catchment extent, planar Euclidean distance was adopted. A common catchment boundary and spatial mask were applied to all methods. Radar-derived areal precipitation supplied by the local meteorological authority was temporally matched to the synchronized gauge series and used only as an external comparative reference; gauge observations remained the inputs to all interpolation methods.

The external reference was the operational quantitative precipitation estimation (QPE) product generated from the Nanchang S-band Doppler weather radar (CINRAD/SA; Beijing Metstar Radar Co., Ltd., Beijing, China) and supplied by the Jiangxi Provincial

Meteorological Service. The delivered product had a nominal Cartesian grid spacing of 1 km and a native update interval of 6 min. Radar cells intersecting the catchment were clipped, and boundary cells were weighted according to their fractional overlap with the catchment before spatial averaging. Native radar estimates were then accumulated to the common 15 min intervals by time-overlap weighting. The provider-side product had undergone its routine quality-control processing. No adjustment using RG1–RG3 was applied.

Radar and gauge timestamps were converted to a common local-time convention before matching. The radar product is used as a gauge-independent comparative reference rather than error-free ground truth. This distinction is important because the 0.719 km^2 catchment is smaller than the nominal 1 km radar grid, and sampling-volume mismatch, reflectivity–rainfall conversion, beam geometry, attenuation, partial-pixel coverage, and subpixel rainfall variability can influence catchment-scale agreement [24]. These uncertainties are considered when interpreting the generally low radar correlations and the very small differences between the leading methods [25–30].

After quality control and temporal synchronization, the final dataset contained 240 common 15 min intervals for each gauge, corresponding to 720 gauge-interval observations. Table 1 summarizes the core datasets and the two evaluation settings.

Table 1. Core datasets and evaluation structure of the event-scale study.

Dataset	Description
Study area	Engineered small catchment in Jiangxi Province (0.719 km^2)
Rain gauges	3 automatic rain gauges
Observation period	Continuous 60 h period in August 2025
Temporal structure	240 synchronized, non-overlapping 15 min intervals per gauge
Temporal resolution	15 min
Gauge observations	720 gauge-interval records (240 intervals × 3 gauges)
Auxiliary data	UAV-derived DOM and DEM
External reference	Operational radar QPE (1 km grid; 6 min native update)
Evaluation procedures	Station-wise LOOCV and radar-referenced full-network comparison

2.3. Data Preprocessing

A three-step preprocessing procedure was applied (Figure 3). First, the three gauge records were screened for negative increments, duplicate timestamps, isolated spikes, missing observations, and inconsistent cumulative changes. Duplicate records were consolidated after checking the instrument logs, and short internal gaps were treated by linear interpolation only when the immediately preceding and following observations were valid; no extrapolation was applied. Second, the independently recorded series were aligned to a common 15 min timeline while preserving accumulated rainfall. Third, the synchronized rainfall values were integrated with station coordinates, elevations, the catchment boundary, the DOM, and the DEM. The same preprocessing sequence was applied before any validation fold or method-specific calculation.

Quality control was completed before the stations were assigned to training and validation subsets, preventing method-specific data cleaning. Coordinates and elevations were checked against the UAV-derived DOM and DEM, and all distances were calculated in a projected planar coordinate system. For methods producing a continuous rainfall surface, predictions were evaluated over the common catchment mask and aggregated using cell-area weights. This consistent spatial support was also used for the radar averaging.

