https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

GraphIDW: Incorporating spatial autocorrelation in satellite-gauge precipitation merging using graph neural networks over a tropical region

Nadee Peiris[1,2], Chamal Perera[1,2], Nimal Wijayaratna [1,2], Lalith Rajapakse[1,2], Ajith Wijemannage[3]

[1]UNESCO-Madanjeet Singh Centre for South Asia Water Management (UMCSAWM), University of Moratuwa, Moratuwa 10400, Sri Lanka
[2]Department of Civil Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka
[3]Department of Meteorology, Bauddhaloka Mawatha, Colombo-07, Sri Lanka

Correspondence to: Nadee Peiris (peirismnnk.24@uom.lk); Chamal Perera (chamalp@uom.lk)

ABSTRACT

Ground-based rain gauges remain the benchmark for accurate precipitation measurement; however, their sparse spatial distribution limits the representation of rainfall heterogeneity. Satellite-based Precipitation Products (SPPs) provide consistent spatial coverage but are often affected by retrieval errors and regional biases, restricting their direct use in local-scale hydrological applications. To overcome these limitations, Precipitation Data Merging (PDM) techniques integrating gauge and satellite observations have gained prominence. This study introduces a novel Machine Learning (ML) framework, GraphIDW, which combines Graph Neural Networks (GNNs) with Inverse Distance Weighting (IDW) interpolation to explicitly incorporate spatial autocorrelation into the merging process, addressing a major limitation of traditional ML-based PDM approaches. The framework was evaluated across the Wet Zone of Sri Lanka from 2001 to 2015 using two state-of-the-art SPPs (IMERG and CHIRPS) together with ground observations. IMERG data (0.1°) were first downscaled to 0.05° using CHIRPS, after which the downscaled product was merged with gauge observations through GraphIDW. A total of 60 gauges (70%) were used for training and 28 (30%) for validation. Results show that GraphIDW outperforms conventional ML algorithms, including Random Forest, Artificial Neural Network, Support Vector Regression, and XGBoost. It achieved the highest probability of detection (0.97) and reduced root mean square error (RMSE) and mean absolute error (MAE) by 13%–41% and 9%–36%, respectively, compared with the original SPPs. The results demonstrate that explicitly accounting for spatial dependence through graph-based learning significantly improves precipitation estimation, particularly in regions characterized by strong spatial heterogeneity. By embedding spatial autocorrelation directly into the merging process, GraphIDW provides a robust and computationally efficient framework for generating high-resolution rainfall datasets that are better suited for hydrological analysis in complex climatic and topographic settings.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

1. Introduction

High-quality precipitation estimation is fundamental for hydrological analysis, water resource assessment, and extreme event monitoring. In-situ rain gauge networks are considered the most reliable source of precipitation measurements (Tian et al., 2018), although their point-based nature represents only a small area surrounding the instrument (Beikahmadi et al., 2023). Precipitation exhibits strong spatial heterogeneity, which exerts a significant influence on meteorological dynamics and hydrological responses (Niu et al., 2017). Consequently, in regions with pronounced spatiotemporal variability driven by convective precipitation, relying solely on ground-based measurements can introduce substantial uncertainties (Bertoncini & Pomeroy, 2025; Avanzi et al., 2021).

Satellite-based Precipitation Products (SPPs) have increasingly been adopted as a viable solution to overcome the spatiotemporal limitations of gauge-based observations; however, their direct application remains restricted by marked systematic retrieval biases and random uncertainties (Wehbe et al., 2020). Over recent decades, considerable research attention has been directed toward integrating ground-based measurements with SPPs to produce precipitation datasets that combine high accuracy with extensive spatial and temporal coverage, a process commonly termed Precipitation Data Merging (PDM) (Kossieris et al., 2024). A wide range of PDM techniques has been developed, which can be broadly classified into global and local correction approaches (Wu et al., 2020). Global correction methods typically adjust precipitation fields without explicitly accounting for spatial variability, whereas local correction techniques incorporate spatial autocorrelation characteristics of precipitation patterns (Wu et al., 2020). Early local correction-based PDM studies largely relied on geostatistical frameworks, including Kriging-based methods, geographically weighted regression, and optimal interpolation (Shen et al., 2014; Verdin et al., 2015; Chao et al., 2018). Nevertheless, these conventional regression-driven approaches often depend on restrictive assumptions that may not be satisfied under complex real-world precipitation regimes.

In recent years, Machine Learning (ML) and Deep Learning (DL) methods have achieved remarkable success across data-driven disciplines due to their capacity to represent complex, nonlinear relationships without imposing rigid mathematical assumptions, making them an appealing alternative to conventional geostatistical techniques (Zhang et al., 2021; Lyu & Yong, 2024). Accordingly, a diverse set of ML algorithms has been employed for PDM, including Support Vector Machines (SVM) (Zhang et al., 2021; Ghosh et al., 2024), Random Forest (RF) models (Baez-Villanueva et al., 2020; Assiri & Qureshi, 2022), DL architectures (Le et al., 2023; Nan et al., 2023), and Extreme Gradient Boosting (XGBoost) approaches (Lei et al., 2022; Papacharalampous et al., 2023). Despite their demonstrated predictive capability, many of these models implicitly assume that training samples are independent and identically distributed, an assumption that contradicts the spatial and temporal autocorrelation intrinsic to precipitation fields (Li & Heap, 2014; Hengl et al., 2018). Consequently, such models often fail to adequately represent the spatial dependence structures that control regional rainfall variability.

Most ML-based PDM frameworks, therefore, incorporate additional covariates alongside precipitation estimates, including meteorological variables, topographic attributes, and spatial location information. However, when these auxiliary predictors are unavailable, conventional ML approaches frequently exhibit limited transferability and poor performance at ungauged

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

locations. As a result, although traditional ML methods can provide localized improvements, their effectiveness for large-scale precipitation merging remains constrained, highlighting the need for hybrid or advanced spatially explicit modeling strategies that directly account for spatial autocorrelation (Shen et al., 2018; Wu et al., 2020). Notably, most previous studies on SPP merging have overlooked the spatial autocorrelation that inherently exists among gauge observations during the merging process (Lei at al., 2022).

To address the aforementioned limitations, this study introduces GraphIDW, a novel PDM framework that integrates Graph Neural Networks (GNNs) with Inverse Distance Weighting (IDW). The proposed hybrid framework is specifically designed for regions characterized by high spatial heterogeneity in precipitation and the absence of auxiliary predictor variables. GraphIDW explicitly accounts for the spatial dependence of rainfall variability across regions. The framework is applied to the Wet Zone of Sri Lanka, a region well known for its high spatiotemporal variability in precipitation, over the period 2001–2015. The specific objectives of this study are to: (i) develop a novel PDM methodology that explicitly incorporates spatial autocorrelation, (ii) assess the performance of the proposed framework using both categorical and continuous evaluation metrics, and (iii) benchmark its performance against existing ML-based PDM approaches. Overall, GraphIDW is anticipated to offer a robust solution for producing more reliable precipitation estimates for hydrological applications, particularly in data-scarce regions.

2. Study area and materials

2.1 Study area

The Wet Zone of Sri Lanka (Fig. 1), covering an extent of 12,665 km², was chosen as the study region. Sri Lanka experiences a typical tropical monsoonal climate and is commonly classified into three principal climatic regions, namely the Wet Zone, Intermediate Zone, and Dry Zone, based on long-term mean annual rainfall patterns and associated biophysical parameters (Punyawardena, 2020). The southwestern region, including the central highlands, constitutes the Wet Zone, with elevations ranging from sea level to 2,530 m above the mean sea level. This zone receives high mean annual rainfall varying between 1,750 mm and 5,500 mm, and does not exhibit a clearly defined dry season. The Wet Zone is characterized by pronounced spatiotemporal variability in rainfall, which poses significant challenges for water resources management. This heterogeneity, coupled with complex topography, often leads to substantial discrepancies in satellite-based precipitation estimates, thereby justifying the selection of this region as a testbed for evaluating spatially explicit ML models for the present analysis.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

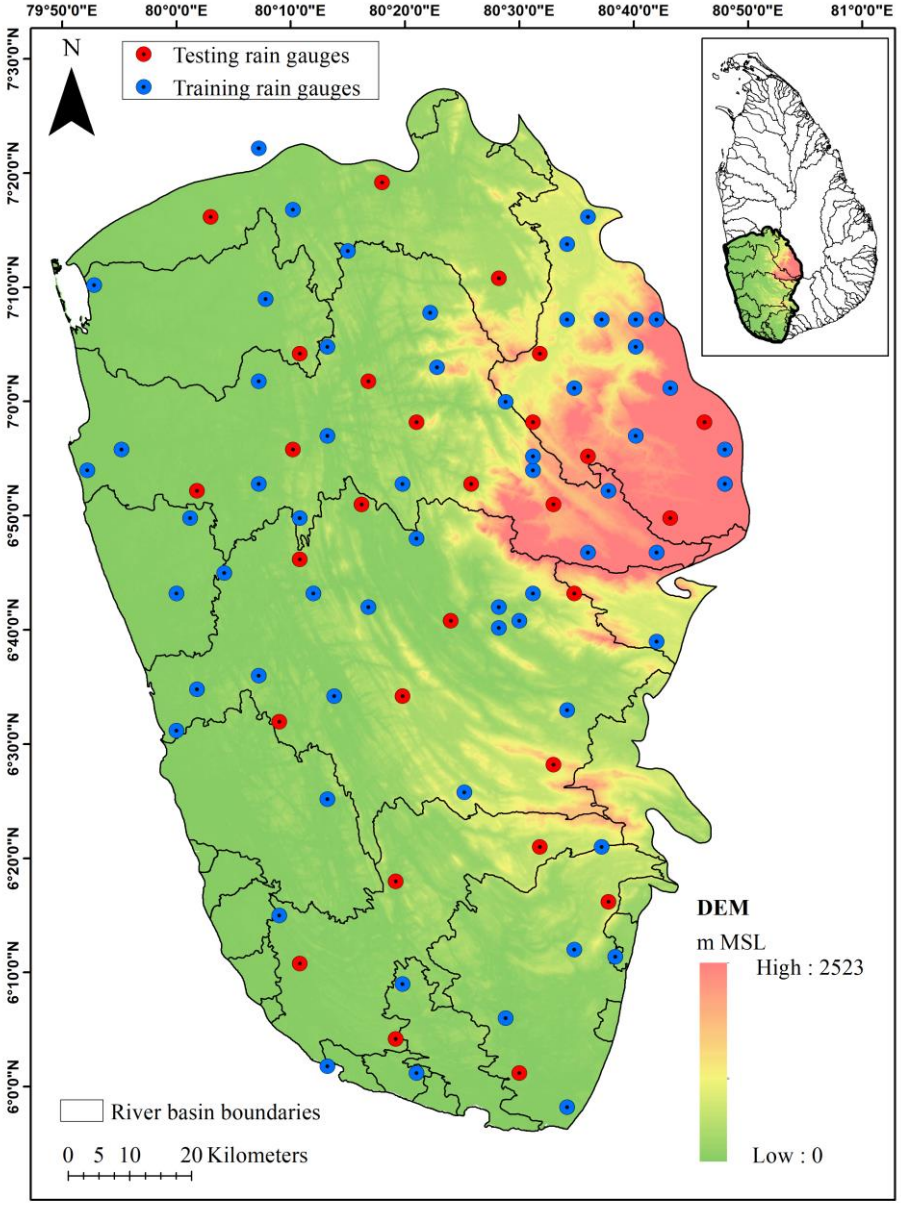


Figure 1. The topography of the Wet Zone in Sri Lanka and the configuration of the rain gauge network

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

2.2 Materials

2.2.1 Observation data

Daily rainfall data from 88 rain gauges located within the Wet Zone are used in the merging process of SPPs. The observed rainfall data, spanning from 2001 to 2015, are obtained from the Department of Meteorology, Sri Lanka. The selected rain gauges provide comprehensive spatial coverage across the entire study area, allowing for effective merging of SPPs with ground-based observations in regions characterized by heterogeneous topography.

2.2.2 Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG)

The IMERG dataset is generated using observations from the Global Precipitation Measurement (GPM) Core Observatory, which was launched in 2014. IMERG delivers precipitation estimates at a spatial resolution of 0.1° × 0.1° with near-global coverage extending from 60°S to 60°N. The product combines measurements from a constellation of Passive Microwave (PMW) and Infrared (IR) sensors to produce high-accuracy precipitation estimates (Ramadhan et al., 2023).

IMERG offers Early, Late, and Final Run products with release latencies of roughly 4 hours, 14 hours, and 3.5 months, respectively. The latest release, IMERG Version 06, extends coverage back to June 2000 through retrospective processing, thereby facilitating the assessment of long-term precipitation patterns (Huffman et al., 2019). For the present study, daily IMERG Early Run data (Table 1) are obtained from NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC, https://disc.gsfc.nasa.gov/) for the period 2001–2015.

2.2.3 Climate Hazards Group InfraRed Precipitation (with Station data) (CHIRP(S))

The CHIRPS dataset integrates long-term climatology (CHPclim), high-resolution satellite observations (CHIRP), and in-situ rain gauge measurements to generate gridded precipitation time series. The satellite-only component, CHIRP, is derived from geostationary Infrared (IR) imagery, where cloud-top temperatures are used as proxies for rainfall intensity (Funk et al., 2015). To produce the final CHIRPS product, CHIRP estimates are merged with gauge observations using a modified inverse distance weighting approach.

