https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

From Points to Images: Deep-Learning Enhanced Spatial-Temporal Rainfall Modelling from Point Measurements

Bing-Zhang Wang[1], Li-Pen Wang[1], and Auguste Gires[2]

[1]Department of Civil Engineering, National Taiwan University, Taipei, Taiwan
[2]HM&Co, École nationale des ponts et chaussées, Institut Polytechnique de Paris, Champs-sur-Marne, France

Correspondence: Li-Pen Wang (lpwang@ntu.edu.tw), Bing-Zhang Wang (bzw1123@gmail.com)

Abstract.

High-resolution rainfall fields are essential for hydrometeorological applications such as flood forecasting and urban drainage modelling. In practice, however, observations are often limited to sparse and irregular rain gauge networks, making it difficult to reconstruct spatial–temporal rainfall structures and maintain temporal continuity. Conventional interpolation-based approaches struggle under these conditions, particularly when observations are highly sparse. This study proposes a data-driven framework, termed P2I-GAN, to reconstruct rainfall fields directly from irregular point measurements. Inspired by the concept of video inpainting in computer vision, the method learns spatial–temporal rainfall structures from radar observations and applies this knowledge to infer rainfall fields from sparse gauge data. This allows spatial organisation of rainfall to be recovered in a temporally consistent manner, even when observations are limited. Evaluation results show that the proposed approach produces realistic rainfall structures while maintaining strong performance in standard statistical metrics, outperforming conventional interpolation methods and remaining competitive with existing learning-based approaches. The framework provides a practical pathway for reconstructing high-resolution rainfall fields from sparse observation networks.

1 Introduction

Accurate rainfall information at high spatial and temporal resolutions is essential for a wide range of hydrological applications, including urban stormwater management and flood risk assessment, where the spatial and temporal variability of rainfall has been shown to strongly influence the performance of subsequent hydrological modelling (Gires et al., 2012; Ochoa-Rodriguez et al., 2015; Cristiano et al., 2017; McCabe et al., 2017; Zhou et al., 2021).

Ground-based rain gauges have long served as the principal source of rainfall observations for hydrological applications (Sun et al., 2018). By providing direct, point-scale measurements with high accuracy, rain gauges continue to serve as the standard reference for validating hydrological models and assessing the accuracy of remote-sensing precipitation products (He et al., 2013; Maggioni et al., 2016). However, operational rain gauge network generally provides insufficient spatial coverage to well represent given rainfall system (Quirmbach and Schultz, 2002; Ochoa-Rodriguez et al., 2015). In addition, as highlighted by Ochoa-Rodriguez et al. (2015), establishing and maintaining dense rain gauge networks across large areas is logistically and financially challenging, particularly for areas with complex topography (Zeng et al., 2018). As a result, their lack of spatial

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

representativeness could affect the performance of operational hydrological applications (Gires et al., 2012; Ochoa-Rodriguez et al., 2015).

To address the limitation of rain gauge measurements, spatial interpolation techniques have been widely employed to estimate rainfall at ungauged locations and to reconstruct spatially continuous rainfall fields (Grimes and Pardo-Igúzquiza, 2010; Ly et al., 2013). Among the available methods, geostatistical approaches –most notably Kriging and its variants– have achieved broad application due to their capacity to explicitly incorporate spatial correlation structures and to provide a theoretical framework for quantifying interpolation uncertainty (Matheron, 1971). Despite their well-established theoretical basis and extensive use in hydrometeorological studies, these techniques tend to produce relatively smooth representations of rainfall variability, face challenges in incorporating temporal continuity, and remain highly sensitive to the spatial distributions of rain gauge observations (Ochoa-Rodriguez et al., 2019). Moreover, their underlying assumptions of spatial stationarity and normality are inconsistent with the complex, non-stationary nature of real-world precipitation fields (Lebrenz and Bárdossy, 2019).

A specific way to address the aforementioned limitations is to incorporate additional data sources that provide complementary spatial information. In contrast to point-scale rain gauges, weather radars are capable of providing spatially and temporally (nearly) continuous observations of precipitation across extensive areas (Sokol et al., 2021). Their utilisation within hydrological applications has increased markedly in recent years, particularly in applications such as flood forecasting, rainfall-runoff modeling, and urban stormwater management (Thorndahl et al., 2017; Shehu and Haberlandt, 2021; Haensler and Weiler, 2022). However, radar-derived estimates are inherently indirect and remain affected by a variety of error sources (van de Beek et al., 2016; Ochoa-Rodriguez et al., 2019). These errors may introduce inconsistencies and hinder the operational use of radar products in hydrological modeling (Shehu and Haberlandt, 2021; Haensler and Weiler, 2022).

To take advantage of the complementary skills of radar and gauge data, data merging (or gauge-based adjustment) techniques have been increasingly utilised to estimate spatially distributed rainfall fields (Hu et al., 2019; Ochoa-Rodriguez et al., 2019). These approaches statistically model the differences between radar and gauge measurements, accounting for their spatial associations, and have been implemented through several geostatistical formulations (Ahmed and de Marsily, 1987; Hudson and Wackernagel, 1994; Matheron, 1973; Myers, 1982; Odeh et al., 1995). Among these, Co-Kriging (CoK), Kriging with External Drift (KED), and Kriging with Radar-based Error Correction (KRE) are the most commonly used.

Despite their advances, the performance of such approaches remains highly influenced by the density of the rain gauge network and by the level of consistency between radar-derived and collocated gauge estimates (Goudenhoofdt and Delobbe, 2009; Hu et al., 2019; Ochoa-Rodriguez et al., 2019). In addition, incorporating temporal modeling within such statistical frameworks could be computationally expensive, making real-time implementation challenging (Nag et al., 2023). As a result, these approaches typically neglect explicitly incorporating temporal continuity and may produce abruptly varying outputs between successive time steps.

In recent years, a range of ML (Machine Learning) and DL (Deep Learning) based methods have been proposed for rainfall-field reconstruction. These approaches can be broadly grouped into three categories according to their primary objectives.

The first group focuses on spatial and/or temporal interpolation using neural-network-based models, including regression-based approaches, DeepKriging, and graph-based methods. These models have demonstrated strong capability in capturing

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

complex spatial dependencies beyond traditional interpolation techniques (Kopczewska, 2022; Zhang et al., 2018; Chen et al., 2024; Nwaila et al., 2024; Sun et al., 2024; Tavassoli et al., 2022; Nag et al., 2023; Chen et al., 2022; Appleby et al., 2020; Wu et al., 2020). However, most studies primarily evaluate point-wise prediction accuracy, with limited assessment of their ability to reconstruct spatially and temporally continuous rainfall fields.

The second group comprises learning-based radar–gauge merging methods, which aim to improve rainfall estimation by incorporating radar observations as additional spatial information (Chiang et al., 2007; Chen et al., 2019; Sarafian et al., 2025). While effective, these approaches require access to radar data during operation, introducing additional operational constraints and differing from the objective of this study, which focuses on reconstructing rainfall fields directly from sparse gauge observations.

The third group includes stochastic downscaling and super-resolution approaches, which aim to enhance the spatial resolution of rainfall fields by recovering fine-scale structures (Ferraris et al., 2003; Rodriguez-Iturbe et al., 1987; Bell, 1987; Perica and Foufoula-Georgiou, 1996; Rebora et al., 2006; Geiss and Hardin, 2020; Shi et al., 2024). However, these methods typically assume spatially continuous input fields and are therefore not directly applicable to reconstruction from sparse point-based observations.

Alternatively, the rainfall reconstruction problem can be viewed from a different perspective. From a computer vision standpoint, spatial–temporal rainfall interpolation can be formulated as an image (or video) inpainting task, where rainfall values at gauge locations correspond to observed pixels, and the remaining regions are treated as missing areas to be reconstructed. Recent advances in image and video inpainting have demonstrated strong capability in recovering missing spatial–temporal structures from partially observed image sequences (Quan et al., 2024; Elharrouss et al., 2025). However, a key difference lies in the degree of data availability. In typical inpainting applications, a substantial proportion of pixels (e.g., 50–60%) are known (Xiang et al., 2023), whereas rain gauge observations are extremely sparse and irregular, often representing less than 1% of the spatial domain. Such extreme sparsity poses significant challenges for directly applying existing inpainting models, as it limits their ability to learn and recover meaningful spatial–temporal structures.

To address this gap, this study proposes a generative spatial–temporal interpolation framework, **P2I-GAN**, to reconstruct spatially and temporally continuous rainfall fields directly from sparse rain gauge observations. The framework learns spatial–temporal structural information from historical radar data and applies it to improve rainfall reconstruction under severe data sparsity. Its performance is evaluated against classical interpolation methods and radar–gauge merging approaches under both radar-based and gauge-only settings.

The remainder of this paper is structured as follows. Section 2 describes the rainfall datasets and their preparation for model training and validation. Section 3 presents the proposed framework, including its architecture and learning objectives. Section 4 outlines the experimental setup and evaluation procedures. Section 5 discusses the results, organised according to two practical usage scenarios. Finally, Sect. 6 summarises the main findings and outlines directions for future work.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

2 Data and preparation

2.1 Rainfall datasets

The study domain is a 128×128 km^2 square region centred on Birmingham, UK (Fig. 1). The region is characterised by diverse precipitation patterns typical of mid-latitude climates, making it suitable for evaluating rainfall reconstruction methods under varying conditions.

Two complementary precipitation datasets are used in this study. These include:

- High-resolution C-band radar composites from the UK Met Office Nimrod system (Golding, 1998): The Nimrod radar dataset provides spatially distributed rainfall estimates at 1-km spatial resolution and 5-min temporal resolution across the UK (Yu et al., 2020). For this study, radar fields corresponding to the selected 128×128 km^2 domain are extracted to represent spatial–temporal rainfall structures at high resolution.

- Rain gauge measurements from the Environment Agency (EA): Rain gauge observations from 99 EA stations within the study domain are used (see both blue and red markers in Fig. 1). The original measurements are in time-of-tips and are further discretised into 5-min resolution to be consistent with the radar data. The spatial distribution of the gauges corresponds to an average density of approximately one station per 13×13 km^2, which is a widely-seen gauge network density across UK.

Both datasets have been widely used in hydrometeorological studies (Kidd et al., 2012; Yu et al., 2020; King et al., 2021; McClean et al., 2020).

2.2 Data preparation

Radar data are organised into sequences to capture the temporal evolution of rainfall fields. Each sequence consists of 16 consecutive frames (corresponding to 80 minutes). To ensure that the training samples contain meaningful precipitation structures, only sequences satisfying the following criteria are retained: (i) at least one pixel exceeds 2.5 mm h^{-1}, and (ii) at least one frame contains more than 20% wet (non-zero) pixels.

A total of 3,885 sequences from 2016 to 2020 are used for training, with an additional 971 sequences from the same period reserved for validation. For testing, 20 representative storm events from 2021 to 2022 (Table 1) are selected, each comprising 100–300 frames. These events span a range of precipitation conditions and are selected to ensure that the test set covers different seasons, rainfall types, and intensity levels.

The 99 rain gauge stations within the study domain are partitioned into two groups. A subset of 79 gauges is used as input locations, while the remaining 20 gauges are held out for independent evaluation. Due to gauge data availability, only the 10 storm events in 2021 are used for testing, although the radar-based event set includes 20 events from 2021 to 2022.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

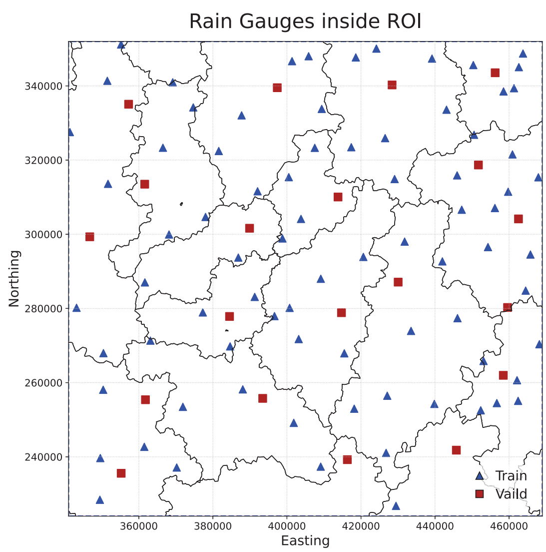


Figure 1. Study area and rain gauge distribution. Blue triangles mark the gauges used for training, red squares indicate those used for validation, and the shaded region shows the radar coverage. The British National Grid (BNG) coordinate system is used here.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Table 1. Selected storm events for result verification and summary statistics from MIDAS rain gauges (RG) and Nimrod radar (RD). RG statistics are calculated from 99 gauge locations, whereas RD statistics are calculated over the full radar field (16,384 pixels). For the 2021 events (Events 1–10), both RG and RD statistics are available; for the 2022 events (Events 11–20), only RD statistics are available.