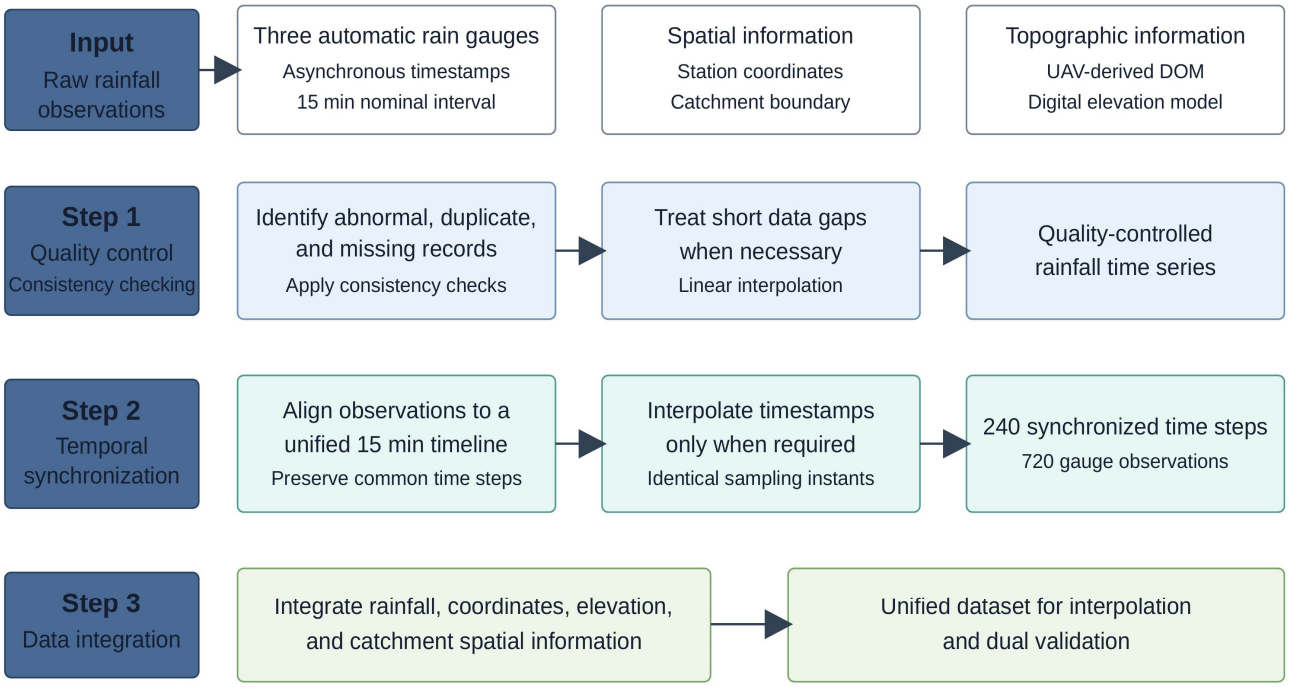

Figure 3. Screening of the gauge records, synchronization to a common 15 min accumulation timeline, and integration with station coordinates, elevation, the catchment boundary, the digital orthophoto map (DOM), and the digital elevation model (DEM).

2.4. Interpolation Methods

Six representative methods were implemented within the same data and spatial-processing workflow. AM, TP, and IDW provide transparent deterministic baselines [1,4–7]; ELR and MLR introduce elevation and planar-coordinate information [8–10]; and a lightweight MLP provides a nonlinear comparison [11–13]. The following subsections state the equations and model settings used in the present analysis.

2.4.1. Arithmetic Mean (AM)

The arithmetic mean estimates areal rainfall by assigning equal weight to all available gauges:

$$P = \frac{1}{n}\sum_{i=1}^{n} P_i \quad (1)$$

where P is the estimated areal rainfall, P_i is the rainfall observed at the i-th station, and n is the number of rain gauges. Because all stations are assigned identical weights, AM serves as the simplest benchmark for evaluating interpolation performance.

2.4.2. Thiessen Polygon (TP)

The Thiessen polygon method represents each gauge by its associated polygon area [4] and estimates areal rainfall as:

$$P = \sum_{i=1}^{n} \omega_i P_i \quad (2)$$

where

$$\omega_i = \frac{A_i}{A} \quad (3)$$

and A_i is the Thiessen polygon area associated with station i, while A is the total catchment area. During LOOCV, omission of one station modifies the polygon topology. Therefore, polygon weights were recalculated and normalized in each validation iteration to ensure

$$\sum_{i=1}^{n} \omega_i = 1 \quad (4)$$

2.4.3. Inverse Distance Weighting (IDW)

IDW estimates rainfall at location x using distance-decaying weights [5–7]:

$$P(x) = \frac{\sum_{i=1}^{n} \omega_i(x) P_i}{\sum_{i=1}^{n} \omega_i(x)} \quad (5)$$

where

$$\omega_i(x) = \frac{1}{d_i^p} \quad (6)$$

where d_i is the planar Euclidean distance between prediction location x and station i, and p is the distance exponent. A fixed exponent of p = 2 was used for every LOOCV fold and for the full-network calculation. The value was selected as a conventional IDW setting and was not tuned to the selected event.

2.4.4. Elevation Linear Regression (ELR)

ELR assumes a linear rainfall–elevation relation [8–10]:

$$P = a + bH \quad (7)$$

where H is elevation and a and b are fitted coefficients. ELR was refitted for each 15 min interval. In each LOOCV fold, the two retained stations define a two-point elevation relation; accordingly, the result is interpreted only as the predictive contribution of elevation for the selected event and network, rather than as a general physical measure of orographic precipitation.

2.4.5. Multiple Linear Regression (MLR)

MLR extends ELR by incorporating planar coordinates [9]:

$$P = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 H \quad (8)$$

where X and Y are planar coordinates, H is elevation, and β_0–β_3 are regression coefficients. With only three gauges—and two independent spatial locations in each LOOCV fold—the design matrix is rank-deficient. MLR was therefore fitted using the minimum-norm least-squares solution. The coefficients are not uniquely identifiable and are treated as event-specific interpolation parameters rather than stable physical rainfall gradients.

2.4.6. Multi-Layer Perceptron (MLP)

The MLP represents a nonlinear mapping from the explanatory variables to rainfall [11–13]:

$$y = f(\mathrm{W}_2 g(\mathrm{W}_1 \mathrm{x} + \mathrm{b}_1) + \mathrm{b}_2) \quad (9)$$

where $g(\cdot)$ is the hidden-layer activation function and $f(\cdot)$ is the output function. A lightweight architecture with one hidden layer and two neurons was used to limit model complexity under the three-gauge spatial design. The purpose of the MLP comparison is to examine whether a small nonlinear model changes the event-scale estimates, not to infer a generally transferable spatial relation from two or three locations.

MLR and MLP used planar coordinates X and Y and elevation H as spatial inputs, with rainfall at the available stations as the response. Models were re-estimated independently for each 15 min time step so that the spatial field could evolve through the event. X, Y, and H were centered and standardized to zero mean and unit variance using parameters calculated only from the training stations in each LOOCV fold; the fitted transformation was then applied to the omitted station and to the catchment grid. For external validation,

scaling parameters were calculated from all three gauges. MLR was fitted by ordinary least squares using the minimum-norm solution for rank-deficient systems. The MLP comprised one hidden layer with two neurons, rectified-linear-unit activation, and a linear output. It was optimized with Adam using mean squared error, an initial learning rate of 0.001, and an L2 penalty of 0.0001 [31,32].

The maximum number of MLP iterations was 2000. Training stopped earlier when the loss improvement was smaller than 10^{-6} for 50 consecutive iterations. A fixed random seed of 42 was used for weight initialization and sample ordering. MLR, feature scaling, and MLP were implemented in Python 3.11 using scikit-learn 1.4.2 (Linear Regression, Standard Scaler, and MLP Regressor, respectively) [31]. Preprocessing, model fitting, prediction, and metric calculation were executed through the same scripted workflow for all LOOCV folds and for the full-network radar comparison. Because only two or three unique spatial locations were available, these settings were selected to constrain flexibility and improve reproducibility rather than to optimize a broad hyperparameter space.

