Response letter

Title: Design and trial implementation of a continental-scale, kilometre-resolution hourly precipitation analysis for Australia using satellite, radar and gauges

MS No.: egusphere-2026-666

Dear Editor and Reviewers,

Thank you for your careful evaluation and constructive comments on our manuscript. We have carefully considered all comments and revised the manuscript accordingly. Our point-by-point responses are provided below, with the reviewer comments shown in **black**, our responses in **blue**, and the corresponding revisions to the manuscript in **purple** for clarity and ease of reference.

Reply to Editor

Dear Authors,

This study presents the design and trial implementation of BRAIN (blended rainfall), a continental-scale, kilometre-resolution hourly precipitation analysis for Australia. An existing method "Optimal Interpolation" was used to blend gauged precipitation, a satellite precipitation estimates and precipitation estimates from radar. The topic is of international interest, and the paper is well structured, clearly written, and easy to follow.

Three reviewers evaluated the manuscript, and raised several major comments concerning mainly: (i) the originality and novelty of the approach which needs to be justified; (ii) the methodology used which can be improved; (iii) the interpretation of the results which needs to be clarified and extended. The authors provided detailed responses to all the points raised by reviewers.

As the topic can be can be of great interest for the hydrologic community, I would like to give a chance for the Authors to reinforce the paper. I recommend "Major revision", and ask the Authors to improve drastically the manuscript responding to all points raised by reviewers.

Kind regards,

Roger Moussa

Response: We sincerely thank you for your efficient coordination of the review process, your positive assessment of our study, and the opportunity to revise and strengthen the manuscript. We also greatly appreciate the detailed and constructive comments provided by the three reviewers. In response to the major issues raised, we have substantially revised the manuscript, particularly in the following areas:

- **Originality and novelty**: We have further clarified the methodological contributions of BRAIN and its distinction from previous studies, including a dedicated subsection (Section 5.1) discussing the key methodological advances and their significance.
- **Methodology**: We have strengthened the methodological framework and improved the description of the bias-correction procedures, the configuration and role of IDW, and the treatment of different data sources. We have also expanded the gauge-based interpolation comparisons and conducted additional sensitivity analyses to examine different methodological configurations.

- **Evaluation and interpretation**: We have substantially expanded the evaluation, including additional analyses across rainfall intensities and gauge-network densities, sensitivity analyses of different IDW configurations, and comparisons with an additional gauge-based interpolation method. We have also repositioned the case events as illustrative examples and added time-series analyses at selected key locations during extreme rainfall events. Based on these extended analyses, we have further strengthened the interpretation and discussion of the results and clarified several important issues raised by the reviewers.

Detailed point-by-point responses to all reviewer comments, together with the corresponding revisions made in the manuscript, are provided below. We sincerely appreciate the time and effort devoted by you and the reviewers to evaluating our work, and we hope that the revised manuscript has adequately addressed the concerns raised and has been substantially strengthened.

Reply on RC1 - Anonymous Referee #1

General Comments:

This manuscript describes the development of a new operational hourly, km scale precipitation analysis for the entire Australian continent, titled 'BRAIN'. The authors adapt an existing method - Optimal Interpolation - to blend gauged precipitation, a satellite precipitation estimate and precipitation estimates from radar. The use of a robust existing method in OI is in my view a sensible choice for an operational system, and the authors make well-judged simplifications to enable its deployment at continental scale (e.g. the use of single radar grid cells to estimate error characteristics). The authors use stringent verification measures, including strict cross-validation and comparisons against a range of existing products that are in wide use in Australia. These are high hurdles to clear, yet the authors are able to demonstrate that BRAIN clearly outperforms existing alternatives. In addition, the manuscript is concise and clear, and in general a pleasure to read. No continentally consistent precipitation product at this temporal or spatial resolution currently exists for Australia, and when BRAIN is operationalised it promises to be a game-changing innovation for a wide range of scientific, industrial and environmental applications. I essentially recommend that the manuscript be published as is, subject to technical corrections.

Response: Thank you for your thorough and highly positive evaluation of our manuscript. We greatly appreciate your recognition of the significance of this work and are encouraged by your assessment that BRAIN has the potential to become a game-changing precipitation product for Australia. We also appreciate your positive comments on our methodological choices, the evaluation strategy, and the potential value of the product for a wide range of scientific, operational, and environmental applications.

I leave it to the authors to determine which of my recommendations they wish to act on, of which I will note only one here: I thought too much emphasis was put on the performance of BRAIN shown in Figure 7, which shows only two isolated events. I prefer the authors to refer to their robust assessments of performance in Figures 5, 6 and 8 (and in the very useful Supplementary materials) than on Figure 7.

I congratulate the authors on what I believe will be a highly impactful product.

Response: Thank you for this constructive suggestion and for your positive assessment of the potential impact of BRAIN. We agree that the overall performance of BRAIN should be supported primarily by the more comprehensive evaluations rather than by the two individual events shown in Figure 7. Accordingly, we have reduced the emphasis placed on these event-based results and renamed the corresponding section as "Section 4.5, Demonstration of BRAIN in selected rainfall events", to clarify that these events are intended to illustrate the spatial behaviour and characteristics of BRAIN rather than provide the primary evidence of its overall performance.

Specific Comments:

R1-SC1

Introduction is lucid, economical and comprehensive. It clearly outlines 1) the promise of this product, 2) the challenges in establishing a sub-daily rainfall product in Australia 3) the range of methods available 4) a clear justification for the authors' use of optimal interpolation for their BRAIN product and 5) how the authors will adapt OI for their product.

Response: Thank you for your positive assessment of the Introduction section.

R1-SC2

L242 "the "true value"" suggest "the "true value", T,"

Response: Thank you for this suggestion. We have revised the sentence as below.

The analysis error at the target gird cell k, denoted by E_k, is then determined as the difference between the theoretical "true value", T_k and the analysis value, A_k:

$$E_k^2 = (T_k - A_k)^2 = [T_k - B_k - \sum_{i=1}^{n} w_i \cdot (O_i - B_i)]^2 \quad (2)$$

Here, T_k denotes the unknown theoretical true value at the target grid cell k, which cannot be directly observed in practice and is introduced for error definition and derivation purposes.

R1-SC3

L243 "the "true value"" - it's not clear from the equations below or the prose how the true values (T_k or T_i) are defined; accordingly the equations for errors are not explicit. Please describe how the 'true value' is estimated or defined (in prose is fine). I appreciate that OI methods are well known, but for those of us (like me) who are not very familiar with them, it would be nice to have this spelled out.

Response: Thank you for pointing this out. The T_k or T_i represents the theoretical true values at the target grid cell k or observation location i. These true values are not directly known in practice. They are introduced only as conceptual variables for deriving the optimal interpolation equations.

To improve clarity, we have revised the manuscript to explicitly state this definition.

The analysis error at the target gird cell k, denoted by E_k, is then determined as the difference between the theoretical "true value", T_k and the analysis value, A_k:

$$E_k^2 = (T_k - A_k)^2 = [T_k - B_k - \sum_{i=1}^{n} w_i \cdot (O_i - B_i)]^2 \quad (2)$$

Here, T_k denotes the unknown theoretical true value at the target grid cell k, which cannot be directly observed in practice and is introduced for error definition and derivation purposes.

R1-SC4

L251 "in normalised form" should this be "in standardised form"? Normalisation can imply a normalising transformation.

Response: Thank you very much for pointing this out. We agree that the term “normalised form” may imply a normalisation transformation, which is not the intended meaning here. In this context, "standardised form" is more appropriate. The manuscript has been revised accordingly.

By expanding and expressing the Eq. 4 in standardised form:

$$\sum_{j=1}^{n}\left(\mu_{ij}+\tau_{ij}\sigma_j+\tau_{ji}\sigma_i+\rho_{ij}\sigma_i\sigma_j\right)w_i'=\mu_{ki}+\tau_{ki}\sigma_i,\ i=1,2,\ldots,n \quad (5)$$

R1-SC5

L255-267 "This system of linear equations is solved to determine the optimal weights w_i, which are used to obtain the best precipitation estimate at each target grid cell by optimally combining satellite, radar, and gauge data." The equations above use a single term (O) to describe observations from gauges or radar, and weights vary with i (location), not data source. What happens when both radar and gauges are available? I assume these have quite different error characeristics, so I couldn't follow how they can be all thrown in together. Oh - never mind, this is explained below (Lines 271-317). It might be good to either avoid the statement "are used to obtain the best precipitation estimate at each target grid cell by optimally combining satellite, radar, and gauge data" at this point, or foreshadow the coming explanation of how you actually combine different observational data sources below, or both. The authors can add a version of this statement at the end of Section 3.2 if they wish.

Response: We thank you for this helpful comment. To improve clarity and avoid confusion, we revised the text to describe the estimation process in a more general way and added a forward reference indicating that the treatment of different observation sources (satellite, radar, and gauge) and their associated error characteristics is introduced in the following section.

This system of linear equations is solved to determine the optimal weights w_i, which are used to obtain the best precipitation estimate at each target grid cell by combining information from the background field and available observations (e.g., radar or gauge). The treatment of different observation sources is described below, while the corresponding source-specific error characteristics are introduced in Section 3.4.

R1-SC6

L272 "while background-to-observation error correlations are assumed to be zero, as different sensor systems are considered independent" no change here, merely a comment: this is interesting, as it's possible to envisage cases where these are not independent (though I think the use of satellite as background vs gauges/radar means the authors are reasonably safe in this assumption).

Response: Thank you for this thoughtful comment. Indeed, complete independence between different data sources cannot always be guaranteed in practice. For example, some satellite retrieval algorithms may incorporate gauge observations during calibration or bias correction, while radar products may also undergo gauge-based adjustment, potentially introducing some dependency between datasets. Explicitly accounting for both within-source and cross-source error correlations, however, would substantially increase the complexity of error covariance estimation and the subsequent statistical interpolation solution. In practical implementations, some simplification of the error covariance structure is therefore generally necessary. In this study, we assume zero background-to-observation error correlations to maintain a tractable framework, while the overall validity of these assumptions is ultimately assessed through independent evaluation of the resulting precipitation estimates.

R1-SC7

L285 "only the nearest radar grid cell to each target grid cell is retained as a supplementary observation" Might be good to (very briefly) note how 'nearest' is defined here: is it possible to have more than one 'nearest' cell? Never mind, this is also covered below!

Response: Thank you for this comment. We agree that a brief clarification at this point improves readability. We have therefore revised the text to indicate that the nearest radar grid cell is identified based on the minimum

Euclidean distance to the target grid cell. We have also added a brief explanation of how cases involving multiple equidistant nearest grid cells are handled.

To address this issue, an error-informed localisation strategy is adopted within the statistical interpolation framework. Specifically, a highly efficient implementation is introduced in which only the nearest radar grid cell within a 200 km search radius, identified as the grid cell with the minimum Euclidean distance to the target grid cell, is retained as a supplementary observation. If no valid radar grid cell is available within this radius, no radar observation is included for that target grid cell. The 200 km radius defines the maximum search range, while the effective influence of the selected radar observation is further controlled by the distance-dependent error correlation function. In cases where multiple grid cells have the same minimum distance, the grid cell with the lower error variance is retained. This nearest-point approximation reflects a deliberate selection strategy that prioritises the most locally relevant radar information while avoiding dilution by redundant contributions. Consequently, each local analysis contains at most one radar observation, and no radar–radar error covariance terms arise in the resulting observation-error covariance matrix. The suitability of this nearest-point approximation has been evaluated in our previous study (Zhang et al., 2026). As a result, this approach enables scalable, grid-cell-wise blending suitable for continental-scale precipitation analysis at kilometre spatial resolution and hourly temporal resolution under practical computational constraints.