Event	Start time	End time	Duration	Summary statistics ($\mathrm{mm\,h^{-1}}$)			
				Max.		Mean	
			(hours)	RG	RD	RG	RD
1	2021-01-15 22:00	2021-01-16 18:00	20	16.8	48.0	0.313	0.424
2	2021-03-09 22:00	2021-03-11 00:00	26	17.4	13.8	0.220	0.285
3	2021-05-03 08:00	2021-05-04 05:00	21	47.3	45.8	0.722	0.642
4	2021-05-08 01:00	2021-05-08 21:00	20	13.4	21.2	0.878	0.670
5	2021-05-20 09:00	2021-05-22 02:00	41	21.0	32.7	0.368	0.386
6	2021-05-23 11:00	2021-05-24 07:00	20	34.8	43.6	0.325	0.288
7	2021-07-05 17:00	2021-07-06 13:00	20	22.6	85.4	0.530	0.432
8	2021-10-02 05:00	2021-10-03 02:00	21	13.0	23.4	0.377	0.334
9	2021-10-31 03:00	2021-11-01 07:00	28	89.5	94.0	0.627	0.526
10	2021-12-07 08:00	2021-12-08 05:00	21	39.2	29.7	0.293	0.355
11	2022-01-08 04:00	2022-01-08 22:00	18	N/A	31.2	N/A	0.221
12	2022-03-16 11:00	2022-03-17 05:00	18	N/A	16.7	N/A	0.603
13	2022-04-12 06:00	2022-04-12 19:00	13	N/A	119.4	N/A	0.337
14	2022-05-11 02:00	2022-05-11 20:00	18	N/A	85.4	N/A	0.417
15	2022-05-15 20:00	2022-05-16 11:00	15	N/A	89.6	N/A	0.287
16	2022-06-18 15:00	2022-06-19 06:00	15	N/A	55.4	N/A	0.459
17	2022-08-25 05:00	2022-08-25 19:00	14	N/A	45.8	N/A	0.232
18	2022-09-30 09:00	2022-10-01 03:00	18	N/A	31.2	N/A	0.474
19	2022-10-31 18:00	2022-11-01 12:00	18	N/A	77.6	N/A	0.580
20	2022-11-23 03:00	2022-11-23 19:00	16	N/A	85.4	N/A	0.506

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

3 Methodology

This section presents the proposed framework for reconstructing high-resolution spatial-temporal rainfall fields from point observations. Before introducing the methodology, a key assumption must be clarified. In this study, we assume that historical radar observations contain structural information that is representative of the underlying spatial-temporal relationships in rainfall fields. Based on this assumption, the proposed network learns these structural features from radar data first and applies them to reconstruct rainfall fields from co-located gauge records. Therefore, during model training, only radar data is involved.

3.1 Problem formulation and the proposed framework

Let $R \subset \mathbb{R}^2$ be a rectangular gridded domain of size $H \times W$, and let $\{\mathbf{x}_n\}_{n=1}^{N} \subset R$ denote coordinates of the irregular spatial locations where point measurements are available. Each location $\mathbf{x}_n$ provides a scalar rainfall value $y_{t,n} \in \mathbb{R}$ at each time step $t \in \{1, \ldots, T\}$ over the time period of interest. As illustrated in Figure 2, these observations are placed at their collocated grid locations to form partially observed rainfall fields, where only pixels associated with point observation sites contain valid values and all other pixels remain masked. Note that, during model training, the difference in point and pixel (areal-averaged) measurements is neglected and will be discussed in Sect. 5.

Let $Y_{in} \in \mathbb{R}^{T \times H \times W}$ denote the collection of all available rainfall observations over time and space, embedded into a grid of size $T \times H \times W$ with missing values at unobserved pixels. Given these sparse and irregularly located measurements, the objective is to learn a mapping, enabling reconstructing a sequence of rainfall images from limited known observations:

$$\left(\hat{Y}_{pred}\right) = f\left(Y_{in}\right) \tag{1}$$

where $\hat{Y}_{pred} \in \mathbb{R}^{T \times H \times W}$ represents an estimate of the reconstructed rainfall fields; and $f(\cdot)$ denotes the mapping to be learned.

A spatial-temporal reconstruction framework is proposed here to progressively transform sparsely distributed gauge observations into continuous rainfall fields. As illustrated in Figure 3, the pipeline (from left to right) begins with an IDW-based *Interpolator* that receives the spatially-spare inputs Y_{in} and produces a 'coarse' estimate of the target rainfall fields $\hat{Y}_{idw}$. This estimate is then refined by a U-Net-based *Generator*, which enhances fine-scale spatial structure and incorporates an orography-aware module that yields the reconstructed field $\hat{Y}_{pred}$. Finally, a pair of *Discriminators* is employed to ensure the preservation of spatial features and temporal consistency across the reconstructed sequence.

In the following sections, we first explain three key components of the proposed framework, namely, Interpolator, Generator and Discriminators. We then detail the design of the loss function for model training.

3.2 Interpolator: IDW-based interpolation with temporal modulation

The *Interpolator* component is designed to address the practical setting in which rain gauge observations are spatially sparse and irregular. While the gauge density in this study (approximately one observation per 13×13 grid cells) is representative of typical operational networks, it corresponds to an extremely sparse sampling condition from an image inpainting perspective,

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

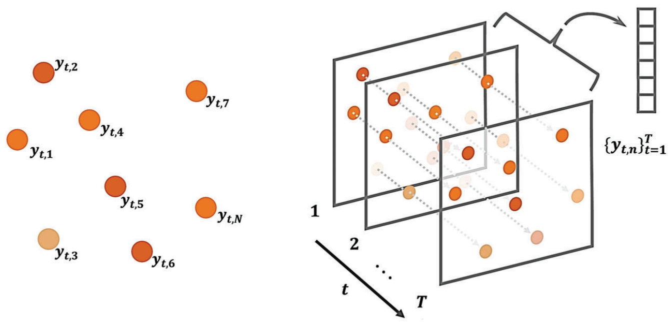


Figure 2. Illustration of the problem setup and sparse rainfall observations. Point measurements at irregular gauge locations ($n = 1, \ldots, N$) across multiple time steps ($t = 1, \ldots, T$). The objective is to reconstruct a spatially and temporally continuous rainfall field from these limited observations.

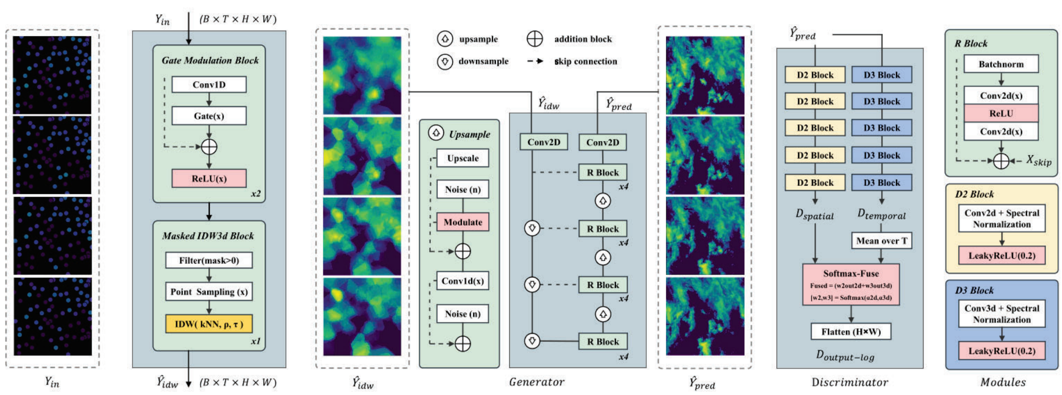


Figure 3. Overall architecture of the proposed spatial-temporal rainfall reconstruction framework. Sparse gauge observations are first transformed into coarse rainfall fields through an IDW-based interpolator with temporal modulation. The preliminary fields are then refined by a U-Net generator that enhances multi-scale spatial structures. Finally, spatial and temporal discriminators jointly improve structural realism and temporal consistency in the reconstructed rainfall sequences. The P2I-GAN+ model incorporates a modulation process at each upsampling layer to account for orographic effects.

where only a very small fraction of pixels are observed. Under such conditions, standard convolutional layers struggle to extract meaningful spatial features due to the limited availability of local information. In addition, gauge measurements often exhibit strong temporal variability, leading to large frame-to-frame fluctuations that can hinder the learning of temporally coherent representations.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

To address these challenges, we involve two modules in this component:

1. **Gating-based temporal modulation**, which adaptively re-weights gauge observations across adjacent time steps, allowing the model to form a more stable temporal representation prior to spatial reconstruction;

2. **Inverse distance weighting (IDW) interpolation**, which generates an initial coarse rainfall field from sparse gauge inputs, providing improved spatial context for the downstream generator.

This component produces an intermediate gridded rainfall field $\hat{Y}_{\text{idw}} \in \mathbb{R}^{T\times H\times W}$ from the sparse gauge observations. For each time step t (time index omitted in the notation below), the interpolated value at a given grid cell $\mathbf{x}_o \in R$ is computed as:

$$\hat{Y}_{\text{idw}}(\mathbf{x}_o) = \frac{\sum_{n=1}^{N} w_n(\mathbf{x}_o)\, y_n}{\sum_{n=1}^{N} w_n(\mathbf{x}_o)}, \quad w_n(\mathbf{x}_o) = \left(\frac{1}{\|\mathbf{x}_o - \mathbf{x}_n\|}\right)^{\rho}. \tag{2}$$

Here, ρ controls the distance-decay rate, and $\rho = 2.0$ (a commonly used default value in IDW) is adopted in all experiments.

Together, these two modules produce a sequence of preliminary rainfall fields $\hat{Y}_{\text{idw}}$ with improved spatial coverage and temporal consistency, which serves as the input to the subsequent generator network.

3.3 Generator: U-Net generator with orography-aware enhancement

To further refine the preliminary IDW fields $\hat{Y}_{idw}$, a U-Net-based architecture is utilised to reconstruct spatial-temporal rainfall structures across multiple scales. The U-Net adopts a coarse-to-fine (or, in ML jargon, downsampling-upsample) design in which broad storm features are first captured at lower resolutions and subsequently enhanced with finer details during upsampling. At each level, residual blocks composed of two 3×3 convolutional layers, following the design in Lim et al. (2017), are used to enhance local feature extraction and support stable model training. Skip connections (see the dashed line arrows in Figure 3) link the encoder and decoder stages together, enabling the direct transfer of spatial information and thereby facilitating the preservation of fine-scale features throughout reconstruction. Together, these multi-scale mechanisms enhance the model's ability to infer rainfall structures from sparse observations.

Although the U-Net architecture effectively refines fine-scale spatial patterns, it does not inherently account for topography-driven rainfall variability. In many regions, radar-derived rainfall estimates exhibit systematic orographic signatures, including windward enhancement, uplift-induced intensification, and lee-side rain shadowing (KITCHEN et al., 1994; Harrison et al., 2000; Houze, 2012; Roe, 2005; Smith, 2006). These processes arise from interactions between prevailing airflow and complex topography, and they introduce spatial patterns that are not explicitly represented by a standard convolutional architecture.

To incorporate these effects within the reconstruction process, we further integrate an orography-aware modulation into each upsampling layer of the generator (see the Noise module in the *Upsample* block in Figure 3). This modulation perturbs the intermediate feature maps using learnable spatial fields that encode topographic influences. Specifically, each feature map is modulated by (i) a spatially-varying scaling factor that amplifies (or suppresses) network responses depending on local topographic characteristics, and (ii) an additive offset that introduces systematic topographic-related variations. Together, this modulation allows the network to adjust reconstructed rainfall patterns in a way consistent with known orographic processes, thereby improving the physical realism of the generated fields.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

3.4 Discriminators: Spatial-temporal discriminator

To ensure that the reconstructed rainfall fields $\hat{Y}_{\text{pred}}$ effectively resemble the observed spatial-temporal structure, we introduce a dual-path discriminator. As illustrated in Figure 3, the *Discriminator* separates its assessment into two pathways, focusing on spatial and temporal aspects of rainfall structure, respectively. Two separate outputs are then combined to form a unified discriminator score, enabling the adversarial loss to spontaneously enforce spatial and temporal consistency.

The spatial pathway adopts a PatchGAN-style design (Isola et al., 2017), treating the input sequence as a multi-channel tensor. A series of Conv2D layers with LeakyReLU activations (see the *D2 Block*) progressively downsample (or upscale) the input while increasing the number of feature channels. This pathway is responsible for identifying spatial structure deficiencies in the generated fields, such as unrealistic smoothness, missing convective cores, or distorted rainfall patterns that pixel-level losses typically fail to detect.