2.5. Complementary Evaluation Procedures

The study used two complementary evaluation procedures (Figure 4). They address different operational questions and differ in station availability, prediction target, and reference data. Accordingly, their results are compared to characterize method behavior under station omission and full-network operation, not to isolate a single validation factor.

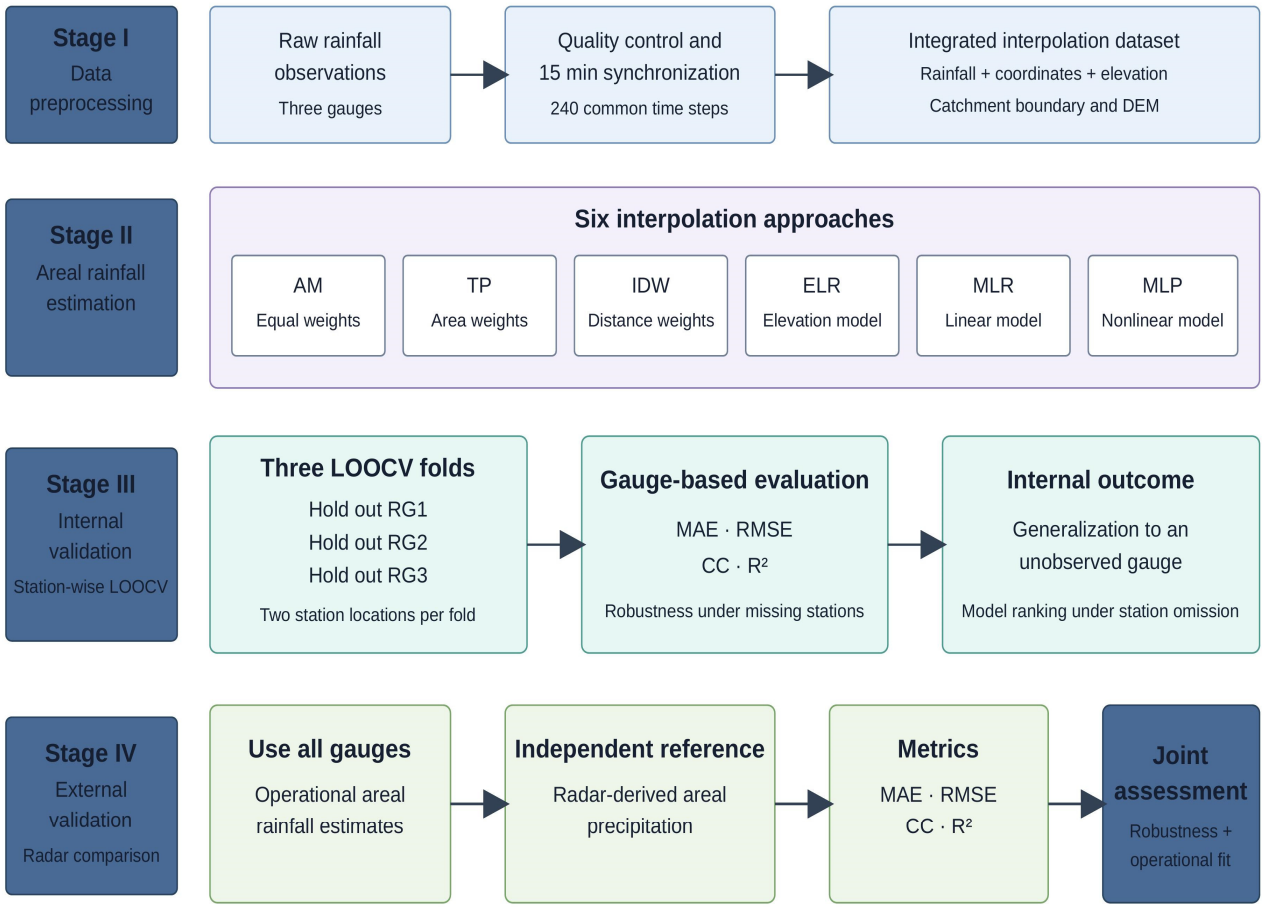

Figure 4. Complementary gauge-referenced station-wise LOOCV and radar-referenced full-network evaluation procedures used in this study. AM, arithmetic mean; TP, Thiessen polygon; IDW, inverse distance weighting; ELR, precipitation–elevation linear regression; MLR, multiple linear regression; MLP, multi-layer perceptron.

Stage I. Data preprocessing. Rainfall observations were quality-controlled and temporally synchronized before constructing the interpolation dataset.

Stage II. Areal rainfall estimation. The six interpolation methods described in Section 2.4 were applied to estimate areal rainfall using identical input datasets.

Stage III. Gauge-referenced station-wise evaluation. In each of three LOOCV folds, one gauge and its complete 240-interval series were withheld, while the other two station locations and their aligned series were used for model construction. Predictions at the omitted gauge were compared with its observations. Thus, the pooled evaluation contains 720 predicted–observed pairs, while each spatial fit is based on only two independent station locations.

Stage IV. Radar-referenced full-network evaluation. All three gauges were used to produce catchment-average rainfall estimates over the same 240 intervals, and the resulting series were compared with the temporally matched radar QPE. Radar data were not used to fit any interpolation method.