Both CHIRP and CHIRPS are provided at a fine spatial resolution of 0.05°×0.05°, offering quasi-global coverage between 50°N–50°S and 180°W–180°E. The datasets are available at daily, pentadal, and monthly temporal scales (Funk et al., 2014). The core computational time step of the products is the pentad; longer time scales (e.g., monthly) are generated through temporal aggregation, while daily precipitation estimates are obtained via disaggregation. Specifically, daily rainfall is derived by redistributing pentadal CHIRP totals using the temporal variability of daily fields from the Coupled Forecast System (CFS), which are first rescaled to the 0.05° grid. At each grid cell, the pentadal precipitation total is proportionally allocated according to the relative daily CFS precipitation fractions (Funk et al., 2015).

The CHIRP and CHIRPS products span the period from 1981 to near-real time (Table 1), making them particularly valuable for long-term climate variability and hydrological analyses. CHIRP is released with a latency of approximately two days after

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

each pentad, while the gauge-integrated CHIRPS product is available with a delay of about three weeks. In this study, monthly CHIRPv2 and CHIRPSv2 datasets for the period 2001-2015 were obtained from https://data.chc.ucsb.edu/.

Table 1: Summary of datasets employed in this study, including their spatial and temporal resolutions

Product	Spatial Resolution	Temporal Resolution
IMERG	0.1° × 0.1°	Daily
CHIRP	0.05° × 0.05°	Monthly
CHIRPS	0.05° × 0.05°	Monthly
SRTM DEM	30 m	-

3. Methodology

Figure 2 presents a schematic overview of the methodological workflow employed in this study, which is structured into four main stages. In the first stage, the required datasets, including SPPs, ground-based rainfall observations, and spatial parameters, are collected. In the second stage, the collected data undergo preprocessing. Specifically, the daily IMERG dataset is downscaled from its native 0.1° spatial resolution to 0.05°, and ground-based observations are split into training and testing subsets. In the third stage, precipitation data merging is performed, where precipitation is predicted using a per-time-step training strategy that formulates the task as a spatial prediction problem. For this purpose, a GNN-IDW-based hybrid framework is employed to estimate pixel-level rainfall at each time step. Finally, in the fourth stage, the performance of the proposed framework is evaluated at both point and areal scales against the original SPPs and several benchmark ML-based merging methods.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

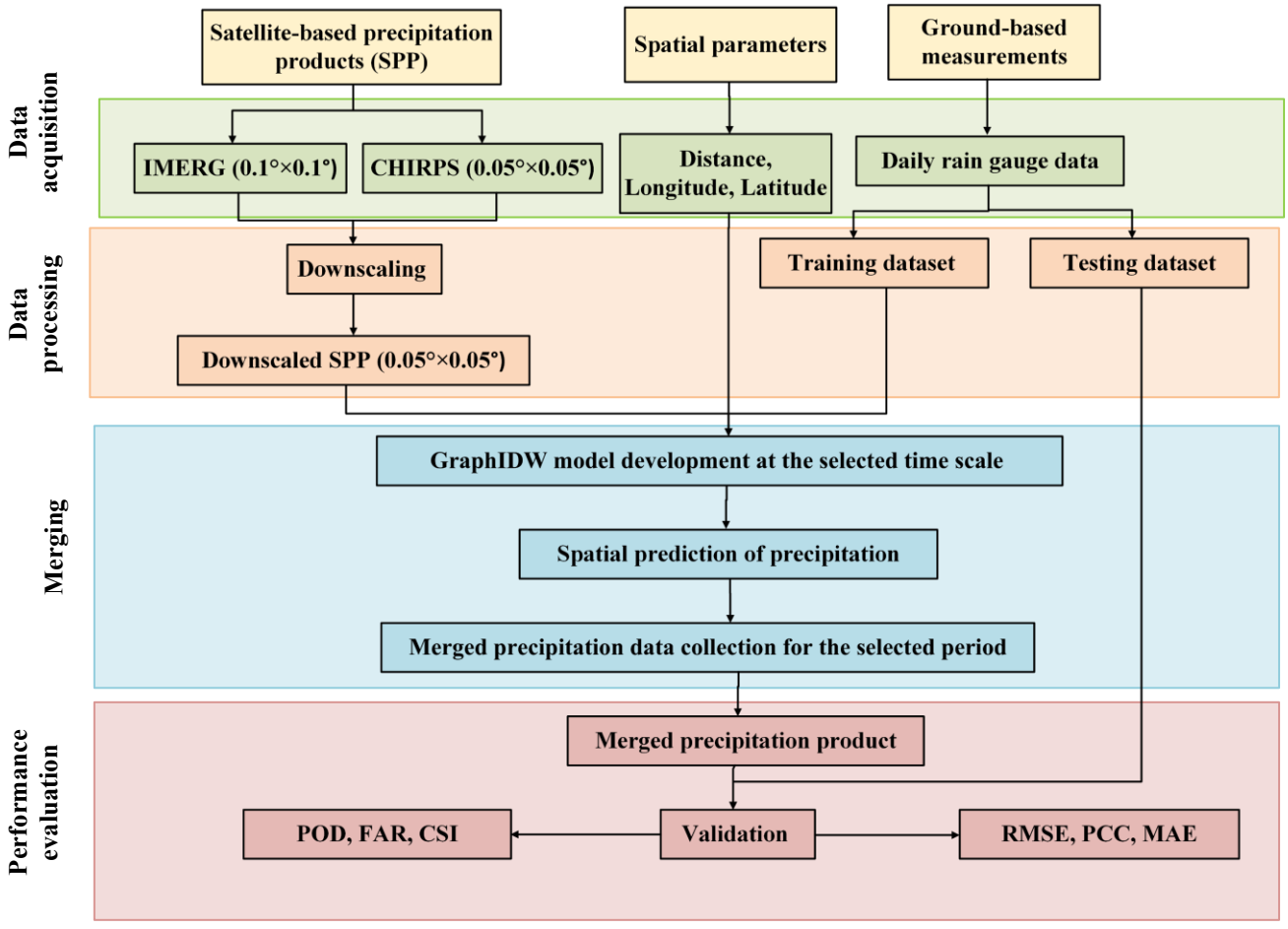


Figure 2. Methodology flowchart of the merging framework used in this study

3.1 Satellite-based precipitation products downscaling

The original spatial resolution of IMERG (0.1° × 0.1°) is too coarse for local-scale analyses in regions characterized by high spatiotemporal variability in precipitation (Tu & Duan, 2024). Traditional statistical downscaling approaches often exploit correlations between precipitation and environmental covariates such as Land Surface Temperature (LST) and the Normalized Difference Vegetation Index (NDVI) (Zhu et al., 2025). However, in the present study area, the correlation between precipitation and these covariates was found to be weak, indicating limited suitability for effective downscaling.

As an alternative approach, monthly CHIRPS precipitation data were integrated to downscale IMERG rainfall estimates to a finer spatial resolution of 0.05°. Initially, the selected SPPs (CHIRP, CHIRPS, and IMERG) were evaluated against gauge observations at both daily and monthly temporal scales. The evaluation results indicate that IMERG outperforms the other SPPs at the daily scale (Fig. 3 (a)), whereas CHIRPS demonstrates superior accuracy at the monthly scale, particularly under monsoon-dominated climatic conditions (Fig. 3(b)). In contrast, the CHIRP product exhibits relatively weaker performance at

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

both temporal scales when compared to CHIRPS. These findings underscore the enhanced accuracy of satellite-based precipitation products when satellite estimates are blended with ground-based gauge observations.

To leverage the strengths of both CHIRPS and IMERG products, CHIRPS monthly data are first disaggregated to the daily scale using the ratio between $IMERG_{daily,0.1°}$, and $IMERG_{monthly,0.1°}$. This ratio (Eq. 1) is then applied to downscale IMERG daily precipitation to a 0.05˚ spatial resolution.

$$IMERG_{daily,0.05°} = CHIRPS_{monthly,0.05°} \times \frac{IMERG_{daily,0.1°}}{IMERG_{monthly,0.1°}} \quad (1)$$

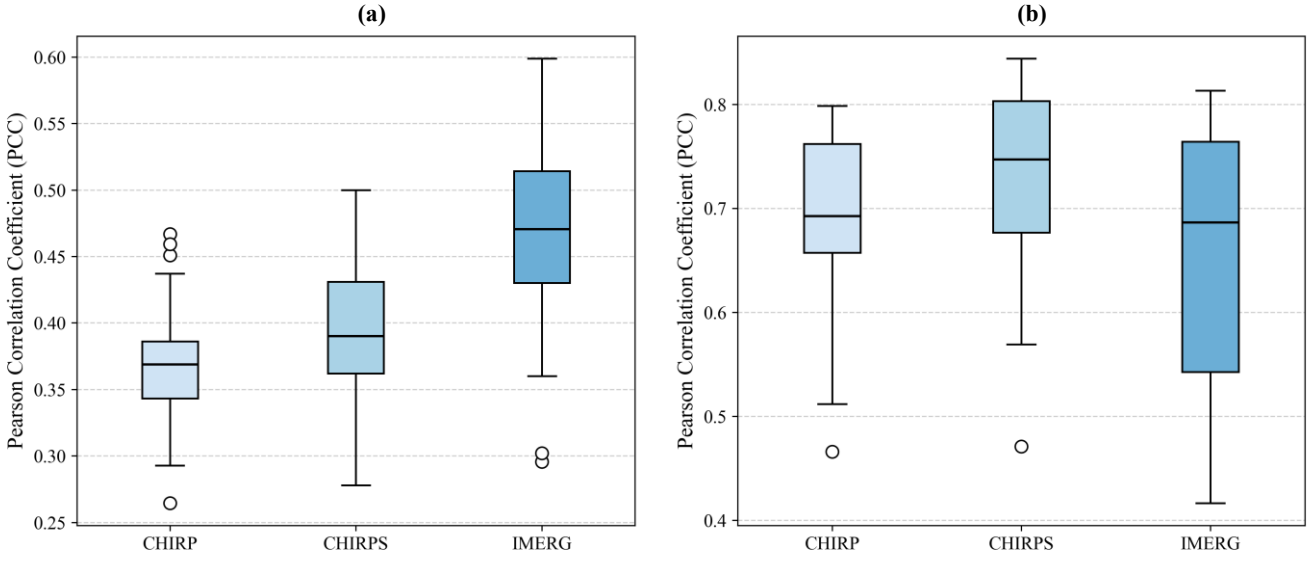


Figure 3. Box plots of variation in the Pearson correlation coefficient for CHIRP, CHIRPS and IMERG products at (a) daily and (b) monthly temporal scale

3.2 Precipitation data merging (PDM)

The primary objective of PDM is to generate gridded precipitation products with high spatiotemporal coverage and improved accuracy (Kossieris et al., 2024). In practice, PDM is typically framed as a supervised learning problem, where a model is developed to relate independent variables (predictors), such as longitude, latitude, and SPPs, to a dependent or target variable, usually gauge-based measurements, which are considered more accurate. This task is commonly formulated as a spatial prediction problem focused on estimating precipitation at locations where observations are unavailable, thereby emphasizing the spatial dimension while not explicitly addressing temporal variability (Kossieris et al., 2024).

The present study introduces GraphIDW, a novel hybrid framework that integrates GNN with IDW for precipitation data merging. This approach leverages the spatial learning capability of GNNs while incorporating IDW, aiming to enhance the reliability of gridded precipitation estimates in regions with heterogeneous rainfall and complex topography.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

3.2.1 Graph Neural Network (GNN)

Graph Neural Networks are DL methods specifically developed to operate on graph-structured data (Zhou et al., 2020). In recent years, they have attracted considerable interest due to their strong empirical performance across a broad spectrum of graph-related learning tasks. Most GNN architectures are built upon an iterative neighborhood aggregation mechanism, whereby each node refines its representation by assimilating feature information from adjacent nodes (Xu et al., 2018). After k iterations, the feature vector of a node encodes structural information from its k-hop neighbourhood, effectively capturing both local topology and node attributes. To derive a holistic representation of the entire graph, node embeddings are typically combined through a pooling operation (Ying et al., 2018).

A graph is formally defined as $G = (V, E)$, where V represents the set of nodes and E denotes the set of edges. The hidden feature representation of a node v at the l-th layer of a GNN is denoted as $h_v^{(l)}$. The neighbourhood of node v, denoted by $N(v) = \{u \in V \mid (v,u) \in E\}$ consists of all nodes directly connected to v. In the neighbourhood aggregation (or message-passing) framework, node representations are updated iteratively across layers. For a k-layer GNN, the update rule at the l-th layer ($l=1,\ldots,k$) for each node $v \in V$ can be written as Eq. (2).

$$h_v^{(l)} = \sigma\left(W_l . AGGREGATE\left(\left\{h_v^{(l-1)}, \forall u \in N(v)\right\}\right)\right) \quad (2)$$

where $AGGREGATE$ is a permutation-invariant function (e.g., mean, sum, or max) defined by the specific GNN variant, $W_{(l)}$ is a trainable weight matrix shared across all nodes at layer l, and σ indicates a non-linear activation function.