The temporal pathway complements this by examining the short-term evolution of rainfall. Using Conv3D layers (the *D3 Block*), it examines cross-time dependencies directly, enabling the model to detect issues such as flickering between frames, abrupt intensity changes, or physically implausible motion. These temporal artifacts are common failure modes when reconstructing rainfall sequences from sparse observations, and the temporal discriminator is specifically designed to detect them.

To stabilise training, spectral normalisation is applied to all convolutional layers in both pathways (Miyato et al., 2018b). By constraining the Lipschitz constant of each layer, spectral normalisation prevents the discriminator from becoming overly dominant, mitigates mode collapse, and leads to more stable gradients, which is particularly important for Conv3D layers since they typically have higher parameter complexity.

These two pathways produce logits D_{spatial} and D_{temporal}, respectively. Because the temporal pathway outputs an additional temporal dimension, its logits are averaged over time to match the spatial pathway. A softmax-normalised fusion mechanism then combines these two ouptuts using learnable weights α_{2d} and α_{3d}, which adaptively determine the level of importance to assign to each structure. Finally, the fused logits D_{output} is used in the adversarial loss, aiming to enforce both spatial and temporal consistency during rainfall reconstruction.

3.5 Loss function design

In this work, a multi-objective loss framework is utilised to jointly optimise pixel-level reconstruction accuracy, spatial structural consistency, temporal coherence, and adversarial discrimination, as defined below:

$$\mathcal{L}_{\text{G}} = \mathcal{L}_{\text{rec}} + \alpha_{\text{pm}}\mathcal{L}_{\text{pm}} + \alpha_{\text{tc}}\mathcal{L}_{\text{tc}} + \alpha_{\text{adv}}\mathcal{L}_{\text{adv}}, \tag{3}$$

which is a weighted sum of four loss components corresponding to reconstruction ($\mathcal{L}_{\text{rec}}$), probability matching ($\mathcal{L}_{\text{pm}}$), temporal consistency ($\mathcal{L}_{\text{tc}}$), and adversarial learning ($\mathcal{L}_{\text{adv}}$), respectively, where α_{pm}, α_{tc}, and α_{adv} are weighting coefficients. The definition of each loss component is detailed in the following sections.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

3.5.1 Reconstruction loss

The reconstruction loss $\mathcal{L}_{\text{rec}}$ measures the pixel-wise discrepancy between the predicted and reference precipitation fields across all time steps. To mitigate the underestimation bias commonly observed in regression-based precipitation generation, we apply an exponential weighting curve to emphasise high-intensity rainfall regions:

$$\mathcal{L}_{\text{rec}} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{HW}\sum_{p} w(Y_{obs}(p))\left|\hat{Y}_{pred}(p) - Y_{obs}(p)\right|, \tag{4}$$

where the weight function $w(\cdot)$ is defined as:

$$w(x) = \begin{cases} a\,e^{bx} + c, & x \le x_{\max}, \\ a\,e^{bx_{\max}} + c, & x > x_{\max}, \end{cases} \tag{5}$$

with $a = 0.50$, $b = 5.14$, $c = 0.12$, and $x_{\max} = 0.70$. These constants were chosen by fitting an exponential weighting curve to the empirical distribution of training data, while $x_{\max}$ was introduced as an upper bound to prevent excessive overweighting of rare extreme values.

3.5.2 Temporal consistency loss

To align the short-term evolution of generated rainfall with that of the observed sequence, we introduce a KL (Kullback-Leibler) based regularisation term that compares normalized temporal differences (Kullback and Leibler, 1951). Specifically, given the temporal gradient $\Delta Y = Y_{t+1} - Y_t$, we compute softmax-normalised distributions and penalise their divergence:

$$\mathcal{L}_{\text{tc}} = \text{KL}\left(\text{softmax}\left(\Delta\hat{Y}_{pred}/\tau\right)\middle\|\text{softmax}\left(\Delta Y_{obs}/\tau\right)\right) \tag{6}$$

where τ is a hyperparameter controlling distribution sharpness. This encourages the generator to replicate realistic temporal variation patterns, such as growth, decay, and displacement, without enforcing explicit temporal smoothing.

3.5.3 Probability-matching loss

The concept of probability matching was first introduced by Ebert (2001), aiming to preserve the spatial pattern of the ensemble mean while adjusting precipitation intensities to better match the ensemble distribution. More recently, Cao et al. (2025) reformulated this idea as a differentiable loss function by introducing a probability-matching (PM) term and combining it with the conventional mean squared error loss, i.e., $\mathcal{L} = \mathcal{L}_{\text{MSE}} + \omega\mathcal{L}_{\text{PM}}$.

In this study, we further modify the PM term to compare only values above the 99th percentile to focus specifically on the upper tail of the rainfall distribution. The resulting loss is defined as:

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

$$\mathcal{L}_{\text{pm}} = \frac{1}{N_q} \sum_{i=1}^{N_q} \left(z^{(99)}_{\text{pred},i} - z^{(99)}_{\text{obs},i} \right)^2, \tag{7}$$

Here, $z^{(99)}_{\text{pred},i}$ and $z^{(99)}_{\text{obs},i}$ denote the sorted predicted and observed rainfall values retained above the 99th percentile, N_q is the number of retained upper-tail samples.

3.5.4 Adversarial loss

245 A hinge-based adversarial loss is incorporated to enforce spatial structural consistency and physically plausible temporal dynamics (Miyato et al., 2018a). The discriminator first produces two logits: D_{spatial} evaluating spatial structure, and D_{temporal} evaluating short-term evolution, respectively These are then combined through softmax-normalised aggregation weights:

$$D(\cdot) = w_2\, D_{\text{spatial}}(\cdot) + w_3\, D_{\text{temporal}}(\cdot) \tag{8}$$

where w_2 and w_3 are softmax-normalized aggregation weights.

250 The generator's adversarial objective is then:

$$\mathcal{L}_{adv} = -\mathbb{E}_{\hat{Y}_{pred}}[D(\hat{Y}_{pred})]. \tag{9}$$

The discriminator is optimized with the hinge formulation:

$$\mathcal{L}_D = \mathbb{E}\big[\max(0, 1 - D(Y_{obs})) + \max(0, 1 + D(\hat{Y}_{pred}))\big] \tag{10}$$

This softmax-fused scheme adaptively balances spatial and temporal discrimination, enabling the generator to produce rainfall fields that maintain both spatial structural consistency and temporally coherent evolution.

4 Experimental setting

4.1 Overview

The experimental design comprises two stages, each targeting a different aspect of model performance.

Stage 1 aims to assess the intrinsic capability of the models to reconstruct spatial–temporal rainfall fields from point observations under controlled conditions. In this setting, only Nimrod radar data are used, thereby avoiding discrepancies between radar and rain gauge measurements. Specifically, radar rainfall values at gauge locations are sampled at each time step and treated as known inputs, from which the full radar fields are interpolated. Classical interpolation methods (e.g. Ordinary Kriging and Radial Basis Function), DL-based interpolation approaches (DeepKriging and Spatio-Temporal DeepKriging), and the proposed models (P2I-GAN and P2I-GAN+) are compared here under the same settings.

Stage 2 aims to evaluate model performance under realistic operational conditions, where only rain gauge observations are available as inputs for our model. In this setting, two categories of methods are compared. In the first category, DL-based models (DeepKriging, Spatio-Temporal DeepKriging, P2I-GAN, and P2I-GAN+) trained using historical radar data are utilised, and only point rain gauge observations at the 79 training stations are used to reconstruct rainfall field. In the second category, radar–gauge merging approaches (Kriging with External Drift and Kriging with Radar Estimates, also known as conditional merging) are utilised. These methods use both radar data and point gauge observations at the 79 training stations as inputs. The reconstructed rainfall fields are then evaluated against observations from 20 independent validation gauges that are not used as inputs.

In both stages, two versions of the proposed model are assessed: P2I-GAN, which reconstructs rainfall fields without topographic modulation, and P2I-GAN+, which incorporates the proposed orography-aware enhancement.

This experimental design enables the evaluation of (1) the reconstruction capability of the models to be assessed under controlled conditions, and (2) their robustness to be evaluated under sparse, gauge-only input scenarios. In particular, it allows us to examine whether spatial–temporal information learned from radar data can improve rainfall reconstruction when only rain gauge observations are available, which typically exhibit stronger variability and more extreme intensities than radar-derived fields.

Table 2 further summarises the experimental design, including the methods, input and output data, and evaluation datasets for each stage.

4.2 Reference methods

Here, we provide a brief explanation of each reference method used to compare with our proposed models. These methods are included to ensure a comprehensive evaluation.

For spatial interpolation, two classical methods, Radial Basis Function (RBF) (HARDY, 1971) and Ordinary Kriging (OK) (Matheron, 1971), are selected as baseline approaches, representing contrasting approaches to modelling spatial dependence, with and without an explicit covariance structure.

For radar–gauge merging, Kriging with External Drift (KED) (Hudson and Wackernagel, 1994) and Kriging with Radar Estimates (KRE) (Ehret et al., 2008) are adopted, as they are among the most widely used methods that incorporate radar rainfall as auxiliary spatial information to improve rainfall field estimation from rain gauge observations (Ochoa-Rodriguez et al., 2019). Although the proposed framework does not require radar inputs during inference, including KED and KRE enables comparison with methods that explicitly combine radar and gauge information.

For deep-learning-based interpolation, DeepKriging (DK) (Chen et al., 2022) and its spatio-temporal extension STDK (Nag et al., 2023) are included as representative nonlinear models for reconstructing rainfall fields from irregular observations.

The definition of each reference method used in this study is given below.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Table 2. Summary of evaluation design for the two experimental stages. Referring to Fig. 1, 79 rain gauges (blue triangles) are used as training locations in both Stage 1 and Stage 2. In Stage 2, 20 rain gauges (red squares) are selected for validation.

Stage	Description	Methods	Input	Output	Evaluated data
Stage 1	Spatial–temporal interpolation capability evaluated under controlled conditions using radar-sampled gauge locations.	OK RBF DK STDK P2I-GAN P2I-GAN+	Radar-derived rainfall values sampled at training gauge locations (79 stations per time step)	Reconstructed rainfall field	Nimrod data (20 events); all pixels except training gauge locations in each frame
Stage 2	Radar–gauge merging methods using radar as second input.	KED KRE	Full radar rainfall field + gauge observations at training locations	Merged rainfall field	20 validation gauges per time step across 20 events
	Gauge-only reconstruction under sparse observations using the same model trained in Stage 1.	DK STDK P2I-GAN P2I-GAN+	Gauge observations at training locations (79 stations per time step)	Reconstructed rainfall field	20 validation gauges per time step across 20 events; radar used for visual comparison

4.2.1 Ordinary Kriging (OK)

Ordinary Kriging (OK) (Matheron, 1971) estimates rainfall at an unknown location as a weighted linear combination of observed values under the assumption of an unknown but locally constant mean:

$$\hat{y}(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i \, y(\mathbf{s}_i), \qquad \sum_{i=1}^{N} \lambda_i = 1, \tag{11}$$

where $\hat{y}(\mathbf{s}_0)$ is the target location, $y(\mathbf{s}_i)$ is the observed rainfall at location $\mathbf{s}_i$, and λ_i are the kriging weights. These weights are obtained by minimising the estimation variance subject to the unbiasedness constraint, based on a fitted variogram model that characterises the 2_{nd}-order spatial dependence of the rainfall field as a function of separation distance.

4.2.2 Radial Basis Function (RBF)

Radial Basis Function (RBF) (HARDY, 1971) interpolation reconstructs the rainfall field as a weighted sum of radial kernels centred at the observation locations:

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

$$\hat{y}(\mathbf{s}_0) = \sum_{i=1}^{N} w_i \, \phi(\|\mathbf{s} - \mathbf{s}_i\|), \tag{12}$$

where w_i are interpolation coefficients, and $\phi(\cdot)$ is a radial basis function. In this study, a thin-plate spline kernel is adopted:

$$\phi(r) = r^2 \log(r). \tag{13}$$

where $r = \|\mathbf{s}_0 - \mathbf{s}_i\|$ denotes the Euclidean distance between the prediction location and the observation location.