The LOOCV separation unit was the station rather than an individual time record. All observations from an omitted gauge were excluded together, preventing repeated records from the same location from entering both the training and evaluation sets. Scaling and model fitting were repeated within each fold. The radar-referenced procedure instead used the complete gauge network and a different prediction target. The two procedures are therefore complementary but not directly interchangeable.

Together, the two procedures provide evidence on gauge-omission behavior and agreement of the complete-network areal estimates with an independent radar product. Their results are interpreted within the scope of the selected event and station geometry.

2.6. Accuracy Assessment

Model performance was quantified using root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and Pearson correlation coefficient (CC):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{10}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{12}$$

$$CC = \frac{\sum_{i=1}^{n}(P_i - \overline{P})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2 \sum_{i=1}^{n}(O_i - \overline{O})^2}} \tag{13}$$

where P_i and O_i are the predicted and observed rainfall values for sample i; $\overline{P}$ and $\overline{O}$ are their means; and n is the number of validation samples. RMSE and MAE quantify error magnitude, R^2 measures the fraction of observed variance explained, and CC measures linear association. Lower RMSE and MAE together with higher R^2 and CC indicate better agreement.

2.7. Comparative Interpretation and Rank Synthesis

Performance was interpreted primarily from MAE, RMSE, CC, and R^2. The percentage reduction relative to the single-gauge reference was calculated for an error metric E as Improvement = $100 \times (E_{single} - E_{method})/E_{single}$. In addition, methods were ranked from 1 (best) to 6 (worst) for each metric, exact ties received their average rank, and the four ranks were averaged without additional weighting. Rankings were calculated before values

were rounded for display. Because MAE and RMSE are related error measures and *CC* and R^2 are related association measures, the mean rank is used only as a descriptive summary. Absolute metric values and practical similarity take precedence when numerical differences are very small.

3. Results

3.1. Overall Performance Under Internal Validation

Table 2 and Figure 5 summarize the pooled station-wise LOOCV results across the three omitted-gauge folds. The table is ordered by descriptive mean rank, while all four original metrics are retained for direct interpretation.

Table 2. Pooled station-wise LOOCV performance of the six methods, ordered by descriptive mean rank. Values are reported to three decimal places.

Interpolation Method	MAE (mm)	RMSE (mm)	CC	R^2
Multiple Linear Regression	0.254	0.433	0.777	0.604
Inverse Distance Weighting	0.262	0.492	0.644	0.414
Multi-layer Perceptron	0.280	0.536	0.542	0.294
Thiessen Polygon	0.278	0.530	0.529	0.280
Arithmetic Mean	0.280	0.536	0.541	0.294
Elevation Linear Regression	1.098	2.659	0.343	0.118

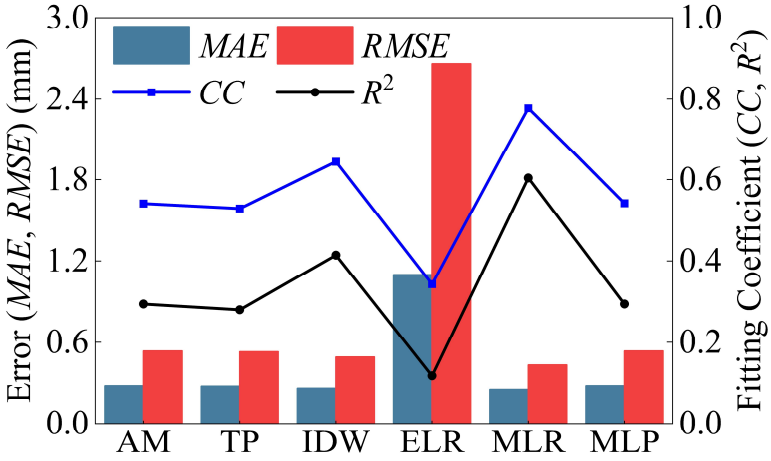

Figure 5. Pooled station-wise LOOCV performance. AM, arithmetic mean; TP, Thiessen polygon; IDW, inverse distance weighting; ELR, precipitation–elevation linear regression; MLR, multiple linear regression; MLP, multi-layer perceptron; MAE, mean absolute error; RMSE, root mean square error; CC, correlation coefficient; R^2, coefficient of determination.

MLR produced the smallest errors (MAE = 0.254 mm; RMSE = 0.433 mm) and the highest CC (0.777) and R^2 (0.604) for the selected event. Each LOOCV fold, however, contained only two independent training locations, as specified in Section 2.5.

IDW ranked second, with MAE = 0.262 mm, RMSE = 0.492 mm, CC = 0.644, and R^2 = 0.414. TP, MLP, and AM formed a closely spaced group in terms of error magnitude, whereas ELR produced the largest errors.

AM and TP had similar RMSE values of 0.536 and 0.530 mm, respectively. MLP yielded the same displayed MAE and RMSE as AM (0.280 and 0.536 mm) and a nearly identical CC.

ELR produced MAE = 1.098 mm and RMSE = 2.659 mm, with CC = 0.343 and R^2 = 0.118. Its errors were substantially larger than those of the other five methods in the LOOCV comparison.

The numerical results therefore place MLR first and IDW second for this event, followed by a closely spaced group of TP, MLP, and AM, with ELR last. Interpretation of these differences is provided in Section 4.

3.2. Effective Spatial Sample Size

Each LOOCV fold retained 240 time intervals for two gauges but only two independent spatial locations. The pooled metrics in Table 2 therefore summarize predictions over time at the three omitted stations; they do not increase the number of independent station locations available to fit the spatial relation.

This distinction is especially relevant for MLR and MLP. MLR used a minimum-norm solution for the rank-deficient spatial design, whereas the lightweight MLP used the fixed architecture and training settings reported in Section 2.4.6. The implications for robustness and transferability are discussed in Section 4.2.