R1-SC8

L321 "except for gauges located outside the radar range where no radar information is available" I think this is saying that gauges outside the radar range are used to correct satellite data. Is this correct? (And if not, why not?)

Response: Thank you for pointing this out. We acknowledge that the original wording was unclear and could lead to this interpretation.

Our intention was not to indicate that gauges outside radar coverage were excluded from satellite bias correction. Rather, satellite and radar bias correction used different gauge sets:

- Satellite bias correction used all available gauges.
- Radar bias correction only used gauges located within radar coverage, because radar values cannot be extracted at locations outside the radar range.

We have revised the manuscript to clarify this distinction.

3.3 Local bias correction of satellite and radar precipitation estimations

Bias correction for satellite and radar data followed a Kalman filtering approach similar to Chumchean et al. (2006), but was adapted from the original mean-field framework to estimate spatially varying local bias fields at the target hourly timescale. Bias estimation was based on co-located satellite/radar–gauge pairs, obtained by extracting satellite or radar estimates at gauge locations and pairing them with the corresponding gauge observations. All available gauges were included in the satellite bias correction, whereas only gauges within radar coverage were included in the radar bias correction because radar estimates are unavailable outside the radar coverage area. Bias at each gauge location was estimated using a Kalman filter with an AR(1) model to capture temporal dependence, and the resulting point estimates were interpolated using ordinary kriging to generate a continuous spatially varying bias field. The corrected product was then obtained by applying the bias field multiplicatively to the raw product.

R1-SC9

L335 "For gauged locations, this task is relatively straightforward, as gauge observations can be used as ground truth to estimate the error characteristics of the background and observational inputs" Excepting that gauges are

point data and satellite/radar are areal averages, which will be smoother in time and space. How did the authors account for areal averaging to compute errors against gauges?

Response: Thank you for this comment. We agree that representativeness differences exist between point-based gauge observations and grid-based satellite or radar estimates because gridded products represent areal averages that are generally smoother in space and time.

In principle, if a sufficient number of gauges were available within a 2 km × 2 km grid cell, their observations could be combined to estimate an areal-average rainfall representative of the grid cell, although such an estimate would itself introduce additional uncertainty. In practice, however, the gauge network is not dense enough to support reliable areal-average estimates at this spatial resolution over most of the study domain. Consequently, point gauge observations were used directly as the reference for error estimation.

Given the relatively high spatial resolution of the analysis grid (2 km), we assumed that point-to-area representativeness differences are comparatively small for the purposes of this framework. Moreover, the same assumption was applied consistently to both satellite and radar products. Therefore, the estimated error characteristics should be interpreted primarily as relative measures of the uncertainties among different data sources within the framework, rather than absolute estimates of their true errors.

We have revised the manuscript to explicitly state this assumption and its implications.

3.4 Error characterisation

…

For gauged locations, this task is relatively straightforward, as gauge observations can be used as the reference for estimating the error characteristics of the background and observational inputs. Although gridded satellite and radar products represent areal averages while gauges provide point measurements, the gauge network is generally too sparse to derive reliable areal-average estimates at the 2 km grid scale. Therefore, point gauge observations were used directly as the reference. At the 2 km spatial resolution adopted in this study, point-to-area representativeness differences were assumed to be sufficiently small and were not explicitly considered. Consequently, the estimated error characteristics should be interpreted as relative measures of uncertainty among different data sources rather than absolute estimates of their true errors.

R1-SC10

L352 "Specifically, the Australian domain was divided into tropical, subtropical, and temperate subregions." It would be nice to see these regions on a map; could simply be added to Figure 3a or similar.

Response: Thank you for this helpful suggestion. We have added the boundaries of the tropical, subtropical, and temperate subregions to both Figure 1 and Figure 3 to clearly show their spatial extent.

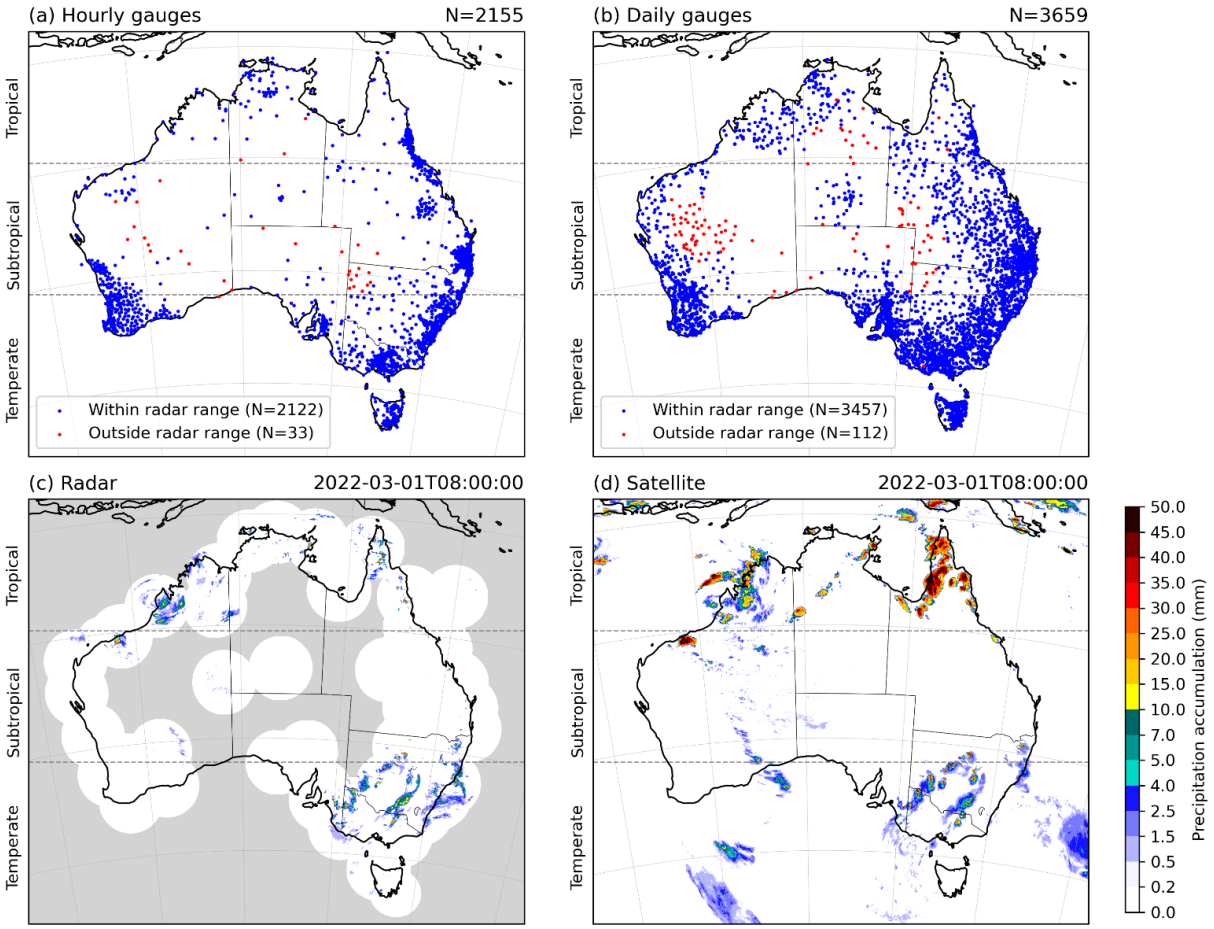


Figure 1. Spatial distribution and coverage of the data sources used in this study: (a) hourly gauges (aggregated from sub-daily) (N=2,155) categorised as within (N=2,122, blue) or outside (N=33, red) the maximum radar range, (b) daily gauges (N=3,659) similarly categorised as within (N=3,457, blue) or outside (N=112, red) the maximum radar range, (c) radar-based hourly precipitation accumulation at 0800 UTC on 1 March 2022, with a maximum radar range of approximately 300 km and a spatial resolution of 1 km, and (d) satellite-based precipitation (CRRPH) for the same time, with seamless spatial coverage at a 2-km resolution. The tropical, subtropical, and temperate subregions used in subsequent analyses are indicated by the horizontal dashed lines.

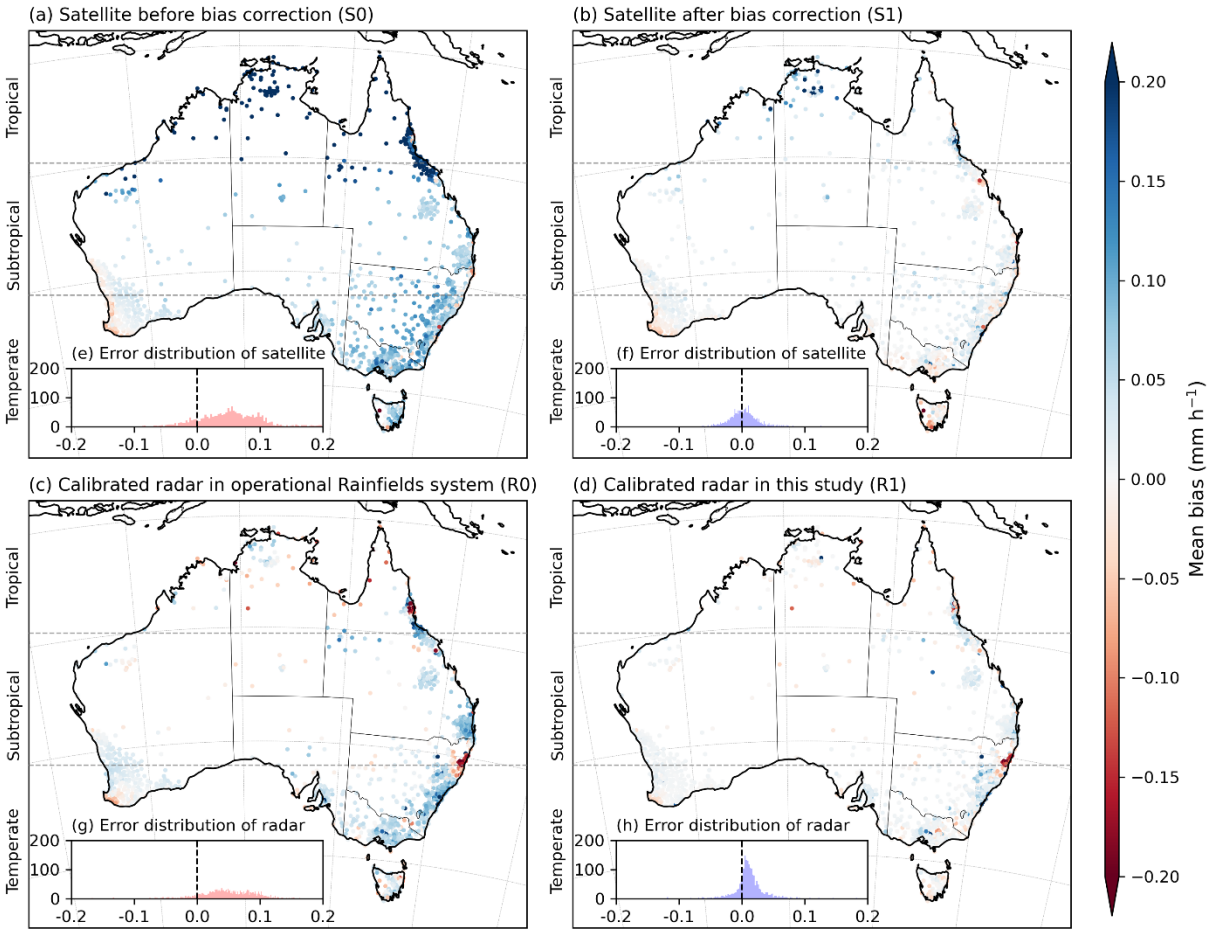


Figure 3. Spatial distribution of mean bias (mm h-1) for (a, b) satellite precipitation before (S0) and after (S1) bias correction, and (c, d) radar precipitation calibrated using the operational Rainfields system (R0) and the method in this study (R1). The Rainfields system calibration (R0) utilizes sub-daily gauges within a 150-km radius and a 15-min time window, whereas the calibration in this study (R1) is expanded to a 300-km radius and a 1-hour time window. The analysis includes all available hourly gauges from 2022 to 2023. Histograms illustrate the error distributions across all hourly gauges for the corresponding datasets, with the dashed line marking zero bias. The boundaries of the tropical, subtropical, and temperate subregions used in the regional analysis are indicated in each panel. Note that fewer gauges are shown for Radar because gauges outside radar coverage are excluded. An extended analysis in different subregions is provided in Fig. S1.

