Reply to RC3 - Anonymous Referee #3

General Comments:

This paper presented a multi-sensor quantitative precipitation estimation (QPE) product intended for operational implementation in Australia. It incorporates a geostationary satellite QPE from Himawari (10-min, 2-km), radar precipitation rates (5-min, 1-km), and gauge observations (15/30-min), aggregates the data to the hourly accumulations separately, and then merges them into a 1-hr, 2-km national QPE product via a statistical interpolation scheme. The satellite QPE, given its full coverage for the continent, served as a background field and the radar and gauge data are treated as observations in the interpolation scheme. To assure computational efficiency for operational applications, the authors limited 1 nearest radar data and 20 nearest gauges for a given grid point and developed analytical functions for the various error characteristics. The new product was evaluated on the 1- and 24h scales using 2-years (2022-2023) data, and the impact of each input data source was analyzed. The new QPE showed an overall improvement over the existing operational products in resolution, accuracy, and/or coverage.

Overall the paper is well-written, and the data, methodology and results were presented clearly. However, the interpretation of the results needs some clarifications. I have two major comments: 1) There is a lack of description of the "IDW" hourly gauge-based QPE, even though it's a key reference; 2) The interpretation of the hourly and daily validation statistics needs clarification and some artifacts in the product should be discussed (please see attached document). I'd recommend a major revision of the paper before it can be accepted for publication.

Response: Thank you for your thorough and constructive review, and for your positive assessment of our work. We appreciate your recognition that the manuscript is well written and that the data, methodology, and results are presented clearly.

We also agree that the two major issues raised require further clarification. To address these comments, we plan to make the following revisions in the manuscript:

- Add a clearer description of the IDW hourly gauge-based QPE, including its implementation details, parameter settings, and relevant references. We will also clarify that IDW is used as a simple gauge-only reference rather than as the primary benchmark for evaluating BRAIN.
- Revise the interpretation of the hourly and daily validation statistics to make the comparison clearer and avoid overstatement.
- Add further discussion of potential artefacts in the blended product, including their possible causes and implications for interpretation and future operational applications.
- Address the detailed points raised in the following specific comments throughout the revised manuscript.

Specific Comments:

Line 47-61: There is a high-resolution multi-sensor QPE in the US Multi-Radar Multi-Sensor (MRMS) system that has been operational on the national scale for years, although not mentioned in the discussion:

> Martinaitis et al. 2020: A Physically-Based Multi-Sensor Quantitative Precipitation Estimation Approach forGap-filling radar coverage. J. Hydromet., 21, 1485-1511. https://doi.org/10.1175/JHM-D-19-0264.1

Response: Thank you for pointing out this important reference. We agree that the MRMS multisensor QPE (MSQPE) system represents an important example of an operational national-scale multisensor precipitation analysis framework. We will add this important reference in our introduction section.

Line 285: "Only the nearest radar grid cell to each target grid cell...": What is the distance limit for the nearest radar grid cell?

Response: Thank you for this helpful comment. We agree that the search range for selecting the nearest radar grid cell should be stated explicitly. In our implementation, the maximum search distance is 200 km. If no radar grid cell is available within this distance, no radar observation is used in the interpolation for that target grid cell.

The 200 km distance was selected to balance the use of nearby radar information with the need to avoid introducing weakly related distant radar observations. Although 200 km is used as the maximum search distance, the actual influence of a radar observation is controlled by the distance-dependent error correlation function. As shown in the manuscript (Fig. 4b), the error correlation decreases rapidly with distance, so radar observations located far from the target grid cell receive very small weights and have limited impact on the final analysis.

To address this comment, we plan to make the following revisions in the manuscript:

- Explicitly state the 200 km maximum search distance used to identify the nearest radar grid cell.
- Clarify that radar observations are only incorporated when a radar grid cell is available within this search distance.
- Add a brief justification for this choice, explaining that the 200 km distance is a maximum search radius rather than an effective influence distance, and that the actual contribution of radar observations decreases with distance according to the error correlation function.