3.2.2 GNN-IDW model (GraphIDW) structure

The proposed hybrid model framework integrates GNN with IDW to improve spatial rainfall estimation. The graph is constructed by treating each IMERG grid cell as a node, with edges defined between neighbouring cells (shown in Fig. 4). Each node is characterized by its geographic coordinates (latitude, longitude) and the corresponding satellite-based rainfall estimate. For each date, these form the node attribute vector, which is given by Eq. (3).

$$x_v = [Latitude, Longitude, Satellite\ Rainfall] \in \mathbb{R}^3 \quad (3)$$

Ground-measured rainfall values are used as learning targets for training nodes. The workflow is designed in a transductive setting, where both training and evaluation nodes are embedded within the same spatial graph, but learning is restricted to training nodes.

Edge weights are assigned inversely proportional to inter-node distance, such that closer nodes exert a stronger influence. The edge weight can be written as Eq. (4).

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

$$w_{uv} = \frac{1}{d_{uv}+\varepsilon} \quad (4)$$

where d_{uv} is the Euclidean distance between nodes u and v, and ε is a small positive constant to avoid division by zero. The predictive model consists of a two-layer Graph Convolutional Network (GCN) followed by a fully connected output layer. At each GCN layer, node embeddings are updated by aggregating weighted information from their neighbours. The hidden representation of node v at layer l is given by Eq. (5).

$$h_v^{(l)} = \sigma\left(W_l \cdot \sum_{u \epsilon N(v)} w_{uv}\, h_u^{(l-1)}\right), \quad h_v^{(l)} \epsilon\, \mathbb{R}^H \quad (5)$$

where $h_u^{(l-1)}$ is the embedding of neighbour u from the previous layer, W_l is the trainable weight matrix, w_{uv} is the inverse-distance edge weight, and $\sigma(\cdot)$ denotes the ReLU activation function.

The final GCN layer L produces a high-dimensional embedding for each node. A linear head then maps this embedding into a scalar rainfall prediction (p_v), which is expressed in Eq. (6).

$$p_v = w^\top h_v^{(L)} + b \quad (6)$$

where $w \epsilon\, \mathbb{R}^H$ is a learnable weight vector, and $b \epsilon\, \mathbb{R}$ is a learnable bias term. Training is performed by minimizing the Root Mean Square Error (RMSE) loss on observed nodes only.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

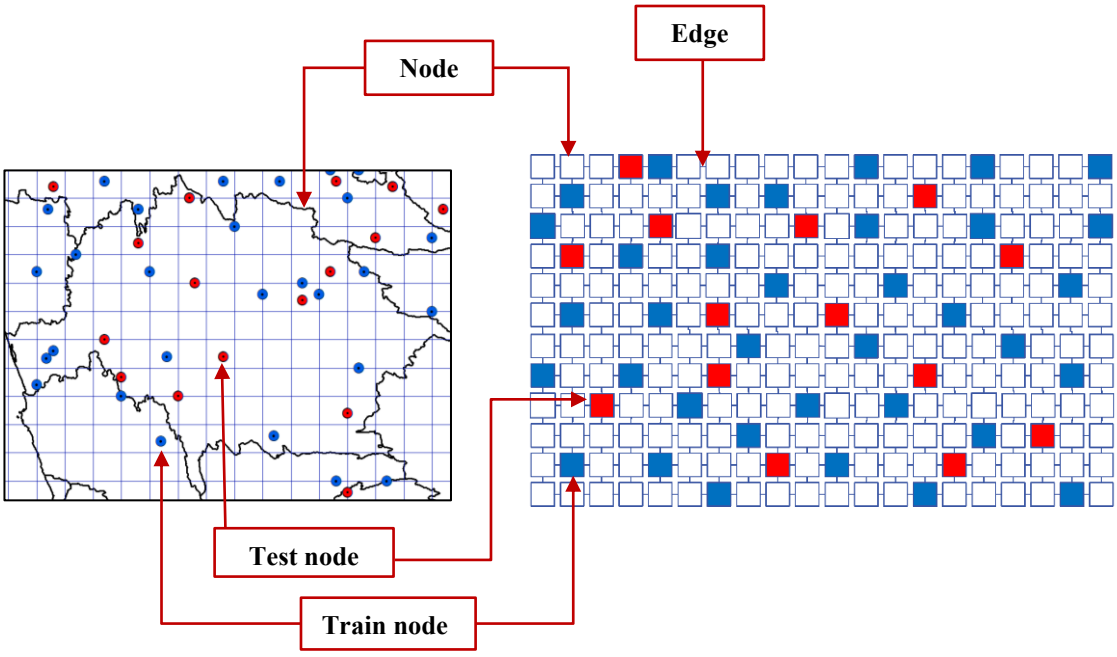


Figure 4. Graph construction in the GNN model

As a post-processing step, residuals between predicted and observed rainfall at training nodes are computed and interpolated to all nodes using IDW. The corrected rainfall estimate (P_{corr}) at each node is obtained by Eq. (7).

$$P_{corr} = P_{GNN} + r \quad (7)$$

where P_{GNN} is the direct GNN prediction, and r is the IDW-interpolated residual.

3.2.3 Merging procedure

The available rain gauge stations are split into two categories for model development and evaluation. For model training, 60 gauges (approximately 70% of the network) are randomly selected, while the remaining 28 gauges are reserved for validation. The study period spanning 2001-2015 is adopted to ensure the simultaneous availability of both ground-based observations and SPP data. At each daily time step, an individual GraphIDW model is constructed to generate grid-level precipitation estimates. Model training is performed using precipitation measurements from the training gauges as the target variable, whereas predictor inputs consist of grid-cell values of the selected covariates corresponding to the same locations. GraphIDW requires the specification of three key hyperparameters: the number of hidden layers, the learning rate, and the number of training epochs.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

3.3 Traditional ML algorithms

To evaluate the performance of the proposed merging framework, four traditional ML models are also developed for comparison. Recent studies have shown that the superiority of ML-based PDM approaches over the traditional merging methods (Baez-Villanueva et al., 2020; Nguyen et al., 2021; Zhang et al., 2021; Lei et al., 2022). Consistent with the proposed approach, a separate ML model is trained for each daily time step to estimate precipitation at the grid scale. In these models, ground-based rainfall observations from the training dataset serve as the dependent variable, while the predictors include longitude, latitude, rainfall values of the nearest neighbouring stations within the training set, and their corresponding Euclidean distances to the target grid cell.

3.3.1 Support Vector Machines (SVM)

The SVM is a widely used supervised ML algorithm applicable to both classification and regression tasks (Vapnik, 2013). Initially developed for classification problems, SVM was later extended to handle regression tasks through a formulation known as Support Vector Regression (SVR) (Suárez Sánchez et al., 2011). This generalization introduces an ε-insensitive zone around the regression function, referred to as the ε-tube, which defines an error tolerance within which deviations are ignored (Awad & Khanna, 2015). SVMs effectively address nonlinear relationships by employing kernel functions that map input features into higher-dimensional spaces where linear separation becomes possible. In this study, the Radial Basis Function (RBF) kernel is utilized, as it behaves similarly to a weighted nearest-neighbour model and significantly influences the classification or regression of new observations. The performance of the SVM model is optimized using two key hyperparameters: the regularization parameter (C) and the kernel coefficient (γ).

3.3.2. Random Forest (RF)

The Random Forest algorithm is an ensemble machine learning approach that enhances predictive accuracy and model stability by combining the outputs of multiple decision trees (Breiman, 2001). Owing to its flexibility and robustness, RF has been extensively utilized for both regression and classification applications (Nguyen et al., 2021; Wijayaweera et al., 2024). The method is based on bootstrap aggregation (bagging), whereby multiple subsets are randomly sampled from the training data to construct individual decision trees. By aggregating the outputs of these trees, the ensemble generally outperforms any single decision tree model. In regression settings, RF estimates are derived by averaging the predictions of all constituent trees, whereas in classification problems, the final class label is determined through majority voting among the trees (Ho, 1998). In this study, key hyperparameters of the RF model include the number of trees in the ensemble, the maximum depth of each tree, and the minimum number of samples required to split an internal node.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

3.3.3 Extreme Gradient Boosting (XGBoost)

The XGBoost is an ensemble learning method that constructs base learners in a sequential manner, with each successive model trained to reduce the residual errors of its predecessors (Lei et al., 2022). Compared to traditional boosting algorithms, XGBoost offers significantly improved computational efficiency and predictive performance (Karunarathna et al., 2026). The algorithm employs gradient descent optimization to minimize a specified loss function and incorporates advanced regularization techniques to mitigate overfitting. In this study, the XGBoost model was optimized using key hyperparameters, including the number of boosting iterations, learning rate, maximum tree depth, and the class weight parameter.

3.3.4 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of biological neural systems (Hassoun, 1995) and have been extensively adopted as alternatives to conventional modeling approaches across a wide range of scientific fields. A typical ANN architecture comprises three fundamental components: an input layer, one or more hidden layers, and an output layer. The input and output layers correspond to the explanatory variables and the target variable, respectively. The hidden layers consist of interconnected neurons that apply nonlinear activation functions, allowing the network to learn complex relationships. Within each hidden neuron, inputs are combined through weighted summation and subsequently transformed by an activation function, enabling the representation of nonlinear dependencies between inputs and outputs.

In this study, a Multilayer Perceptron (MLP) model trained using the backpropagation algorithm is adopted. This learning framework is recognized for its computational efficiency, simplicity, and strong predictive capability (Beikahmadi et al., 2023). MLP architecture is particularly well-suited for capturing highly nonlinear patterns without imposing prior assumptions regarding the functional form of the underlying relationships.

3.4 Evaluation Matrices

3.4.1 Point-scale evaluation

The performance of the merged precipitation framework is assessed using quantitative and categorical evaluation metrics. Quantitative measures include Root Mean Squared Error (RMSE), Pearson's Correlation Coefficient (PCC), Relative Bias (RB%), and Mean Absolute Error (MAE), which are defined in Eq. (8) to Eq. (11).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_{obs,i}-P_{merged,i}\right)^2}{n}} \quad (8)$$

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

$$PCC = \frac{\sum_{i=1}^{N}(P_{merged,i}-P_{merged,m})(P_{obs,i}-P_{obs,m})}{\sqrt{\sum_{i=1}^{n}(P_{merged,i}-P_{merged,m})^2}\sqrt{\sum_{i=1}^{n}(P_{obs,i}-P_{obs,m})^2}} \quad (9)$$

$$RB\% = \frac{\sum_{i=1}^{N}(P_{merged,i}-P_{obs,i})}{\sum_{i=1}^{N}(P_{obs,i})} \times 100\% \quad (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_{obs,i} - P_{merged,i}| \quad (11)$$

where $P_{obs,i}$ and $P_{merged,i}$ denote the observed and merged values, respectively, and $\bar{P}_{obs}$ denotes the mean of the observed values, and n is the total number of data points.

To evaluate the capability of the merged satellite precipitation products in detecting rainfall occurrence and non-occurrence, four categorical performance metrics are employed: Probability of Detection (POD), Success Ratio (SR), Bias Score (BS), and Critical Success Index (CSI). In this context, a *hit* denotes a precipitation event correctly predicted by the model, a *miss* represents an observed event that was not predicted, and a *false alarm* indicates a predicted event that did not occur. The POD measures the proportion of observed precipitation events that were successfully detected. The SR expresses the proportion of correctly predicted events among all predicted events (hits and false alarms). The BS indicates the model's tendency to overestimate or underestimate rainfall frequency, where BS > 1 suggests overprediction and BS < 1 suggests underprediction. The CSI provides an overall measure of detection skill by simultaneously considering hits, misses, and false alarms. The categorical metrics are mathematically defined in Eq. (12) to Eq. (15).

$$POD = \frac{Hits}{Hits+Misses} \quad (12)$$

$$SR = \frac{Hits}{Hits+False\ Alarms} \quad (13)$$

$$BS = \frac{Hits+\ False\ Alarms}{Hits+Misses} \quad (14)$$

$$CSI = \frac{Hits}{Hits+Flase\ Alarms+Misses} \quad (15)$$

The four categorical metrics are jointly visualized using the performance diagram proposed by Roebber (2009), which provides an integrated representation of detection skill. In this diagram, points located closer to the upper-right corner indicate higher accuracy and better detectability of precipitation events. An ideal performance is achieved when POD = SR = CSI = BS = 1. The categorical metrics are computed for each test gauge at precipitation thresholds of 1, 5, 10, and 15 mm d^{-1}, respectively. These thresholds were selected to capture different levels of rainfall intensity detection. This selection also aligns with

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

thresholds commonly used in satellite precipitation validation and hydrological applications worldwide (Paredes-Trejo et al., 2019; Zhang et al., 2021).

3.4.2 Areal-scale evaluation

Rain gauges provide point-based measurements, whereas satellite products represent areal averages over grid cells. This fundamental difference in spatial representation introduces a scale mismatch between observed rainfall and merged satellite estimates. The performance of the merged satellite precipitation product is therefore affected when it is validated against only the nearest rain gauge, as localized rainfall events may lead to large discrepancies between the point measurement and the grid-averaged estimate.

To address this limitation, in addition to point-wise evaluation, the comparison of Mean Areal Precipitation (MAP) between observations and the merged satellite product has been introduced. The MAP for the rain gauge network was derived using the Thiessen polygon method, which ensures that gauge measurements are weighted according to their representative catchment area, thereby providing a more consistent basis for comparison with the areal satellite estimates.

