Response Sheet for Referee 02 Comments (egusphere-2025-6551)

We thank Referee #2 for their careful review and constructive suggestions, which have helped improve the manuscript. In the following sections, we respond to each comment and describe the corresponding revisions made to the manuscript.

The reviewer's comments are presented in bold red text, our responses are provided in black text, and the revised manuscript text is shown in blue. All line numbers and references correspond to the originally submitted manuscript.

Comment 01

One thing I think is important when employing machine learning techniques is to justify why these techniques are needed when studying/addressing the issue identified here. I think the paragraph starting at line 51 does a nice job of addressing some of these improvements and different techniques that have been imposed. My guess is that with the relatively dense station network available for this paper's study region, an intelligent IDW approach might be all that is required to achieve very similar results as the ML techniques used in the paper. Some of this seems to be born out later in the paper when it is shown that performance is about the same when using only 40% of the available stations compared to 70% that was used in the study.

Thank you for this insightful comment. We agree that the use of machine learning techniques should be more clearly justified, particularly given the relatively dense gauge network available in the study region.

We acknowledge that IDW can perform competitively in regions with relatively dense gauge coverage, as it directly exploits the spatial proximity of gauge observations. The results show comparable performance under 40% and 70% station availability suggest that the dominant spatial rainfall patterns in the Wet Zone can be partly captured by the existing gauge network. Therefore, in the revised manuscript, IDW is explicitly positioned as a baseline method and evaluated alongside the ML-based approaches. This allows the added value of the proposed GraphIDW framework to be assessed relative to a simple and widely used distance-based interpolation method.

The motivation for using ML in this study is not solely to outperform IDW under the current dense-gauge condition, but to evaluate whether data-driven models can provide a flexible framework for learning nonlinear rainfall relationships and spatial dependencies from satellite–gauge information. In particular, the proposed GraphIDW framework is designed to explicitly represent spatial autocorrelation through graph-based connectivity while also applying residual IDW correction to reduce remaining local errors. This is especially relevant for regions with strong spatial rainfall heterogeneity, complex topographic gradients, and potential data limitations.

The updated Figs 5,6 and 7 have been updated as follows by adding the results of IDW method.