4.2.3 Kriging with External Drift (KED)

Kriging with External Drift (KED) (Hudson and Wackernagel, 1994) incorporates radar rainfall as an auxiliary variable to guide the spatial interpolation of rain gauge observations. The rainfall field is modelled as the sum of a deterministic drift term related to radar rainfall and a spatially correlated residual term:

$$y(\mathbf{s}) = m(\mathbf{s}) + \varepsilon(\mathbf{s}), \qquad m(\mathbf{s}) = \beta_0 + \beta_1 R(\mathbf{s}), \tag{14}$$

where $R(\mathbf{s})$ denotes the radar rainfall estimate at location $\mathbf{s}$, β_0 and β_1 are drift coefficients, and $\varepsilon(\mathbf{s})$ is a zero-mean residual random field. Rain gauge observations $y(\mathbf{s}_i)$ are treated as the primary data to be interpolated, while the radar field provides auxiliary information through the drift term.

The rainfall at an unobserved location $\mathbf{s}_0$ is then estimated using a kriging estimator:

$$\hat{y}(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i \, y(\mathbf{s}_i), \tag{15}$$

where the weights λ_i are determined by solving the kriging system under the constraint that the estimator is unbiased with respect to the external drift. In this way, the radar field guides the large-scale spatial structure of the interpolation, while the gauge observations ensure local consistency.

4.2.4 Kriging with Radar Estimates (KRE)

Kriging with Radar Estimates (KRE) (Ehret et al., 2008) combines a smooth gauge-based interpolation with a radar-derived correction field to recover spatial detail resolved by the radar data. The merged estimate is given by

$$\hat{y}(\mathbf{s}) = G_{\mathrm{OK}}(\mathbf{s}) + [R(\mathbf{s}) - R_{\mathrm{OK}}(\mathbf{s})], \tag{16}$$

where $G_{\mathrm{OK}}(\mathbf{s})$ is the ordinary kriging interpolation of rain gauge observations, $R(\mathbf{s})$ is the original radar rainfall field, and $R_{\mathrm{OK}}(\mathbf{s})$ is the ordinary kriging interpolation of radar values sampled at the gauge locations. The term $R(\mathbf{s}) - R_{\mathrm{OK}}(\mathbf{s})$ represents

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

a radar-derived correction field, which is added to the gauge-based estimate to reintroduce spatial variability resolved by the radar observations.

This formulation follows the principle of conditional merging, in which the gauge interpolation preserves the point observations while the radar-derived correction term restores spatial patterns unresolved by the sparse gauge network.

4.2.5 DeepKriging (DK)

DeepKriging (DK) (Chen et al., 2022) reformulates spatial interpolation as a regression problem in which spatial dependence is represented through a set of basis functions rather than an explicit covariance model. In this framework, rainfall observations at gauge locations are modelled as a function of spatial coordinates encoded via basis expansions.

Following Chen et al. (2022), the positional encoding is constructed using multi-resolution compactly supported radial basis functions, which are then used as inputs to a multilayer perceptron (MLP) to predict rainfall values. The basis functions are defined as

$$\phi_j^*(\mathbf{s}) = \phi(\|\mathbf{s} - \mathbf{u}_j\|/\theta), \tag{17}$$

where $\mathbf{u}_j$ denotes the j-th knot location and θ is a scale parameter. Let $r = \|\mathbf{s} - \mathbf{u}_j\|/\theta$. The compactly supported Wendland basis function is given by

$$\phi(r) = \begin{cases} (1-r)^6(35r^2 + 18r + 3)/3, & r \in [0,1], \\ 0, & \text{otherwise}. \end{cases} \tag{18}$$

4.2.6 Spatio-temporal DeepKriging (STDK)

Spatio-temporal DeepKriging (STDK) (Nag et al., 2023) extends DK by incorporating both spatial and temporal information through basis expansions. Rainfall values are predicted using a neural network that takes spatial and temporal encodings as inputs:

$$\hat{y}(\mathbf{s},t) = f_\theta\left(\mathbf{x}(\mathbf{s},t), \boldsymbol{\phi}^{(S)}(\mathbf{s}), \boldsymbol{\phi}^{(T)}(t)\right), \tag{19}$$

where $\boldsymbol{\phi}^{(S)}(\mathbf{s})$ and $\boldsymbol{\phi}^{(T)}(t)$ denote the spatial and temporal basis expansions, respectively, and $\mathbf{x}(\mathbf{s},t)$ represents additional input features (e.g. gauge observations). The spatial component follows the multi-resolution Wendland basis construction, while the temporal component is represented using radial basis functions over time. This formulation enables the model to capture joint spatio-temporal dependence without specifying a parametric covariance model.

It is worth noting that the original network architectures of DK and STDK are preserved in this study. However, the positional encoding is modified to ensure computational feasibility for high-resolution grids. In addition, rain gauge observations are

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

incorporated as inputs to enable point-to-field reconstruction, and the training loss is computed over the spatial field rather than only at gauge locations. To ensure a fair comparison, DK and STDK are trained using the same loss functions as P2I-GAN, excluding the adversarial components. Therefore, the performance differences primarily reflect differences in model architecture rather than loss formulation or data usage. The impact from the above changes in DK and STDK is explained in the Supplement S1, and we demonstrate that our modifications can improve both the efficiency and quality of DK and STDK in generating rain fields.

4.3 Evaluation metrics

The evaluation metrics are grouped according to the aspects of model performance they assess. The first group consists of continuous metrics that quantify the agreement between reconstructed and reference rainfall fields. These include image-based measures commonly used in computer vision, as well as statistical metrics such as the Perkins Skill Score (PSS) and Nash-Sutcliffe Efficiency (NSE). Together, these metrics evaluate numerical accuracy, distributional consistency, and the ability of the models to reproduce spatial rainfall structures.

The second group consists of categorical, threshold-based skill scores, including the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Heidke Skill Score (HSS), evaluated at multiple rainfall thresholds. In addition, the Fraction Skill Score (FSS) is computed over different spatial neighbourhood scales to assess scale-dependent performance. These metrics are particularly relevant for hydrometeorological applications, where accurate detection of rainfall occurrence and exceedance at different intensity levels is crucial.

4.3.1 Continuous indices

Reconstruction performance is evaluated using continuous metrics that quantify differences between reconstructed and reference rainfall fields in terms of magnitude, variance, and distribution. Measures used include:

– Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) measure the average magnitude of pixel-wise errors:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_{\mathrm{obs}}(i) - \hat{Y}(i)\right|, \quad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{\mathrm{obs}}(i) - \hat{Y}(i)\right)^2}, \tag{20}$$

where $Y_{\mathrm{obs}}(i)$ and $\hat{Y}(i)$ denote the observed and reconstructed rainfall values at grid cell i, respectively.

– Nash-Sutcliffe Efficiency (NSE) Nash and Sutcliffe (1970) evaluates predictive skill relative to the variability of the observations:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}\left(Y_{\mathrm{obs}}(i) - \hat{Y}(i)\right)^2}{\sum_{i=1}^{N}\left(Y_{\mathrm{obs}}(i) - \bar{Y}_{\mathrm{obs}}\right)^2}, \tag{21}$$

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

where $\bar{Y}_{\mathrm{obs}}$ is the mean of the observed rainfall field. A value closer to 1 indicates better agreement, while values below 0 indicate performance worse than using the observed mean.

– Perkins Skill Score (PSS, see Perkins et al. (2007)) is used to assess distributional similarity. For each time step t, it is defined as

$$S_{\mathrm{PSS}}(t) = \sum_{i=1}^{n} \min\bigl(Z_{\mathrm{pred},i}(t), Z_{\mathrm{obs},i}(t)\bigr), \tag{22}$$

where $Z_{\mathrm{pred},i}(t)$ and $Z_{\mathrm{obs},i}(t)$ denote the relative frequencies of predicted and observed rainfall values in the i-th bin, respectively. The PSS measures the overlap between the two distributions, with a value of 1 indicating identical distributions.

4.3.2 Categorical indices

Following Shi et al. (2017), the performance is converted into binary events based on a given threshold R_{thr}, such that rainfall values exceeding the threshold are considered as rainfall occurrences. This yields four outcomes: hits (H), misses (M), false alarms (F), and correct negatives (C). Based on these outcomes, a series of categorical measures are computed. These include:

$$\mathrm{POD} = \frac{H}{H+M}, \quad \mathrm{FAR} = \frac{F}{H+F}, \tag{23}$$

$$\mathrm{CSI} = \frac{H}{H+M+F}, \tag{24}$$

$$\mathrm{HSS} = \frac{2(H \cdot C - M \cdot F)}{(H+M)(M+C)+(H+F)(F+C)}. \tag{25}$$

These metrics evaluate the ability of the models to detect rainfall occurrence and intensity exceedance at different thresholds. However, they are sensitive to spatial displacement errors, as small location shifts can lead to missed or falsely detected events.

To account for spatial displacement, the Fraction Skill Score (FSS) is included, which compares fractional rainfall coverage within neighbourhood windows of radius r (Roberts and Lean, 2008):

$$\mathrm{FSS}(r) = 1 - \frac{\sum (f_o - f_p)^2}{\sum f_o^2 + \sum f_p^2}, \tag{26}$$

where f_o and f_p denote the fractions of observed and predicted rainfall occurrences within each neighbourhood window, respectively. The summations are taken over all grid cells in the domain. FSS evaluates spatial agreement across scales and is less sensitive to small spatial displacements.

All categorical metrics are computed for rainfall thresholds $R_{\mathrm{thr}} \in \{0.5, 2.0, 4.0, 8.0\}\ \mathrm{mm\,h^{-1}}$.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Table 3. Training Configuration and Model Parameters

Parameter	Value
Input size $H \times W \times T \times C$	$128 \times 128 \times 16 \times 1$
Optimizer	Adam ($\beta_1 = 0$, $\beta_2 = 0.99$)
Learning rate	1×10^{-4}
Learning rate decay	0.99995
Batch size	12
Total training iterations	2×10^5
Warm-up and steady phases	2×10^4, 5×10^4
Adversarial loss weight α_{adv}	0.01
KL divergence loss weight α_{tc}	0.05
Probability-matching loss weight α_{pm}	10.0
Training epochs	600

4.4 Model training setup

All experiments are conducted using input tensors of size $128 \times 128 \times 16 \times 1$, corresponding to the spatial, temporal, and channel dimensions. Model training is implemented in PyTorch using the Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.99$. The initial learning rate is set to 1×10^{-4} and decays exponentially with a factor of 0.99995. The batch size is 12.

Training is conducted for 600 epochs, comprising an initial warm-up phase of 2×10^4 iterations followed by a steady training phase. All experiments are performed on a workstation equipped with an NVIDIA RTX 3080 GPU. A summary of the training configuration is provided in Table 3.

All rainfall values are expressed in mm h^{-1}. To reduce skewness and stabilise training, a logarithmic transformation is applied to the raw rainfall intensity R:

$$X = \left(\frac{\log_{10}(R/0.036)}{0.0625} \right) \times 3, \tag{27}$$

where X denotes the transformed rainfall intensity. This transformation compresses high-intensity values while enhancing variability at lower intensities, resulting in a distribution that is closer to Gaussian. The transformed values are subsequently clipped and linearly normalised to the range $[0, 1]$ for model training.

5 Results & Discussion

The performance of the proposed framework is evaluated under two complementary settings, corresponding to controlled and operational scenarios. Stage 1 examines the ability of the models to reconstruct spatial–temporal rainfall fields from sparse observations using radar-derived data, providing a controlled benchmark for interpolation performance. Stage 2 evaluates model performance under more realistic conditions, where only rain gauge observations are available as inputs.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

It should be noted that several training and implementation adjustments were made to DK and STDK to ensure a fairer comparison with the proposed framework, including modifications to the input conditioning and loss formulation. These details are provided in Sect. S1 of the Supplement. In addition, further event-based visual examples are included in Sect. S2 of the Supplement to show representative results beyond the cases presented in this study.

5.1 Stage 1: Assessing interpolation using radar measurements only

In the first evaluation stage, we assess the interpolation skill of the proposed P2I-GAN and P2I-GAN+ using only radar measurements, focusing purely on models' spatial–temporal reconstruction capability. That is, radar data collocated at gauge locations are used to reconstruct the complete radar images, which are then evaluated using the observed radar data. Comparisons are made against classical interpolation methods (OK, RBF) and deep-learning baselines (DK, STDK).

5.1.1 Continuous metrics

Table 4 summarises the results of the continuous measures for all interpolation methods. As seen, the proposed P2I-GAN and P2I-GAN+ achieve the most favorable performance across all four metrics. P2I-GAN+ obtains the best MAE (0.173) and PSS (0.768), while P2I-GAN results in the lowest RMSE (0.585) and the highest SSIM (0.632). The difference between these two P2I-GAN variants is small. Compared with the best baseline results, the proposed models reduce MAE from 0.195 to 0.173–0.174 (corresponding to an improvement of about 10.8–11.3%) and increase SSIM from 0.563 to 0.631–0.632 (corresponding to an improvement of about 12.1–12.3%).