4. Results

4.1 Overall performance

Figure 5 illustrates the boxplots of the PCC, RMSE, and MAE values for the original $IMERG_{0.05^\circ}$ (IM_CH) and the merged SPPs obtained using the five ML algorithms over the Wet Zone of Sri Lanka. The merged products achieve noticeably higher median PCC values (Fig. 5(a)) compared to the original $IMERG_{0.05^\circ}$, indicating a substantial improvement in correlation with ground-based observations. Among all models, the proposed GraphIDW framework demonstrates the best overall performance, with median PCC values exceeding 0.70 and a narrow interquartile range (IQR), signifying both high accuracy and consistent performance across stations. The RF, XGBoost, and SVR models yield relatively similar median PCC values (0.66, 0.67, and 0.67, respectively), while the ANN model shows a slightly higher median PCC (0.70) than these three models but exhibits a wider IQR, reflecting greater variability in its performance. A single outlier with a lower PCC is observed in the results of the GraphIDW, RF, XGBoost, and SVR models, which may correspond to stations with localized rainfall anomalies or limited gauge representativeness.

Figure 5(b) presents the boxplots of RMSE for the different precipitation products. All merged SPPs obtained from the ML models exhibit a markedly better agreement with ground-based precipitation, as indicated by lower median RMSE values ranging from 13.1 to 14.3 mm d^{-1} compared to the original $IMERG_{0.05^\circ}$, which shows the median RMSE of 17.8 mm d^{-1}, confirming the effectiveness of the merging process. Among the tested algorithms, the GraphIDW framework once again outperforms the others, achieving the lowest median RMSE value (13.1 mm d^{-1}) and demonstrating the most consistent performance across stations. Following the merging process, the RMSE between the merged SPPs and observed rainfall decreases by 13%–40% compared to the original $IMERG_{0.05^\circ}$, confirming the effectiveness of the GraphIDW framework in

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

reducing estimation errors. A single outlier with relatively low RMSE appears in both the original and merged SPPs, which likely corresponds to a station experiencing comparatively low rainfall variability or minimal gauge-satellite discrepancies. As presented in Fig. 5(c), similar to the RMSE results, all merged SPPs obtained from the ML models demonstrate lower MAE values compared to the original IMERG$_{0.05^\circ}$, indicating a clear reduction in daily estimation bias. Interestingly, the SVR algorithm exhibits the best agreement with observed precipitation, achieving the lowest median MAE of 6.6 mm d^{-1}, while the other ML models show comparable median values ranging between 6.8 and 7.1 mm d^{-1}. A few isolated outliers are also observed in the MAE boxplots of some ML models. Overall, the results highlight that the GraphIDW framework delivers the most accurate and consistent rainfall estimates across the Wet Zone of Sri Lanka, achieving a 9%–36% reduction in MAE compared to the original IMERG$_{0.05^\circ}$.

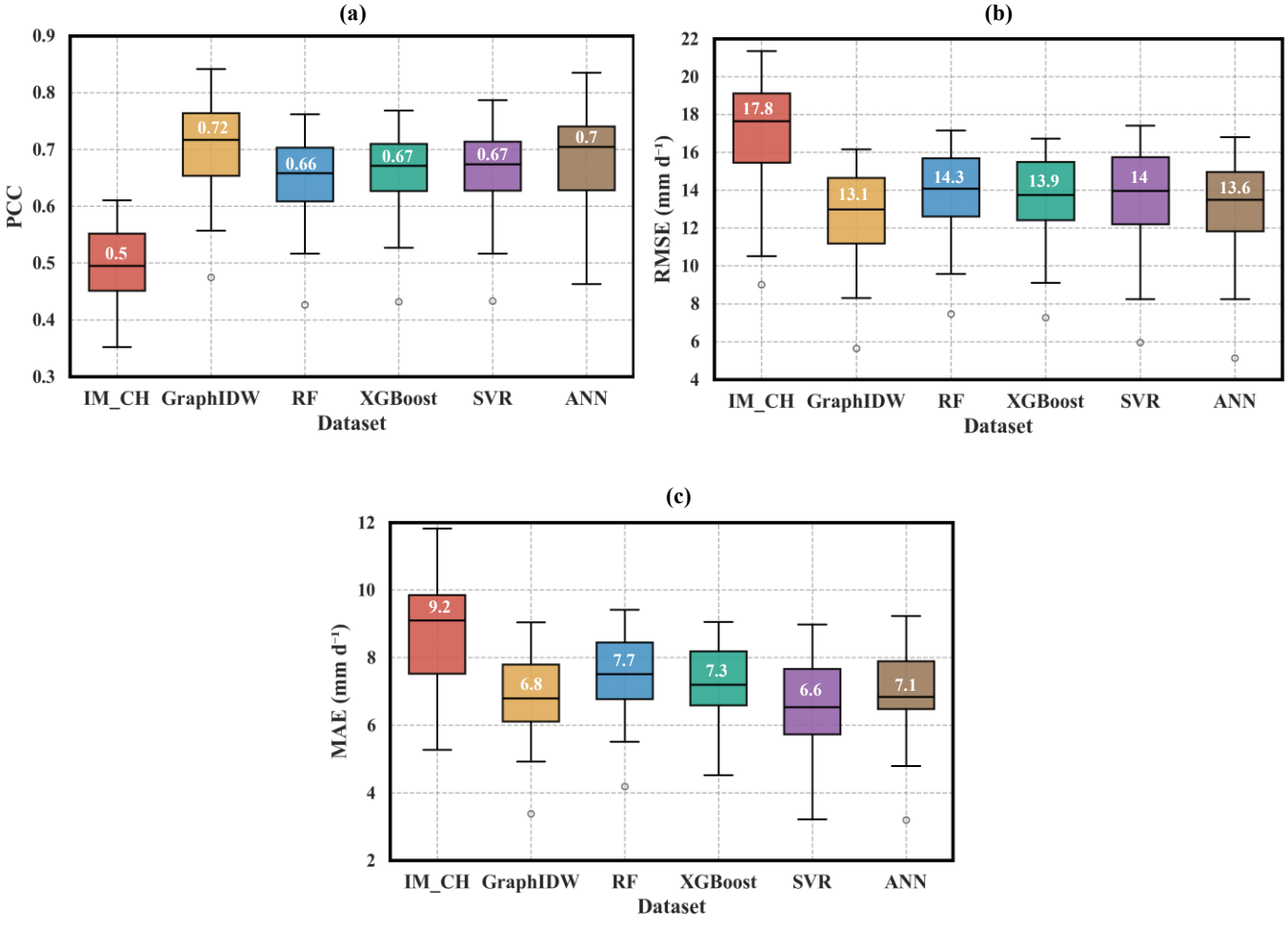


Figure 5. Boxplots of three quantitative metrics (PCC (a), RMSE (b), and MAE (c)) for six products, including original IMERG$_{0.05^\circ}$ (IM_CH) and ML-based merged products

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Figure 6 presents Roebber's performance diagrams for the original $IMERG_{0.05^\circ}$ and the merged products of the ML algorithms under different precipitation thresholds across the study area.

At the 1 mm d^{-1} threshold (Fig. 6(a)), all ML-based products achieve high POD and SR values, demonstrating strong capability in detecting light rainfall events. The GraphIDW model shows the best overall detectability, with CSI values approaching 0.6, followed by RF, ANN, and XGBoost, while the SVR model deviates from the others with a comparatively lower POD value (0.85). All products, including the original $IMERG_{0.05^\circ}$, tend to overestimate light precipitation, as indicated by Bias Score (BS) > 1.

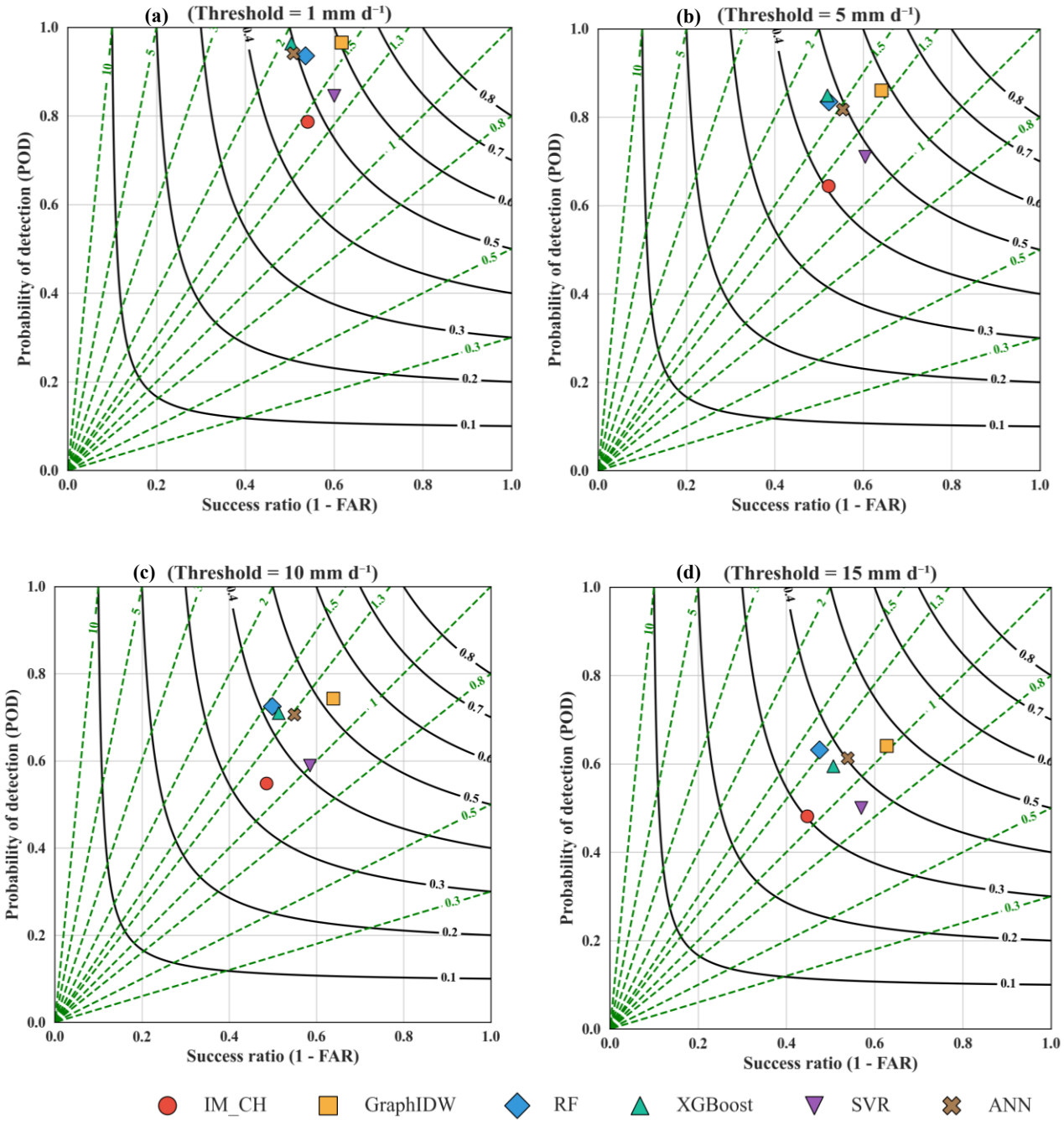
At the 5 mm d^{-1} threshold (Fig. 6(b)), a slight reduction in detection accuracy is observed for all models as rainfall intensity increases. Nevertheless, GraphIDW maintains the highest performance, exhibiting the highest POD, CSI, and SR values. The RF, ANN, and XGBoost models perform comparably, whereas SVR continues to show relatively weaker detection capability.

At the 10 mm d^{-1} threshold (Fig. 6(c)), the POD decreases to around 0.7 for most ML models, except for SVR, which achieves a comparatively lower value of 0.6. Even under this higher threshold, GraphIDW sustains its superior performance with balanced detection accuracy and higher SR.

At the 15 mm d^{-1} threshold (Fig. 6(d)), the performance of all models declines further. The reduced frequency of intense rainfall events could be a contributing factor. Nonetheless, GraphIDW retains the highest detection capability, closely followed by RF, XGBoost, and ANN, confirming its robustness under varying rainfall conditions.

Overall, GraphIDW demonstrates the best performance across all four thresholds. Although its precipitation detection capability decreases with increasing rainfall intensity, its Success Ratio remains nearly constant (SR ≈ 0.63). The model tends to overpredict light precipitation events (higher BS values), which gradually decrease with increasing thresholds. All ML-based merged products outperform the original $IMERG_{0.05^\circ}$ in rainfall event detection, though their overall detection capability diminishes as rainfall intensity increases.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026



400

Figure 6. Roebber's performance diagram for the original IMERG$_{0.05°}$ (IM_CH) and the merged precipitation products. The green dashed lines represent the bias score (BS), while the black contour lines indicate the critical success index (CSI). The four diagrams present the detection performance for the precipitation thresholds of 1, 5, 10, and 15 mm d^{-1}, respectively.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

4.2 Intensity evaluation

In the previous evaluation, the overall performance of the merged products was assessed without differentiating between rainfall intensities. However, precipitation intensity varies substantially across the complex terrain of the study area, necessitating a more detailed assessment across different rainfall regimes. To address this, precipitation events are categorized into three intensity ranges: light rain (1–15 mm d^{-1}), moderate rain (15–50 mm d^{-1}), and heavy rain (>50 mm d^{-1}).

Figure 7 presents the histograms of three evaluation metrics (PCC, RMSE, and RB%) for both the original and merged SPP products. As shown in Fig. 7(a), the GraphIDW model achieves the highest PCC values across all intensity categories, followed closely by the ANN model. Notably, all ML-based merged products demonstrate stronger correlations with gauge observations compared to the original $IMERG_{0.05°}$, highlighting their improved ability to reproduce observed rainfall variability.