3.3. External Validation Using Radar Precipitation

Table 3 and Figure 6 report the radar-referenced comparison of the catchment-average series produced with all three gauges. This procedure uses a different prediction target and reference dataset from LOOCV and is therefore presented separately.

Table 3. Radar-referenced full-network performance of the six methods, ordered by descriptive mean rank. Values are reported to three decimal places.

Interpolation Method	MAE (mm)	RMSE (mm)	CC	R^2
Multi-layer Perceptron	0.371	0.668	0.343	0.118
Multiple Linear Regression	0.378	0.669	0.341	0.116
Thiessen Polygon	0.375	0.678	0.280	0.079
Arithmetic Mean	0.378	0.702	0.260	0.068
Elevation Linear Regression	0.378	0.702	0.260	0.068
Inverse Distance Weighting	0.421	0.796	0.201	0.040

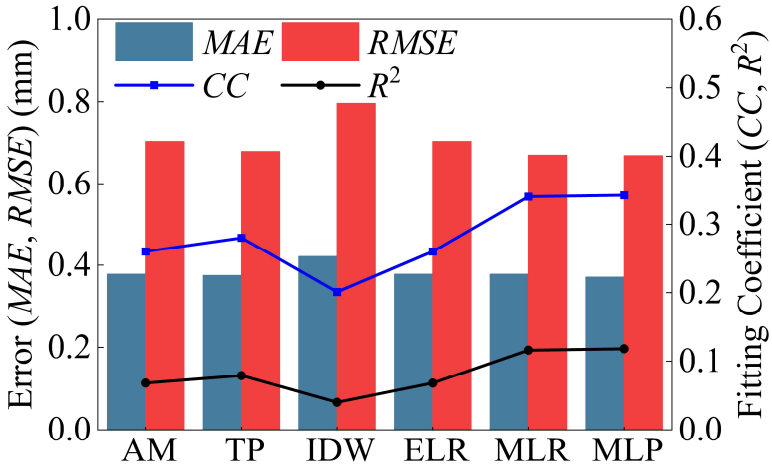

Figure 6. Radar-referenced full-network performance. AM, arithmetic mean; TP, Thiessen polygon; IDW, inverse distance weighting; ELR, precipitation–elevation linear regression; MLR, multiple linear regression; MLP, multi-layer perceptron; MAE, mean absolute error; RMSE, root mean square error; CC, correlation coefficient; R^2, coefficient of determination.

MLP produced MAE = 0.371 mm and RMSE = 0.668 mm, while MLR produced MAE = 0.378 mm and RMSE = 0.669 mm. Their RMSE values differed by only 0.001 mm. MLP and MLR also had similar CC and R^2 values.

TP produced MAE = 0.375 mm and RMSE = 0.678 mm. AM and ELR had identical values at the displayed precision in Table 3 (MAE = 0.378 mm; RMSE = 0.702 mm; CC = 0.260; R^2 = 0.068).

IDW produced the largest radar-referenced errors among the six methods, with MAE = 0.421 mm and RMSE = 0.796 mm. Its CC and R^2 were also the lowest in Table 3.

Relative to the single-gauge reference used in the original comparison, every three-gauge estimate reduced both MAE and RMSE during the selected event. The corresponding percentage reductions are reported in Figure 7.

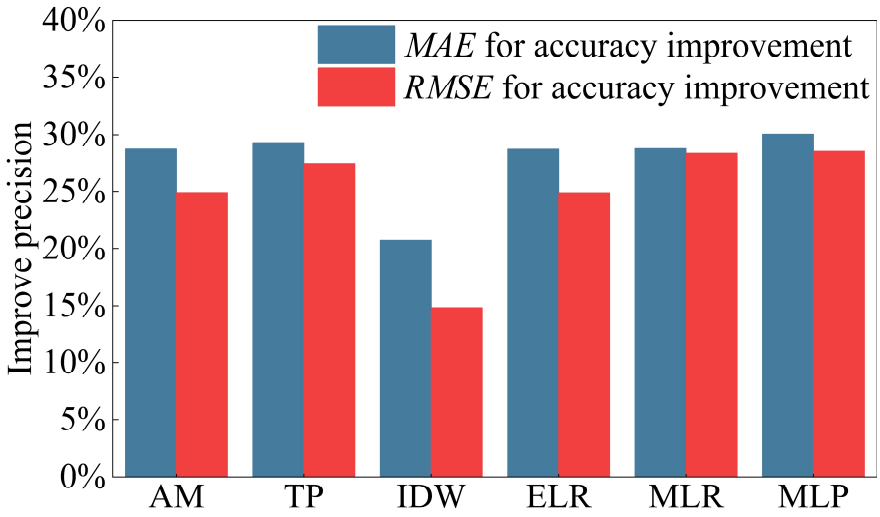

Figure 7. Event-specific reductions in MAE and RMSE relative to the single-gauge reference. AM, arithmetic mean; TP, Thiessen polygon; IDW, inverse distance weighting; ELR, precipitation–elevation linear regression; MLR, multiple linear regression; MLP, multi-layer perceptron.

The radar-referenced CC values ranged from 0.201 to 0.343 and the R^2 values from 0.040 to 0.118. These values show that the external association was modest for every method and provide important context for the small numerical differences among the leading estimates.

Accordingly, MLP and MLR are treated as practically similar in the radar comparison; the 0.001 mm RMSE difference is not used to claim substantial superiority.

3.4. Comparison of Gauge- and Radar-Referenced Results

Percentage reductions in MAE and RMSE relative to the single-gauge reference were calculated using the equation in Section 2.7. Figure 7 summarizes these event-specific reductions for the six three-gauge estimates.