Line 318 - 324: The "mean bias correction" in the title of this section is confusing. The description (line319-324) indicates that the radar (satellite) vs. gauge biases are estimated at each gauge location, and then the biases at gauge locations are interpolated to the radar/(satellite) grid. The interpolated bias is then applied to the original radar/satellite QPE pixel-by-pixel. To me, this is a localized bias correction instead of a mean bias correction because the latter implies a domain-wide bias correction (Chumchean etal. 2006) rather than pixel-by-pixel bias correction. I think the local bias correction is a better approach than the mean bias correction, but the description should be clarified.

Response: Thank you for this valuable comment. We agree that the term "local bias correction" more accurately describes the methodology implemented in this study. Although the general bias correction framework follows Chumchean et al. (2006), our implementation estimates the bias at gauge locations, interpolates the bias field to the radar or satellite grids, and then applies the correction locally to each grid cell. Therefore, it is more appropriately described as a local bias correction rather than a domain-wide mean bias correction.

To address this comment, we plan to make the following revisions in the manuscript:

- Revise the section title by replacing "mean bias correction" with "local bias correction".
- Revise the related text to clarify that the bias is estimated at gauge locations, interpolated to the radar/satellite grids, and then applied locally to each grid cell.
- Clarify the distinction between our implementation and the domain-wide mean bias correction described by Chumchean et al. (2006).

Line 326-328: 150-km (R0) vs. 300-km (R1) radius: Is there a limit on the number of gauges used for the bias estimate with such a large radius of influence? With a 300 km radius, there must be a huge number of gauges in some well populated areas (e.g., S.E. Coast).

Response: Thank you for this insightful question.

There is no predefined limit on the number of gauges used for the bias estimation within the search radius. The purpose of increasing the search radius from 150 km (R0) to 300 km (R1) is to increase the number of gauges available for estimating the bias, particularly in regions between 150 km and 300 km.

It is also important to note that two modifications were introduced in R1 compared with R0. First, the temporal window was extended from 15 minutes to 1 hour, allowing additional gauge reports with longer reporting latencies to be included. Second, the spatial search radius was increased from 150 km to 300 km, enabling additional gauges to contribute to the bias estimation, particularly between 150 km and 300 km from the radar. Together, these two modifications increase the number of available observations and improve the robustness of the bias estimation.

The estimated point-scale bias is subsequently interpolated to the radar domain using ordinary kriging. During this process, the spatial distribution of the gauges is naturally accounted for through the kriging interpolation. We acknowledge that gauge density may still influence the quality of the bias estimates, and this is an inherent characteristic of gauge-based bias correction methods.

To address this comment, we plan to make the following revisions in the manuscript:

- Clarify that no predefined limit is imposed on the number of gauges used for the bias estimation.
- Expand the description of the differences between R0 and R1, including both the larger spatial search radius and the longer temporal window.
- Add a brief explanation that the point-scale bias is interpolated using ordinary kriging, which accounts for the spatial distribution of the available gauges.

Line 352: "… divided into tropical, subtropical, and temperate subregions": it would be helpful to the readers to mark these regions in Fig. 1.

Response: Thank you for this very useful suggestion. We agree that marking these subregions would improve the readability of the manuscript and help readers better interpret the regional evaluation results.

To address this comment, we plan to make the following revisions in the manuscript:

- Mark the tropical, subtropical, and temperate subregions in Figure 1 or Figure 3.
- Add a brief description of these subregions in the figure caption and the main text where they are first introduced.

Line 380: "IDW-based interpolation": since this is the key reference QPE for the evaluation of the new product, it is important to provide a clear description. For instance, what is the weighting function (e.g.,1/d? 1/d2), the radius of influence (e.g., 200km, 300km) and the number of gauges threshold (e.g, 5, 20)? It mentioned in the supplemental material that the IDW only uses the nearest 5 gauges. That information should be included in the paper.

Response: Thank you for this helpful suggestion. We agree that the description of the IDW interpolation is insufficient in the current manuscript, particularly given that it serves as a key reference in the evaluation of the proposed framework.