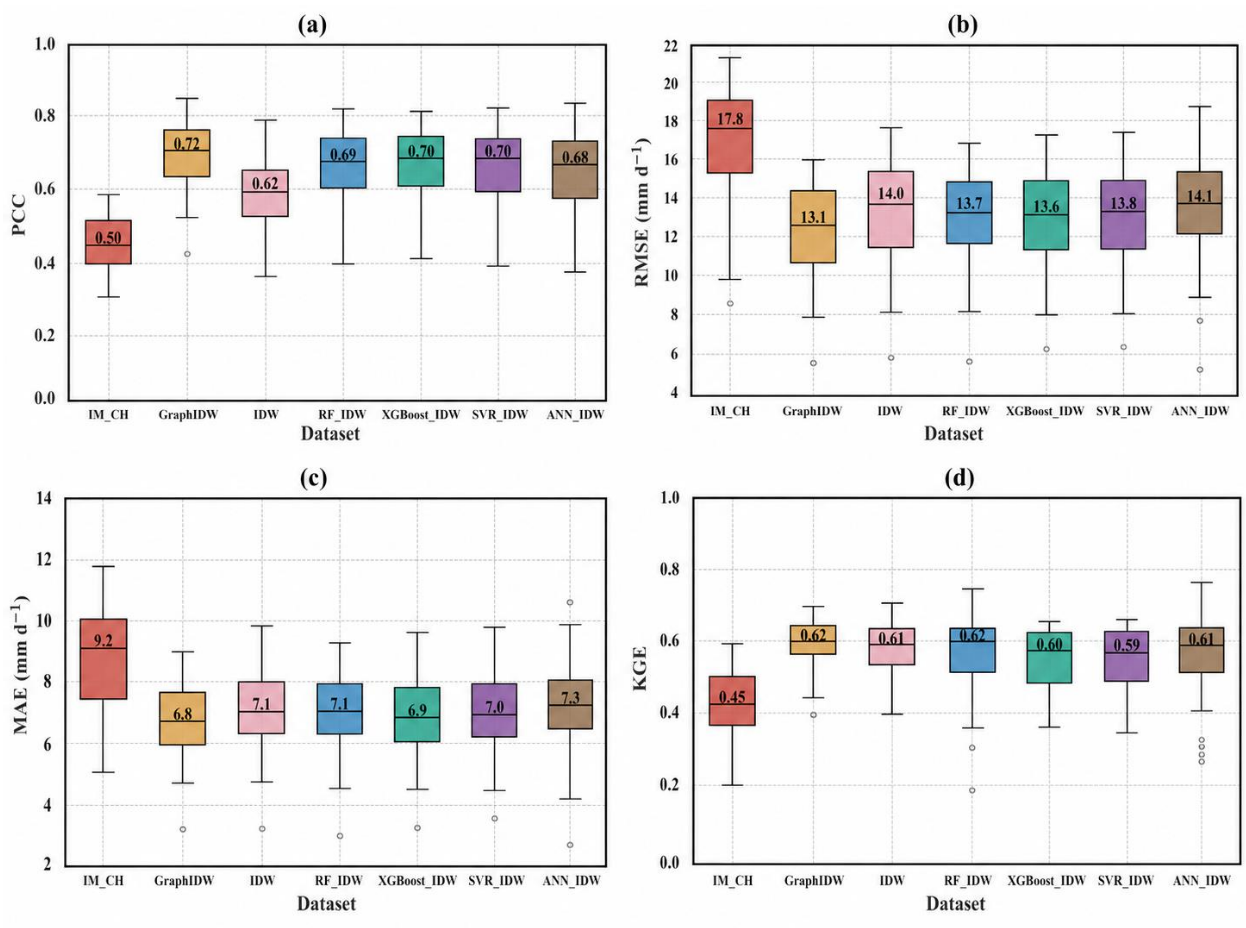

Figure 5. Boxplots of four quantitative metrics (PCC (a), RMSE (b), MAE (c), and (d) KGE) for six products, including original (IM_CH) product

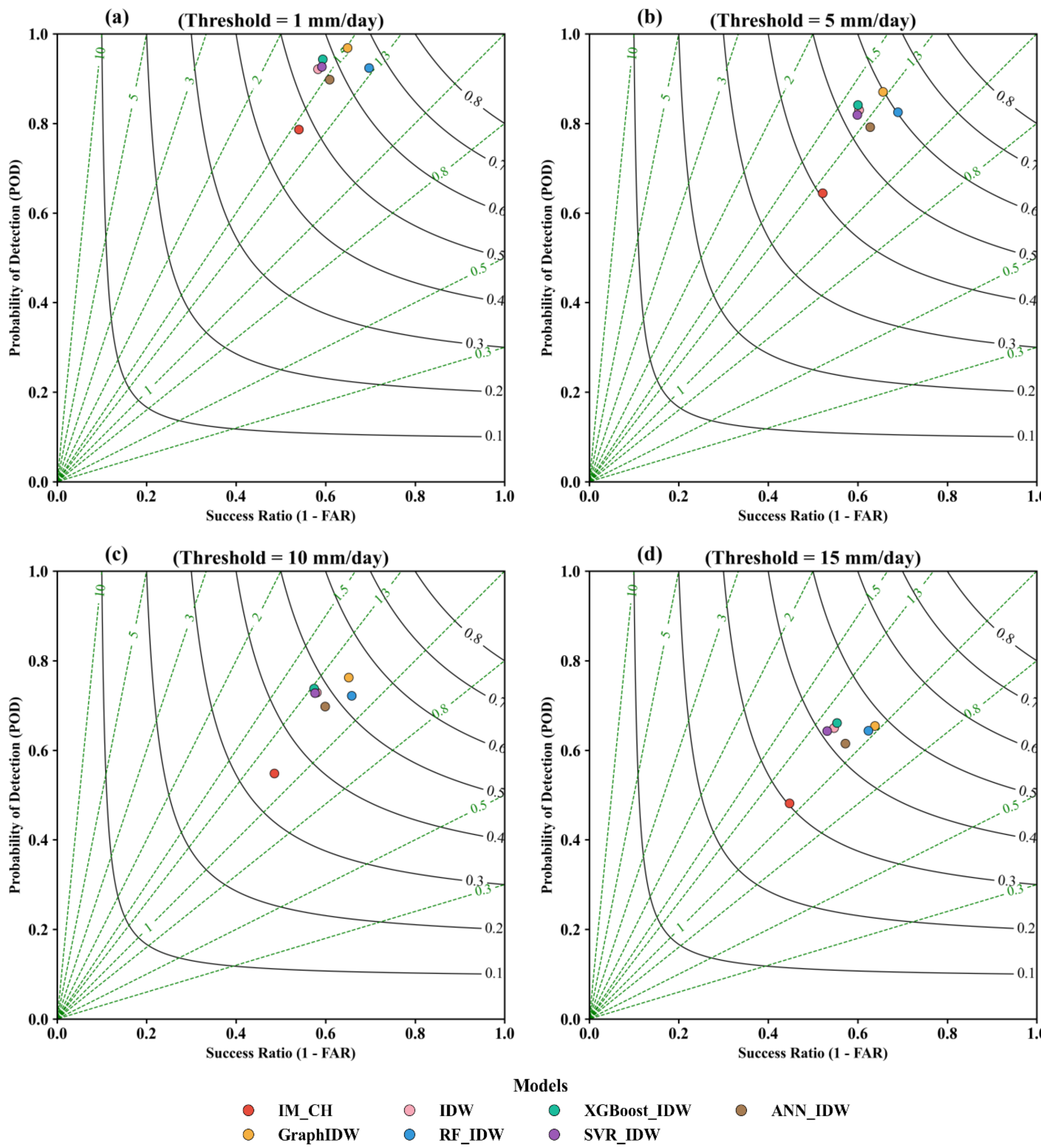

Figure 6. Roebber's performance diagram for the original (IM_CH) and the merged precipitation products. The green dashed lines represent the bias score (BS), while the black contour lines indicate the critical success index (CSI). The four diagrams present the detection performance for the precipitation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