According to Fig. 7(b), the performance divergence becomes more evident at higher intensities (>50 mm d^{-1}), where the GraphIDW model shows the greatest reduction in RMSE relative to the original $IMERG_{0.05°}$, indicating superior skill in capturing heavy rainfall events. All ML models exhibit comparable accuracy in estimating light and moderate rainfall. As illustrated in Fig. 7(c), moderate and heavy rainfall events are generally underestimated by all ML models, whereas light rainfall tends to be overestimated, as reflected in the relative bias patterns.

Overall, the GraphIDW model consistently demonstrates the most balanced performance across all rainfall thresholds, achieving the highest correlation, lowest RMSE, and smallest bias magnitude, confirming its robustness and suitability for spatial precipitation merging in complex tropical terrains.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

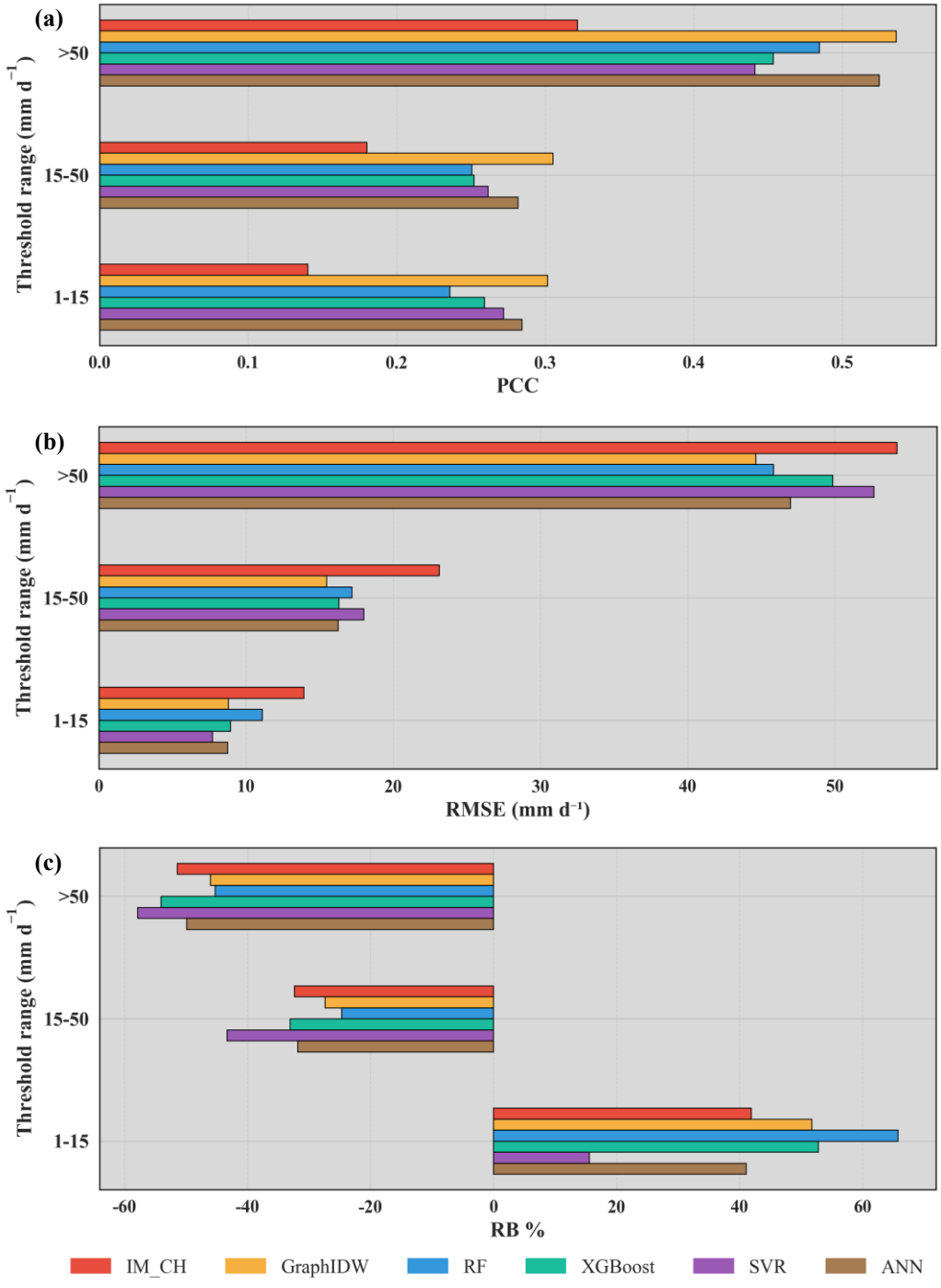


Figure 7. Histogram comparison of (a) PCC, (b) RMSE, and (c) RB% of six precipitation products against ground observations under different rain intensities. (Heavy rain: >50 mm d^{-1}; Moderate rain: 15–50 mm d^{-1}; Light rain: 1–15 mm d^{-1}).

435

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

4.3 Impact of localized events

Since the RMSE values for heavy rainfall events (>50 mm d^{-1}) are considerably higher in the ML-based merged products, an additional analysis is carried out to investigate the influence of localized or isolated rainfall events on RMSE estimation. The Kalu and Kelani River Basins, located entirely within Sri Lanka's Wet Zone, are selected for this purpose. First, all dates on which the rainfall at each selected test gauge exceeded 50 mm are identified, and the corresponding rainfall values at other gauges are extracted for those dates. Boxplots are then generated for each date using these values. If the rainfall value at the selected gauge appears as an outlier in its respective boxplot, it is classified as a *localized event* for that gauge. In this study, outliers are defined as data points that fall below the lower fence (Q1 − 1.5 × IQR) or above the upper fence (Q3 + 1.5 × IQR), where Q1, Q3, and IQR denote the first quartile, third quartile, and interquartile range, respectively. Subsequently, RMSE values are computed separately for localized and non-localized events at each gauge location. The summary of localized events in each test gauge in the basins is shown in Table 2. The average proportion of localized rainfall events exceeding 50 mm during the period 2001–2015 is approximately 30% in both river basins. This relatively high percentage has contributed to the notable increase in RMSE values observed in the ML-based merged products. Figure 8 presents boxplots of RMSE values for localized, non-localized, and all heavy rainfall events (>50 mm). The results reveal no substantial improvement in RMSE for localized events across the ML-based merged products, indicating persistent difficulty in accurately capturing highly isolated rainfall patterns. In contrast, all ML models exhibit clear improvements in estimating non-localized events, where rainfall is more spatially coherent.

Among the tested models, the GraphIDW model demonstrates the best overall performance in reproducing non-localized heavy rainfall, closely followed by the RF model. This finding aligns with the lower RMSE values observed in GraphIDW and RF models for events exceeding 50 mm in Fig. 7(b). Specifically, in the GraphIDW model, RMSE for non-localized events decreases markedly from the range of 43–60 mm to 26–55 mm, while for localized events the reduction is more limited from 57–74 mm to 49–73 mm. Overall, the GraphIDW framework achieves a reduction in total RMSE from 48–61 mm to 31–57 mm, highlighting its superior capability in mitigating errors under spatially consistent rainfall conditions. Notably, the original $IMERG_{0.05^\circ}$ also performs better for non-localized events than for localized ones, emphasizing the inherent challenge of representing highly localized rainfall through satellite-based or merged approaches.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Table 2. Percentage of localized events in test gauges in the Kelani and Kalu River Basins

Basin	Gauge Location	No. of Events Exceeding 50 mm	No. of Localized Events	Localized Events (%)
Kelani	Dunedin Estate	119	36	30
	Elston	290	75	26
	Maliboda	305	102	33
	Maussakelle	133	22	17
	Meegoda-Nawalamulla	189	54	29
	Udabage	173	52	30
Kalu	Agalawatta	291	89	31
	Alupolla Group	294	104	35
	Depedena Group	194	75	39
	Eheliyagoda S.P	294	88	30
	Gonapenigala Iranganie	48	17	35
	Ratnapura	229	45	20
	Rayigama	223	44	20

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

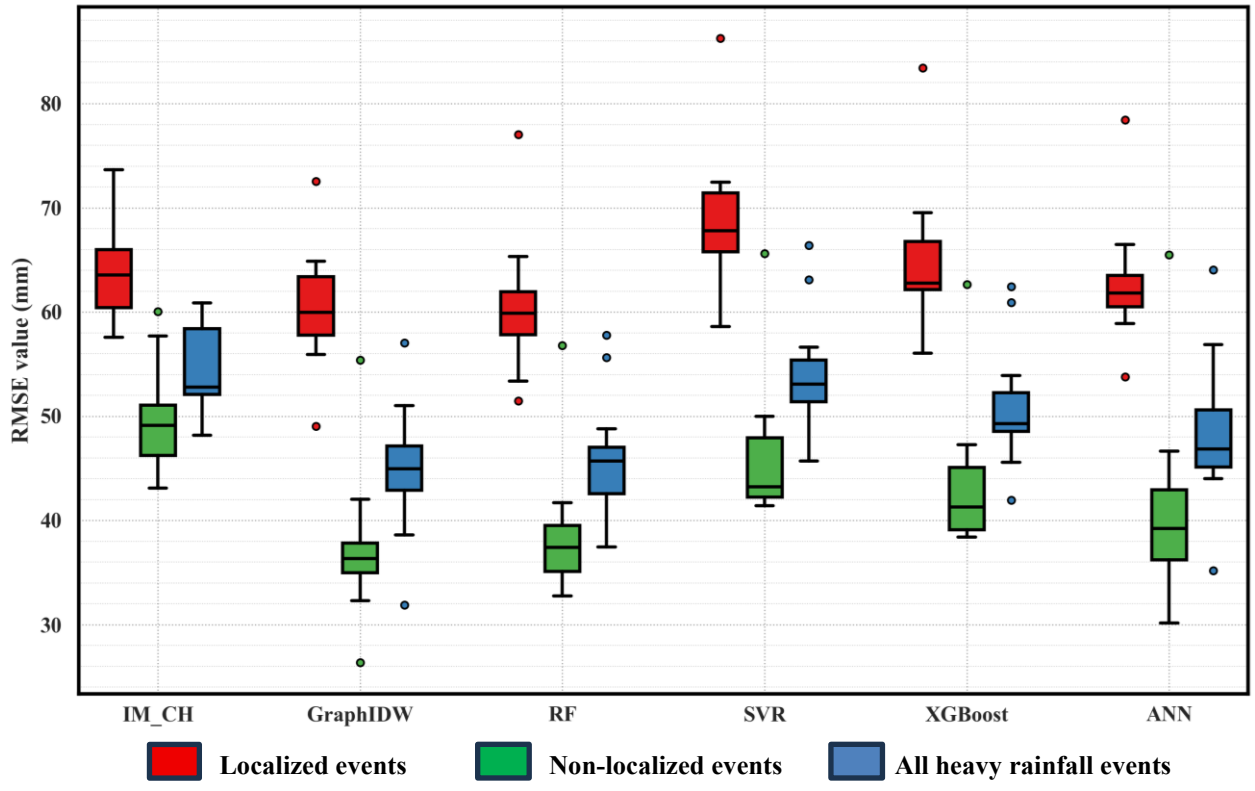


Figure 8. Box Plots of RMSE variation in localized, non-localized, and all heavy rainfall events (>50 mm) across all the test gauges

4.4 Performance during flood events

According to the previous analysis, the GraphIDW framework demonstrated superior performance in PDM compared to other ML-based approaches. Figure 9 and Figure 10 illustrate the performance of the GraphIDW model during major flood events. For this evaluation, eight flood events that occurred in the Kelani and Kalu River Basins between 2001 and 2015 have been selected, with their details summarized in Table 3.

Compared to the original IMERG$_{0.05^\circ}$, the GraphIDW-merged product shows improved accuracy in estimating rainfall magnitudes during all flood events. Notably, during these events, the entire basin experienced widespread heavy rainfall, indicating that spatially distributed rainfall plays a more dominant role in flood generation than localized precipitation events. The GraphIDW model successfully reproduces the observed spatial rainfall patterns associated with these floods. However, minor discrepancies are observed in certain locations where extremely high rainfall occurs; in such instances, GraphIDW tends to slightly underestimate the precipitation intensity relative to the observed values.

