

Response to Reviewer #1

We sincerely thank the reviewer for the positive assessment of our manuscript and for the constructive comments. We appreciate the reviewer’s recognition of the novelty and practical relevance of the proposed framework. We have carefully addressed all comments and revised the manuscript accordingly. The main changes include additional explanation of the temporal-modulation mechanism, clarification of the role of topography in P2I-GAN+, additional evaluation of bias and gauge-density sensitivity, and improved descriptions of the preprocessing and evaluation metrics.

General comments

Reviewer comment:

The authors provide a new machine learning approach to reconstruct rainfall fields from irregular point measurements. Inspired by video-inpainting from computer vision spatial-temporal rainfall structures are derived from historical radar observations and applied to infer rainfall fields from sparse gauge only data. The validation is done in two stages with comparisons against various reference methods. The new approach outperforms the reference methods well in stage 1 (radar-sampled interpolation) and is comparable in stage 2 (real gauge driven reconstruction). The presented approach is novel and shows good performance. The application is useful in cases if no radar data are available. The procedures are plausible and the manuscript is well written. The reader however needs to have some knowledge of machine learning technics to be able follow the text. I have some comments for improvement of the manuscript. Publication is recommended after minor revisions.

Response:

We sincerely thank the reviewer for the positive assessment of our work and for recognising the novelty, performance, and practical relevance of the proposed framework. We also appreciate the constructive suggestions for improving the clarity and completeness of the manuscript. We have carefully considered all comments and revised the manuscript accordingly, as detailed below.

Comment 1: Figure 1

Reviewer comment:

Fig. 1: I cannot see a shaded area there as indicated in the figure caption.

Response:

Thank you for pointing this out. The reviewer is correct. The shaded area mentioned in the original caption was left from an earlier version of the figure and is not present in the submitted figure. We will therefore correct the caption to remove this incorrect description in the revised manuscript.

Comment 2: Gating-based temporal modulation

Reviewer comment:

Lines 155 ff.: “Gating-based temporal modulation ...”. How this technically works is not quite clear to me and needs some more explanation here.

Response:

We agree that the original description focused mainly on the purpose of the temporal modulation but did not sufficiently explain its implementation.

In brief, the gating-based temporal modulation uses two sets of learnable 1×1 convolutional weights to generate temporal gates. This gate is multiplied element-wise with the intermediate input features, allowing the contribution of information from adjacent time steps to be adjusted adaptively. This mechanism is introduced because sparse rainfall observations can exhibit substantial frame-to-frame variability at individual locations. By adaptively modulating the temporal information before spatial interpolation, the model can reduce the influence of abrupt local fluctuations and provide a more stable temporal representation for the subsequent IDW interpolation and rainfall-field reconstruction.

We will expand Sect. 3.2 accordingly to describe the construction of the gates, how they operate on the intermediate features, and how the temporally modulated information is subsequently passed to the IDW interpolation. We will also add the corresponding mathematical formulation to make the implementation and data flow clearer.

Comment 3: Orography and transferability

Reviewer comment:

Section 3.3: It seems that no DEM is used to consider orography. This should be more clearly stated and discussed; if fixed local structures are learned here this also has implications for a transfer to other regions.

Response:

We thank the reviewer for this important observation. The reviewer is correct that P2I-GAN+ does not use a DEM or other explicit topographic variables as model inputs. Instead, the spatial modulation fields are learned from the historical radar training data and are intended to represent persistent location-dependent rainfall characteristics, which may include rainfall patterns associated with local orography. We will revise Sect. 3.3 to clarify this distinction and have avoided implying that elevation information is explicitly provided to the model.

We also agree that this design has implications for transferability. Because the learned modulation fields are location dependent, they cannot be assumed to transfer directly to a geographically different domain. We will therefore add this point to the discussion as a limitation of the current

P2I-GAN+ implementation. Future extensions could explicitly condition the modulation on DEM-derived information, which would provide a physically interpretable treatment of orographic effects and potentially improve transferability across regions.

Comment 4: Bias metric

Reviewer comment:

Section 4.3.1: Why is there no bias included in the continuous metrics? Later I found out the scatterplots show graphically the bias. But there are no real quantitative figures. May be a bias metric can still be added?

Response:

We agree with the reviewer that a quantitative bias measure provides useful complementary information to MAE and RMSE, particularly because the scatterplots indicate systematic underestimation at higher rainfall intensities. We will therefore include the multiplicative bias (BIAS) as an additional continuous evaluation metric and report it for both experimental stages.

The metric is defined as the ratio of the total reconstructed rainfall to the total reference rainfall:

$$\text{BIAS} = \frac{\sum_{i=1}^N \hat{R}_i}{\sum_{i=1}^N R_i}, \quad (1)$$

where $\hat{R}_i$ and R_i denote the reconstructed and reference rainfall intensities, respectively. A value of 1 indicates no aggregate bias, while values above and below 1 indicate overall overestimation and underestimation, respectively.