A direct comparison between predicted and the observed rainfall estimates (see Fig. 5) show substantial dispersion across all methods, particularly at higher rainfall intensities. P2I-GAN+ and P2I-GAN achieve the highest R^2 values (0.493 and 0.495, respectively), while retaining slopes comparable to RBF. Although RBF gives a slightly higher slope (0.538), its much lower R^2 (0.248) indicates larger scatter and lower robustness. By contrast, DK and STDK show both lower slopes and lower R^2 values, indicating weaker performance in reproducing the observed rainfall distribution. Nevertheless, all methods remain well below the 1:1 line at higher intensities, indicating systematic underestimation of heavy rainfall, likely because sparse samples at gauge locations cannot fully capture localized convective cores.

Figure 4 further confirms the advantage of the proposed models in terms of both NSE level and event-to-event stability across the 20 test events. P2I-GAN and P2I-GAN+ perform higher central tendency, with median NSE values around 0.6, and their mean values are also consistently higher than those of the benchmark methods. By comparison, DK and STDK have substantially lower medians, with distributions concentrated around 0.3, suggesting limited reconstruction skill in many events. RBF shows a wider spread, implying larger event-dependent variability, while OK achieves comparatively competitive NSE values but still remains below the proposed models in terms of central performance. This suggests that the performance gains of P2I-GAN and P2I-GAN+ are consistently observed across different rainfall events.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Table 4. Summary of Stage 1 (interpolation) verification results. Performance metrics were computed using radar gridded data. Lower MAE and RMSE indicate smaller errors, while higher PSS indicate higher distribution similarity (Boldface = best, underline = second best).

MODELS	MAE ↓	RMSE ↓	PSS ↑
	($\mathrm{mm\,h^{-1}}$)	($\mathrm{mm\,h^{-1}}$)	(-)
OK	0.195	0.641	0.619
RBF	0.220	0.923	<u>0.756</u>
DK	0.259	0.834	0.586
STDK	0.261	0.777	0.645
P2I-GAN	<u>0.174</u>	**0.585**	0.754
P2I-GAN+	**0.173**	<u>0.587</u>	**0.768**

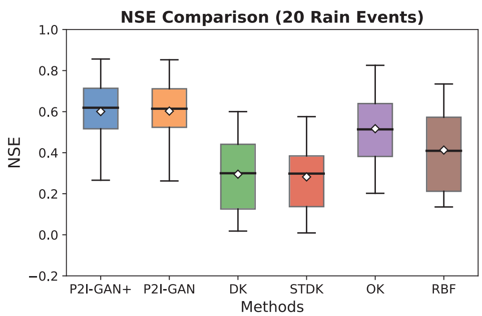


Figure 4. Comparison of NSE values for 20 rainfall events across different reconstruction methods.

5.1.2 Categorical Metrics

Table 5 summarises POD, FAR, CSI, and HSS scores across four rainfall thresholds. As seen, for all cases, P2I-GAN variants achieve either the best or the second best performance. It is particularly evident in CSI and HSS at moderate and heavy rainfall thresholds. P2I-GAN+ constantly show slight improvements over P2I-GAN at 2.0–8.0 $\mathrm{mm\,h^{-1}}$. This demonstrates

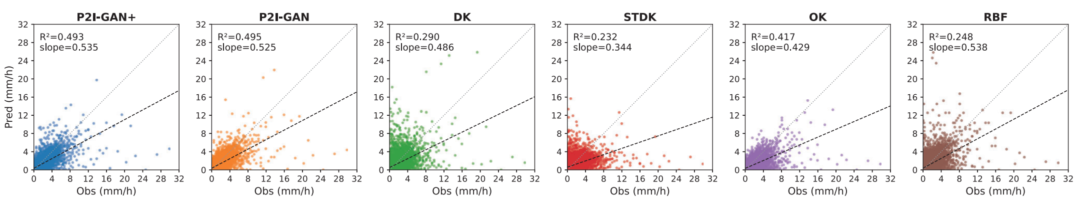


Figure 5. Scatter plots between predicted and true rainfall intensities across different interpolation models.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

that the orography-aware modulation provides a measurable, though limited, benefit in the interpolation-only setting. At the highest threshold (8.0 $\mathrm{mm\,h^{-1}}$), P2I-GAN+ achieves the best CSI and HSS values, reaching 0.137 and 0.241, respectively, corresponding to improvements by about 16.1% and 15.3% over RBF (the best-performing class interpolation method).

The Fractions Skill Score (see Table 6) shows that P2I-GAN and P2I-GAN+ consistently outperform all reference methods across all rainfall thresholds and all scales. The difference in performance becomes more evident at higher rainfall thresholds, where accurate reconstruction of spatial structure is more challenging. In particular, both proposed models retain the highest FSS values at thresholds of 4.0 and 8.0 $\mathrm{mm\,h^{-1}}$ across all spatial scales. By contrast, DK and especially STDK show a marked decline as the threshold exceeds 4.0 $\mathrm{mm\,h^{-1}}$, where the substantially low scores indicate their poor ability to preserve spatial structure of heavy rainfall.

Table 5. Categorical verification scores across four rainfall thresholds (0.5, 2.0, 4.0, 8.0 $\mathrm{mm\,h^{-1}}$) for Stage 1 (Radar-only interpolation). Boldface = best, underline = second best.

MODELS	$\geq$ 0.5 $\mathrm{mm\,h^{-1}}$				$\geq$ 2.0 $\mathrm{mm\,h^{-1}}$				$\geq$ 4.0 $\mathrm{mm\,h^{-1}}$				$\geq$ 8.0 $\mathrm{mm\,h^{-1}}$			
	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS
OK	0.755	**0.161**	0.659	0.738	0.447	**0.258**	0.387	0.540	0.258	**0.326**	0.230	0.370	0.105	**0.465**	0.096	0.175
RBF	0.786	0.198	0.658	0.734	0.555	0.379	0.414	0.565	0.380	0.571	0.252	0.398	**0.206**	0.785	0.118	0.209
DK	0.823	0.257	0.640	0.712	0.559	0.460	0.379	0.524	0.318	0.711	0.178	0.296	0.125	0.884	0.064	0.119
STDK	0.826	0.296	0.613	0.682	0.441	0.465	0.319	0.458	0.122	0.770	0.086	0.154	0.014	0.965	0.010	0.019
P2I-GAN	**0.828**	0.173	0.706	0.777	0.575	0.260	0.478	0.630	0.372	0.393	0.299	0.457	0.157	0.497	0.136	0.239
P2I-GAN+	0.823	0.163	**0.709**	**0.780**	**0.582**	0.263	**0.482**	**0.634**	**0.385**	0.410	**0.304**	**0.462**	0.164	0.541	**0.137**	**0.241**

Table 6. FSS Scores at Multiple Rainfall Thresholds and Spatial Scales. Neighborhood spatial sizes are expressed in kilometers. Higher values indicate better spatial structure prediction. Boldface = best, underline = second best.

THRESHOLDS	$\geq$ 0.5 $\mathrm{mm\,h^{-1}}$				$\geq$ 2.0 $\mathrm{mm\,h^{-1}}$				$\geq$ 4.0 $\mathrm{mm\,h^{-1}}$				$\geq$ 8.0 $\mathrm{mm\,h^{-1}}$			
MODELS	1×1	2×2	4×4	8×8	1×1	2×2	4×4	8×8	1×1	2×2	4×4	8×8	1×1	2×2	4×4	8×8
OK	0.795	0.829	0.860	0.896	0.558	0.617	0.673	0.740	0.373	0.447	0.521	0.610	0.176	0.233	0.297	0.376
RBF	0.794	0.829	0.861	0.898	0.586	0.642	0.696	0.763	0.403	0.467	0.532	0.615	0.210	0.258	0.309	0.374
DK	0.781	0.814	0.843	0.877	0.549	0.604	0.658	0.725	0.303	0.353	0.408	0.485	0.120	0.148	0.180	0.227
STDK	0.760	0.788	0.814	0.847	0.483	0.529	0.576	0.639	0.159	0.190	0.226	0.284	0.020	0.026	0.034	0.050
P2I-GAN	0.828	0.869	0.903	0.937	0.647	0.720	0.783	0.851	0.461	0.549	0.635	0.735	0.240	0.314	0.398	0.504
P2I-GAN+	**0.830**	**0.872**	**0.906**	**0.939**	**0.650**	**0.723**	**0.786**	**0.853**	**0.466**	**0.554**	**0.640**	**0.739**	**0.242**	**0.316**	**0.401**	**0.509**

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

5.1.3 Visual Comparison

We further conduct visual inspection by comparing reconstructed rainfall fields from two selected events representing different storm types (see Figure 6: the top panel is a summer event, while the bottom a winter event). In the summer case, both P2I-GAN+ and P2I-GAN can better recover the rainfall structure across scales. In particular, the observed fine details are preserved by both P2I-GAN+ and P2I-GAN. Both DK and STDK can capture the broad rainfall pattern but without much details. In addition, both DK and STDK evidently show artificial artifacts, with enhanced values repeatedly appearing in similar regions rather than varying with the event structure.

In the winter case, temporal discontinuity becomes evident in those fields generated from the OK and RBF. Abrupt variations in spatial patterns can be clearly identified as time changes. Compared to them, both P2I-GAN+ and P2I-GAN demonstrate a much better ability to maintain temporal coherency of rainfall fields across time. It is also evident that STDK tends to generate overly smooth rainfall fields, largely underestimating high-intensity rainfall values. Additional event-based visual results from 2021–2022 are presented in Sect. S2.1 of the Supplement.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

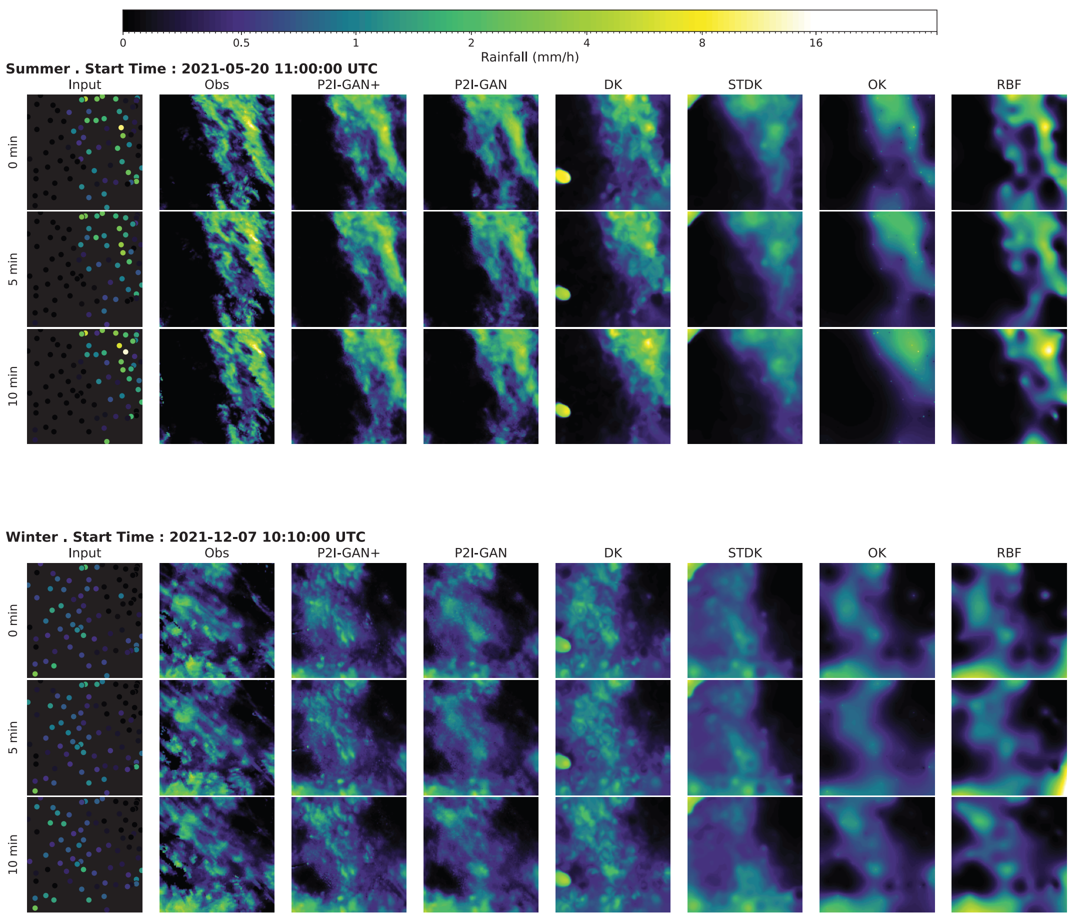


Figure 6. Reconstruction examples for summer and winter rainfall events (selected according to their occurrence times), comparing outputs from multiple interpolation and reconstruction methods over time.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

5.2 Stage 2: Real-world gauge-derived reconstruction

In the second evaluation stage, we evaluate the proposed framework under a realistic operational scenario in which only rain-gauge observations are provided as inputs, while radar fields are used for visual comparison and those gauge observations at locations not used for model training are used for metric calculation.