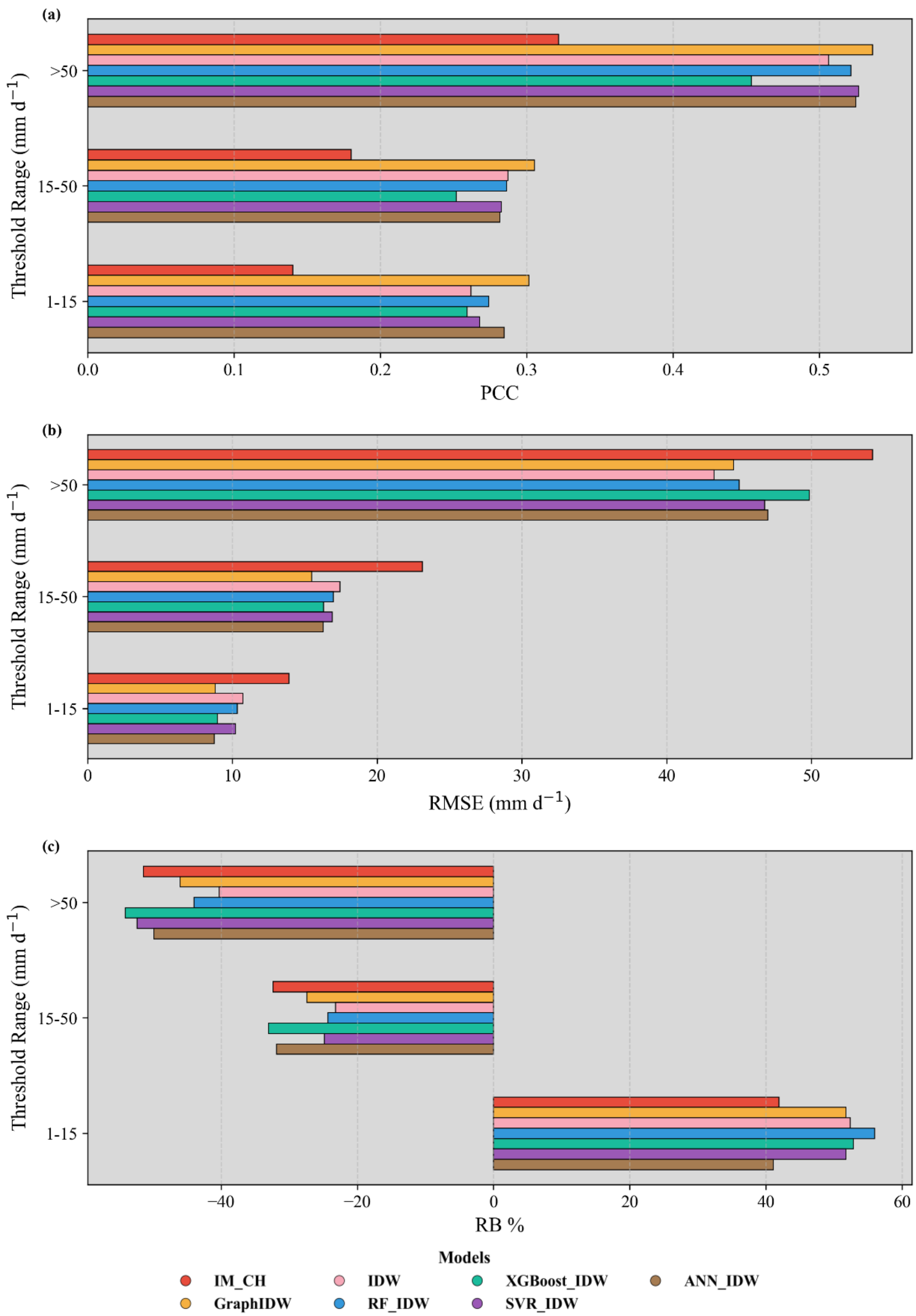

Figure 7. Histogram comparison of (a) PCC, (b) RMSE, and (c) RB% of six precipitation products against ground observations under different rain intensities. (Heavy rain: >50 mm d^{-1}; Moderate: 15-50 mm d^{-1}; Light rain: 1-15 mm d^{-1}).

Comment 02

My biggest criticism of the paper might be in the formulation of the "IMERG-daily,0.05" product put forth in equation 1. If I'm understanding this equation correctly, the fraction on the right side of the equation results in each day's proportion of the monthly total, calculated independently for every 0.1-degree pixel. This daily proportion is then multiplied by the monthly total at the corresponding CHIRPS 0.05-degree pixel. This results in more of a CHIRPS-daily product than an IMERG-daily product, because the sum of the daily values will result in the CHIRPS monthly total. If there are big differences between IMERG and CHIRPS, imagine a monthly IMERG total of 50mm while CHIRPS is 250mm, the result of equation 1 is going to result in a data product that is going to be more like the 250mm rainfall total.

As I was thinking about a better formulation of equation 1, I couldn't figure out how best to use the higher spatial resolution of CHIRPS combined with the better temporal resolution of IMERG. The thing to do may be to look at the fraction of each 0.05-degree CHIRPS monthly pixel relative to a 0.1-degree version of CHIRPS, and use that proportion to downscale the IMERG monthly totals to the 0.05-degree value, and then use that in the equation 1. While I'm not sure the paper needs to be changed, please address this in the response to the reviewer.