R1-SC11

Evaluation strategies: outperforming both IDW and AGCD is a high bar to clear; I commend the authors for this kind of rigour.

Response: Thank you for your positive comment. We appreciate your recognition of the evaluation strategy adopted in this study.

R1-SC12

L393 "The spatial match between the gridded data and gauge locations was established using nearest interpolation." I'm not sure what was being interpolated here - please clarify.

Response: Thank you for pointing this out. We agree that the original wording was ambiguous. The intention was to indicate that values from the gridded precipitation products were extracted at gauge locations using nearest-neighbour interpolation for comparison with gauge observations. We have revised the text to clarify this.

3.5.2 Performance metrics and stratified evaluation

…

For all gauge-based evaluations, gridded estimates were matched to gauge locations using nearest-neighbour interpolation, whereby the value from the grid cell with the minimum Euclidean distance to each gauge location was selected. This approach preserves the values of the original gridded analyses and avoids additional spatial smoothing that may be introduced by interpolation methods such as bilinear interpolation. Nevertheless, comparisons between point-based gauge observations and grid-cell rainfall estimates are subject to inherent point-to-area representativeness differences, which should be considered when interpreting the evaluation results.

R1-SC13

L407 "direct comparison with global precipitation products would be neither fully fair nor particularly informative." Fair enough - I assume the authors are suggesting BRAIN would substantially outperform these global products. As the verification in this study is already very rigorous, and you can't do everything, this is fine. I think it would nonetheless be interesting to compare BRAIN to these global products in a future study if possible, as it would reinforce confidence in BRAIN for Australian users.

Response: Thank you for this valuable suggestion. We agree that a comparison with global precipitation products would provide useful additional context and could further strengthen confidence in BRAIN for Australian applications. We are currently working towards the development of a long-term, operational-quality dataset based on the proposed framework. Once this dataset becomes available, a more systematic and comprehensive comparison with global products will become both feasible and more meaningful.

There are several existing global precipitation products (e.g., Beck et al., 2019; Ma et al., 2025) that could potentially be used for comparison with BRAIN. However, these products differ from BRAIN in their spatial and temporal resolutions, operational objectives, and observational inputs. In particular, BRAIN is designed as a high-resolution operational product for Australia with an update latency of less than one hour and incorporates a large number of hourly gauge observations and regional radar data that are not generally used in global precipitation products. These differences need to be carefully considered when conducting direct product comparisons. The current study therefore focuses on the development and evaluation of the BRAIN framework, including its performance relative to its constituent data sources and different blending configurations. We are currently working towards the development of a long-term BRAIN dataset, which will enable more systematic comparisons with existing global precipitation products. Such comparisons will also provide an important opportunity to assess the potential benefits of incorporating regional observational datasets.

R1-SC14

L547 "BR-SRG clearly outperforms AGCD and IDW in representing fine-scale spaital structure and intensity". I understand the need to show example predictions to illustrate how BRAIN functions in comparison to other products. I do not think it's appropriate to use these figures to describe relative performance, in particular because e.g. BRAIN does not outperform AGCD in out-of-radar regions on average according to Figure 6. Suggest this sentence be reworded to "BR-SRG shows fine-scale structure and intensities not present in AGCD and IDW."

Response: Thank you for this helpful comment. We agree that these selected events should be used to illustrate the characteristics of the different precipitation estimates rather than to draw conclusions about their relative overall performance. Consistent with our response to the previous comment, we have reduced the emphasis on performance comparisons in this section and now use these cases primarily as demonstrations.

4.5 Demonstration of BRAIN in selected rainfall events

Figure 10 demonstrates the spatial characteristics of the hourly blended product (BR-SRG) for two selected extreme precipitation events, alongside the AGCD daily product and hourly gauge-based IDW interpolation: a 24-h accumulation during Cyclone Ellie from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and a 24-h accumulation during Cyclone Jasper from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. The corresponding satellite and radar estimates and hourly and daily gauge observations are shown in Figs. S5–S6, with the highest gauge-recorded accumulations summarised in Tables S1–S2. Both satellite and radar captured the major rainfall systems associated with the two events, with satellite providing spatially continuous information across Australia and radar providing additional high-resolution information within its coverage. By integrating these spatially distributed estimates with gauge observations, BR-SRG captured the broad-scale rainfall patterns while retaining finer-scale spatial structures and variations in precipitation intensity that were less evident in AGCD and IDW.

R1-SC15

L550 Figure 7. A few things to improve about this figure: 1) it would be nice to put the dates for the two rows in the y-axis label of the left-most panel. 2) I could not see the 'x' symbols described in the text; please choose a symbol that is legible. 3) I do not think it's good practice to present additional results on rainfall totals at gauges in the caption. This information could be presented in small text boxes within each panel; alternatively an additional panel could be added to each row with a bar chart showing the values for each product.

Response: Thank you very much for these constructive suggestions. We have revised the original Figure 7 (now Figure 10) accordingly, as shown below.

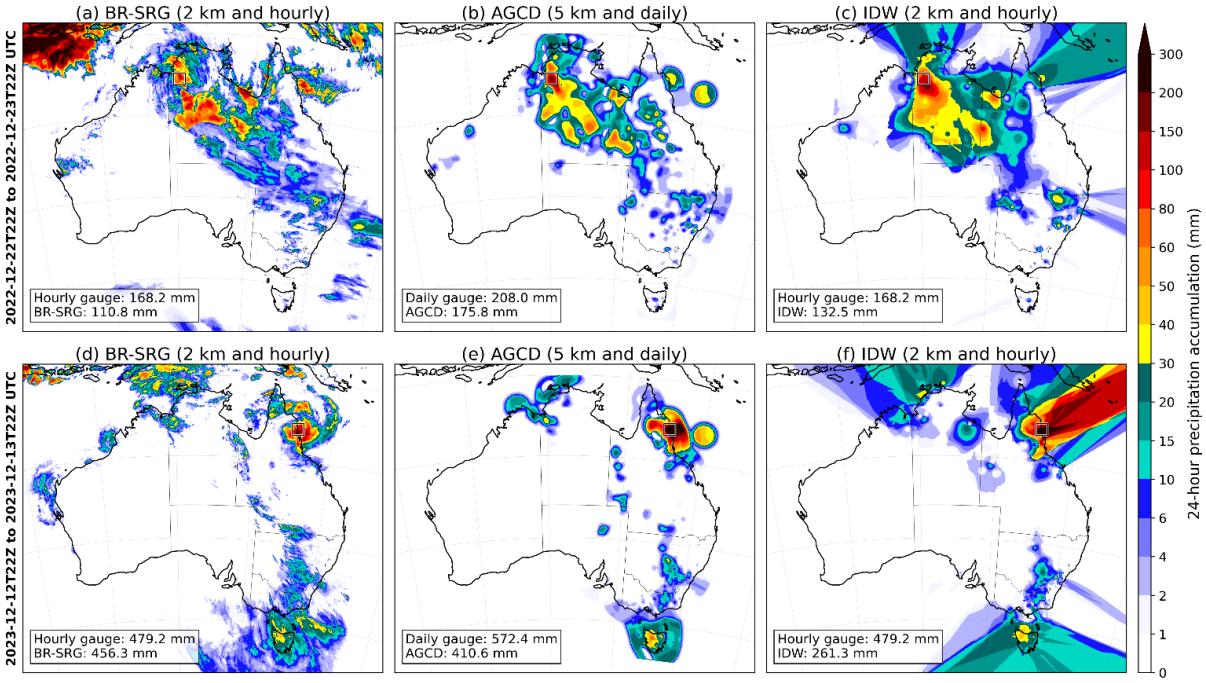


Figure 10. Example rainfall analyses from the three-source blend (BR-SRG; 2-km and hourly), AGCD (5-km and daily), and gauge-based IDW interpolation (2-km and hourly) for two selected 24-h periods during extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. Note that AGCD incorporates the denser daily-gauge network (Fig. S6b, d), whereas BR-SRG and IDW use the comparatively sparser hourly-gauge network (Fig. S6a, c). BR-SRG and IDW are accumulated over the indicated UTC periods, whereas AGCD represents the standard daily gauge accumulation period of approximately 9 am to

9 am local time, ending on 23 December 2022 and 14 December 2023, respectively, with the exact UTC period varying across Australian time zones. The red square indicates the location of the gauge with the highest observed 24-hour precipitation for each event. Further details of the satellite and radar estimates and gauge observations are provided in Figs. S5–S6 and Tables S1–S2.

R1-SC16

L561 "underestimates peaks"; as per the previous comment, this does not reflect e.g. bias statistics shown in Figures 5 & 6. It's my view that readers are drawn to examples like these, and they may use them to infer that these results are generally true, rather than refer to the much more robust statistics presented in Figs 5 & 6. I think that all that can be said is that BRAIN shows much greater spatial detail than the other products. This detail is supported by the verification metrics shown in Figures 5 & 6.

Response: Thank you for this suggestion again. Consistent with our response to R1-SC14, we have therefore revised this section. The event-based analysis is now used primarily to demonstrate the spatial characteristics and finer-scale detail represented by BRAIN.

R1-SC17

L565 "AGCD (Fig. 7e) fails to capture the peak at Cape Tribulation". 1) International readers may not know where Cape Tribuation is and 2) The totals by AGCD on and off the coast around Cape Tribulation are higher than either BR-SRG or IDW, so I do not think this comment accurate reflects what's shown in the figure. Again, I suggest avoiding making inferences about performance from figure 7. Figure 7 is useful in showing (1) the additional spatial detail shown by BRAIN over both AGCD and IDW and (2) that the spatial detail is consistent across the continent, while for IDW it is not. In my view conclusions on performance should be drawn only from the robust assessments shown in Figs 5 & 6.

Response: Thank you for this helpful comment again. We agree with your suggestions. More broadly, in response to this comment and the related comments above, we have substantially rewritten the entire event-based analysis section.

4.5 Demonstration of BRAIN in selected rainfall events

Figure 10 demonstrates the spatial characteristics of the hourly blended product (BR-SRG) for two selected extreme precipitation events, alongside the AGCD daily product and hourly gauge-based IDW interpolation: a 24-h accumulation during Cyclone Ellie from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and a 24-h accumulation during Cyclone Jasper from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. The corresponding satellite and radar estimates and hourly and daily gauge observations are shown in Figs. S5–S6, with the highest gauge-recorded accumulations summarised in Tables S1–S2. Both satellite and radar captured the major rainfall systems associated with the two events, with satellite providing spatially continuous information across Australia and radar providing additional high-resolution information within its coverage. By integrating these spatially distributed estimates with gauge observations, BR-SRG captured the broad-scale rainfall patterns while retaining finer-scale spatial structures and variations in precipitation intensity that were less evident in AGCD and IDW.