Comparisons are made against geostatistical merging methods (KED, KRE) and deep-learning baselines (DK, STDK). Notably, only KED and KRE explicitly use both radar and gauge data during calculation. All other methods –including ours– use only gauge measurements as input, providing a fair test of each method's ability to extrapolate from point observations.

5.2.1 Continuous metrics

Table 7 summarises the results of continuous measures. P2I-GAN and P2I-GAN+ achieve the lowest MAE and RMSE, with P2I-GAN yielding the best overall accuracy (MAE = 0.279; RMSE = 0.934). These results are consistent with the interpolation-only evaluation (Stage 1), suggesting that the proposed architecture generalises well even when applied to the more variable gauge input. In contrast, the highest PSS is obtained by KRE (0.552), followed by KED (0.529), indicating that the current model remains slightly weaker in reproducing the rainfall distribution than merging methods that directly incorporate radar data. However, when compared with Table 4, the gap relative to the radar-based evaluation remains evident. This discrepancy likely arises from the scale mismatch between gauges and radar: gauges measure rainfall at a point, whereas radar pixels represent area-averaged values over roughly 1 km^2 (Wilson and Brandes, 1979; Gires et al., 2014). As shown in Table 1, RD exceeds RG in some 2021 events, likely due to sampling limitations of the gauge network, since rain gauges do not always capture the cores of rain cells. Together with the intrinsic scale mismatch between point measurements and area-averaged radar pixels, this contributes to the discrepancies in gauge–radar comparison. In addition, although rainfall statistics generally exhibit stronger variability and more pronounced extremes at point scale, localised intense cores may also be captured within a radar pixel but missed by the rain gauge network, as gauges do not necessarily sample these small-scale extremes (Wilson and Brandes, 1979; Moreau et al., 2009; Gires et al., 2014).

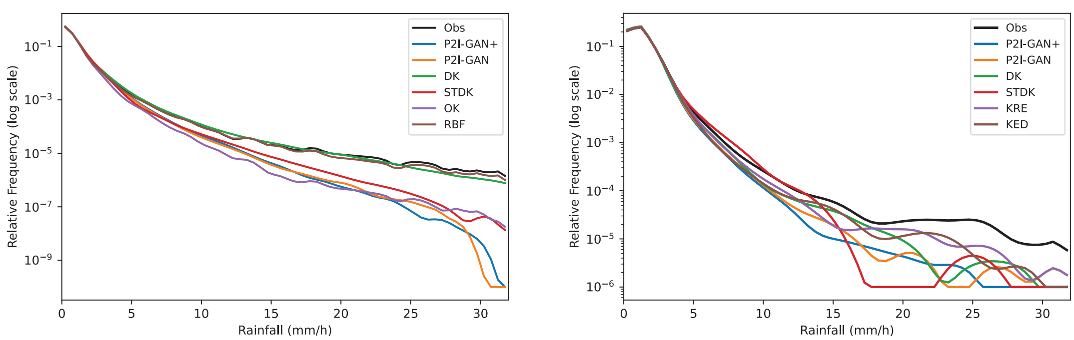
Table 7. Summary of Stage 2 (gauge-based and merging) verification results. Performance metrics were computed using independent gauges not used for model training or data merging. Lower MAE and RMSE indicate smaller errors, while higher PSS indicate higher distribution similarity (Boldface = best, underline = second best).

MODELS	MAE ↓ (mm h^{-1})	RMSE ↓ (mm h^{-1})	PSS ↑ (-)
KED	0.304	0.987	0.529
KRE	0.327	1.014	**0.552**
DK	0.311	1.016	0.302
STDK	0.355	1.120	0.376
P2I-GAN	**0.279**	**0.934**	0.484
P2I-GAN+	0.281	0.970	0.483

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Figure 7 further compares the rainfall-intensity distributions in Stage 1 and Stage 2. In Stage 1, RBF and DK follow the observed distribution more closely at the highest intensities, whereas most other methods show a clearer decay in the tail, indicating underestimation of extreme rainfall. This pattern may appear inconsistent with the PSS results in Table 4, where P2I-GAN and RBF show stronger overall distributional performance. A likely reason is that PSS is dominated by the bulk of the distribution and is therefore less sensitive to tail behaviour at rare extreme values. At the same time, Figures 6 and 10 show consistent artificial artifacts across different frames for DK and STDK, suggesting that these models partly maintain the overall distribution by repeatedly introducing similar artificial patterns in each frame. This implies that, although their overall distributions may appear reasonable, their accuracy in representing extreme rainfall remains limited. In Stage 2, the distributional behaviour changes. KRE and KED shows a slightly better representation of the high-intensity tail, while P2I-GAN retains relatively better performance in the extreme range than most other methods. The advantage of DK becomes less evident in this stage, and STDK shows a rapid decay at higher rainfall intensities, indicating much weaker representation of extremes.

Figure 8 compares the NSE distributions of different methods across the 10 selected rainfall events. Compared with Fig. 4, all DL-based methods show a certain degree of performance degradation. Among these methods, KED achieves the strongest overall performance, with a median NSE value of approximately 0.48, while KRE follows closely with a median NSE value around 0.40. For the proposed DL-based methods, P2I-GAN and P2I-GAN+ show median NSE values around 0.40 with relatively narrow interquartile ranges. This indicates that their outputs remain close to the merging-based methods while maintaining more stable event-to-event performance. These results demonstrate the practical potential of the proposed models, especially in situations where real-time radar observations are unavailable. By contrast, DK and STDK produce lower central NSE values, particularly STDK, whose distribution is concentrated at lower scores, suggesting weaker reconstruction skill for these selected events.



(a) Comparison in Stage 1 using the full radar field.

(b) Comparison in Stage 2 using the 20 validation gauge locations.

Figure 7. Probability distributions of rainfall intensity for the ground truth (Obs) and the reconstructed fields obtained by different models.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

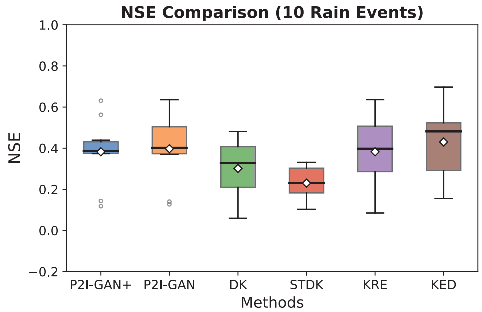


Figure 8. Comparison of NSE values for 10 rainfall events across different reconstruction methods.

The scatter plots in Fig. 9 show substantial dispersion and systematic underestimation across all merging models, consistent with the behaviour already seen in Fig. 5. Among the methods under comparison, P2I-GAN achieves the highest R^2 (0.518), followed closely by P2I-GAN+ (0.509), indicating that the proposed models retain the strongest consistency in the merging setting. In contrast, STDK yields the steepest fitted slope (0.576), followed by KRE (0.558) and KED (0.546), suggesting a relatively stronger response to increasing observed rainfall intensity. KED and KRE also outperform DK and STDK in terms of R^2, showing that these merging methods provide a more balanced relationship between spread and trend than the other traditional baselines. This behaviour is also broadly consistent with Fig. 7, where KED and KRE better preserve the rainfall distribution. Overall, the results indicate that KED and KRE still perform better in terms of distributional preservation and overall merging stability, while the deep learning-based methods show comparable potential in terms of event-wise consistency, as reflected by their higher R^2 values.

5.2.2 Categorical Metrics

Table 8 summarizes POD, FAR, CSI, and HSS for the real-world gauge-derived reconstruction setting (Stage 2). At lower thresholds, KED attains the strongest overall performance, achieving the best CSI and HSS at 0.5 and 2.0 $\mathrm{mm\,h^{-1}}$. STDK

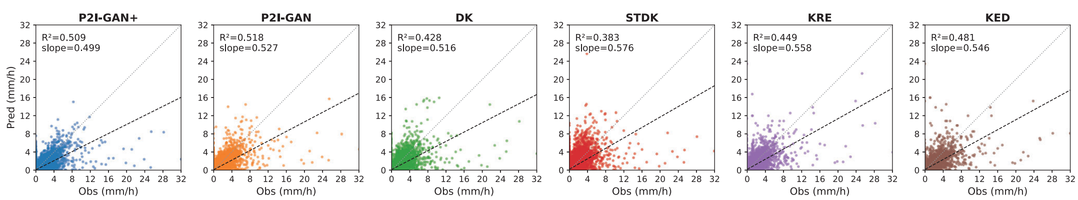


Figure 9. Scatter plots between predicted and observed rainfall for different merging models.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Table 8. Categorical verification scores across four rainfall thresholds (0.5, 2.0, 4.0, 8.0 $\mathrm{mm\,h^{-1}}$) for Stage 2 (gauge-derived reconstruction). Boldface = best, underline = second best.

MODELS	$\geq 0.5\ \mathrm{mm\,h^{-1}}$				$\geq 2.0\ \mathrm{mm\,h^{-1}}$				$\geq 4.0\ \mathrm{mm\,h^{-1}}$				$\geq 8.0\ \mathrm{mm\,h^{-1}}$			
	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR	CSI	HSS
KED	**0.712**	0.292	**0.550**	**0.630**	0.585	0.309	**0.464**	**0.612**	0.322	0.430	0.259	0.404	0.194	0.636	0.145	0.251
KRE	0.677	0.318	0.514	0.592	0.594	0.365	0.443	0.590	0.360	0.491	0.267	0.413	**0.231**	0.598	**0.172**	**0.292**
DK	0.584	0.211	0.505	0.598	0.539	0.322	0.429	0.577	0.363	0.525	0.259	0.401	0.140	0.704	0.105	0.188
STDK	0.634	0.272	0.512	0.597	**0.623**	0.458	0.408	0.549	**0.410**	0.645	0.235	0.368	0.102	0.873	0.060	0.111
P2I-GAN	0.589	0.199	0.514	0.607	0.535	0.285	0.441	0.590	0.377	0.369	**0.309**	**0.465**	0.177	0.554	0.145	0.253
P2I-GAN+	0.559	**0.185**	0.496	0.592	0.502	**0.266**	0.425	0.575	0.350	**0.365**	0.291	0.444	0.177	**0.441**	0.156	0.268

yields the highest POD at 2.0 and 4.0 $\mathrm{mm\,h^{-1}}$, indicating stronger hit rates, but this advantage is accompanied by much higher FAR and weaker overall CSI and HSS. At 4.0 $\mathrm{mm\,h^{-1}}$, P2I-GAN becomes the best-performing method in terms of CSI and HSS, suggesting that it is more effective for moderate rainfall events. At the highest threshold of 8.0 $\mathrm{mm\,h^{-1}}$, however, KRE achieves the best performance, while P2I-GAN+ provides the lowest FAR and the second-best CSI and HSS. Overall, KED and KRE still perform better in terms of overall categorical skill and stability across thresholds, whereas the deep learning-based methods show competitive or even superior potential for moderate to intense events.

The benefit of the orography-aware modulation becomes clearer at the highest threshold. At 8 $\mathrm{mm\,h^{-1}}$, P2I-GAN+ improves FAR from 0.554 to 0.441, CSI from 0.145 to 0.156, and HSS from 0.253 to 0.268 relative to P2I-GAN, corresponding to a markedly lower false alarm rate and gains of about 7.6% and 5.9% in CSI and HSS, respectively. This suggests that, although the advantage of P2I-GAN+ is limited at lower rainfall thresholds, it contributes more clearly to the detection of intense rainfall events.

5.2.3 Visual Comparison

Here we compare the reconstructed fields from six selected methods for representative summer and winter events. As seen in Fig. 10, the radar reference and gauge inputs are not always consistent in either spatial coverage or rainfall intensity at a given time. In this setting, the gauge maps are mainly used to show the sparse input observations, whereas the radar fields serve as a visual reference for the realism of the reconstructed rainfall patterns.