Thank you for this important and valuable suggestion. We agree with the reviewer's interpretation of Eq. (1). In the original formulation, the ratio between daily IMERG rainfall and monthly IMERG rainfall represents the fractional contribution of each day to the monthly total at the 0.1° scale. This IMERG-derived daily fraction is then applied to the corresponding CHIRPS monthly rainfall total at 0.05° resolution. Therefore, the resulting daily estimates are constrained to sum to the CHIRPS monthly total rather than the IMERG monthly total.

The purpose of this formulation is to combine the complementary strengths of the two products identified during the preliminary SPP evaluation. IMERG shows better performance at the daily temporal scale and is therefore used to represent intra-monthly rainfall variability. In contrast, CHIRPS shows better agreement with gauge observations at the monthly scale and provides rainfall estimates at finer 0.05° spatial resolution. Therefore, monthly CHIRPS is used to provide the high-resolution monthly rainfall magnitude, while IMERG provides the daily temporal distribution.

We acknowledge that the original notation $IMERG_{daily,0.05°}$ could incorrectly imply that the generated product is purely IMERG-derived. To avoid this ambiguity, we have revised the notation to IM_CH throughout the manuscript. The revised notation better reflects the hybrid nature of the product, in which IMERG contributes the daily temporal fraction and CHIRPS contributes the monthly rainfall magnitude. We have also clarified in the methodology that IM_CH should be interpreted as a CHIRPS-constrained, IMERG-temporally-disaggregated daily rainfall product, rather than a purely IMERG-derived 0.05° product.

We also appreciate the reviewer's suggested alternative formulation, in which the spatial fraction of each 0.05° CHIRPS pixel relative to the corresponding aggregated 0.1° CHIRPS grid cell is used to downscale IMERG monthly totals before applying the IMERG daily fractions. To assess this suggestion, we additionally tested this alternative formulation (CH_IM product). The resulting downscaled product produced broadly similar performance to the current **IM_CH** formulation.

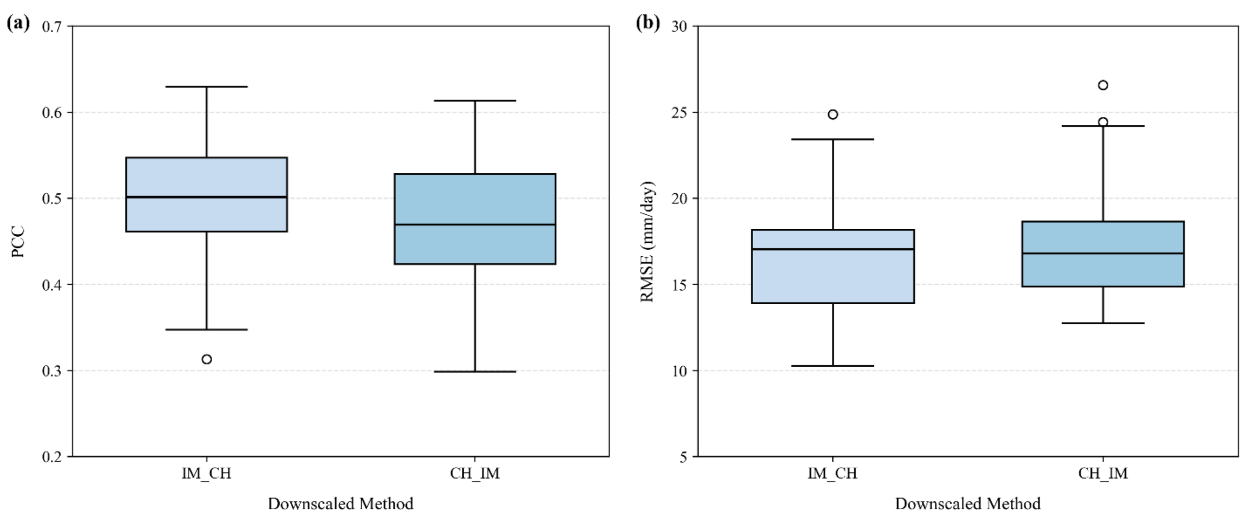

Figure 1. Variation of (a) PCC and (b) RMSE of IM_CH and CH_IM products

The current formulation is retained because its objective is to preserve the monthly rainfall magnitude from CHIRPS, which shows better agreement with gauge observations at the monthly scale, while using IMERG to represent daily rainfall variability. In contrast, the alternative formulation preserves the IMERG monthly rainfall magnitude and uses CHIRPS mainly to introduce finer-scale spatial variability. We have clarified this distinction in the revised manuscript.

Following part will be added to the Section 3.1.

The performance of the generated **IM_CH** product was compared with the original IMERG and CHIRPS products using gauge observations as the reference. As shown in Fig. 4, **IM_CH** produces the highest median PCC among the evaluated products, indicating improved representation of daily rainfall variability compared with the original IMERG and CHIRPS products. The RMSE distribution also indicates that **IM_CH** has the lowest median RMSE, although the improvement relative to IMERG is modest. These results suggest that the proposed downscaling formulation improves the agreement between satellite-based rainfall estimates and gauge observations by combining the stronger daily-scale temporal representation of IMERG with the more reliable monthly rainfall magnitude of CHIRPS.