The representation of the most intense rainfall, however, depended on how well the different observations captured the rainfall maxima. During Cyclone Ellie, satellite (Fig. S5a) and radar (Fig. S5b) both identified substantial rainfall over northern Australia, but the highest rainfall occurred in an area with relatively sparse hourly gauge observations (Fig. S6a). Only one hourly gauge recorded a 24-h accumulation exceeding 100 mm (168.2 mm), while the next highest accumulation was 51.4 mm (Table S1). In comparison, the denser daily gauge network (Fig. S6b) provided multiple observations of high rainfall, including accumulations of 208.0, 184.6, and 141.4 mm (Table S1). The limited sampling of the rainfall maximum by hourly gauges therefore constrained the

observational information available to BR-SRG for representing the highest rainfall centre, whereas the additional daily observations provided stronger constraints for AGCD.

During Cyclone Jasper, satellite (Fig. S5c) and radar (Fig. S5d) clearly captured the intense rainfall system, with several hourly gauges located within the heavy rainfall centre and recording high rainfall accumulations (Fig. S6c). The highest hourly-gauge-derived 24-h accumulations reached 479.2, 456.5, and 378.4 mm (Table S2). The combination of spatially distributed satellite and radar estimates with these high-rainfall gauge observations enabled BR-SRG to better represent the extreme rainfall than in the Cyclone Ellie case.

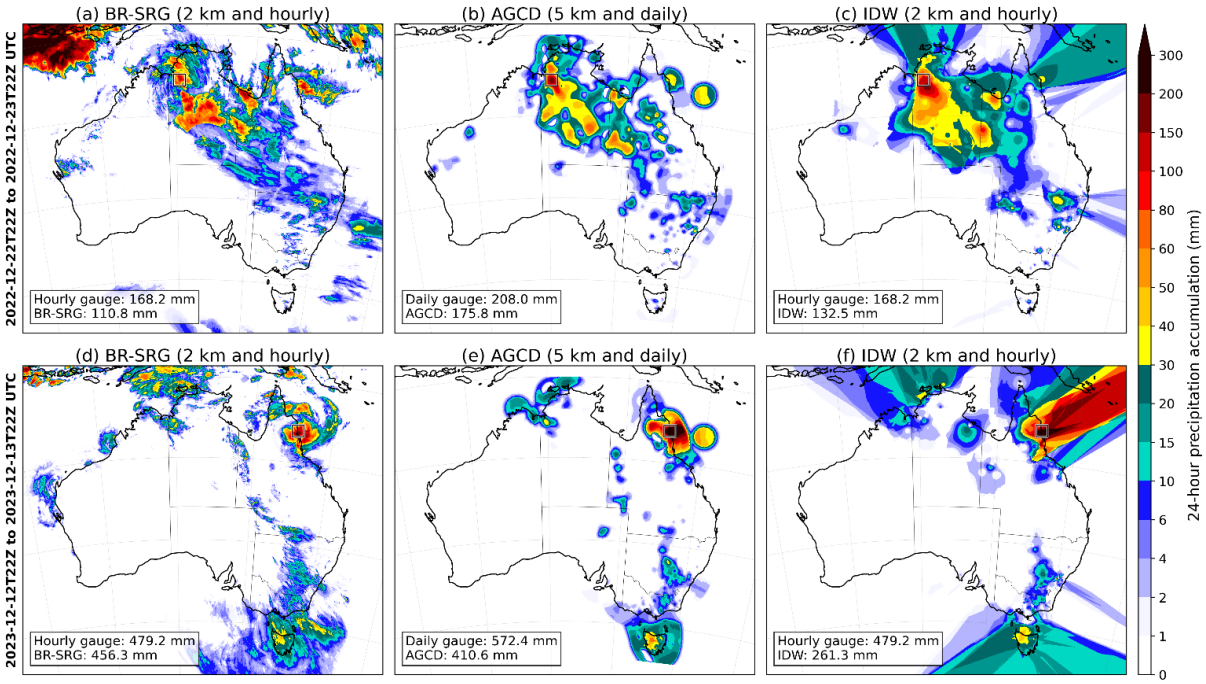


Figure 10. Example rainfall analyses from the three-source blend (BR-SRG; 2-km and hourly), AGCD (5-km and daily), and gauge-based IDW interpolation (2-km and hourly) for two selected 24-h periods during extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. Note that AGCD incorporates the denser daily-gauge network (Fig. S6b, d), whereas BR-SRG and IDW use the comparatively sparser hourly-gauge network (Fig. S6a, c). BR-SRG and IDW are accumulated over the indicated UTC periods, whereas AGCD represents the standard daily gauge accumulation period of approximately 9 am to 9 am local time, ending on 23 December 2022 and 14 December 2023, respectively, with the exact UTC period varying across Australian time zones. The red square indicates the location of the gauge with the highest observed 24-hour precipitation for each event. Further details of the satellite and radar estimates and gauge observations are provided in Figs. S5–S6 and Tables S1–S2.

R1-SC18

L616 "Future development of BRAIN will focus on improving input data quality, refining error characterisation, and enhancing spatial and temporal resolution." (1) It would be nice to give an indication of how computationally demanding a re-estimation of the OI is when new predictor products are used or existing predictor products change. (2) Will the use of e.g. additional satellite products violate the assumption of error independence between products that is crucial to the simplifications pre

Response: Thank you for these forward-looking questions. We have expanded the discussion to briefly address both points.

Future development of BRAIN will focus on improving input data quality, refining error characterisation, and enhancing spatial and temporal resolution. Error characterisation is performed offline using historical data. Therefore, introducing or updating an input product mainly requires repeating this offline estimation and subsequent evaluation, while the operational analysis remains computationally efficient. Anticipated upgrades to the Himawari-based CRRPH product are expected to further improve performance, particularly in tropical and radar-sparse regions, while ongoing radar recalibration efforts are likely to enhance local accuracy. The integration of complementary satellite products (e.g. IMERG), additional ground-based observations (including daily gauge observations and the expanding availability of third-party sub-daily gauge observations), and improved representation of error characteristics in sparsely observed regions will further strengthen the system. Additional products, particularly satellite products, may also introduce cross-source error dependencies due to shared retrieval methods and/or calibration procedures. The current independence assumption is a practical simplification and should therefore be re-evaluated when extending BRAIN to additional data sources.

Typos etc.

L39 "Satellites-based" should be "Satellite-based"

Response: Thank you for your suggestion. We have corrected the text.

L50 "collocation-based approach provides" should be "co-location-based approaches provide"

Response: Thank you for the suggestion. We have retained "triple collocation-based", as "triple collocation" is the established terminology in the remote sensing and hydrological literature. The adjective "collocation-based" is derived directly from this established term rather than from the general word collocation. For example, Pan et al. (2015) use Triple collocation in the title of their paper:

Pan, M., Fisher, C.K., Chaney, N.W., Zhan, W., Crow, W.T., Aires, F., Entekhabi, D., Wood, E.F., 2015. **Triple collocation**: Beyond three estimates and separation of structural/non-structural errors. Remote Sensing of Environment 171, 299–310. https://doi.org/10.1016/j.rse.2015.10.028

L51 Suggest a sentence break before 'However', and suggest "However, both performance-based and co-location-based approaches typically..."

Response: Thank you for your suggestion. The sentence has been revised as follows:

For example, performance-based averaging methods derive weights for non-gauge data sources from gauge-based skill metrics and apply them in a grid-aligned manner, but their applicability may be limited in regions with very sparse gauge coverage. Triple collocation-based approach provides a gauge-independent alternative by estimating weights from independent data sources (e.g., Dong et al., 2022; Koster et al., 2021; Xu et al., 2020). However, both performance-based and triple collocation-based approaches typically require spatially complete datasets and assign fixed source weights over time, limiting their applicability under heterogeneous data coverage, such as mixtures of spatially incomplete gridded fields and point observations.

L58 Suggest a sentence break before 'While', and suggest 'While offering high flexibility, ML-based approaches have...'

Response: Thank you for your suggestion. The sentence has been revised as follows:

More recently, machine learning–based approaches have been explored (e.g., Gavahi et al., 2023; Kossieris et al., 2024; Xu et al., 2024), in which non-gauge data sources are treated as input features and model relationships are learned using gauge observations as targets. While offering high flexibility, ML-based approaches have limited interpretability and substantial computational demands, which have constrained their adoption in operational precipitation analysis.

L59 "In addition, all above" should be "In addition, all the above"

Response: Thank you for your suggestion. The sentence has been revised as follows:

In addition, all the above approaches typically do not assimilate gauge observations directly as a primary data source in the blending step, which can limit performance in gauge-rich regions where point observations provide the most accurate precipitation information.

L75 "data source used for background field" should be "data source used for the background field"

Response: Thank you for your suggestion. The sentence has been revised as follows:

In practice, the achievable spatial and temporal resolution of such systems is largely constrained by the data source used for the background field.

L251 "By expanding and expressing the above equation" suggest "By expanding and expressing Equation 4"

Response: Thank you for your suggestion. The sentence has been revised as follows:

By expanding and expressing Eq. 4 in standardised form:

$$\sum_{j=1}^{n}\left(\mu_{ij}+\tau_{ij}\sigma_j+\tau_{ji}\sigma_i+\rho_{ij}\sigma_i\sigma_j\right)w_i'=\mu_{ki}+\tau_{ki}\sigma_i,\ i=1,2,\dots,n \qquad (5)$$

L264 "is adjusted weight accounting" should be "is the adjusted weight accounting"

Response: Thank you for your suggestion. The sentence has been revised as follows:

$w_i'=w_i\left(f_i^{\,2}/f_k^{\,2}\right)^{\frac{1}{2}}, w_i'\in\mathbb{R}^{n\times 1}$ is the scaled weight accounting for differences in background error across between the *i*th observational location and target grid cell *k*.

L337 suggest deleting "due to the absence of gauge observations" - it's redundant.

Response: Thank you for your suggestion. We have removed the redundant phrase and revised the sentence as follows:

For ungauged grid locations (99.9% of the 2 km land grid cells nationally), however, the process becomes considerably more challenging. A common approach is to incorporate spatial and temporal windows to gather sufficient samples.

L394 "For the trialled implementation" suggest "For the trial implementation"

Response: Thank you for your suggestion. The sentence has been revised as follows:

For the trial implementation, a two-year dataset from 2022 to 2023 was used in this study.

L414 "mean bias error" suggest deleting "error" as it's redundant

Response: Thank you for your suggestion. We agree that "error" is redundant in this context and have removed it. The sentence has been revised as follows:

Figure 3 demonstrates the spatial and statistical improvements in mean bias for satellite and radar precipitation data after bias correction, evaluated against hourly gauge observations across Australia during the 2022–2023 period.

L604 "providing a consistent and flexible rainfall analysis capability" Suggest deleting "capability"

Response: Thank you for your suggestion. We agree that "capability" is redundant in this context and have removed it. The sentence has been revised as follows:

In the longer term, the framework is designed to support both historical reconstruction and near–real-time product generation, providing a consistent and flexible rainfall analysis for a wide range of users.

L607 "for gridded radar source" suggest deleting "source"

Response: Thank you for your suggestion. We agree that "source" is redundant in this context and have removed it. The sentence has been revised as follows:

The error-informed localisation strategy embedded within the statistical interpolation framework is shown to be effective at continental scale. In particular, the nearest-point approximation for gridded radar substantially reduces modelling and computational complexity, enabling multi-source blending at kilometre and hourly resolutions over large spatial domains.