For both events, P2I-GAN+ and KRE produce comparatively more realistic visual structures. DK and STDK again show persistent artificial artifacts across frames, consistent with the behaviour already noted in Stage 1. DK can recover the broad rainfall structure, whereas STDK produces smoother fields. KED yields relatively blurred patterns overall. By contrast, the advantage of P2I-GAN becomes more evident in the case, where the observations are sparse and fragmented, and yet the model maintains relatively stable generation with temporal continuity. Overall, P2I-GAN, using only gauge observations as

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

input, produces rainfall fields with visual realism comparable to KRE while retaining more stable temporal evolution, especially in the convective event with limited observations. More event-based visual comparisons can be found in Sect. S2.2 of the Supplement.

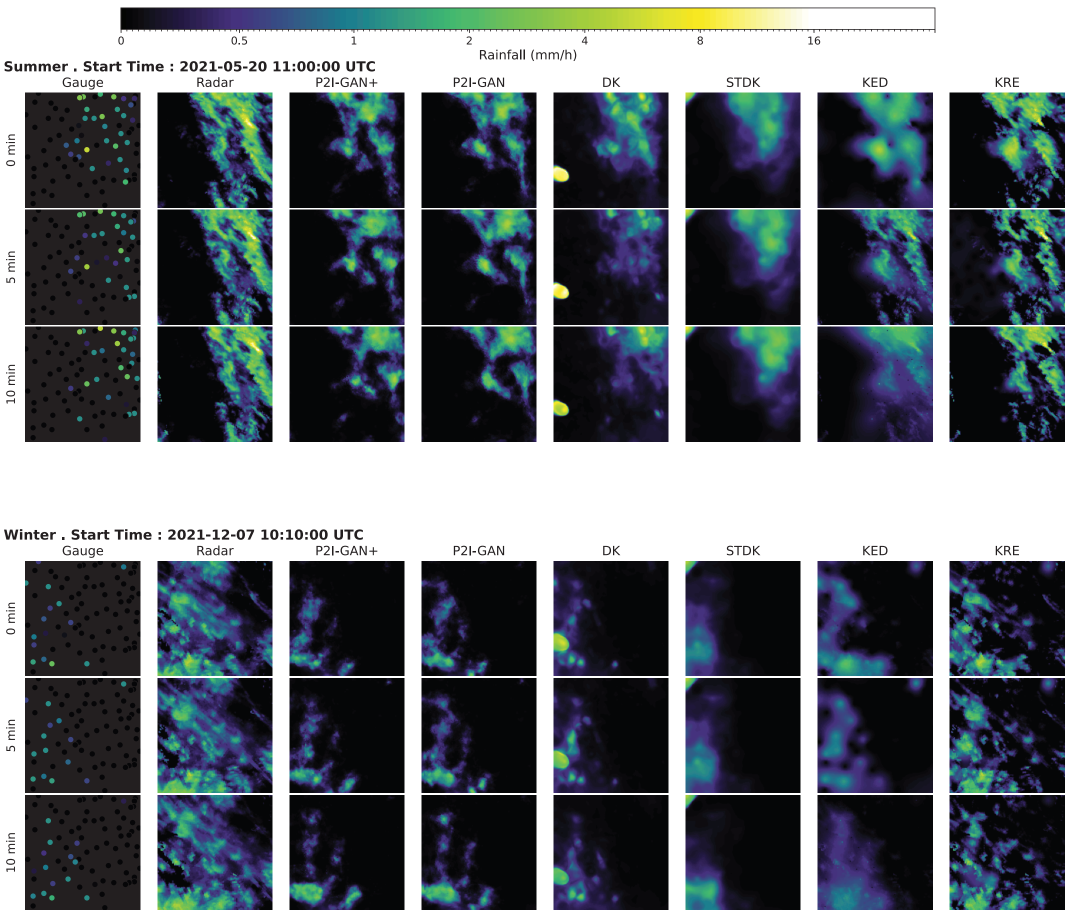


Figure 10. Reconstruction examples for summer and winter rainfall events (selected manually according to their occurrence times), comparing outputs from multiple radar-gauge merging methods over time.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

6 Conclusions

This study presents P2I-GAN, a deep generative framework that reformulates rainfall interpolation as a video-inpainting problem, enabling the reconstruction of spatial–temporal rainfall fields from highly sparse and irregular rain-gauge observations. By integrating an IDW-based initialisation, a U-Net generator, and dual spatial–temporal discriminators, the model learns rainfall evolution patterns purely from historical radar imagery and applies them to gauge-only inputs. A variant of the model, P2I-GAN+, further incorporates orographic modulation to account for topographic-induced rainfall variability.

Through extensive experiments, including both interpolation-only and real-world gauge-driven reconstruction settings, we demonstrate that the proposed models consistently outperform classical interpolation methods and provide competitive results as established radar–gauge merging techniques. Improvements are evident in pixel-wise accuracy, spatial structural similarity, categorical skill scores, and visual consistency. Notably, the model achieves these gains without requiring radar fields at inference, demonstrating that spatial–temporal representations learned from radar can generalise effectively to gauge data. This capability is particularly valuable for operational hydrology, where radar may be unavailable, or too expensive to deploy.

Despite these strengths, limitations remain. The model exhibits mild smoothing and underestimation of extreme intensities, reflecting challenges inherent to sparse inputs and highly skewed rainfall distributions. Future work will explore enhanced decoder architectures, such as cascade modulation, multi-threshold edge conditioning, and multifractal-based constraints, to better capture sharp gradients and localised extremes. Overall, the proposed framework provides a practical and scalable solution for generating high-resolution rainfall fields from sparse observations, with strong potential for real-time flood forecasting, urban hydrology, and applications in data-scarce regions. Validation of the retrieved rainfall fields can also be updated in future work. It could incorporate multifractal analysis to better check if the expected scale invariant features and variability across scales are indeed properly simulated, as well as hydrological validation, to confirm water amount distribution at ground level.

Code and data availability. The code used to train the models, perform rainfall field reconstruction, compute evaluation metrics, and produce the figures and tables in this manuscript is archived at https://doi.org/10.5281/zenodo.19503192(Wang et al., 2026). This release corresponds to the version of the P2I-GAN benchmark framework used for the experiments presented in this study. All scripts are implemented in Python and include model training, inference, interpolation, merging, verification, and visualization components required to reproduce the reported results.

The rainfall datasets used in this study are available at https://doi.org/10.5281/zenodo.17181918 (Lin et al., 2025), which provides 5-min gauge records, including 5-min EA-ST rainfall observations for 2021. The EA-ST dataset in the Zenodo record has been reformatted into CSV (Comma-Separated Values) files. The NIMROD radar data used in this study were produced by the Met Office NIMROD system and were accessed via the Centre for Environmental Data Analysis (CEDA) Archive under the Met Office weather-data access route. Access to these radar data requires user registration with CEDA and approval under the relevant access conditions.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Author contributions. Bing-Zhang Wang (BW) and Li-Pen Wang (LW) jointly conceptualised research idea. BW led the model development and implementation. BW and LW co-designed the experiments to evaluate the result and prepared manuscript, with support from Auguste Gires (AG).

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Acknowledgements. The authors express their gratitude for the financial support received from two National Science and Technology Council research projects (NSTC 113-2923-M-002-001-MY4 and 114-2625-M-002-011-). In addition, they extend their appreciation to the Centre for Environmental Data Analysis for providing access to the NIMROD radar data produced by the Met Office. Authors acknowledge the France-Taiwan Ra2DW project, supported by the French National Research Agency (ANR-23-CE01-0019-01). The authors acknowledge the use of AI-assisted language tools to improve the clarity and readability of the manuscript.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

References

Ahmed, S. and de Marsily, G.: Comparison of Geostatistical Methods for Estimating Transmissivity Using Data on Transmissivity and Specific Capacity, Water Resources Research, 23, 1717–1737, https://doi.org/10.1029/WR023i009p01717, 1987.

Appleby, G., Liu, L., and Liu, L.-P.: Kriging Convolutional Networks, Proceedings of the AAAI Conference on Artificial Intelligence, 34, 3187–3194, https://doi.org/10.1609/aaai.v34i04.5716, 2020.

Bell, T. L.: A space-time stochastic model of rainfall for satellite remote-sensing studies, Journal of Geophysical Research: Atmospheres, 92, 9631–9643, https://doi.org/https://doi.org/10.1029/JD092iD08p09631, 1987.

Cao, Y., Li, S., Feng, J., Chen, L., Li, H., and Zhang, Y.: Enhancing Machine Learning Models for Nowcasting and Short-Term Forecasting of Precipitation With a Novel Probability-Matching Loss Function, Geophysical Research Letters, 52, e2025GL119 442, https://doi.org/https://doi.org/10.1029/2025GL119442, e2025GL119442 2025GL119442, 2025.

Chen, H., Chandrasekar, V., Tan, H., and Cifelli, R.: Rainfall Estimation From Ground Radar and TRMM Precipitation Radar Using Hybrid Deep Neural Networks, Geophysical Research Letters, 46, 10 669–10 678, https://doi.org/https://doi.org/10.1029/2019GL084771, 2019.

Chen, K., Liu, E., Deng, M., Tan, X., Wang, J., Shi, Y., and Wang, Z.: DKNN: deep kriging neural network for interpretable geospatial interpolation, INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 38, 1486–1530, https://doi.org/10.1080/13658816.2024.2347316, 2024.

Chen, W., Li, Y., Reich, B. J., and Sun, Y.: DeepKriging: Spatially Dependent Deep Neural Networks for Spatial Prediction, https://arxiv.org/abs/2007.11972, 2022.

Chiang, Y.-M., Chang, F.-J., Jou, B. J.-D., and Lin, P.-F.: Dynamic ANN for precipitation estimation and forecasting from radar observations, Journal of Hydrology, 334, 250–261, https://doi.org/https://doi.org/10.1016/j.jhydrol.2006.10.021, 2007.

Cristiano, E., ten Veldhuis, M.-C., and van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review, Hydrology and Earth System Sciences, 21, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017, 2017.

Ebert, E. E.: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation, Monthly Weather Review, 129, 2461 – 2480, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2, 2001.

Ehret, U., Götzinger, J., Bárdossy, A., and Pegram, G. G.: Radar-based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany, International Journal of River Basin Management, 6, 323–329, https://doi.org/10.1080/15715124.2008.9635359, 2008.

Elharrouss, O., Damseh, R., Belkacem, A. N., Badidi, E., and Lakas, A.: Transformer-based image and video inpainting: current challenges and future directions, ARTIFICIAL INTELLIGENCE REVIEW, 58, https://doi.org/10.1007/s10462-024-11075-9, 2025.

Ferraris, L., Gabellani, S., Rebora, N., and Provenzale, A.: A comparison of stochastic models for spatial rainfall downscaling, Water Resources Research, 39, https://doi.org/https://doi.org/10.1029/2003WR002504, 2003.

Geiss, A. and Hardin, J. C.: Radar Super Resolution Using a Deep Convolutional Neural Network, Journal of Atmospheric and Oceanic Technology, 37, 2197 – 2207, https://doi.org/10.1175/JTECH-D-20-0074.1, 2020.

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I., and Simoes, N.: Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study, Journal of Hydrology, 442-443, 117–128, http://www.sciencedirect.com/science/article/pii/S0022169412002685, 2012.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Gires, A., Tchiguirinskaia, I., Schertzer, D., Schellart, A., Berne, A., and Lovejoy, S.: Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data, Atmospheric Research, 138, 125–138, https://doi.org/https://doi.org/10.1016/j.atmosres.2013.11.008, 2014.

Golding, B. W.: Nimrod: a system for generating automated very short range forecasts, Meteorological Applications, 5, 1–16, https://doi.org/10.1017/S1350482798000577, 1998.

Goudenhoofdt, E. and Delobbe, L.: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, Hydrology and Earth System Sciences, 13, 195–203, https://doi.org/10.5194/hess-13-195-2009, 2009.

Grimes, D. I. F. and Pardo-Igúzquiza, E.: Geostatistical Analysis of Rainfall, Geographical Analysis, 42, 136–160, https://doi.org/10.1111/j.1538-4632.2010.00787.x, 2010.

Haensler, A. and Weiler, M.: Enhancing the usability of weather radar data for the statistical analysis of extreme precipitation events, HYDROLOGY AND EARTH SYSTEM SCIENCES, 26, 5069–5084, https://doi.org/10.5194/hess-26-5069-2022, 2022.

HARDY, R.: MULTIQUADRIC EQUATIONS OF TOPOGRAPHY AND OTHER IRREGULAR SURFACES, JOURNAL OF GEOPHYSICAL RESEARCH, 76, 1905+, https://doi.org/10.1029/JB076i008p01905, 1971.

Harrison, D., Driscoll, S., and Kitchen, M.: Improving precipitation estimates from weather radar using quality control and correction techniques, METEOROLOGICAL APPLICATIONS, 7, 135–144, https://doi.org/10.1017/S1350482700001468, 2000.