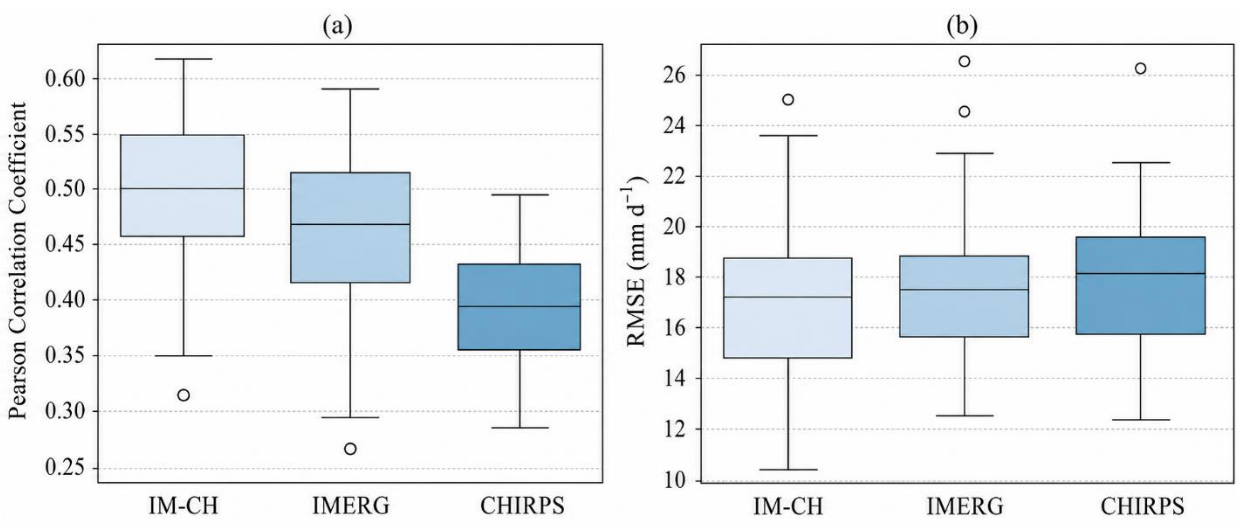

Figure 4. Box plots of spatial variability of (a) Pearson correlation coefficient and (b) root mean square error between IM-CH, IMERG and CHIRPS and observed rainfall

Comment 03

The other downside of the current formulation is that CHIRPS has station data already included in the product, especially before 2007 (if using CHIRPS version 3) or so. As a result you may be normalizing your rainfall estimate to a product that has station data included, and then using that as the "baseline" product in your analysis. Another alternative would be to use CHIRP (no "S") which has the same spatial resolution and a very similar mean, but is made only using satellite data and without the addition of stations. However, none of this

changes the subsequent analysis or methods, but it might mean that your "IMERG" baseline would have some different values in the skill assessments.

Thank you for this important observation. We agree that the use of CHIRPS introduces a degree of dependence on station-informed rainfall information, since CHIRPS incorporates gauge observations in addition to satellite and infrared-based estimates. Therefore, by constraining the generated IM_CH product to CHIRPS monthly totals, the resulting product inherits some characteristics of a gauge-adjusted precipitation dataset. We acknowledge that this should be considered when interpreting comparisons involving CHIRPS or the generated IM_CH product as baseline datasets.

We also appreciate the reviewer's suggestion to consider CHIRP instead of CHIRPS. CHIRP provides the same 0.05° spatial resolution but does not include gauge correction, and therefore it would provide a more satellite-only basis for the downscaling framework. In the present study, CHIRP was initially evaluated together with CHIRPS and IMERG at daily and monthly temporal scales. However, CHIRPS showed better agreement with gauge observations at the monthly scale, whereas CHIRP showed relatively weaker performance. Therefore, CHIRPS was selected as the monthly rainfall constraint in Eq. (1). We have clarified this rationale in the revised manuscript.

We acknowledge that using CHIRP instead of CHIRPS could change the values of the generated baseline product and the resulting skill metrics, although it would not alter the subsequent modelling framework or analysis procedure. This point has been added to the revised Discussion section as a limitation.

Following part will be added to Section **5.2 Limitations and uncertainties**

Additional uncertainty may also arise from the downscaling procedure used to generate the IM_CH product. In this study, IMERG is used to provide the intra-monthly daily rainfall distribution, while CHIRPS is used to constrain the monthly rainfall magnitude at 0.05° resolution. Although this formulation is designed to combine the better daily-scale performance of IMERG with the better monthly-scale performance of CHIRPS, the resulting product is not a purely IMERG-derived satellite product. Since CHIRPS incorporates gauge-adjusted information, the generated IM_CH product may inherit some characteristics of station-informed rainfall data. This should be considered when interpreting model performance and comparisons involving CHIRPS or IM_CH.

Comment 04

I really like the residual interpolation step that is included in GraphIDW. I wonder how much improved the satellite-based estimate would be with the residual IDW added at the end? Similarly, how would the other ML techniques fare with that extra "corrective" step? Given the slim differences in skill metrics across the different ML techniques, I could imagine the residual IDW improving some of the other approaches such that their skill metrics were better than GraphIDW. I would suggest exploring this, and potentially including it in the results.