MLP reduced MAE and RMSE by approximately 30.1% and 28.5%, respectively, while MLR produced corresponding reductions of approximately 28.7% and 28.4%. TP reduced MAE by approximately 29.3% and RMSE by 27.5%. AM and ELR had the same displayed percentage reductions, and IDW showed the smallest reductions among the six methods.

The descriptive order changed between the two procedures. MLR and IDW occupied the first two positions in LOOCV, whereas MLP, MLR, and TP formed the leading group in the radar-referenced comparison. IDW changed from second in LOOCV to sixth in the radar comparison.

These rank changes occur across procedures that differ simultaneously in station availability, prediction target, and reference dataset. They are therefore reported as a comparison of two operational settings rather than as the isolated effect of validation strategy.

Within the radar-referenced setting, the RMSE difference between MLP and MLR was 0.001 mm and the MAE difference was 0.007 mm. The original metrics therefore indicate practical similarity despite their different ordinal ranks.

3.5. Descriptive Rank Synthesis

Table 4 summarizes the mean ranks calculated from the unrounded MAE, RMSE, CC, and R^2 results. MLR had a LOOCV mean rank of 1.00, and MLP had a radar-referenced mean rank of 1.00. IDW changed from 2.00 in LOOCV to 6.00 in the radar comparison, while TP changed from 4.00 to 2.75.

Table 4. Descriptive mean ranks calculated from the unrounded MAE, RMSE, CC, and R^2 values; a positive rank shift indicates a worse position in the radar-referenced comparison.

Interpolation Method	LOOCV Mean Rank	Radar Mean Rank	Rank Shift
Arithmetic Mean	4.13	4.38	+0.25
Thiessen Polygon	4.00	2.75	−1.25
Inverse Distance Weighting	2.00	6.00	+4.00
Elevation Linear Regression	6.00	4.38	−1.62
Multiple Linear Regression	1.00	2.50	+1.50
Multi-layer Perceptron	3.88	1.00	−2.88

Because the four metrics are not statistically independent and very small numerical differences can change an ordinal position, Table 4 is used only to summarize method ordering. The numerical interpretation in Sections 3.1–3.4 is based primarily on the original metric values.

4. Discussion

4.1. Interpretation of the Complementary Evaluation Procedures

The principal comparison is the change in method ordering between station-wise LOOCV and the radar-referenced full-network procedure. MLR gave the lowest LOOCV errors, whereas MLP and MLR produced practically similar errors when all gauges were used. Previous studies likewise show that interpolation performance depends on network density, temporal scale, topography, and evaluation criterion [6,7,9,14–17]. In the present study, however, the two procedures also differ in station availability, prediction target, and reference data. The observed ranking change should therefore be interpreted as evidence that the two operational questions are not interchangeable, rather than as the isolated causal effect of validation strategy.

LOOCV examines the ability to predict at an omitted station and directly exposes sensitivity to gauge loss. The radar comparison examines catchment-average estimates produced by the complete network while also incorporating uncertainty in the radar reference [3,18,19]. Considering both prevents method selection from relying solely on either station transfer or agreement with one spatially continuous reference.

4.2. Effect of Sample Size on Linear and Nonlinear Interpolation Models

The effective sample size of the spatial problem is determined by the number of independent locations rather than by the 720 repeated gauge-interval records alone. Each LOOCV fold contains only two training locations, and the full-network procedure contains three. This severe spatial limitation affects both linear and nonlinear models and explains why conventional parameter interpretation is inappropriate even when the temporal series is comparatively long [11,20].

For MLR, the favorable LOOCV metrics describe event-specific predictions obtained from a minimum-norm solution; they do not demonstrate uniquely identifiable coordinate or elevation gradients. For MLP, regularization, early stopping, a small hidden layer, and a fixed seed reduce but do not eliminate sensitivity to the extremely small spatial design. The 0.001 mm radar-referenced RMSE difference between the two models is therefore practically negligible in this case.

4.3. Engineering Implications for Sparse Rain-Gauge Networks

The results support several practical considerations for rainfall monitoring in small engineering catchments.

First, the intended operating condition should determine the evaluation emphasis. When temporary gauge loss is a central concern, station-wise testing is directly relevant; when the objective is routine catchment-average estimation with the complete network, an areal reference provides complementary information. In both cases, the original error values should take precedence over small differences in ordinal rank.

Second, MLP and MLR should be regarded as one practical performance tier in the radar comparison. A preference between them should therefore consider reproducibility, numerical stability, computational requirements, and ease of implementation rather than the 0.001 mm RMSE difference alone.

Third, AM and TP remain useful low-complexity alternatives. Their transparent weighting and limited parameter requirements are advantageous in operational settings, while the event-specific error reductions relative to the single-gauge reference show the practical value of combining the available observations.

Finally, ELR performance should be interpreted as a predictive result for this event and network: elevation did not provide useful additional information under the tested conditions. Three gauges in one small catchment are not sufficient to infer the general physical importance of orographic forcing.

4.4. Transferability and Decision Implications

The findings provide an event-scale decision framework rather than a universal hierarchy. Transfer to another catchment requires attention to network geometry, because station spacing and boundary position directly affect TP and IDW weights; to covariate relevance, because elevation may not improve prediction in every storm; and to validation equivalence, because record-wise temporal splitting does not test prediction at an unmonitored location [33,34]. These considerations limit direct transfer of the numerical ranking while preserving the usefulness of the evaluation logic.