To address this comment, we plan to make the following revisions in the manuscript:

- Add a dedicated description of the IDW interpolation method in the main text.
- Describe the implementation details, including the weighting function, the search radius, and the maximum number of gauges used for the interpolation.

Line 386: "... we primarily used correlation and root mean square error...": I'm surprised that bias is not stated here since it's an important metric not represented by CC and only partially by RMSE. The authors mentioned "bias" later in line 391, but I'd recommend including it in line 386 to be clear.

Response: Thank you for this helpful comment and recommendation. We agree that bias is an important evaluation metric and should be introduced together with correlation and RMSE in the description of the validation methodology.

To address this comment, we plan to make the following revisions in the manuscript:

- Revise the text to explicitly include bias, together with correlation and RMSE, when introducing the evaluation metrics.
- Clarify that bias is one of the primary metrics used to comprehensively assess the performance of the precipitation products.

Line 394: “...using nearest interpolation”: Please elaborate. For instance, are you taking the four nearest grid points around the gauge and applying a bi-linear interpolation?

Response: Thank you for this helpful comment. We agree that the interpolation method should be described more clearly.

The interpolation method used in this study is nearest-neighbour interpolation rather than bilinear interpolation. This approach was adopted because precipitation often exhibits strong spatial variability and intermittency, and nearest-neighbour interpolation avoids introducing the artificial smoothing that may occur with bilinear interpolation.

To address this comment, we plan to make the following revisions in the manuscript:

- Explicitly state that nearest-neighbour interpolation was used to establish the spatial correspondence between gridded precipitation products and gauge locations.
- Clarify that the precipitation value from the single grid cell with the minimum Euclidean distance to each gauge location was extracted for the comparison.
- Briefly explain the rationale for adopting nearest-neighbour interpolation instead of bilinear interpolation.

Line 398-411: All these arguments support the need for the new operational product for Australia, especially in terms of resolution and latency. It would still be scientifically valuable to compare the new product to the global products to demonstrate improved accuracy with local gauge and radar data. But I understand that may be out of the scope of this paper.

Response: Thank you for this valuable comment and understanding. We agree that comparisons with existing global precipitation products would be valuable and could further demonstrate the benefits of incorporating local gauge and radar observations. However, we feel such comparisons are beyond the scope of the current study, which focuses on the development and evaluation of the proposed framework.

Our plan is to conduct these comparisons once the long-term operational BRAIN dataset has been developed. At that stage, more systematic and comprehensive evaluations against existing global precipitation products will be possible, providing a more robust assessment of the advantages of the proposed framework over a wider range of climatic conditions and rainfall events.

To address this comment, we plan to make the following revisions in the manuscript:

- Clarify the scope of the current study and explain why comparisons with global precipitation products are not included.
- Add a brief discussion highlighting the value of such comparisons as part of future work.
- State that, once the long-term operational BRAIN dataset has been developed, we plan to conduct more systematic and comprehensive comparisons with existing global precipitation products.

Line 419-420: "... the error distribution (Fig.3f ...) becomes narrower and more symmetric": This improvement is pretty significant. Could this be due to the local- (instead of mean-) bias correction in S1? Was the bias correction in S0 the domain-wide mean bias correction? Also see comment about Line318-324.

Response: Thank you for this insightful comment. We agree that the substantial improvement in the error distribution is primarily due to the introduction of the local bias correction in S1.

To clarify, S0 does not apply any bias correction because of its initial implementation. It directly uses the original CRRPH satellite precipitation product from Himawari satellite. In contrast, S1 introduces the local bias correction described in Section 3.2, which estimates the bias at gauge locations, interpolates the bias field to the satellite grid, and applies the correction locally to each grid cell. This substantially reduces the systematic errors in the satellite estimates, leading to the narrower and more symmetric error distribution shown in Figure 3.

To address this comment, we plan to make the following revisions in the manuscript:

- Clarify that S0 uses the original satellite product without any bias correction, whereas S1 introduces the proposed local bias correction.
- Revise the bias correction method to consistently describe the methodology as a local bias correction rather than a mean bias correction.
- Revise the discussion of Figure 3 to explain that the improved error distribution is primarily attributable to the local bias correction implemented in S1.