Thank you for this thoughtful and constructive suggestion. We agree that applying the residual IDW correction only to GraphIDW may make it difficult to determine whether the improved performance is primarily due to the graph-based learning component or the residual interpolation step. We also agree that applying the same corrective step to the other ML models provides a more direct and consistent comparison.

In response, we have revised the analysis by applying the residual IDW correction procedure to all benchmark ML models. For each model, residuals are calculated at the training gauges as the difference between observed rainfall and model-predicted rainfall. These residuals are then spatially interpolated using IDW and added back to the original model estimates to generate residual-corrected precipitation estimates. This procedure allows the contribution of the residual correction step to be evaluated consistently across GraphIDW and the traditional ML models.

The revised results figures and performance comparisons now include the residual-corrected versions of the evaluated models. This provides a more balanced "apple-to-apple" comparison and clarifies whether the residual IDW correction improves the other ML approaches to a level comparable with, or better than, GraphIDW. The Results and Discussion sections have also been updated to describe the effect of the corrective step on model performance and to interpret the remaining differences among the models after residual correction.

The updated figures are shown in response to Comment 01.

Comment 05

Conversely, I think you could look at the estimate from just using the GNN technique, without the IDW, to look at the improvement over the IMERG alone. Then, compare that

with the final GraphIDW output, and it would give you an estimate of the improvement of each of these components to the overall estimate. If you wanted, that GNN-only estimate could be compared to the other ML techniques to get an idea for how the IDW might help other estimates as well.

We appreciate this valuable recommendation, as it helps clarify the individual contributions of the GNN component and the residual IDW correction step. We agree that evaluating the contribution of the residual IDW correction separately is important for determining whether the final GraphIDW performance is mainly due to the graph-based learning component, the residual correction step, or their combination.

In response, we have added an ablation analysis comparing three outputs: the original **IM_CH** product, the **GNN-only** estimate without residual IDW correction, and the final **GNN_IDW** output after residual correction. This allows the contribution of the GNN component and the added value of the residual IDW correction to be assessed separately.

A new subtopic will be added to the Discussion Section.

Added value of GNN learning and residual IDW correction

An ablation analysis was conducted to quantify the individual contribution of the graph-based learning component and the residual IDW correction step. The original IM_CH product, the GNN-only estimate, and the final GNN_IDW output were compared using MAE, RMSE, and PCC. The results show a progressive improvement from IM_CH to GNN and then from GNN to GNN_IDW.

For MAE, the median value decreases from 9.2 mm d^{-1} for IM_CH to 7.4 mm d^{-1} for GNN, representing a 19.6% reduction. After applying residual IDW correction, the median MAE further decreases to 6.8 mm d^{-1}, corresponding to an additional 8.1% reduction relative to GNN-only and an overall 26.1% reduction relative to IM_CH. Similarly, the median RMSE decreases from 17.8 mm d^{-1} for IM_CH to 14.9 mm d^{-1} for GNN, indicating a 16.3% reduction. The final GNN_IDW output further reduces the median RMSE to 13.1 mm d^{-1}, giving an additional 12.1% reduction relative to GNN-only and an overall 26.4% reduction relative to IM_CH.

The PCC results also indicate improved agreement with gauge observations. The median PCC increases from 0.50 for IM_CH to 0.64 for GNN, representing a 28.0% increase. After residual

IDW correction, the median PCC further increases to 0.72, corresponding to an additional 12.5% increase relative to GNN-only and an overall 44.0% increase relative to IM_CH.

These results indicate that the GNN component substantially improves the original satellite-based estimate by learning spatial relationships among neighbouring nodes. The residual IDW correction provides additional improvement by correcting spatially structured errors that remain after GNN prediction. Therefore, the final GNN_IDW output benefits from both graph-based spatial learning and local residual adjustment, confirming the added value of integrating residual IDW correction into the GraphIDW framework.