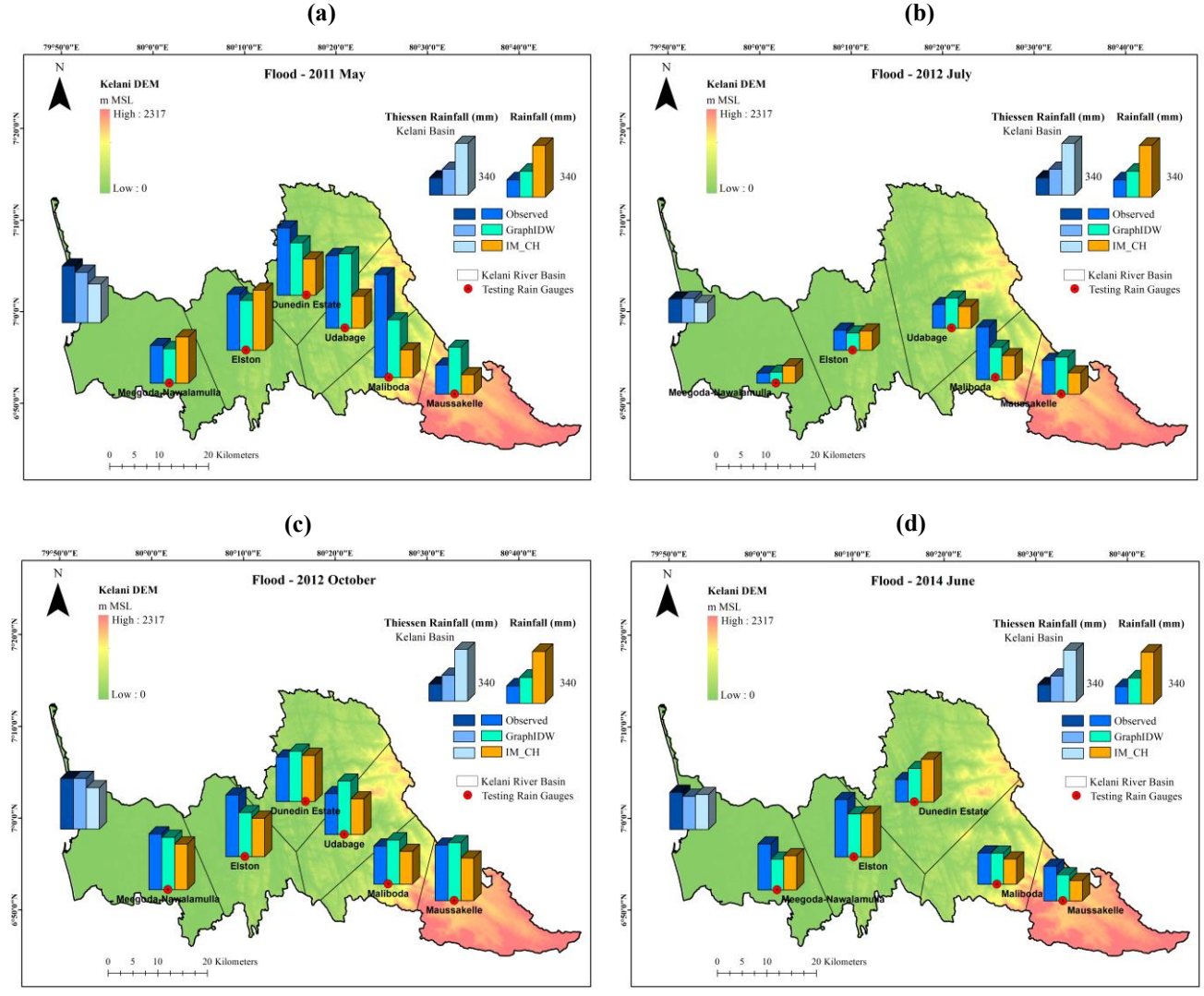
https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Table 3: Summary of the floods in the Kelani and Kalu River Basins

Year	Duration	Basin
2011 May	23 May 2011 to 31 May 2011	Kelani
	22 April 2011 to 01 May 2011	Kalu
2012 July	01 July 2012 to 10 July 2012	Kelani
		Kalu
2012 October	24 October 2012 to 02 November 2012	Kelani
	26 October 20212 to 02 November 2012	Kalu
2014 June	29 May 2014 to 03 June 2014	Kelani
	28 May 2014 to 05 June 2014	Kalu

(Source: Irrigation Department, Sri Lanka)

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026



525 **Figure 9.** Comparison of ground-observed and model-predicted rainfall amounts from the original IMERG$_{0.05^\circ}$ (IM_CH) and GraphIDW-merged products, along with the Thiessen-weighted average rainfall for the Kelani River Basin during the flood events of (a) May 2011, (b) July 2012, (c) October 2012, and (d) June 2014.

530

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

535

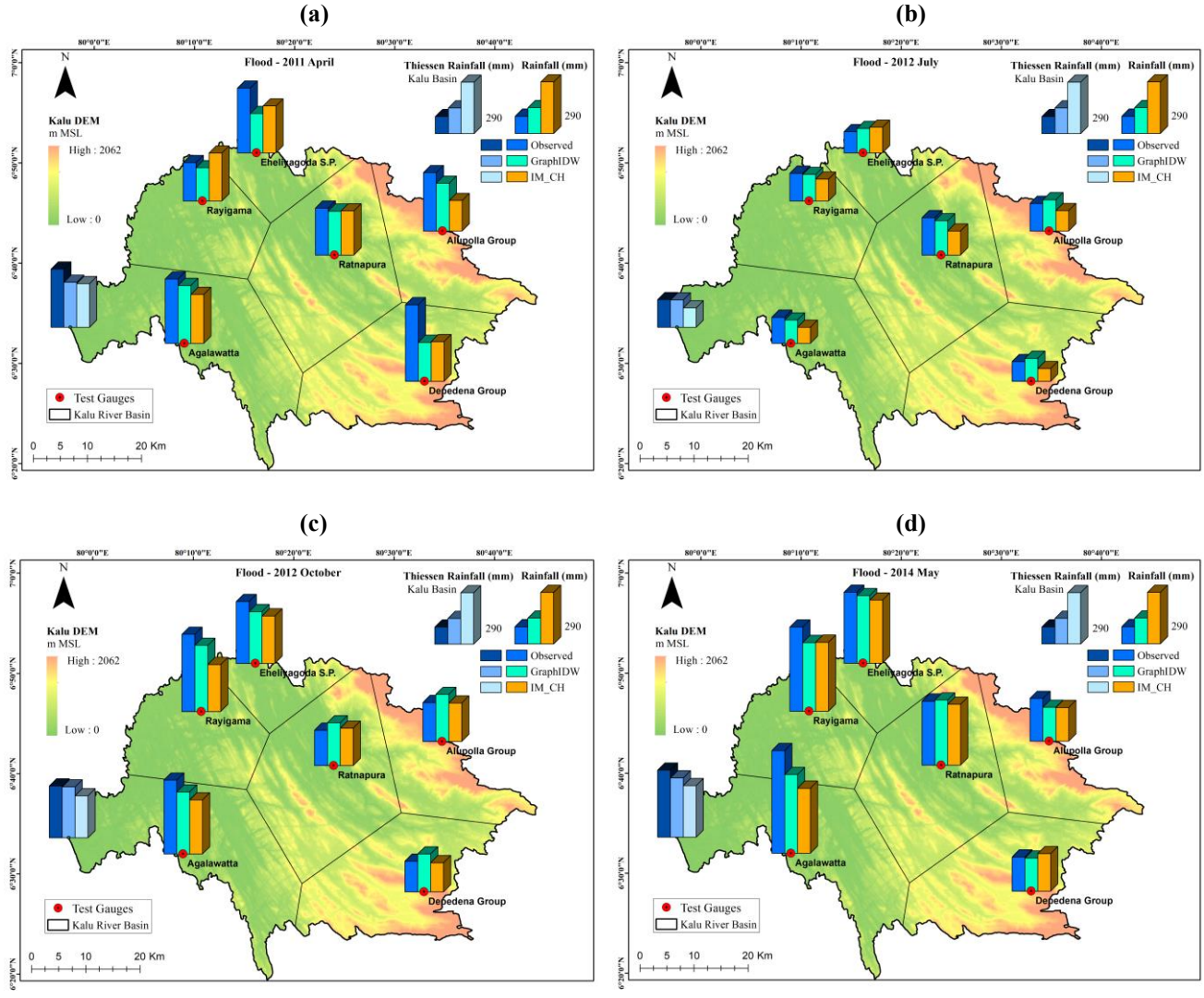


Figure 10. Comparison of ground-observed and model-predicted rainfall amounts from the original IMERG$_{0.05^\circ}$ (IM_CH) and GraphIDW-merged products, along with the Thiessen-weighted average rainfall for the Kalu River Basin during the flood events of (a) April 2011, (b) July 2012, (c) October 2012, and (d) May 2014.

540 **4.5 Spatial Evaluation**

Figure 11 illustrates the spatial variation of the single mass curve (SMC) gradient (indicated by star symbols) and the RMSE between the original IMERG$_{0.05^\circ}$ and observed rainfall (indicated by circular symbols). The results reveal a clear spatial heterogeneity in rainfall distribution across the Wet Zone. Areas located in the central highlands exhibit lower SMC gradient values, indicating relatively low accumulated rainfall. Similarly, coastal and boundary regions also experience comparatively

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

low rainfall amounts. In contrast, the western slopes of the central highlands display higher SMC gradients, signifying intense rainfall influenced primarily by orographic effects.

Notably, the original IMERG$_{0.05^\circ}$ demonstrates poor agreement with ground observations in high-rainfall regions, as reflected by the elevated RMSE values. A positive relationship is observed between rainfall and RMSE magnitude, where errors tend to increase with higher rainfall amounts. The comparative spatial analysis (Fig. 12) further indicates that the GraphIDW-merged product substantially reduces RMSE across most regions compared to the original IMERG$_{0.05^\circ}$. However, some discrepancies remain over the western slopes of the central highlands, where rainfall intensity and spatial variability are highest.

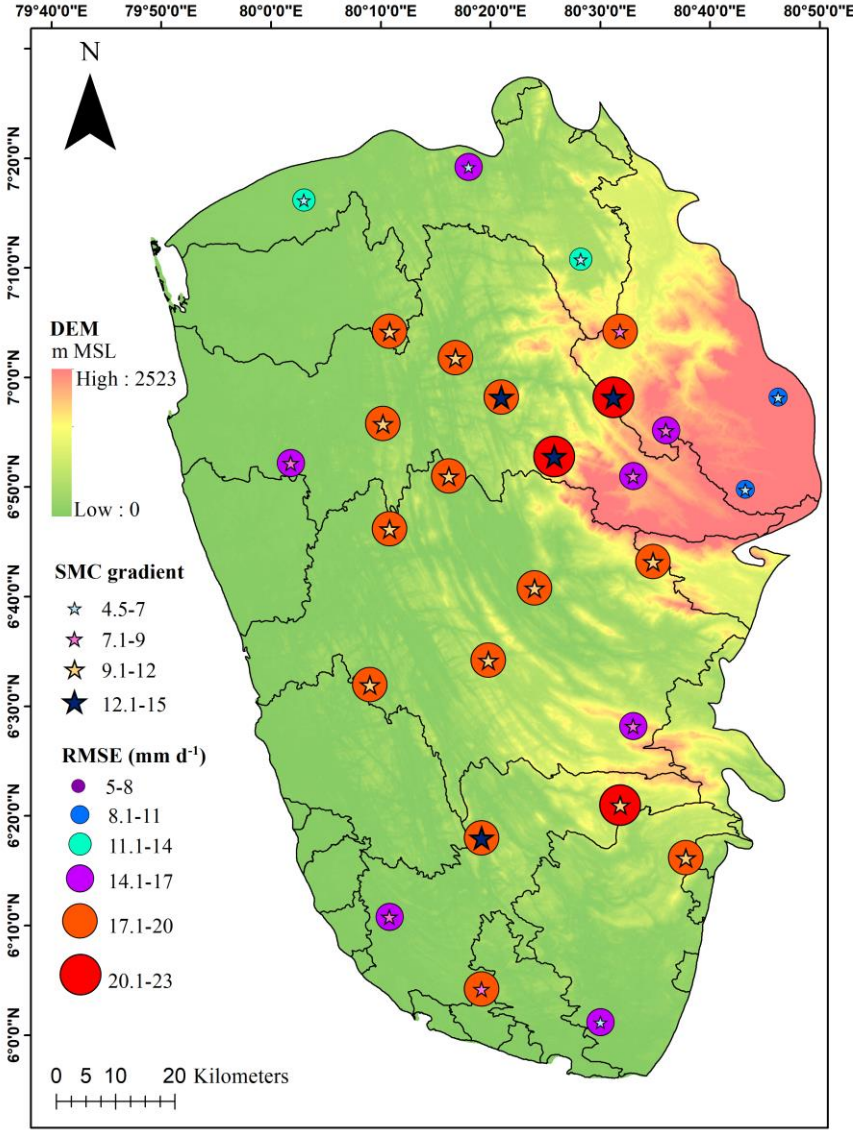


Figure 11. Spatial variability of single mass curve (SMC) gradients and corresponding RMSE distribution of the original IMERG$_{0.05^\circ}$ across the Wet Zone of Sri Lanka.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