Reply to RC2 - Anonymous Referee #2

General Comments:

The manuscript presents an hourly, at 2 km resolution blended rainfall dataset (BRAIN) for Australia, derived from satellite, radar, and gauge observations. Given the current lack of a national hourly gridded rainfall product, this dataset addresses a clear need. The manuscript is generally well written, and the figures are clearly presented.

Response: Thank you for your positive assessment of our work and for recognising the value of the BRAIN dataset for Australia.

However, the methodology used in the study, including the Optimal Interpolation (OI) and Bias Correction, was developed and applied in other studies for the same purpose of generating blended rainfall datasets from satellite, radar, and gauge data (e.g., Yu et al. 2020 (blended data in the UK), Xia et al. 2025 (blended data in China)). There is no significant improvement in this study when the method was applied to Australia.

Response: We acknowledge your concern regarding the originality of the methodology. We agree that bias correction and optimal interpolation have been widely used in previous multi-source rainfall blending studies, including Yu et al. (2020) and Xia et al. (2025), and we do not intend to claim these methods as new. Our methodological choices were informed by our previous assessment of approaches for multi-source rainfall

blending (Zhang et al., 2025), with particular consideration given to computational feasibility and the highly heterogeneous observing network across Australia.

The contribution of this study lies primarily in the implementation and evaluation of a computationally efficient multi-source blending framework for national-scale hourly rainfall analysis. Compared with previous studies, BRAIN adopts a localised three-source OI implementation in which satellite precipitation provides the spatially continuous background, while radar and gauge observations are incorporated separately according to their respective error characteristics. The framework also allows radar information to contribute beyond hard radar-coverage boundaries through distance-dependent localisation. These design choices enable efficient application over a large domain at high spatial and temporal resolutions while accommodating substantial spatial and temporal variations in observational availability.

We have revised the manuscript to more clearly acknowledge previous studies and distinguish the established methodology from the specific methodological contributions and implementation choices of BRAIN. A new subsection, Section 5.1, "Methodological contributions of the BRAIN framework", has been added as follows:

5.1 Methodological contributions of the BRAIN framework

Multi-source blending of satellite, radar, and gauge precipitation observations using bias correction and statistical interpolation is not a new concept (e.g., Gandin, 1965; Xie and Xiong, 2011, Yu et al, 2020, Xia et al., 2025), but the BRAIN framework differs from previous approaches in several aspects of its implementation.

First, BRAIN was designed for computationally efficient application over a continental-scale domain such as Australia at high spatial and temporal resolution. The nearest-radar-point approximation within the localisation strategy substantially reduces the dimensionality of the statistical interpolation problem while retaining the source-specific error characterisation (Zhang et al., 2026), differing from the more computationally intensive formulation used in UKGrsHP (Yu et al., 2020).

BRAIN also uses a single satellite product as the spatially continuous background, with radar incorporated as an additional observation together with gauge observations. This differs from approaches that first combine satellite and radar to construct a blended background field for subsequent statistical interpolation with gauge observations (e.g. Xia et al., 2025). Because statistical interpolation requires characterisation of the spatial error structure of the background field, such pre-blending produces a new background whose error characteristics may differ from those of the original satellite and/or radar products. By retaining satellite and radar as separate sources, BRAIN allows their respective error characteristics and contributions to be explicitly represented within the statistical interpolation framework.

A further distinctive feature is that radar information is not restricted by a hard coverage mask. For each target location, BRAIN searches for the nearest available radar observation within a predefined maximum distance (200 km in this study). Consequently, radar information can still contribute to the blended estimate when the target location lies outside radar coverage, provided that a sufficiently close and reliable radar observation is available. Its effective contribution is determined by both the distance-dependent error correlation (Fig. 4b) and the estimated error variance, with greater weight generally assigned to observations that are closer to the target location and have lower estimated errors. Even when the nearby radar estimate is not more accurate than the local satellite estimate on its own, it can provide complementary information that can improve the resulting estimate after blending.

This formulation allows radar information to contribute both within and near radar-covered regions, providing a gradual transition across radar-coverage boundaries while retaining the satellite estimate as the spatially continuous background. This differs from simple radar extrapolation, in which rainfall information within the radar domain is spatially extended into areas without radar observations. In BRAIN, the radar field itself is not extrapolated. The satellite estimate remains the background at every target location, while available radar and gauge observations contribute through the statistical interpolation according to their estimated error characteristics and spatial correlations.

In addition, the use of inverse distance weighting (IDW) interpolation as a benchmark raises some concerns. Describing IDW as "the operational method in Australia for estimating hourly rainfall" may be somewhat misleading. IDW is a relatively simple interpolation technique, typically used for observing purposes rather than for producing authoritative rainfall datasets. Established Australian products such as AGCD, SILO, and ANUClimate employ more sophisticated interpolation methods, albeit primarily for daily timescales. As such, comparisons against IDW may not provide a sufficiently robust benchmark. The conclusion that BRAIN outperforms gauge data using the IDW interpolation is not convincing and can mislead users of the dataset.

Response: We acknowledge your concern. We agree that our previous description of IDW as "the operational method in Australia for estimating hourly rainfall" was too broad and could be misinterpreted. IDW is not used to produce national gridded rainfall datasets such as AGCD, SILO, or ANUClimate.

We used IDW because it has been, and continues to be, used as the operational methodology for estimating basin and/or sub-basin rainfall from sub-daily gauge observations for national flood forecasting at the Bureau of Meteorology (Malone, 1999; Malone and Carroll, 2022). Our intention was therefore to provide a practical gauge-based reference relevant to operational hydrological applications, rather than to represent IDW as the standard method for producing national gridded rainfall products. We also recognise that IDW alone provides a relatively simple gauge-interpolation comparison.

To address this concern, we have revised the description of IDW and clarified its role in the study. We have also added sensitivity tests using different IDW configurations and a thin-plate spline interpolation method to provide a more robust assessment of gauge-based interpolation. In addition, the corresponding discussion and conclusions have been revised to avoid implying that comparison with IDW alone demonstrates the superiority of BRAIN over gauge-based rainfall products. Please see our detailed responses to the major comments below for further information and the corresponding revisions.

References:

- Malone, T.: Using URBS for Real Time Flood Modelling, in: Joint Congress: 25th Hydrology & Water Resources Symposium, 2nd International Conference on Water Resources & Environment Research, Handbook and Proceedings, 603–608, https://search.informit.org/doi/10.3316/informit.705872257387556, 1999.
- Malone, T. and Carroll, D.: Guidelines for URBS routing parameters, in: Hydrology & Water Resources Symposium 2022, The Past, the Present, the Future, Brisbane, 145–155, https://search.informit.org/doi/10.3316/informit.905985293458115, 2022.

Unfortunately, I suggest rejecting the manuscript for publication in HESS at this stage. I hope that the authors will revise the method and validation processes after the trial. While the dataset has potential value, I suggest submitting the manuscript to another dataset journal.

Response: We acknowledge your concerns regarding the methodology and validation of this study. We have carefully considered these concerns and have substantially strengthened the validation and clarified the methodological choices and contributions of this work.

We respectfully believe that this study is appropriate for Hydrology and Earth System Sciences. Although the study adopts the established concepts of bias correction and optimal interpolation, its main contribution lies in the implementation and evaluation of a computationally efficient, national-scale multi-source precipitation blending framework for Australia. The framework addresses specific observational and operational challenges, including the heterogeneous availability of gauge, radar, and satellite observations across space and time. We also believe that this large-domain application provides a useful reference for the development of similar multi-source precipitation blending systems.

Once a long-term BRAIN dataset has been produced, we plan to prepare a separate data paper for submission to an appropriate data journal. The present manuscript instead focuses on the methodological framework and its implementation.

Major Comments:

R2-MC1

1. The introduction, section 2 and discussion sections would benefit from a more thorough review and discussion of potential errors and artifacts in the satellite and radar datasets. I suggest adding a few paragraphs to that outline the limitations across all three data sources used in the blending process, the issue that these errors may propagate into the final blended product, and the point-area spatial mismatch, even after the bias correction (Ochoa-Rodriguez et al, 2019).

Response: Thank you for this valuable suggestion. We agree that satellite, radar, and gauge observations all contain source-specific uncertainties and limitations. These uncertainties are considered in the OI framework through the error characteristics assigned to the different data sources, but they cannot be completely eliminated. Residual errors may therefore propagate into the final blended rainfall estimates. We also acknowledge the point-area spatial mismatch between gauge observations and gridded satellite and radar estimates.

To address this comment, we have expanded the literature review in introduction, data section and discussion of data uncertainties throughout the revised manuscript:

1 Introduction

Multiple observational sources provide complementary information for characterising precipitation. Rain gauges provide relatively direct and reliable measurements at individual locations but have limited ability to represent spatial variability, particularly in regions with sparse observation networks (Villarini et al., 2008). Weather radars provide precipitation estimates at high spatial and temporal resolutions over broader areas but are typically concentrated in densely populated regions, leaving large remote areas unmonitored (e.g., Zhang et al., 2025). Satellite-based precipitation products provide extensive spatial coverage and frequent temporal sampling, offering valuable precipitation information where ground-based observations are unavailable (e.g., Sun et al., 2018; Zhang et al., 2021b, c). Reanalysis products combine numerical weather prediction models with assimilated observations to provide spatially and temporally complete precipitation fields, although their precipitation estimates depend on model configuration, physical parameterisations, spatial resolutions, and data assimilation systems (e.g., Su et al., 2019, 2021, 2025).

Despite these complementary strengths, each source is subject to distinct uncertainties. Gauge measurements can be affected by instrumental errors (e.g., wind-induced undercatch) and local sampling limitations (e.g., discrete tipping in tipping-bucket gauges) (Ciach, 2003; Dunn et al., 2025; Ochoa-Rodriguez et al., 2019; Segovia-Cardozo et al., 2023), while their point-scale measurements may not fully represent precipitation averaged over a larger grid cell. Radar precipitation estimates are affected by uncertainties associated with reflectivity-to-rainfall conversion (Kirsch et al., 2019), attenuation (Delrieu et al., 1999), beam blockage (McRoberts and Nielsen-Gammon, 2017), ground clutter (Hubbert et al., 2009b, a), and range-dependent sampling (Ochoa-Rodriguez et al., 2019). Satellite precipitation estimates rely on indirect relationships between remotely sensed cloud or hydrometeor properties and surface precipitation and are therefore subject to retrieval and algorithmic uncertainties, including difficulties in detecting or quantifying certain precipitation regimes (Chen et al., 2023; Li et al., 2026; So and Shin, 2018; Upadhyaya et al., 2020; Zhu et al., 2026). Reanalysis precipitation is also subject to uncertainties inherent in its production process, as precipitation is generally diagnosed by the numerical model rather than directly assimilated, with its accuracy therefore depending strongly on model physics and parameterisations (e.g., Su et al., 2019, 2021, 2025). These errors can vary substantially across space, time, and precipitation conditions. Integrating information from multiple sources therefore provides an opportunity to

leverage their complementary information while reducing dependence on limitations of any individual source (e.g., Beck et al., 2017, 2019).

2 Data sources

2.1 In situ observations

Although the quality control procedures remove many identifiable erroneous observations, residual measurement and sampling uncertainties may remain in the gauge data used in this study. The hourly observations are aggregated from sub-hourly measurements, primarily from tipping-bucket gauges, and may therefore retain uncertainties associated with the original sub-hourly measurements and their accumulation to hourly totals. In addition, some isolated gauges have few nearby reference stations available for spatial consistency checks, while random or less readily identifiable errors may not be detected by the quality control procedures. It is also worth noting that rainfall measured by an individual gauge may differ from the corresponding 2-km grid-cell average because of sub-grid precipitation variability, introducing spatial representativeness uncertainty when point observations are used to represent gridded precipitation.