The radar reference introduces an additional scale issue. The catchment area of 0.719 km^2 is smaller than the nominal 1 km radar grid, so the comparison may depend on one or a few fractionally overlapping cells [35]. Low CC and R^2 values across all methods indicate that radar-referenced differences should be interpreted together with radar sampling, retrieval, and temporal-matching uncertainty. The radar product is therefore a complementary reference, not a definitive measure of the true catchment rainfall field.

4.5. Study Scope and Future Perspectives

This study is limited to one continuous 60 h rainfall event, one 0.719 km^2 catchment, and one three-gauge geometry. The 240 time intervals describe temporal evolution within the event but do not constitute independent rainfall events. Consequently, the results do not establish a climatological ranking or a general advantage of one model class.

The rank synthesis is also descriptive. Equal weighting gives related error metrics (MAE and RMSE) and related association metrics (CC and R^2) the same influence, while conversion to ordinal ranks can magnify negligible numerical differences. This is par-

ticularly evident for MLP and MLR in the radar comparison. Absolute errors, practical similarity, and the intended operating condition should therefore remain the primary basis for method selection.

Temporal harmonization and the radar product add further uncertainty to the 15 min comparison. Screening, common-timeline alignment, fractional spatial overlap, and temporal aggregation were applied consistently, but they cannot remove sampling and scale differences between point gauges and radar pixels. The reported external metrics should be read within that measurement context.

A broader assessment across rainfall events, storm types, network configurations, and catchments would allow the stability of the present ranking to be tested. Representative spatial-field checks could additionally examine gradients, negative predictions, boundary extrapolation, and sensitivity to the omitted station.

5. Conclusions

This event-scale study compared six areal-rainfall estimation methods in a 0.719 km^2 catchment monitored by three gauges. Station-wise LOOCV and a radar-referenced full-network comparison were used as complementary procedures addressing different prediction targets. The principal conclusions are as follows.

(1) In the station-wise LOOCV comparison, MLR produced the lowest RMSE (0.433 mm) and the highest CC (0.777), while ELR produced the largest errors. This result is specific to the selected event and does not establish physically identifiable MLR gradients from two training locations.

(2) In the radar-referenced full-network comparison, MLP and MLR produced RMSE values of 0.668 and 0.669 mm, respectively. The 0.001 mm difference is practically negligible, and the two methods should be treated as having similar external performance for this event.

(3) All six three-gauge estimates reduced MAE and RMSE relative to the single-gauge reference during the selected event. These improvements demonstrate the event-specific benefit of combining the available observations, rather than a universal percentage gain for sparse networks.

(4) The two evaluation procedures provide complementary evidence: LOOCV characterizes prediction under station omission, whereas the radar comparison characterizes full-network areal estimates against an uncertain spatial reference. Because station number, target, and reference data change simultaneously, their ranking difference should not be attributed to validation strategy alone.

Overall, method selection should be matched to the intended operational setting and based primarily on the original error metrics rather than small rank differences. The numerical ordering reported here is restricted to one 60 h event, one small catchment, and one three-gauge arrangement. Broader multi-event and multi-catchment evaluation is needed to assess transferability.

Author Contributions: Conceptualization, Y.M. and X.Z. (Xiaobo Zhang); methodology, Y.M., C.C. and X.Z. (Xiaobo Zhang); software, C.C. and Y.L.; validation, H.L. and X.Z. (Xiaojun Zhang); formal analysis, Y.M. and C.C.; investigation, Y.M., C.C. and H.L.; resources, X.Z. (Xiaobo Zhang) and Q.J.; data curation, C.C. and H.L.; writing—original draft preparation, Y.M. and F.X.; writing—review and editing, X.Z. (Xiaobo Zhang), X.Z. (Xiaojun Zhang) and Q.J.; visualization, C.C. and F.X.; supervision, X.Z. (Xiaobo Zhang) and Q.J.; project administration, X.Z. (Xiaobo Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2023YFC3012200), the National Natural Science Foundation of China (Grant

No. 52409130), and the Natural Science Foundation of Jiangxi Province (Grant Nos. 20242BAB25317 and 20232BAB214083).

Data Availability Statement: The rain-gauge data supporting the findings of this study are available from the corresponding author upon reasonable request. Restrictions apply to the radar-derived precipitation data, which were obtained from the local meteorological authority and are available with its permission.

Conflicts of Interest: Author Yanzhi Liu was employed by the company Jiangxi Duchang Jinding Tungsten Molybdenum Mining Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders, if any, had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AM	Arithmetic Mean
TP	Thiessen Polygon
IDW	Inverse Distance Weighting
ELR	precipitation–elevation linear regression (Elevation Linear Regression)
MLR	Multiple Linear Regression
MLP	Multi-Layer Perceptron
LOOCV	Leave-One-Out Cross-Validation
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
CC	Correlation Coefficient
DOM	Digital Orthophoto Maps
DEM	Digital Elevation Model
ANN	Artificial Neural Networks
RF	Random Forest
SVR	Support Vector Regression
PINNs	Physics-Informed Neural Networks
GNNs	Graph Neural Networks
R^2	Coefficient of Determination

References

1. Ly, S.; Charles, C.; Degré, A. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: A review. *Biotechnol. Agron. Soc. Environ.* **2013**, *17*, 392–406.
2. Segond, M.L.; Wheater, H.S.; Onof, C. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *J. Hydrol.* **2007**, *347*, 116–131. [CrossRef]
3. Hwang, S.H.; Kim, K.B.; Han, D. Comparison of methods to estimate areal means of short-duration rainfalls in small catchments using rain-gauge and radar data. *J. Hydrol.* **2020**, *588*, 125084. [CrossRef]
4. Thiessen, A.H. Precipitation averages for large areas. *Mon. Weather Rev.* **1911**, *39*, 1082–1089. [CrossRef]
5. Chen, F.W.; Liu, C.W. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy Water Environ.* **2012**, *10*, 209–222. [CrossRef]
6. Wagner, P.D.; Fiener, P.; Wilken, F.; Kumar, S.; Schneider, K. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data-scarce regions. *J. Hydrol.* **2012**, *464–465*, 388–400. [CrossRef]
7. Hwang, Y.; Clark, M.; Rajagopalan, B.; Leavesley, G.H. Spatial interpolation schemes of daily precipitation for hydrologic modeling. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 295–320. [CrossRef]
8. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129. [CrossRef]