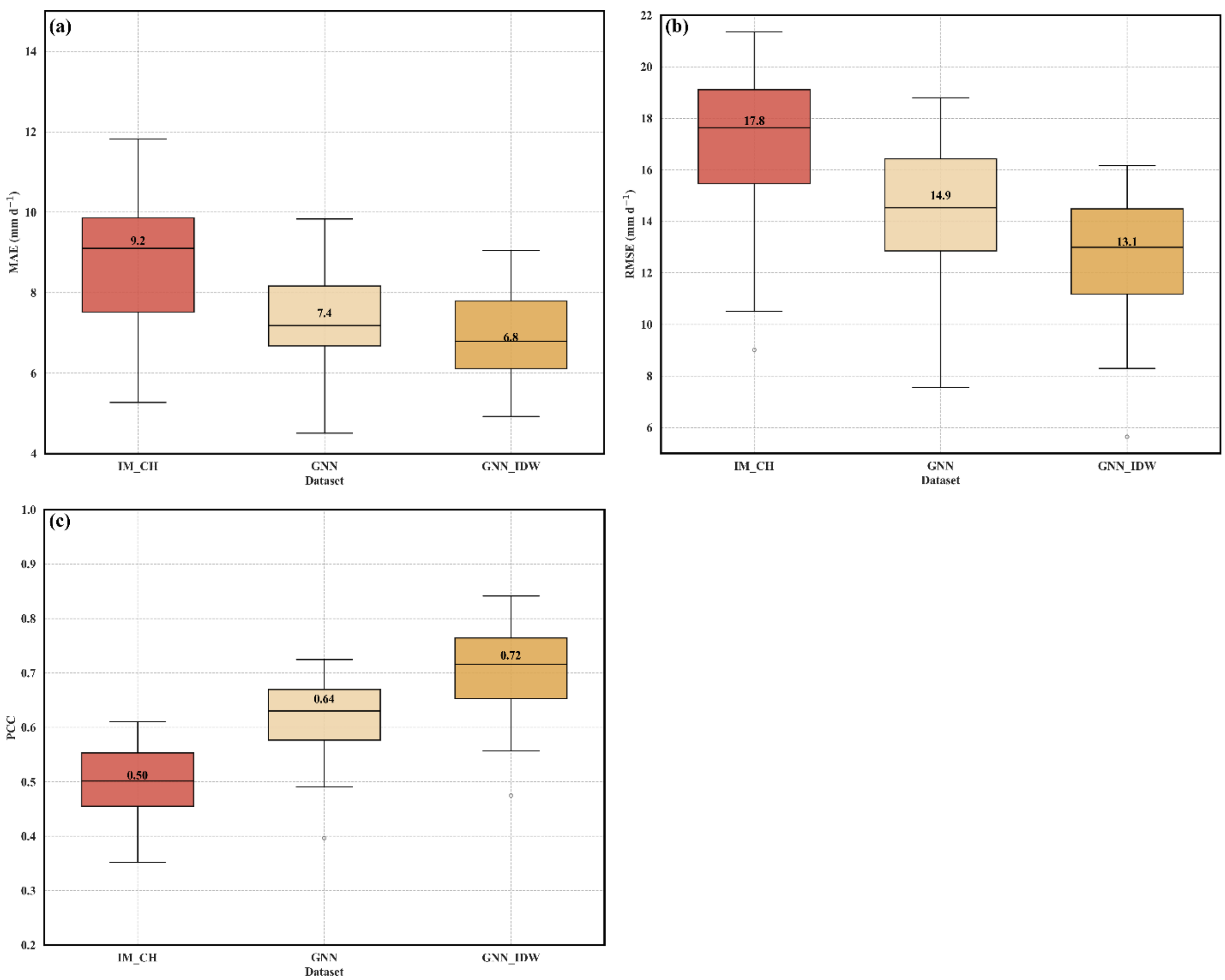

Figure. Boxplots of three quantitative metrics (MAE (a), RMSE (b), and PCC (c)) for GNN, GNN_IDW and original (IM_CH) products

Comment 06

This study is using a relatively large number of stations compared to some regions that are more reliant on satellite rainfall estimates. That is very useful for producing such an accurate result, but it may be that the results are different in a significantly more data sparse region. You touch on this a little bit in the discussion, and I think it would be interesting to see how few stations are needed before there is a notable drop in estimate performance.

Thank you for this insightful comment. We agree that the relatively dense station network available in the present study region contributes to the strong model performance and that the results may differ in more data-sparse regions. The use of a dense gauge network in this study is intentional because the Wet Zone of Sri Lanka is characterized by strong spatial heterogeneity in rainfall, driven by complex topography, monsoonal circulation, and localized convective processes. As shown in the revised **Figures 11 and 12**, annual average rainfall and model error patterns vary considerably across the study region, highlighting the need for sufficient gauge coverage to represent these spatial gradients.

To address the reviewer's comment, we have expanded the discussion on station-density sensitivity and the transferability of the proposed framework.

The revised Figs. 11 and 12 are shown here.