The corresponding definition will be added to Sect. 4.3.1, and the results will be included in Tables 4 and 7. As a representative example, the updated Stage 1 results (corresponding to Table 4 in the original manuscript) are presented in Table 1 below.

Table 1: Stage 1 interpolation results. Arrows indicate the preferred direction of each metric. For BIAS, a value of 1 indicates no aggregate bias, while values above and below 1 indicate overestimation and underestimation, respectively. Boldface and underlining denote the best and second-best results, respectively.

MODEL	MAE ↓ (mm/h)	RMSE ↓ (mm/h)	PSS ↑ (-)	SSIM ↑ (-)	BIAS → 1 (-)
OK	0.195	0.641	0.619	0.506	1.316
RBF	0.220	0.923	<u>0.756</u>	0.563	1.540
DK	0.259	0.834	0.586	0.419	2.250
STDK	0.261	0.777	0.645	0.414	2.047
P2I-GAN	<u>0.174</u>	0.585	0.754	0.632	<u>1.286</u>
P2I-GAN+	0.173	<u>0.587</u>	0.768	<u>0.632</u>	1.249

Comment 5: Logarithmic transformation and back-transformation bias

Reviewer comment:

Lines 410ff.: It is known that interpolation based on transformed data is not unbiased after back-transformation. Please discuss this and may be show the effect of this logarithmic transformation on the bias.

Response:

We thank the reviewer for highlighting this issue. We agree that interpolation in the logarithmically transformed domain does not guarantee unbiased estimates after back-transformation to the original rainfall scale. We will expand the discussion of the preprocessing procedure to acknowledge this effect and examine its influence on the reconstructed rainfall intensities.

To quantify the effect of the interpolation domain, we compared two configurations: (i) interpolation in the logarithmically encoded domain followed by back-transformation and (ii) direct interpolation in the original rainfall domain (mm h^{-1}). For this comparison, we use the additive mean bias, $\text{MB} = \mathbb{E}[\hat{R} - R]$, rather than the multiplicative bias (BIAS) introduced in response to Comment 4. While BIAS measures the ratio of total reconstructed to total reference rainfall, MB expresses the average estimation error in rainfall-intensity units. The difference $\Delta\text{MB} = \text{MB}_{\log} - \text{MB}_{\text{direct}}$ therefore quantifies the change in mean bias associated with interpolation in the logarithmic domain.

For Ordinary Kriging (OK), logarithmic-domain interpolation resulted in a mean bias of $-0.0708 \text{ mm h}^{-1}$, compared with 0.0094 mm h^{-1} for direct interpolation, giving $\Delta\text{MB} = -0.0802 \text{ mm h}^{-1}$. For the Radial Basis Function (RBF) method, the corresponding difference was smaller, at $-0.0090 \text{ mm h}^{-1}$. When evaluated by reference rainfall-intensity class, the additional negative bias generally increased in magnitude with rainfall intensity for both methods. However, substantial underestimation at higher intensities remained even with direct interpolation, indicating that the logarithmic transformation is not the sole source of this behaviour. The sparse observations and the smoothing inherent in spatial interpolation may also contribute.

These results should be distinguished from the aggregate BIAS values reported in response to Comment 4. In particular, the negative changes in MB observed here do not imply that the logarithmic transformation explains the aggregate overestimation indicated by BIAS values above 1. The present comparison isolates the effect of the interpolation domain for OK and RBF; it does not separately quantify that effect in the trained P2I-GAN models.

We will add this analysis and the accompanying discussion to the revised Supplement. The detailed results are presented in Table 2 below.

Table 2: Comparison of interpolation in the logarithmically encoded domain and direct interpolation in the original rainfall domain. Additive mean bias is defined as $MB = \mathbb{E}[\hat{R} - R]$, where negative values indicate underestimation on average. The difference associated with the choice of interpolation domain is defined as $\Delta MB = MB_{\log} - MB_{\text{direct}}$; thus, $\Delta MB < 0$ indicates a more negative mean bias for logarithmic-domain interpolation than for direct interpolation. The analysis uses 200 evenly sampled frames from the 20 test events and includes unobserved grid cells only. All mean bias values are expressed in mm h^{-1} .