Line 423-424: "... the error distribution (Fig. 3h...) becomes more symmetric and centered around zero". Same question as above and see comment for Line 318-324.

Response: Thank you for pointing this out again. We agree that the further improvement in the error distribution from R0 to R1 is primarily due to the local bias correction. Please also see our responses to comment for Line 318-324.

To clarify, R1 extends the local bias correction implemented in R0 by incorporating a larger amount of gauge information through both a broader spatial search radius (300 km instead of 150 km) and a longer temporal window (1 hour instead of 15 minutes). These modifications increase the number of gauges available for bias estimation, leading to a more robust bias field and, consequently, a more symmetric error distribution centred around zero.

To address this comment, we plan to make the following revisions in the manuscript:

- Clarify that the improved error distribution from R0 to R1 is primarily attributable to the local bias correction with more available gauges.
- Revise the discussion of Figure 3 to explain how the larger spatial window and longer temporal window contribute to the improved bias estimation.

Line 491-492: "This behavior is expected, as R1 is calibrated using more available gauges, which favours ...". This statement is confusing as it implies R1 uses more gauges than BR-SRG. But shouldn't both products be using the same gauges? Please clarify. It seems to me that the difference between R1 and BR-SRG reflected the advantages of radar observations to better capture the intensities and spatial variations of small-scale precipitation than satellite and gauge. Therefore, this result may indicate the need to further increase the radar QPE's contributions to the BR-SRG within the radar coverage.

Response: Thank you for this insightful comment. We apologise for the confusing description. The apparent inconsistency arises because two separate processes are involved in the framework: bias correction and blending.

For the local bias correction, all available gauge observations were used to estimate the radar and satellite bias fields. Owing to the substantial computational cost, cross-validation was not applied during this step. In contrast, the blending stage was evaluated using a 20-fold cross-validation framework, where a subset of gauges was withheld from the blending and used only for independent evaluation.

Consequently, some gauges used for evaluating BR-SRG were included in the bias correction of R1 but were withheld from the blending process. At these locations, R1 may therefore exhibit slightly better agreement with the gauge observations than BR-SRG.

To address this comment, we plan to make the following revisions in the manuscript:

- Explicitly state that cross-validation was applied only during the blending stage, whereas all available gauges were used for the bias correction.
- Revise the discussion of the comparison between R1 and BR-SRG to avoid the misleading implication.

Line 496-498: "Outside radar coverage...". It appears that the BR-SRG had lower RMSE than BR-SG in both hourly and daily evaluations. Does this improvement come from the radar influence? Does this mean that the horizontal extrapolation from radar at a distance can provide better precipitation information than the local satellite? Please elaborate.

Response: Thank you for this insightful question.

You are correct that the improvement of BR-SRG over BR-SG outside the radar coverage is due to the influence of radar information through the statistical interpolation framework. However, this does not imply that the extrapolated radar information alone provides better precipitation estimates than the local satellite observations. Rather, the improvement arises from the weighted combination of the local satellite background and the neighbouring radar observations through the optimal interpolation framework. The relative contribution of each source is determined by the estimated error characteristics and their spatial correlations. As the distance from the radar increases, the influence of the radar observations gradually decreases according to the error correlation function, while the satellite background becomes increasingly dominant.

To address this comment, we plan to make the following revisions in the manuscript:

- Clarify that the improved performance outside radar coverage results from the weighted blending of satellite and radar information, rather than from radar extrapolation alone.
- Expand the discussion of how the contribution of radar observations gradually decreases with distance according to the error correlation model.
- Revise the text to avoid implying that extrapolated radar information is inherently more accurate than local satellite observations outside the radar coverage.