9. Chen, T.; Ren, L.; Yuan, F.; Yang, X.; Jiang, S.; Tang, T.; Liu, Y.; Zhao, C.; Zhang, L. Comparison of spatial interpolation schemes for rainfall data and application in hydrological modeling. *Water* **2017**, *9*, 342. [CrossRef]
10. Bárdossy, A.; Pegram, G. Interpolation of precipitation under topographic influence at different time scales. *Water Resour. Res.* **2013**, *49*, 4545–4565. [CrossRef]
11. Li, J.; Heap, A.D.; Potter, A.; Daniell, J.J. Application of machine learning methods to spatial interpolation of environmental variables. *Environ. Model. Softw.* **2011**, *26*, 1647–1659. [CrossRef]
12. Tan, J.; Xie, X.; Zuo, J.; Xing, X.; Liu, B.; Xia, Q.; Zhang, Y. Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with multiple covariates. *J. Hydrol.* **2021**, *598*, 126270. [CrossRef]
13. Li, J.; Shen, Y.; Chen, L.; Ng, C.W.W. SSIN: Self-supervised learning for rainfall spatial interpolation. *Proc. ACM Manag. Data* **2023**, *1*, 1–21. [CrossRef]
14. Berndt, C.; Haberlandt, U. Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density. *J. Hydrol. Reg. Stud.* **2018**, *15*, 184–202. [CrossRef]
15. Liu, D.; Zhao, Q.; Fu, D.; Guo, S.; Liu, P.; Zeng, Y. Comparison of spatial interpolation methods for the estimation of precipitation patterns at different time scales to improve the accuracy of discharge simulations. *Hydrol. Res.* **2020**, *51*, 583–601. [CrossRef]
16. Yang, R.; Xing, B. A comparison of the performance of different interpolation methods in replicating rainfall magnitudes under different climatic conditions in Chongqing Province (China). *Atmosphere* **2021**, *12*, 1318. [CrossRef]
17. Bárdossy, A.; Modiri, E.; Anwar, F.; Pegram, G. Gridded daily precipitation data for Iran: A comparison of different methods. *J. Hydrol. Reg. Stud.* **2021**, *38*, 100958. [CrossRef]
18. Goudenhoofdt, E.; Delobbe, L. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 195–203. [CrossRef]
19. Villarini, G.; Krajewski, W.F. Review of the different sources of uncertainty in single-polarization radar-based estimates of rainfall. *Surv. Geophys.* **2010**, *31*, 107–129. [CrossRef]
20. Li, J.; Heap, A.D. Spatial interpolation methods applied in the environmental sciences: A review. *Environ. Model. Softw.* **2014**, *53*, 173–189. [CrossRef]
21. Verworn, A.; Haberlandt, U. Spatial interpolation of hourly rainfall—Effect of additional information, variogram inference and storm properties. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 569–584. [CrossRef]
22. Ly, S.; Charles, C.; Degré, A. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Amblève catchments, Belgium. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2259–2274. [CrossRef]
23. Dirks, K.N.; Hay, J.E.; Stow, C.D.; Harris, D. High-resolution studies of rainfall on Norfolk Island, Part II: Interpolation of rainfall data. *J. Hydrol.* **1998**, *208*, 187–193. [CrossRef]
24. Faurès, J.M.; Goodrich, D.C.; Woolhiser, D.A.; Sorooshian, S. Impact of small-scale spatial rainfall variability on runoff modeling. *J. Hydrol.* **1995**, *173*, 309–326. [CrossRef]
25. Haberlandt, U. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.* **2007**, *332*, 144–157. [CrossRef]
26. Moulin, L.; Gaume, E.; Obled, C. Uncertainties on mean areal precipitation: Assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 99–114. [CrossRef]
27. Schiemann, R.; Erdin, R.; Willi, M.; Frei, C.; Berenguer, M.; Sempere-Torres, D. Geostatistical radar–rain-gauge combination with nonparametric correlograms: Methodological considerations and application in Switzerland. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1515–1536. [CrossRef]
28. Thorndahl, S.; Einfalt, T.; Willems, P.; Nielsen, J.E.; ten Veldhuis, M.-C.; Arnbjerg-Nielsen, K.; Rasmussen, M.R.; Molnar, P. Weather radar rainfall data in urban hydrology. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1359–1380. [CrossRef]
29. McKee, J.L.; Binns, A.D. A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology. *Can. Water Resour. J.* **2016**, *41*, 186–203. [CrossRef]
30. Ochoa-Rodriguez, S.; Wang, L.-P.; Willems, P.; Onof, C. A review of radar–rain gauge data merging methods and their potential for urban hydrological applications. *Water Resour. Res.* **2019**, *55*, 6356–6391. [CrossRef]
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
32. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2015**, arXiv:1412.6980. [CrossRef]
33. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]

34. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [CrossRef] [PubMed]
35. Peleg, N.; Ben-Asher, M.; Morin, E. Radar subpixel-scale rainfall variability and uncertainty: Lessons learned from observations of a dense rain-gauge network. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2195–2208. [CrossRef]

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.