He, X., Sonnenborg, T. O., Refsgaard, J. C., Vejen, F., and Jensen, K. H.: Evaluation of the value of radar QPE data and rain gauge data for hydrological modeling, Water Resources Research, 49, 5989–6005, https://doi.org/https://doi.org/10.1002/wrcr.20471, 2013.

Houze, R. A.: Orographic effects on precipitating clouds, Reviews of Geophysics, 50, RG1001, https://doi.org/10.1029/2011RG000365, 2012.

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y., and Li, L.: Rainfall Spatial Estimations: A Review from Spatial Interpolation to Multi-Source Data Merging, Water, 11, 579, https://doi.org/10.3390/w11030579, 2019.

Hudson, G. and Wackernagel, H.: Mapping Temperature Using Kriging with External Drift: Theory and an Example from Scotland, International Journal of Climatology, 14, 77–91, https://doi.org/10.1002/joc.3370140107, 1994.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976, https://doi.org/10.1109/CVPR.2017.632, 2017.

Kidd, C., Bauer, P., Turk, J., Huffman, G., Joyce, R., Hsu, K.-L., and Braithwaite, D.: Intercomparison of high-resolution precipitation products over northwest Europe, Journal of Hydrometeorology, 13, 67–83, https://doi.org/10.1175/JHM-D-11-042.1, 2012.

King, J. A., Engelstaedter, S., Washington, R., and Munday, C.: Variability of the Turkana low-level jet in reanalysis and models: Implications for rainfall, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 154, https://doi.org/10.1029/2020JD034429, 2021.

KITCHEN, M., BROWN, R., and DAVIES, A.: REAL-TIME CORRECTION OF WEATHER RADAR DATA FOR THE EFFECTS OF BRIGHT BAND, RANGE AND OROGRAPHIC GROWTH IN WIDESPREAD PRECIPITATION, QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY, 120, 1231–1254, https://doi.org/10.1002/qj.49712051906, 1994.

Kopczewska, K.: Spatial Machine Learning: New Opportunities for Regional Science, The Annals of Regional Science, 68, 713–755, https://doi.org/10.1007/s00168-021-01101-x, 2022.

Kullback, S. and Leibler, R. A.: On Information and Sufficiency, Annals of Mathematical Statistics, 22, 79–86, https://doi.org/10.1214/AOMS/1177729694, 1951.

Lebrenz, H. and Bárdossy, A.: Geostatistical interpolation by quantile kriging, Hydrology and Earth System Sciences, 23, 1633–1648, https://doi.org/10.5194/hess-23-1633-2019, 2019.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M.: Enhanced Deep Residual Networks for Single Image Super-Resolution, https://arxiv.org/abs/1707.02921, 2017.

Lin, H.-M., Wang, L.-P., and Han, J.-Y.: Gauge Rainfall Datasets from MIDAS-open (hourly) and EA-ST (5-min), Southern UK, 2016–2021, https://doi.org/10.5281/zenodo.17181918, 2025.

Ly, S., Charles, C., and Degre, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: A review, Biotechnologie, Agronomie, Société et Environnement, 17, 392–406, 2013.

Maggioni, V., Meyers, P. C., and Robinson, M. D.: A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era, Journal of Hydrometeorology, 17, 1101 – 1117, https://doi.org/10.1175/JHM-D-15-0190.1, 2016.

Matheron, G.: The Theory of Regionalized Variables and Its Applications, École Nationale Supérieure des Mines de Paris, Paris, France, https://cg.ensmp.fr/bibliotheque/public/MATHERON_Ouvrage_00167.pdf, cahiers du Centre de Morphologie Mathématique de Fontainebleau, No. 5, 1971.

Matheron, G.: The Intrinsic Random Functions and Their Applications, Advances in Applied Probability, 5, 439–468, https://scispace.com/pdf/the-intrinsic-random-functions-and-their-applications-45zyo5v3ck.pdf, 1973.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The future of Earth observation in hydrology, Hydrology and Earth System Sciences, 21, 3879–3914, https://doi.org/10.5194/hess-21-3879-2017, 2017.

McClean, F., Dawson, R., and Kilsby, C.: Implications of using global digital elevation models for flood risk analysis in cities, Water Resources Research, 56, e2020WR028 241, https://doi.org/10.1029/2020WR028241, 2020.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y.: Spectral Normalization for Generative Adversarial Networks, CoRR, abs/1802.05957, http://arxiv.org/abs/1802.05957, 2018a.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y.: Spectral Normalization for Generative Adversarial Networks, https://arxiv.org/abs/1802.05957, 2018b.

Moreau, E., Testud, J., and Le Bouar, E.: Rainfall spatial variability observed by X-band weather radar and its implication for the accuracy of rainfall estimates, Advances in Water Resources, 32, 1011–1019, https://doi.org/https://doi.org/10.1016/j.advwatres.2008.11.007, weather Radar and Hydrology, 2009.

Myers, D. E.: Matrix Formulation of Co-kriging, Mathematical Geology, 14, 249–257, https://doi.org/10.1007/BF01032887, 1982.

Nag, P., Sun, Y., and Reich, B. J.: Spatio-temporal DeepKriging for interpolation and probabilistic forecasting, Spatial Statistics, 57, 100 773, https://doi.org/https://doi.org/10.1016/j.spasta.2023.100773, 2023.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nwaila, G. T., Zhang, S. E., Bourdeau, J. E., Frimmel, H. E., and Ghorbani, Y.: Spatial Interpolation Using Machine Learning: From Patterns and Regularities to Block Models, NATURAL RESOURCES RESEARCH, 33, 129–161, https://doi.org/10.1007/s11053-023-10280-7, 2024.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., and ten Veldhuis, M.-C.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, Journal of Hydrology, 531, 389–407, https://doi.org/https://doi.org/10.1016/j.jhydrol.2015.05.035, hydrologic Applications of Weather Radar, 2015.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Ochoa-Rodriguez, S., Wang, L.-P., Willems, P., and Onof, C.: A Review of Radar-Rain Gauge Data Merging Methods and Their Potential for Urban Hydrological Applications, Water Resources Research, 55, 6356–6391, https://doi.org/https://doi.org/10.1029/2018WR023332, 2019.

Odeh, I. O. A., McBratney, A. B., and Chittleborough, D. J.: Further Results on Prediction of Soil Properties from Terrain Attributes: Heterotopic Cokriging and Regression-Kriging, Geoderma, 67, 215–226, https://doi.org/10.1016/0016-7061(95)00007-B, 1995.

Perica, S. and Foufoula-Georgiou, E.: Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, J. Geophys. Res., 101, 26,347–26,361, https://doi.org/10.1029/96JD01870, 1996.

Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J.: Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions, Journal of Climate, 20, 4356 – 4376, https://doi.org/10.1175/JCLI4253.1, 2007.

Quan, W., Chen, J., Liu, Y., Yan, D.-M., and Wonka, P.: Deep Learning-Based Image and Video Inpainting: A Survey, INTERNATIONAL JOURNAL OF COMPUTER VISION, 132, 2367–2400, https://doi.org/10.1007/s11263-023-01977-6, 2024.

Quirmbach, M. and Schultz, G.: Comparison of rain gauge and radar data as input to an urban rainfall-runoff model, Water Science and Technology, 45, 27–33, https://doi.org/10.2166/wst.2002.0023, 2002.

Rebora, N., Ferraris, L., von Hardenberg, J., and Provenzale, A.: RainFARM: Rainfall Downscaling by a Filtered Autoregressive Model, Journal of Hydrometeorology, 7, 724 – 738, https://doi.org/10.1175/JHM517.1, 2006.

Roberts, N. M. and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Monthly Weather Review, 136, 78 – 97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: Some models for rainfall based on stochastic point processes, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 410, 269–288, https://doi.org/10.1098/rspa.1987.0039, 1987.

Roe, G. H.: Orographic precipitation, Annual Review of Earth and Planetary Sciences, 33, 645–671, https://doi.org/10.1146/annurev.earth.33.092203.122541, 2005.

Sarafian, R., Barron, M., Nathan, S., Nissenbaum, D., Vadislavsky, E., Levi, Y., and Rudich, Y.: RadarGaugeNet: Radar-Rain-Gauge-Based AI Architecture for Real-Time Quantitative Precipitation Estimation, IEEE Transactions on Geoscience and Remote Sensing, 63, https://doi.org/10.1109/TGRS.2025.3640078, publisher Copyright: © 2025 IEEE., 2025.

Shehu, B. and Haberlandt, U.: Relevance of merging radar and rainfall gauge data for rainfall nowcasting in urban hydrology, JOURNAL OF HYDROLOGY, 594, https://doi.org/10.1016/j.jhydrol.2020.125931, 2021.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and WOO, W.-c.: Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2017/file/a6db4ed04f1621a119799fd3d7545d3d-Paper.pdf, 2017.

Shi, Z., Geng, H., Wu, F., Geng, L., and Zhuang, X.: Radar-SR3: A Weather Radar Image Super-Resolution Generation Model Based on SR3, Atmosphere, 15, https://doi.org/10.3390/atmos15010040, 2024.

Smith, R. B.: Progress on the theory of orographic precipitation, Proceedings of the Royal Society A, 462, 2703–2727, https://doi.org/10.1098/rspa.2006.1712, 2006.

Sokol, Z., Szturc, J., Orellana-Alvear, J., Popova, J., Jurczyk, A., and Celleri, R.: The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review, Remote Sensing, 13, 351, https://doi.org/10.3390/rs13030351, 2021.

https://doi.org/10.5194/egusphere-2026-2406
Preprint. Discussion started: 23 July 2026

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, Reviews of Geophysics, 56, 79–107, https://doi.org/https://doi.org/10.1002/2017RG000574, 2018.

Sun, T., Yan, N., Zhu, W., and Zhuang, Q.: Assessing a machine learning-based downscaling framework for obtaining 1km daily precipitation from GPM data, Heliyon, 10, e36 368, https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e36368, 2024.

Tavassoli, A., Waghei, Y., and Nazemi, A.: Hybrid MLP-IDW approach based on nearest neighbor for spatial prediction, COMPUTATIONAL STATISTICS, 37, 1943–1962, https://doi.org/10.1007/s00180-021-01186-0, 2022.

Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., ten Veldhuis, M.-C., Arnbjerg-Nielsen, K., Rasmussen, M. R., and Molnar, P.: Weather radar rainfall data in urban hydrology, Hydrology and Earth System Sciences, 21, 1359–1380, https://doi.org/10.5194/hess-21-1359-2017, 2017.

van de Beek, C. Z., Leijnse, H., Hazenberg, P., and Uijlenhoet, R.: Close-Range Radar Rainfall Estimation and Error Analysis, Atmospheric Measurement Techniques, 9, 3837–3850, https://doi.org/10.5194/amt-9-3837-2016, 2016.

Wang, B.-Z., Wang, L.-P., and Auguste, G.: P2I-GAN Benchmark: Deep Generative Framework for Spatio-Temporal Rainfall Reconstruction from Sparse Gauges, https://doi.org/10.5281/zenodo.19503192, 2026.

Wilson, J. W. and Brandes, E. A.: Radar Measurement of Rainfall—A Summary, Bulletin of the American Meteorological Society, 60, 1048 – 1060, https://doi.org/10.1175/1520-0477(1979)060<1048:RMORS>2.0.CO;2, 1979.

Wu, Y., Zhuang, D., Labbe, A., and Sun, L.: Inductive Graph Neural Networks for Spatiotemporal Kriging, https://arxiv.org/abs/2006.07527, 2020.

Xiang, H., Zou, Q., Nawaz, M. A., Huang, X., Zhang, F., and Yu, H.: Deep learning for image inpainting: A survey, PATTERN RECOGNITION, 134, https://doi.org/10.1016/j.patcog.2022.109046, 2023.

Yu, J., Li, X.-F., Lewis, E., Blenkinsop, S., and Fowler, H. J.: UKGrsHP: A UK high-resolution gauge–radar–satellite merged hourly precipitation analysis dataset, Climate dynamics, 54, 2919–2940, https://doi.org/10.1007/s00382-020-05144-2, 2020.

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., Chen, J., Guo, S.-L., and Liu, J.: The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach, Journal of Hydrology, 563, 106–122, https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.05.058, 2018.

Zhang, G., Rui, X., and Fan, Y.: Critical Review of Methods to Estimate PM2.5 Concentrations within Specified Research Region, ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION, 7, https://doi.org/10.3390/ijgi7090368, 2018.

Zhou, W., Zhu, Z., Xie, Y., and Cai, Y.: Impacts of rainfall spatial and temporal variabilities on runoff quality and quantity at the watershed scale, Journal of Hydrology, 603, 127 057, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.127057, 2021.