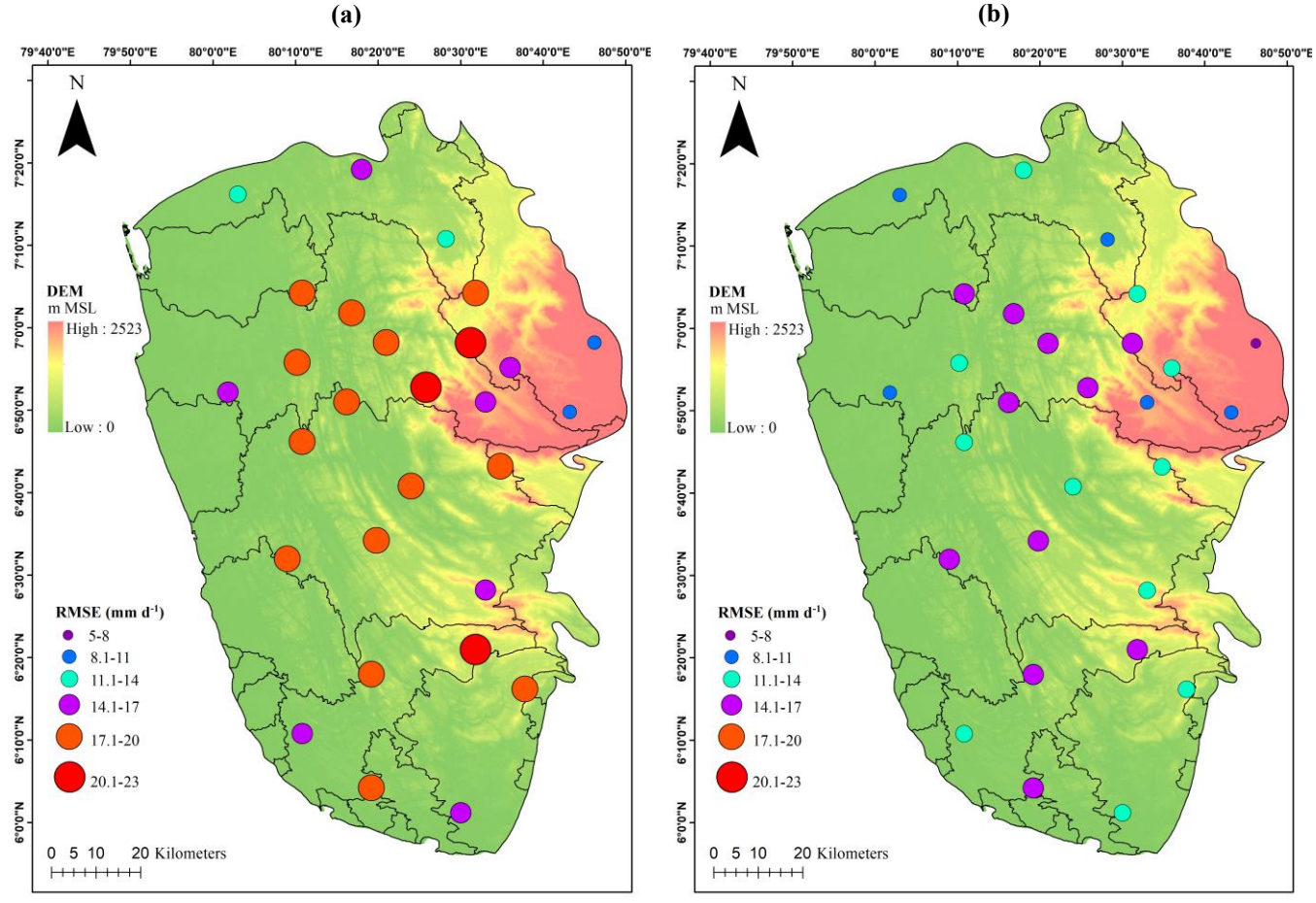


Figure 12. Spatial variation of RMSE in the (a) original IMERG$_{0.05^\circ}$ and (b) GraphIDW-predicted rainfall across the Wet Zone of Sri Lanka.

5. Discussion

As illustrated in Fig. 12(a), the largest errors occur along the western slopes of the Central Highlands, where the original IMERG$_{0.05^\circ}$ product systematically underestimates precipitation. This bias indicates a potential limitation of IMERG in representing orographic rainfall, consistent with previous findings, which demonstrate that SPPs derived from combined IR and PMW observations tend to perform poorly over complex mountainous regions (Derin & Yilmaz, 2014; Tan et al., 2015; Huang et al., 2018).

This study proposes a novel GraphIDW (GNN-IDW), precipitation merging framework that explicitly incorporates the spatial autocorrelation of rainfall to enhance satellite precipitation accuracy. In addition to the proposed approach, results from five conventional ML algorithms were also presented for comparison. Following the approach of Baez-Villanueva et al. (2020) and Zhang et al. (2021), the models were trained on a daily timescale, which has been shown to improve precipitation estimation accuracy. All ML-based merging techniques demonstrated improved performance over the original IMERG$_{0.05^\circ}$. However, the

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

GraphIDW framework consistently outperformed the other ML-based methods, emphasizing the significance of incorporating spatial autocorrelation into rainfall merging algorithms. These results are in agreement with earlier studies that explicitly incorporated spatial dependence into ML models (Baez-Villanueva et al., 2020; Sekulić et al., 2020; Nguyen et al., 2021). Figure 13 illustrates the computational training time of the five ML models using a training sample of 60 rain gauges. Among them, the RF model required the longest training time, primarily due to its ensemble structure and parallel tree construction process (Lei et al., 2022). In contrast, the GraphIDW model demonstrated the shortest runtime, indicating superior computational efficiency. When jointly evaluating model complexity, predictive performance, and computational efficiency, the proposed GraphIDW framework demonstrates the most favorable performance for daily precipitation estimation across the study region.

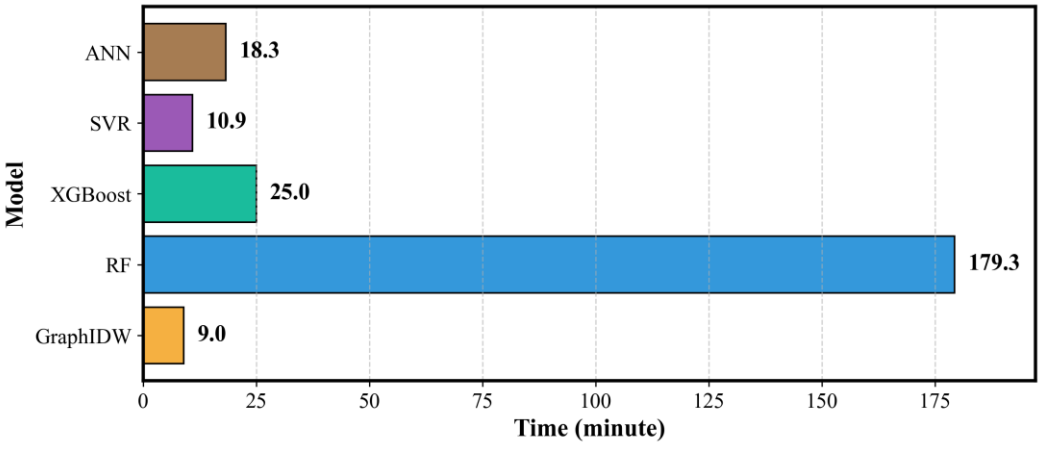


Figure 13. Computation time of five ML models for PDM

5.1 Impact of rain gauge density

The density of rain gauges significantly influences the performance of ML-based merged precipitation products, as all models rely exclusively on gauge observations. To examine this sensitivity, the models were trained using three subsets of the gauge network corresponding 40%, 50%, and 70% of the total network. Figure 14 presents boxplots of the RMSE values for the ML-based merged products at test gauges under different training sample sizes. For comparison, the RMSE range of the original $IMERG_{0.05^{\circ}}$, which remains independent of training sample size, is also shown.

Model performance exhibited a clear reduction as the number of training gauges decreased, in line with observations reported in earlier studies (Nguyen et al., 2021; Zhang et al., 2021; Lei et al., 2022). This reduction in sample size limits the amount of information available for model learning, thereby affecting accuracy (Chang et al., 2019). Among the models, GraphIDW and ANN exhibited the smallest increase in RMSE with decreasing training density, indicating higher robustness. In contrast, the RF model showed the greatest sensitivity to changes in input data, resulting in reduced accuracy. Overall, these results confirm that the proposed GraphIDW framework is both effective and resilient, making it a promising approach for improving precipitation estimates in data-scarce regions.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

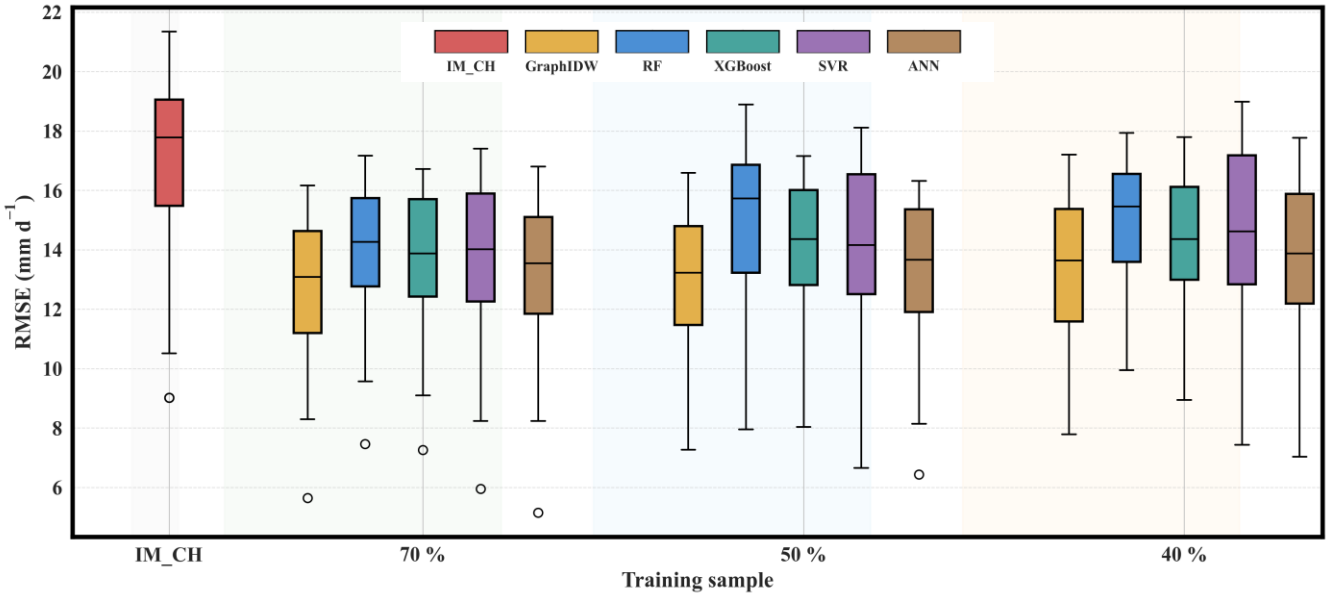


Figure 14. Performance comparison of the original IMERG$_{0.05°}$ and ML-based merged products across varying training sample sizes. The percentages represent the proportion of training samples relative to the total available dataset.

5.2 Limitations and uncertainties

Consistent with previous studies (Baez-Villanueva et al., 2020; Zhang et al., 2021; Kossieris et al., 2024), we assumed that point-scale gauge observations are representative of grid-scale precipitation values. However, this assumption may introduce uncertainties, particularly in regions with complex topography and strong spatial rainfall gradients.

In addition, the downscaling process itself could also contribute additional uncertainty to the merged results. Since the proposed GraphIDW framework relies solely on SPPs and ground-based observations, achieving optimal performance requires high-quality and well-calibrated rainfall data.

6. Conclusions

This study presents a novel GraphIDW framework for merging SPPs with ground-based observations, specifically designed to enhance rainfall estimation over spatially heterogeneous tropical terrains. The proposed approach integrates GNNs with IDW interpolation, enabling the model to account for the spatial autocorrelation inherent in rainfall patterns.

The performance of the GraphIDW framework was rigorously evaluated using 28 independent rain gauges located in the Wet Zone of Sri Lanka, covering the period 2001–2015. Its accuracy and reliability were compared against the original SPPs and four widely used ML algorithms (RF, ANN, XGBoost, and SVR) through a comprehensive set of statistical and categorical metrics. The key findings are summarized below:

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

1. All ML algorithms substantially reduced discrepancies between SPPs and gauge observations, improving rainfall detection capability. GraphIDW achieved the highest POD, SR, and CSI values (0.97, 0.62, and 0.60, respectively) for the 1 mm d^{-1} threshold.
2. Error statistics revealed that GraphIDW achieved the greatest accuracy, reducing RMSE by 13%–41% and MAE by 9%–36% relative to the original SPPs.
3. Across rainfall intensity classes, GraphIDW exhibited consistently superior performance in detecting light, moderate, and heavy rainfall, although limited improvements were observed for localized extreme rainfall (>50 mm d^{-1}), likely due to the high spatial variability of convective storms.
4. The model effectively reproduced observed spatial rainfall distributions during major flood events, though minor underestimations persisted at certain high-rainfall locations.
5. Quantitative and categorical evaluations consistently confirmed the superiority of GraphIDW over other ML-based merging approaches, emphasizing the critical role of spatial autocorrelation in improving precipitation merging accuracy.
6. Model performance improved with increasing gauge density, highlighting the importance of spatially representative training data.
7. In addition to its accuracy, GraphIDW demonstrated high computational efficiency, requiring the least training time among the tested ML algorithms.

In summary, the major contributions of this study are twofold. First, it introduces a novel ML framework (GraphIDW) for precipitation data merging that effectively integrates spatial dependencies without relying on auxiliary variables. Second, to the best of our knowledge, this represents the first application of GNNs for PDM, demonstrating their potential for data-scarce tropical regions. The proposed framework offers a scalable and computationally efficient solution for generating high-resolution, bias-corrected precipitation datasets, thereby providing a valuable foundation for hydrological modeling and climate impact assessments in complex terrains. Given its efficiency and spatial learning capability, GraphIDW also holds promise for future applications such as real-time rainfall forecasting and long-term climate analyses.

Code availability. The codes used to generate the results of this study are available from the corresponding author upon request.