2.2 Radar-based observations

The R1 radar estimates used for blending benefit from both the operational Rainfields calibration and the additional hourly gauge-based recalibration applied in this study. Nevertheless, residual errors and artefacts present in the underlying radar observations may remain after these corrections. For example, local effects associated with beam blockage, anomalous echoes, and other retrieval or sampling issues may not always be fully identified or removed during operational quality control and processing. In addition, the gauge-based bias correction estimates local radar biases from available gauge observations and interpolates these estimates to construct spatially continuous bias fields. However, this correction does not necessarily remove biases across different precipitation intensities, and residual biases may remain for specific rainfall intensity ranges. Its effectiveness also depends on the availability and representativeness of nearby gauges and on how well the interpolated bias field represents the spatial variability of the actual radar bias. The correction may therefore be less effective in areas far from available gauges or with sparse gauge coverage. Details of the bias-correction procedure are provided in Section 3.3.

2.3 Satellite-based observations

As an infrared satellite precipitation product, CRRPH infers surface rainfall indirectly from cloud properties observed by the Himawari AHI. Retrieval errors may therefore occur when the observed cloud characteristics do not adequately represent the occurrence or intensity of surface precipitation (Karagiannidis et al., 2023; So and Shin, 2018). In addition, the calibrated lookup tables used to convert retrieved cloud properties into rainfall rates represent empirical relationships and may not fully capture variations in the cloud–rainfall relationship under different precipitation conditions. To reduce systematic discrepancies with gauge observations, a spatially varying local bias correction was applied as described in Section 3.3. S0 denotes the original CRRPH product without bias correction, while S1 denotes the bias-corrected product used as the background field for blending. The correction is designed to reduce the overall systematic bias of CRRPH relative to gauge observations towards zero, but does not necessarily remove biases across different precipitation intensities. Residual biases may therefore remain for specific rainfall intensity ranges. Moreover, the correction cannot fully address errors in rainfall occurrence or spatial structure inherent in the satellite retrieval. As the estimated bias field is derived from available gauges and spatially interpolated, its effectiveness depends on the availability and representativeness of nearby gauges and may be reduced in areas far from available gauges or with sparse gauge coverage.

5 Discussion

5.3 Limitations and future directions

One limitation of the current framework is the propagation of residual errors from input observations into the blended analysis. Residual errors that are not identified through quality control, particularly for gauge observations, or are not fully removed by bias correction, particularly for satellite and radar estimates, can propagate into the blended analysis. Although the statistical interpolation weighting accounts for the intensity-dependent error characteristics of each source, an individual source generally does not receive complete weight, as this would require the error of the other sources to become overwhelmingly large in comparison. Consequently, residual biases in sources retaining non-zero weights can propagate into the final analysis.

For example, during Cyclone Jasper, underestimation in the satellite estimates contributed to underestimation in BR-SRG even when radar closely represented the observed accumulation (Fig. 14a). Conversely, during Cyclone Ellie, radar substantially underestimated the observed accumulation, while the satellite estimates were higher and partly compensated for the radar underestimation in the blended analysis (Fig. 14b). These contrasting cases demonstrate the value of retaining complementary information from multiple data sources, rather than simply increasing the contribution of a single source.

Together with the relative differences among S1, R1, BR-SG, and BR-SRG in the rainfall-intensity analysis (Fig. 7), these cases suggest that residual underestimation in the satellite estimates may contribute to the negative bias observed in BR-SRG, although the intensity-stratified bias is also affected by observation-based stratification. Further improvements should therefore focus on bias correction, particularly in better accounting for systematic biases across different rainfall intensities.

R2-MC2

2. IDW should not be used an operational benchmark for hourly gauge rainfall data in the validation. In the paper, the authors claim that "the gauge-based interpolation approach currently used in flood operations". In practice, IDW is a simplistic method, typically suitable for preliminary visualisation of rainfall patterns rather than for operational flood applications. To my knowledge, IDW is not used to generate regional or national daily or sub-daily rainfall products. In Australia, widely used datasets such as AGCD, SILO, and ANUClimate rely on more advanced interpolation techniques. At the hourly scale, more robust approaches have also been explored—for example, thin-plate spline interpolation at 1 km resolution (Nguyen et al., 2026). Given this context, using IDW as the primary benchmark is not appropriate. Comparing the BRAIN dataset against such a basic method does not provide a rigorous assessment of performance, and the conclusion that the blended product outperforms gauge-based interpolation is therefore not sufficiently justified. his framing may mislead readers, particularly for sensitive applications such as flood modelling, where reliable rainfall estimates are critical.

Response: Thank you for this important comment. We agree that the role of IDW in our study requires clearer explanation.

First, IDW remains the current official methodology used by the Bureau of Meteorology for estimating basin and/or sub-basin rainfall from sub-daily gauge observations for operational flood forecasting and hydrological modelling (Malone, 1999; Malone and Carroll, 2022). However, it is not used to generate national gridded rainfall products such as AGCD, SILO, or ANUClimate. We have revised the manuscript to clearly distinguish these different applications.

Second, IDW was not intended to serve as the primary benchmark for BRAIN, but rather as a simple gauge-only reference to assess the added value of incorporating satellite and radar observations. We agree that our previous wording placed too much emphasis on this comparison and could lead to an overly broad interpretation of the results.

To address this comment, we have made the following revisions:

- We have clarified the role of IDW and moderated the wording throughout the manuscript to avoid implying that BRAIN has been comprehensively demonstrated to outperform gauge-only rainfall products or more advanced gauge-based interpolation methods.
- We have examined additional IDW configurations, including imposing a 200 km maximum search distance and varying the number of neighbouring gauges, to assess the sensitivity of the interpolated fields to these settings. These additional analyses help distinguish limitations associated with the original IDW configuration from more general characteristics of gauge-only interpolation.
- We have also applied thin-plate spline interpolation to selected case studies as an additional, more sophisticated gauge-based comparison. These comparisons further illustrate challenges associated with gauge-only interpolation, particularly in areas far from available gauges and the tendency of interpolation to produce spatially smooth rainfall fields.
- We have reduced the emphasis on the IDW comparison in the case study and now present it as an illustrative gauge-only reference rather than as a primary benchmark for establishing the superiority of BRAIN.

3.5 Evaluation strategies

The evaluation was designed primarily to assess the added value of integrating satellite, radar, and gauge observations within the BRAIN framework. The three-source satellite–radar–gauge analysis was therefore evaluated against individual-source fields and reduced-source blending configurations (i.e., satellite-gauge blending) to determine how the incorporation of additional observational information affects the analysis performance.

BRAIN was also compared with rainfall products and approaches currently used operationally by the Bureau to provide additional operational context. Gauge-based IDW interpolation was included as a simple gauge-only reference because it is currently used in the Bureau's operational flood forecasting system to estimate catchment- and/or sub-catchment-scale rainfall from sub-daily gauge observations for hydrological modelling (Malone, 1999; Malone and Carroll, 2022), rather than to generate a national sub-daily gridded rainfall product. AGCD was included as an existing national operational daily gridded rainfall product (Jones et al., 2009).

3.5.3 IDW implementation and comparison with AGCD

Gauge-based IDW interpolation was included as a simple gauge-only reference. The primary IDW configuration used the five nearest available gauges, selected according to Euclidean distance, with an inverse-distance power of 2 and no maximum search distance. This configuration is referred to as IDW-K5 and is shortened to IDW in the main evaluation. To examine the sensitivity of the comparison to the IDW configuration, two additional configurations were considered. IDW-K5-L retained the five-gauge configuration of IDW-K5 but imposed a maximum search distance of 200 km, allowing the effect of the distance constraint to be examined. IDW-K20-L used up to 20 gauges within a 200-km radius, consistent with the gauge-selection configuration used in BRAIN. The sensitivity of the results to these IDW configurations is examined in Section 5.2.

5.2 Behaviour and sensitivity of gauge-based interpolation

…

The sensitivity of IDW (Fig. 12) to its interpolation configuration was further examined using the same rainfall cases (Fig. 10). IDW-K5 provides a continuous field but can extend gauge influence over long distances in gauge-sparse regions. Applying a maximum search distance (IDW-K5-L) avoids such long-range interpolation but produces isolated rainfall patterns and areas without estimates, while increasing the number of gauges (IDW-K20-L) provides only limited improvement. Despite these differences, all configurations strongly smooth the rainfall field and poorly represent its spatial structure, particularly in gauge-sparse regions. These limitations are more evident in the spatial case studies than in the point-based validation above, which evaluates performance only at gauge locations. More sophisticated gauge-based interpolation methods, such as thin-plate spline (TPS) (Nguyen

et al., 2026) (Fig. 13), can improve spatial continuity but remain constrained by sparse gauge observations, with local rainfall features being overly smoothed and estimates becoming increasingly unreliable far from available gauges.

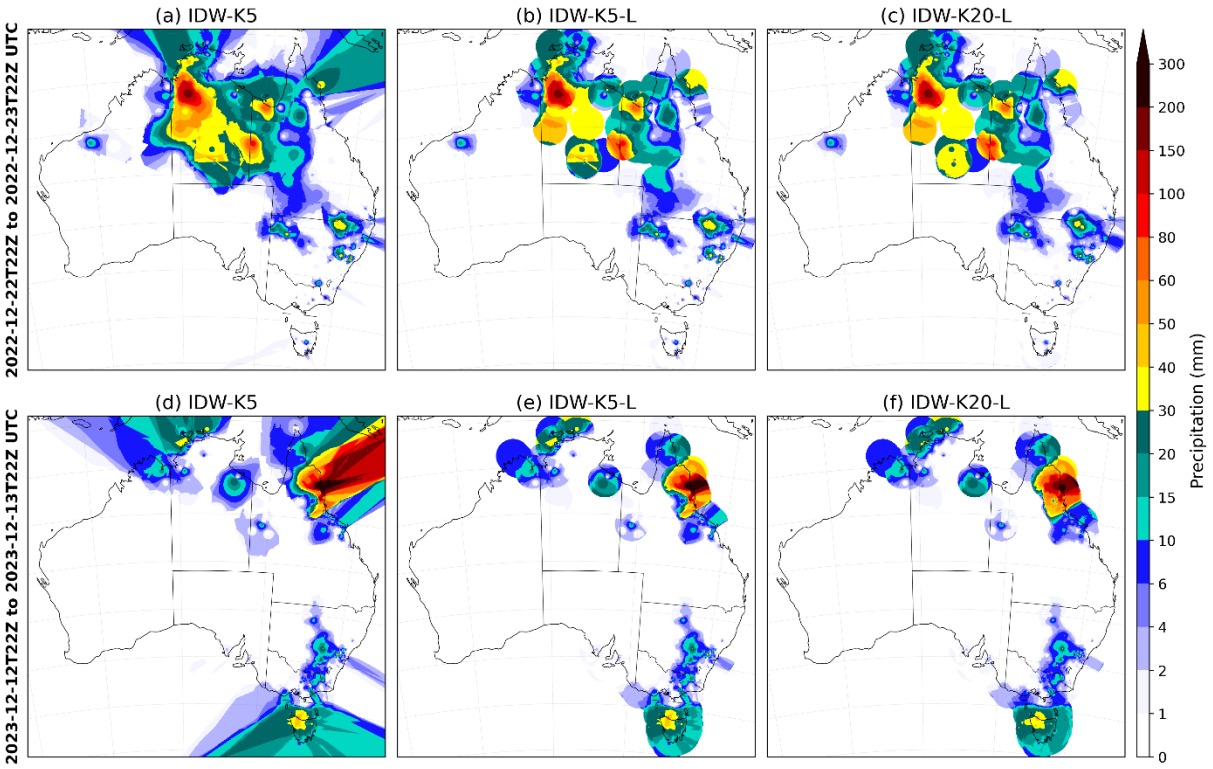


Figure 12. Sensitivity of 24-h accumulated precipitation estimates to different IDW configurations for two selected extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. The three configurations are IDW-K5 (a, d), using the five nearest gauges without a maximum search distance; IDW-K5-L (b, e), using up to five nearest gauges within a 200-km search radius; and IDW-K20-L (c, f), using up to 20 nearest gauges within the same search radius. IDW-K5 corresponds to the IDW configuration used in Fig. 10.