Observed rainfall (mm/h)	N	OK			RBF		
		$Bias_{\log}$	$Bias_{\text{direct}}$	$\Delta Bias$ (mm/h)	$Bias_{\log}$	$Bias_{\text{direct}}$	$\Delta Bias$
All	3,261,000	-0.0708	0.0094	-0.0802	-0.0014	0.0076	-0.0090
$R = 0$	1,959,672	0.0379	0.0633	-0.0255	0.0459	0.0392	0.0067
$0 < R < 0.1$	122,262	0.0711	0.1708	-0.0997	0.1265	0.1354	-0.0089
$0.1 \leq R < 0.5$	505,622	0.0222	0.1546	-0.1325	0.1098	0.1289	-0.0190
$0.5 \leq R < 2$	507,084	-0.1896	-0.0232	-0.1664	-0.0242	0.0098	-0.0340
$2 \leq R < 5$	147,079	-0.9578	-0.7025	-0.2552	-0.5631	-0.4947	-0.0684
$5 \leq R < 10$	16,575	-3.2909	-2.8384	-0.4525	-2.4432	-2.2628	-0.1804
$R \geq 10$	2,706	-12.4120	-11.2684	-1.1436	-11.0357	-10.4893	-0.5465

Comment 6: Clipping of transformed rainfall values

Reviewer comment:

Lines 415ff: What means clipped? Are the data outside the range just cutted away?
Isn't that further smoothing the reconstructed field?

Response:

We thank the reviewer for pointing out that the term “clipped” can be misleading and was insufficiently explained in the original manuscript. We will therefore replace this term with “range constraint” and revise Sect. 4.4 to clarify how it is applied. Specifically, this constraint is not applied during data preprocessing or model training and therefore does not affect the training samples, optimisation process, or learned model parameters. It is applied only to the model output during inference, whereby predicted values outside the predefined valid range are bounded to that range to prevent anomalous or physically implausible rainfall intensities.

We also clarify that this operation does not involve any spatial averaging or filtering. It acts independently on individual predicted values and therefore does not spatially smooth the reconstructed rainfall fields. We will add these clarifications to the revised manuscript.

Comment 7: Definition of SSIM

Reviewer comment:

Lines 4.3.5ff: What is SSIM. This metric is never defined and also not contained in table 4?

Response:

We thank the reviewer for pointing this out. SSIM refers to the Structural Similarity Index, which evaluates the similarity between two spatial fields in terms of local intensity, contrast, and structural information. While SSIM was used in the Stage 1 analysis, its definition was not provided in Sect. 4.3.1, and the corresponding results were not included in Table 4. We will therefore add the definition and reference for SSIM to Sect. 4.3.1 and include the corresponding SSIM results in Table 4 in the original manuscript. The related discussion in Sect. 5.1.1 will also be revised accordingly.

Moreover, we included the SSIM index in Table 1. The results show that P2I-GAN and P2I-GAN+ both achieve an SSIM of 0.632 after rounding, outperforming the traditional interpolation methods. Compared with RBF, which achieves the highest SSIM among the traditional methods, the proposed methods increase SSIM by approximately 0.069, corresponding to a relative improvement of about 12.3%.

Comment 8: Contribution of individual model components**Reviewer comment:**

General question 1: As far as I see the framework has at least seven components. To which degree do these components contribute to the final performance i.e. which importance have the different components?

Response:

We thank the reviewer for this useful suggestion. We agree that the proposed framework contains several components designed to address different aspects of the reconstruction problem, including sparse spatial input, temporal continuity, spatial structure, and rainfall-intensity distribution. To better quantify their individual contributions, we will add an ablation analysis in the Supplement.

Rather than evaluating every possible combination of components, we focus on controlled comparisons that isolate the major conceptual additions to the framework. The planned ablation study includes: (i) a comparison between the basic reconstruction backbone and the GAN-based framework; (ii) removal of the Masked IDW3D input, gating mechanism individually from the complete architecture; and (iii) removal of $\mathcal{L}_{tc}$ and $\mathcal{L}_{pm}$ individually to assess their respective contributions. We will provide the detailed comparison in Supplement Sect.3.

Comment 9: Influence of gauge density**Reviewer comment:**

General questions 2: For me it would be interesting to know how much the gauge density is influencing the results?

Response:

We thank the reviewer for this valuable suggestion. To quantify the influence of gauge density, we will conduct a controlled gauge-density sensitivity analysis covering a range of observation densities and report the results in the revised Supplement. Specifically, input configurations will be generated with approximately one observation within every $n \times n$ -pixel area, where $n = 3, 5, 7, 9, \dots$. Increasing n therefore represents progressively sparser observation conditions. Separate models will be trained for each gauge-density configuration while keeping the model architecture, data partition, training settings, and independent validation data unchanged.

To account for potential sensitivity to the spatial arrangement of observations, multiple spatial configurations will be considered at each gauge density, and performance will be evaluated across multiple rainfall events. The resulting performance-density relationships will quantify how reconstruction accuracy changes with observation density and indicate the range of sparsity over which the proposed framework remains effective. A summary of this analysis will be added to Supplement Sect. 4.