Data availability. IMERG V07 can be accessed from the NASA Global Precipitation Measurement (GPM) website (https://gpm.nasa.gov/data). CHIRPS V2 and CHIRP V2 are available via the University of California, Santa Barbara, Climate Hazards Center (CHC) website (https://www.chc.ucsb.edu/data/chirps/). The observed rainfall data for this study are available from the Department of Meteorology, Sri Lanka, upon request.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Author contribution. NP: conceptualisation, methodology, data curation, software, visualisation, writing (original draft preparation). CP: conceptualisation, methodology, supervision, writing (review and editing). NW: conceptualisation, methodology, supervision, writing (review and editing). LR: conceptualisation, methodology, supervision, writing (review and editing). AW: conceptualisation, resources, writing (review and editing).

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors would like to express their gratitude to the Department of Meteorology, Sri Lanka for providing data free of charge for this research. Additionally, the support extended by the UNESCO Madanjeet Singh Centre for South Asia Water Management (UMCSAWM) and the Department of Civil Engineering, University of Moratuwa are greatly acknowledged.

During the preparation of this manuscript, the authors used ChatGPT (version 4, OpenAI) and Grammarly AI tools to refine English grammar and sentence structure. Following the use of these tools, the authors carefully reviewed and edited the text as necessary and take full responsibility for the content of the publication.

Financial support. This research was financially supported by the Climate Action for Advanced Inclusive Education (CAFIED) Project, funded under the UNESCO Participation Programme 2024–2025, whose support made this work possible.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

References

Assiri, M. E. and Qureshi, S.: A Multi-Source Data Fusion Method to Improve the Accuracy of Precipitation Products: A Machine Learning Algorithm, Remote Sensing, 14, 6389, https://doi.org/10.3390/rs14246389, 2022.

Avanzi, F., Ercolani, G., Gabellani, S., Cremonese, E., Pogliotti, P., Filippa, G., Morra di Cella, U., Ratto, S., Stevenin, H., Cauduro, M., and Juglair, S.: Learning about precipitation lapse rates from snow course data improves water balance modeling, Hydrology and Earth System Sciences, 25, 2109–2131, https://doi.org/10.5194/hess-25-2109-2021, 2021.

Awad, M. and Khanna, R.: Support Vector Regression, in: Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, edited by: Awad, M. and Khanna, R., Apress, Berkeley, CA, 67–80, https://doi.org/10.1007/978-1-4302-5990-9_4, 2015.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., and Xuan Thinh, N.: RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements, Remote Sensing of Environment, 239, 111606, https://doi.org/10.1016/j.rse.2019.111606, 2020.

Beikahmadi, N., Francipane, A., and Noto, L. V.: Smart Data Blending Framework to Enhance Precipitation Estimation through Interconnected Atmospheric, Satellite, and Surface Variables, Hydrology, 10, 128, https://doi.org/10.3390/hydrology10060128, 2023.

Bertoncini, A. and Pomeroy, J. W.: Quantifying spatiotemporal and elevational precipitation gauge network uncertainty in the Canadian Rockies, Hydrology and Earth System Sciences, 29, 983–1000, https://doi.org/10.5194/hess-29-983-2025, 2025.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.

Chang, T. K., Aziz, N. A. A., Hussain, A., and Sulaiman, J.: The impact of training data sequence on the performance of neuro-fuzzy rainfall–runoff models with online learning, *Water*, 11, 52, https://doi.org/10.3390/w11010052 , 2018.

Chao, L., Zhang, K., Li, Z., Zhu, Y., Wang, J., and Yu, Z.: Geographically weighted regression based methods for merging satellite and gauge precipitation, Journal of Hydrology, 558, 275–289, https://doi.org/10.1016/j.jhydrol.2018.01.042, 2018.

Department of Irrigation, Sri Lanka.: Hydrological Annual 2010–2011, Department of Irrigation, Colombo, Sri Lanka, 2011.

Department of Irrigation, Sri Lanka.: Hydrological Annual 2011–2012, Department of Irrigation, Colombo, Sri Lanka, 2012.

Department of Irrigation, Sri Lanka.: Hydrological Annual 2012–2013, Department of Irrigation, Colombo, Sri Lanka, 2013.

Department of Irrigation, Sri Lanka.: Hydrological Annual 2013–2014, Department of Irrigation, Colombo, Sri Lanka, 2014

Derin, Y. and Yilmaz, K. K.: Evaluation of Multiple Satellite-Based Precipitation Products over Complex Topography, https://doi.org/10.1175/JHM-D-13-0191.1, 2014.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., and Verdin, A. P.: A quasi-global precipitation time series for drought monitoring, Data Series, U.S. Geological Survey [dataset], https://data. chc.ucsb.edu/products/CHIRPS-2.0/ (last access: 14 November 2025), 2014.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Sci Data, 2, 150066, https://doi.org/10.1038/sdata.2015.66, 2015.

Hassoun, M. H.: Fundamentals of Artificial Neural Networks, MIT Press, 546 pp., 1995.

Hengl, T., Nussbaum, M., Wright, M., Heuvelink, G., and Graeler, B.: Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, https://doi.org/10.7287/peerj.preprints.26693v2, 2018.

Ho, T. K.: The Random Subspace Method for Constructing De cision Forests (PDF), IEEE. T. Pattern. Anal., 20, 832–844, https://doi.org/10.1109/34.709601, 1998.

Huang, W.-R., Chang, Y.-H., and Liu, P.-Y.: Assessment of IMERG precipitation over Taiwan at multiple timescales, Atmospheric Research, 214, 239–249, https://doi.org/10.1016/j.atmosres.2018.08.004, 2018.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., and Joyce, R.: Algorithm theoretical basis document (ATBD) NASA global precipitation measurement (GPM) integrated multi-satellitE Re trievals for GPM (IMERG), Nasa, 29 December 2019.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., and Tan, J. K.: Integrated Multi-satellitE Retrievals for GPM (IMERG), NASA [data set], https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGDE_07/ (last access: 12 November 2025), 2019.

Karunarathna, S., Dissanayake, B. C., Gunawardhana, L., and Rajapakse, L.: High-Resolution Downscaling of GRACE-Derived Groundwater Storage Anomalies using Stacking Ensemble Machine Learning in the Data-Scarce Tropical Catchments, Earth Syst Environ, https://doi.org/10.1007/s41748-026-01022-9, 2026.

Kossieris, P., Tsoukalas, I., Brocca, L., Mosaffa, H., Makropoulos, C., and Anghelea, A.: Precipitation data merging via machine learning: Revisiting conceptual and technical aspects, Journal of Hydrology, 637, 131424, https://doi.org/10.1016/j.jhydrol.2024.131424, 2024.

Le, X.-H., Nguyen Van, L., Hai Nguyen, D., Nguyen, G. V., Jung, S., and Lee, G.: Comparison of bias-corrected multisatellite precipitation products by deep learning framework, International Journal of Applied Earth Observation and Geoinformation, 116, 103177, https://doi.org/10.1016/j.jag.2022.103177, 2023.

Lei, H., Zhao, H., and Ao, T.: A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China, Hydrology and Earth System Sciences, 26, 2969–2995, https://doi.org/10.5194/hess-26-2969-2022, 2022.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Li, J. and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: A review, Environmental Modelling & Software, 53, 173–189, https://doi.org/10.1016/j.envsoft.2013.12.008, 2014.

Lyu, Y. and Yong, B.: A Novel Double Machine Learning Strategy for Producing High-Precision Multi-Source Merging Precipitation Estimates Over the Tibetan Plateau, Water Resources Research, 60, e2023WR035643, https://doi.org/10.1029/2023WR035643, 2024.

Nan, T., Chen, J., Ding, Z., Li, W., and Chen, H.: Deep learning-based multi-source precipitation merging for the Tibetan Plateau, Sci. China Earth Sci., 66, 852–870, https://doi.org/10.1007/s11430-022-1050-2, 2023.

Nguyen, G. V., Le, X.-H., Van, L. N., Jung, S., Yeon, M., and Lee, G.: Application of Random Forest Algorithm for Merging Multiple Satellite Precipitation Products across South Korea, Remote Sensing, 13, 4033, https://doi.org/10.3390/rs13204033, 2021.

Niu, J., Chen, J., Wang, K., and Sivakumar, B.: Coherent modes in multi-scale variability of precipitation over the headwater catchments in the Pearl River basin, South China, Hydrological Processes, 31, 948–955, https://doi.org/10.1002/hyp.11078, 2017.

Papacharalampous, G., Tyralis, H., Doulamis, A., and Doulamis, N.: Comparison of Machine Learning Algorithms for Merging Gridded Satellite and Earth-Observed Precipitation Data, Water, 15, 634, https://doi.org/10.3390/w15040634, 2023.

Paredes-Trejo, F., Barbosa, H., and dos Santos, C. A. C.: Evaluation of the Performance of SM2RAIN-Derived Rainfall Products over Brazil, Remote Sensing, 11, 1113, https://doi.org/10.3390/rs11091113, 2019.

Punyawardena, B. V. R.: Climate, in: The Soils of Sri Lanka, edited by: Mapa, R. B., Springer International Publishing, Cham, 13–22, https://doi.org/10.1007/978-3-030-44144-9_2, 2020.

Ramadhan, R., Marzuki, M., Yusnaini, H., Muharsyah, R., Tangang, F., Vonnisa, M., and Harmadi, H.: A Preliminary Assessment of the GSMaP Version 08 Products over Indonesian Maritime Continent against Gauge Data, Remote Sensing, 15, 1115, https://doi.org/10.3390/rs15041115, 2023.

Roebber, P. J.: Visualizing Multiple Measures of Forecast Quality, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., and Bajat, B.: Random Forest Spatial Interpolation, Remote Sensing, 12, 1687, https://doi.org/10.3390/rs12101687, 2020.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., and Tsai, W.-P.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrology and Earth System Sciences, 22, 5639–5656, https://doi.org/10.5194/hess-22-5639-2018, 2018.

Shen, Y., Zhao, P., Pan, Y., and Yu, J.: A high spatiotemporal gauge-satellite merged precipitation analysis over China, Journal of Geophysical Research: Atmospheres, 119, 3063–3075, https://doi.org/10.1002/2013JD020686, 2014.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

Suárez Sánchez, A., García Nieto, P. J., Riesgo Fernández, P., del Coz Díaz, J. J., and Iglesias-Rodríguez, F. J.: Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain), Mathematical and Computer Modelling, 54, 1453–1466, https://doi.org/10.1016/j.mcm.2011.04.017, 2011.

Tan, M. L., Ibrahim, A. L., Duan, Z., Cracknell, A. P., and Chaplot, V.: Evaluation of Six High-Resolution Satellite and Ground-Based Precipitation Products over Malaysia, Remote Sensing, 7, 1504–1528, https://doi.org/10.3390/rs70201504, 2015.

Tian, F., Hou, S., Yang, L., Hu, H., and Hou, A.: How Does the Evaluation of the GPM IMERG Rainfall Product Depend on Gauge Density and Rainfall Intensity?, https://doi.org/10.1175/JHM-D-17-0161.1, 2018.

Tu, L. and Duan, L.: Spatial downscaling analysis of GPM IMERG precipitation dataset based on multiscale geographically weighted regression model: a case study of the Inner Mongolia Reach of the Yellow River basin, Front. Environ. Sci., 12, https://doi.org/10.3389/fenvs.2024.1389587, 2024.

Vapnik, V.: The Nature of Statistical Learning Theory, Springer Science & Business Media, 324 pp., 2013.

Verdin, A., Rajagopalan, B., Kleiber, W., and Funk, C.: A Bayesian kriging approach for blending satellite and ground precipitation observations, Water Resources Research, 51, 908–921, https://doi.org/10.1002/2014WR015963, 2015.

Wehbe, Y., Temimi, M., and Adler, R. F.: Enhancing Precipitation Estimates Through the Fusion of Weather Radar, Satellite Retrievals, and Surface Parameters, Remote Sensing, 12, 1342, https://doi.org/10.3390/rs12081342, 2020.

Wijayaweera, N., Gunawardhana, L., Bamunawala, J., Sirisena, J., Rajapakse, L., Patabendige, C. S., and Karunaweera, H.: Use of Machine Learning and Indexing Techniques for Identifying Industrial Pollutant Sources: A Case Study of the Lower Kelani River Basin, Sri Lanka, Water, 16, 2766, https://doi.org/10.3390/w16192766, 2024.

Wu, H., Yang, Q., Liu, J., and Wang, G.: A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China, Journal of Hydrology, 584, 124664, https://doi.org/10.1016/j.jhydrol.2020.124664, 2020.

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., and Jegelka, S.: Representation Learning on Graphs with Jumping Knowledge Networks, in: Proceedings of the 35th International Conference on Machine Learning, 5453–5462, 2018.

Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., and Leskovec, J.: Hierarchical Graph Representation Learning with Differentiable Pooling, in: Advances in Neural Information Processing Systems, 2018.

Zhang, L., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., and Yingchun, G.: Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach, Journal of Hydrology, 594, 125969, https://doi.org/10.1016/j.jhydrol.2021.125969, 2021.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M.: Graph neural networks: A review of methods and applications, AI Open, 1, 57–81, https://doi.org/10.1016/j.aiopen.2021.01.001, 2020.

https://doi.org/10.5194/egusphere-2025-6551
Preprint. Discussion started: 13 February 2026

790 Zhu, H., Zhou, Q., and Krisp, J. M.: Exploring machine learning approaches for precipitation downscaling, Geo-spatial Information Science, 1–17, https://doi.org/10.1080/10095020.2025.2477547, 2025.