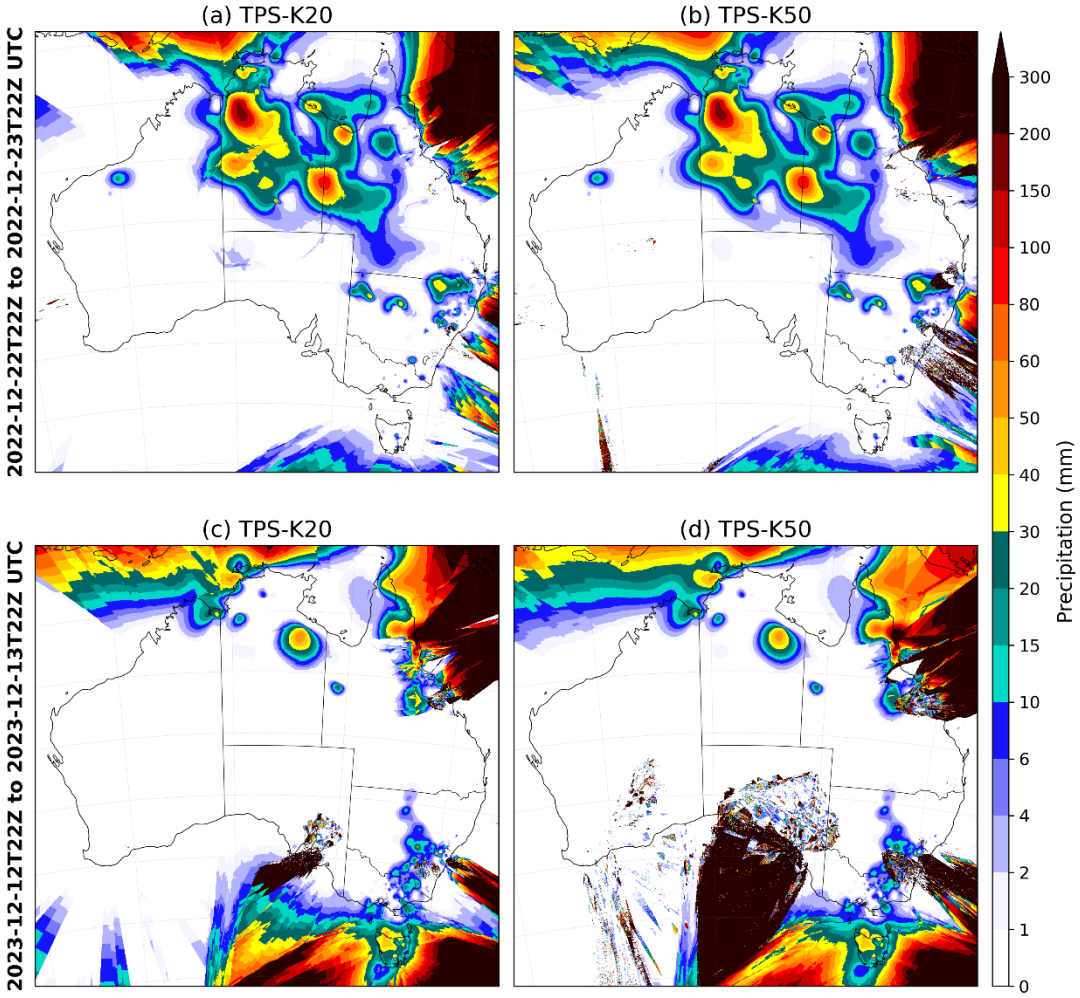


Figure 13. Gauge-based precipitation estimates obtained using locally fitted two-dimensional thin-plate spline (TPS) interpolation for two selected extreme precipitation events: (a, b) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (c, d) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. The two configurations are TPS-K20 (a, c), fitted locally using the 20 nearest gauges, and TPS-K50 (b, d), fitted locally using the 50 nearest gauges. These estimates provide additional gauge-based interpolation examples for comparison with the IDW estimates in Figs. 10 and 12.

References:

- Malone, T.: Using URBS for Real Time Flood Modelling, in: Joint Congress: 25th Hydrology & Water Resources Symposium, 2nd International Conference on Water Resources & Environment Research, Handbook and Proceedings, 603–608, https://search.informit.org/doi/10.3316/informit.705872257387556, 1999.
- Malone, T. and Carroll, D.: Guidelines for URBS routing parameters, in: Hydrology & Water Resources Symposium 2022, The Past, the Present, the Future, Brisbane, 145–155, https://search.informit.org/doi/10.3316/informit.905985293458115, 2022.

R2-MC3

3. What interpolation distance was used in the IDW interpolation (apologies if this was specified and I overlooked it)? Figures 7c and 7f suggest that the IDW-based interpolation of hourly gauge data performs poorly, even when compared with AGCD. In particular, the method appears to over-smooth the observations and can produce

unrealistically broad rainfall patterns. For example, while AGCD (Fig. 7e) indicates no rainfall within the highlighted red-box area below, the IDW result shows a substantial area with 6-20 mm/day. This discrepancy further highlights the limitations of IDW and reinforces that it is not a reliable benchmark for validation.

Response: Thank you for this helpful comment and for carefully examining the comparison figures. We acknowledge that the implementation details of the IDW interpolation were not sufficiently described in the original manuscript. We address the two issues raised in this comment separately below.

IDW interpolation distance and sensitivity to the interpolation configuration

In the original implementation, IDW used the five nearest available hourly gauges, with weights based on inverse Euclidean distance with a power of 2. No maximum search distance was imposed. This configuration was adopted to ensure that an interpolated rainfall estimate could be generated for every grid cell across the target domain. Imposing a strict distance limit can leave grid cells without an estimate in regions where the hourly gauge network is sparse.

We agree that the absence of a distance constraint can contribute to overly broad and spatially smooth rainfall patterns, particularly where the nearest available gauges are relatively distant from the target grid cell. To examine the sensitivity of the IDW results to this configuration, we have therefore added two additional IDW implementations.

- **IDW-K5-L**: The first retains the five-nearest-gauge configuration but introduces a maximum search distance of 200 km, allowing us to assess the effect of restricting interpolation to nearby observations.
- **IDW-K20-L**: The second uses up to 20 gauges within the same distance limit (200 km), consistent with the maximum number of hourly gauge observations considered in the BRAIN blending configuration.

The additional comparisons show that restricting the interpolation distance reduces the spatial spreading of rainfall into areas far from available gauges. Nevertheless, the IDW fields remain relatively smooth because IDW estimates rainfall solely through the spatial weighting of gauge observations and does not incorporate information on rainfall spatial structure from other observing sources.

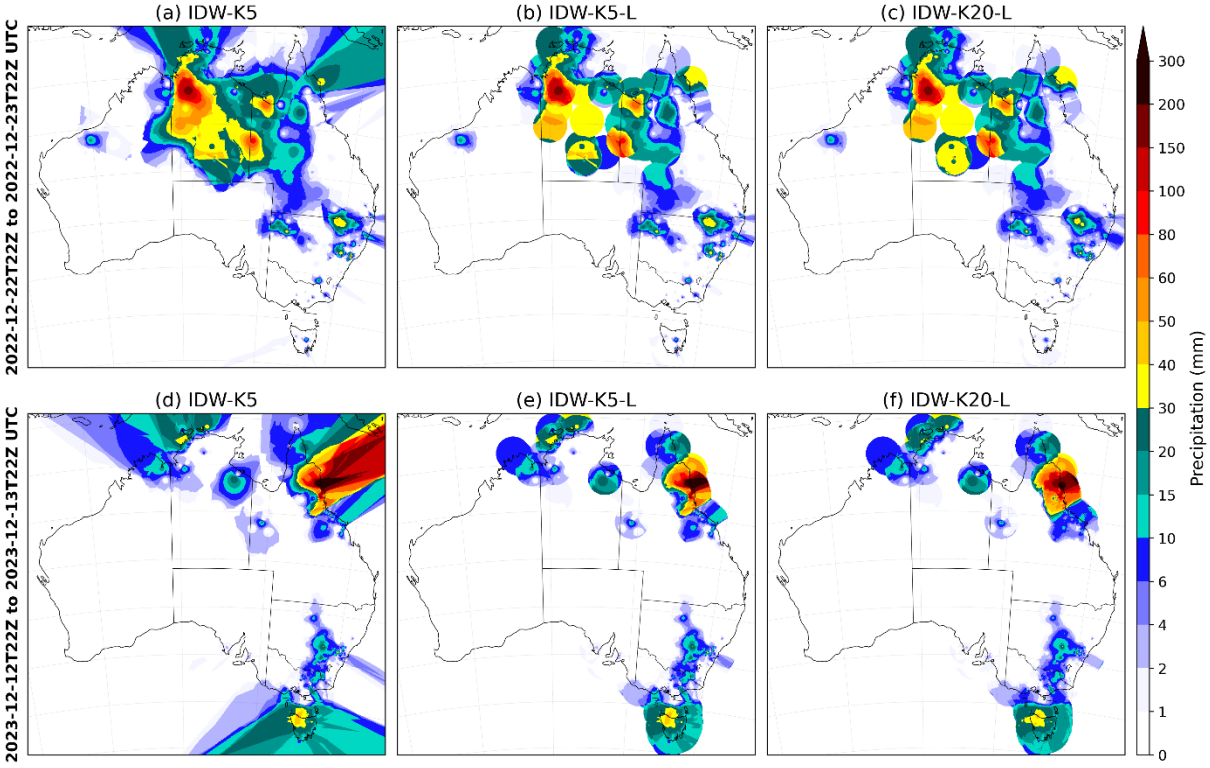


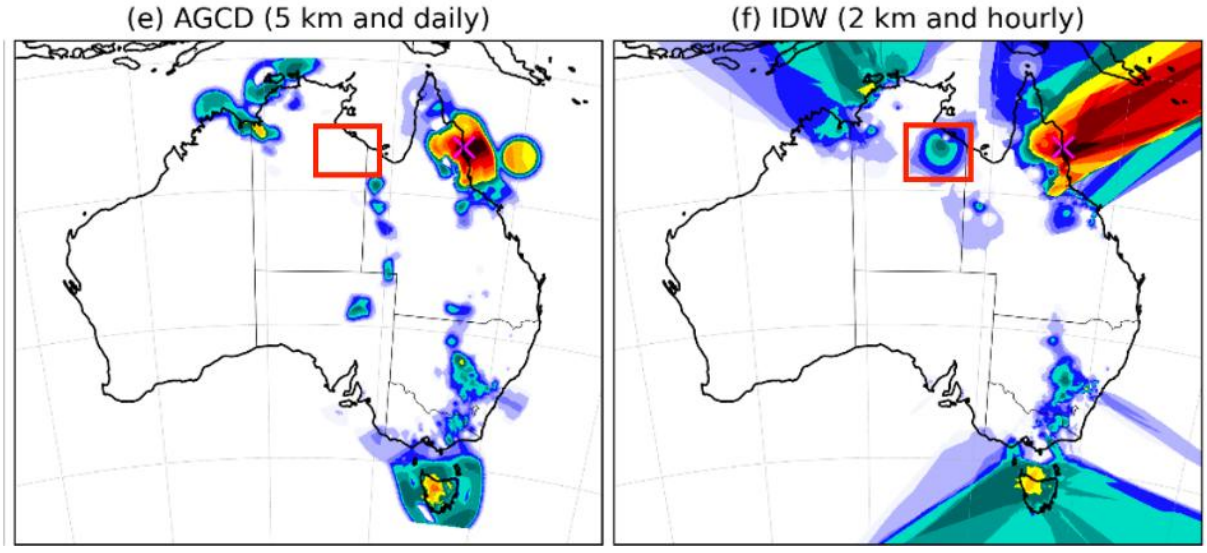
Figure 12. Sensitivity of 24-h accumulated precipitation estimates to different IDW configurations for two selected extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. The three configurations are IDW-K5 (a, d), using the five nearest gauges without a maximum search distance; IDW-K5-L (b, e), using up to five nearest gauges within a 200-km search radius; and IDW-K20-L (c, f), using up to 20 nearest gauges within the same search radius. IDW-K5 corresponds to the IDW configuration used in Fig. 10.