Line 500-510: The discussion of the BR-SRG vs. IDW performance outside the radar coverage was quite confusing to me, partially due to the lack of info about the IDW product (see comment on Line 380). The authors attributed the worse RMSE in BR-SRG to 1) BR-SRG uses a smaller radius of influence than IDW; 2) lack of sub-daily gauges; 3) sub-daily errors in IDW canceled each other when aggregated to daily; 4) very small validation samples. Below are my questions regarding the contributing factors 1) – 3):

1) Does BR-SRG use a 200 km radius of influence while IDW uses an infinite? Please clarify. If this argument is valid, then it implies that the gauge observations 200+ km away can provide better precipitation estimates than the local satellite pixel. Is that the case sometimes?
2) Same as 1)
3) Would this same argument work for satellite QPE? If not, please explain why.

Response: Thank you for this thoughtful comment. We are sorry that our discussion in this section was confusing and appreciate you pointing this out. We agree that the interpretation of the differences between BR-SRG and IDW outside the radar coverage requires further clarification.

To address this comment, we plan to make the following revisions in the manuscript:

- Add a clearer description of the IDW implementation in the main text, including its interpolation settings and assumptions.
- Add representative precipitation time series at selected locations to illustrate the behaviour of the different products during individual events. These examples will help readers better understand the strengths and limitations of each product and provide a more intuitive explanation of the differences observed in the statistical evaluation.
- Revise the discussion of the comparison between BR-SRG and IDW to provide a clearer interpretation of the evaluation results and avoid over-speculation regarding the contributing factors.

Line 505: “(see Fig. S3 and Text S2b)”. I couldn’t seem to find Text S2b in the supplemental material. Did you mean “Text S2.2”?

Response: Thank you for pointing this out. This was a typo. We will correct the reference from "Text S2b" to "Text S2.2" in the revised manuscript.

Line 509: “(see Fig. 1a, Fig. S2, and Text S2a)”. I couldn’t find Text S2a.

Response: Thank you for pointing this out again. We will correct the reference from "Text S2b" to "Text S2.2" in the revised manuscript.

Line 516: “It performs best within radar coverage” - please see comment on Line 491-492.

Response: Thank you for this comment. Please see our detailed response to the comment on Lines 491–492. We will revise the manuscript accordingly to clarify the interpretation and avoid overstatement.

Line 535: “...framework over short distances” - what are the short distances? Also see comment on Line285.

Response: Thank you for this question. We agree that the term "short distances" is ambiguous and should be described more clearly.

Similar to our response to the comment on Line 285, the 200 km value represents the maximum search radius used to identify candidate radar observations rather than the effective influence distance. In the optimal interpolation framework, the influence of each observation is determined by the error correlation function, which decreases gradually with distance. Consequently, nearby radar observations have relatively high error correlations and therefore exert a stronger influence on the analysis, whereas the influence decreases progressively with increasing distance and becomes negligible at larger distances.

To address this comment, we plan to make the following revisions in the manuscript:

- Revise the description of "short distances" to explain that the effective influence is controlled by the distance-dependent error correlation and gradually decreases with increasing distance.
- Clarify that the 200 km value refers to the maximum search radius rather than the effective influence distance.
- Add a cross-reference to the error correlation analysis (Figure 4) to illustrate how the radar influence diminishes with distance.

Fig. 7a: There appear to be some arc-shaped discontinuities in the BR-SRG product (see a copy of Fig. 7a below with the artifacts marked by red lines). What could be causing those?

Response: Thank you for pointing this out. We agree that the arc-shaped discontinuities shown in Figure 7a are artefacts in the blended product and should be discussed more explicitly.

These discontinuities are likely caused by the sharp contrast between the satellite and radar precipitation estimates. Within radar-covered areas, the blended product relies more strongly on radar information, whereas outside the radar coverage it relies primarily on the satellite background field. When the radar and satellite estimates differ substantially near the radar boundary, discontinuities can appear in the blended field.

To address this comment, we plan to make the following revisions in the manuscript:

- Add a discussion of these arc-shaped artefacts in the interpretation of Figure 7.
- Explain that they are likely associated with differences between radar-dominated and satellite-dominated regions near radar coverage boundaries.
- Discuss this as a limitation of the current blending framework under cases where the satellite and radar estimates differ markedly, and identify smoother transitions across radar coverage boundaries as an area for future improvement.