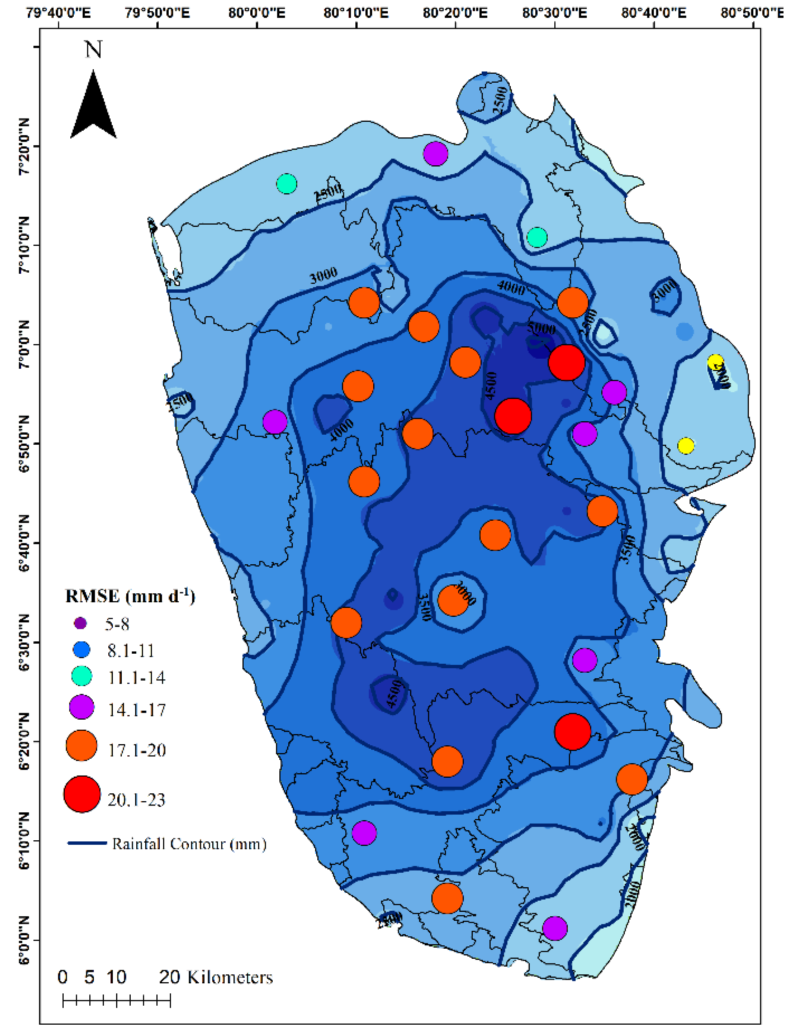

Figure 11. Spatial distribution of RMSE for the original IM_CH product overlaid on the IDW-interpolated observed annual average rainfall across the Wet Zone of Sri Lanka.

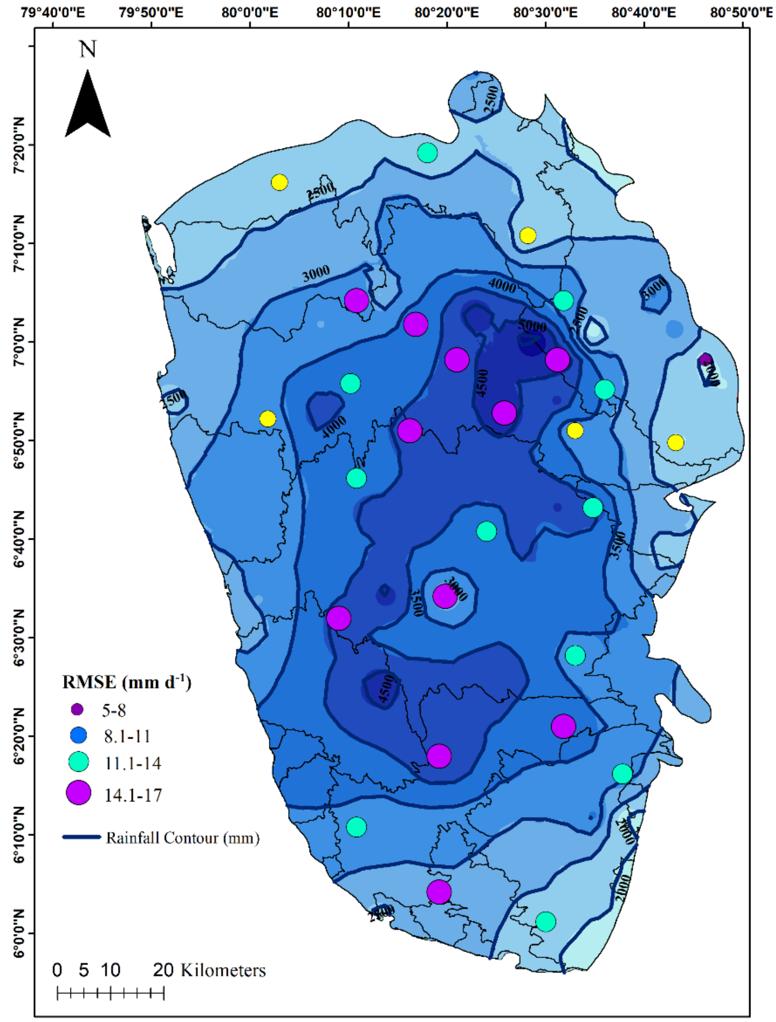

Figure 12. Spatial distribution of RMSE for GraphIDW-predicted rainfall overlaid on the IDW-interpolated observed annual average rainfall across the Wet Zone of Sri Lanka.

Section **5.1 Impact of rain gauge density,** Lines 580-584 will be updated as follows:

The density of rain gauges plays an important role in the performance of ML-based merged precipitation products, as model training and residual correction rely directly on gauge observations. In the present study, the use of a relatively dense gauge network is particularly important because the Wet Zone of Sri Lanka exhibits strong spatial rainfall heterogeneity, as illustrated by the spatial variability of annual average rainfall and model error patterns in Figs. 11 and 12. Therefore, sufficient gauge coverage is necessary to capture the spatial rainfall gradients across the region and to provide reliable information for model learning.

To examine the sensitivity of model performance to gauge availability, the models are trained using three subsets of the gauge network, corresponding to 40%, 50%, and 70% of the total stations. Figure 14 presents boxplots of RMSE values for the ML-based merged precipitation products at the test gauges under different training sample sizes. For comparison, the RMSE distribution of the original IM_CH product, which is independent of training sample size, is also shown.

.