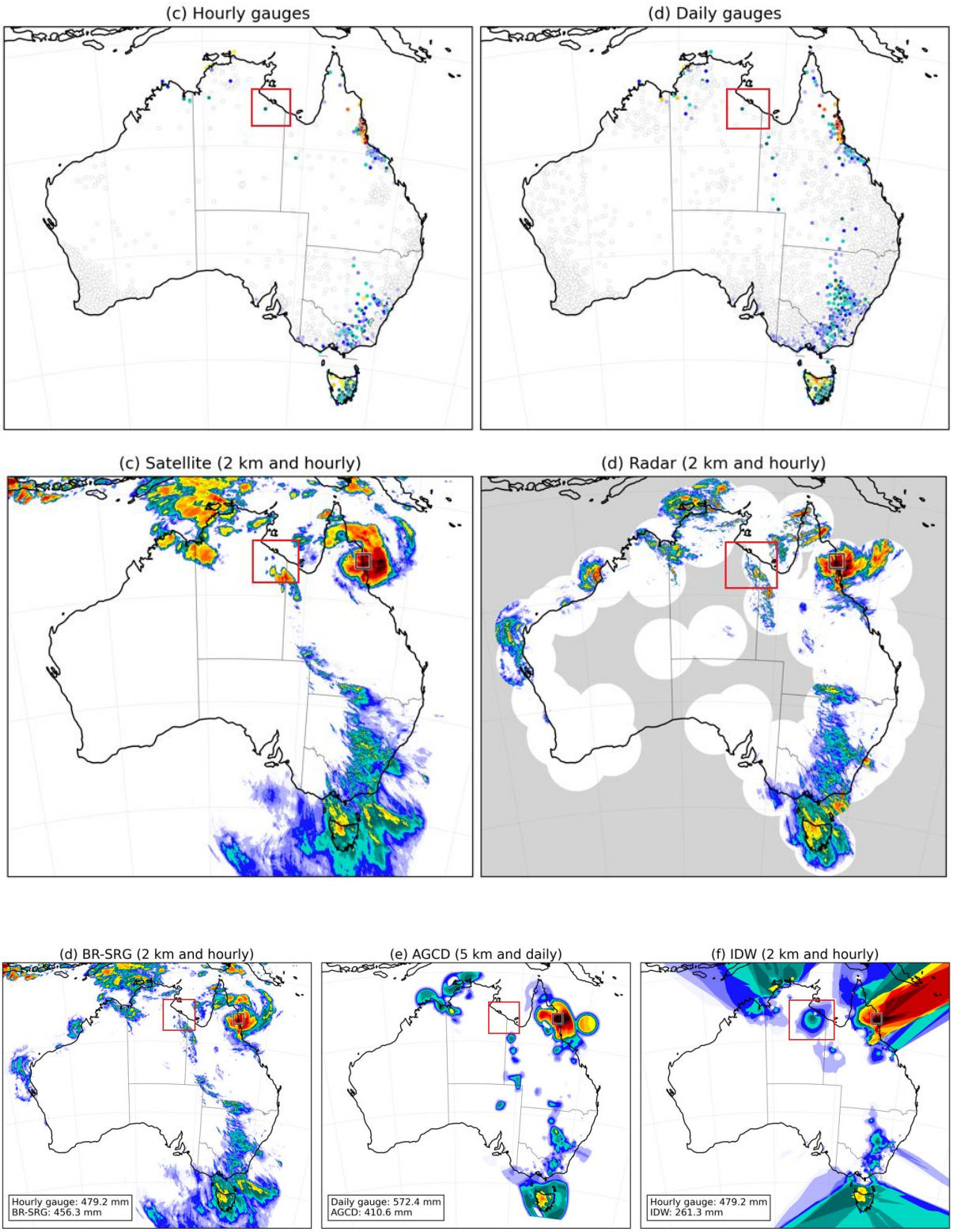
Investigation of the specific discrepancy between IDW and AGCD

We also investigated the specific discrepancy identified, where AGCD indicates no rainfall within the red-box area, whereas the IDW interpolation shows a relatively broad area of approximately 6–20 mm day^{-1}. Examination of the underlying gauge observations showed that an hourly gauge within this area reported rainfall during the accumulation period, and this observation was therefore used in the IDW interpolation. The corresponding daily gauge observation also reported rainfall at this location. In contrast, this rainfall feature is not represented in AGCD. This difference may be related to differences in the gauges used for the AGCD analysis and for IDW analysis and/or their associated quality-control procedures.

This case illustrates a broader challenge in the quality control of hourly gauge observations, particularly in gauge-sparse regions where there may be insufficient neighbouring observations to independently verify an isolated rainfall report. In this case, both the satellite and radar estimates show little or no rainfall around this gauge, providing additional spatial information that is unavailable to gauge-only interpolation. Consequently, while IDW propagates the rainfall reported by this gauge to the surrounding area, its influence is substantially constrained in BRAIN-SRG because the gauge observation differs considerably from the satellite and radar estimates.

We therefore agree that this example demonstrates a limitation of IDW as a gauge-only interpolation method. At the same time, it also highlights the value of incorporating complementary observations into hourly rainfall analysis. Satellite and radar observations can provide additional spatial context for isolated gauge measurements and constrain their influence on the final analysis when substantial discrepancies exist among the different data sources.





To address these issues, we have made the following revisions:

- We have expanded the methodology section to provide a complete description of the original IDW implementation, including the weighting function, number of neighbouring gauges, and interpolation distance settings.

- We have added the distance-limited IDW configurations described above to examine the sensitivity of the IDW rainfall fields to the interpolation settings.
- We have investigated the specific discrepancy identified in the red-box area using the underlying hourly and daily gauge observations together with the corresponding satellite and radar estimates, and have expanded the discussion to explain the different rainfall patterns.
- We have clarified the figure caption and associated text and revised the interpretation of the case study to better explain the differences among IDW, AGCD, and BRAIN.
- Consistent with our response to R2-MC2, we have reduced the emphasis on IDW as a benchmark and now present this comparison as an illustrative case study of the characteristics of different rainfall analysis approaches.

R2-MC4

4. Please justify the influence radius of 200 km in the OI. Is there any sensitivity analysis on that distance? It is unlikely that the rainfall value at one point in the grid is correlated with a point 200 km away. In other studies, the distance used is around 100 km. Figure 4 also shows that the error correlation is less than 0.1 if the distance is more than 100 km. Similarly, why is the calibration radius in R1 expanded to 300 km compared to 150 km in R0? Increasing the interpolation distance can generate the issue of over-smoothing the blended dataset.

Response: Thank you for this important comment. We agree that the rationale for the spatial distance parameters requires further clarification. The 200 km influence radius used in the OI and the radius used for radar bias correction serve different purposes and should be interpreted separately. We address these two issues below.

The 200 km influence radius used in OI

We thank you for raising this point. The 200 km radius represents the maximum localisation boundary for selecting observations in the local OI calculation, rather than the distance over which observations are assumed to have substantial influence. Within this boundary, the effective contribution of an observation is governed by the error-correlation function and decreases with distance. As shown in Figure 4, the estimated error correlation decreases to below approximately 0.1 beyond 100 km. Therefore, observations located more than 100 km away may still be included in the calculation, but their contribution to the analysis is generally small.

To further assess the sensitivity to this parameter, we repeated the OI analysis using a maximum localisation radius of 100 km instead of 200 km. The results show only limited differences in the overall performance of the blended rainfall fields between the two configurations. This is consistent with the rapid decay of the error correlation shown in Figure 4 and indicates that observations between 100 and 200 km have only limited influence on the resulting analysis. We therefore retained 200 km as a conservative choice to avoid unnecessarily excluding potentially useful observations, although this comes at the expense of greater computational demand compared with using a 100 km radius.

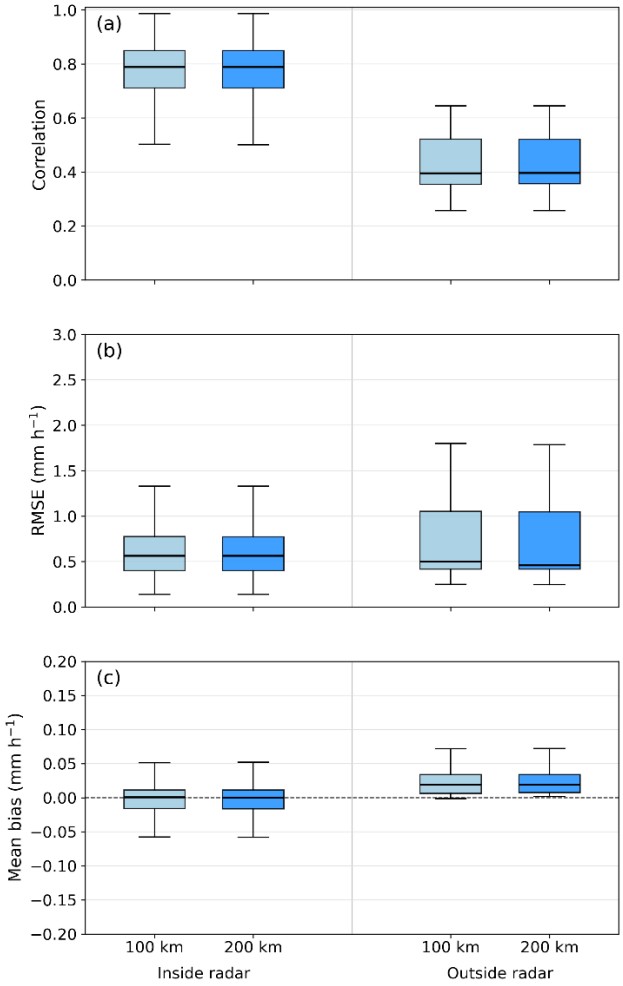


Figure S19. Sensitivity of BR-SRG performance to the influence radius in the hourly-gauge 20-fold cross-validation. Results are shown for influence radii of 100 and 200 km, separately for evaluation gauges within and outside radar coverage. Panels show (a) correlation, (b) RMSE, and (c) mean bias. Each box represents the distribution of station-based metrics across gauges.

We have revised the manuscript to clarify the role of the 200 km localisation radius and indicate that the additional sensitivity analysis shows limited differences between the 100 and 200 km configurations.

A sensitivity test using an alternative influence radius of 100 km showed only minor differences from the 200-km configuration (Fig. S19). We therefore retained the 200-km radius as a conservative choice to avoid excluding potentially useful observations, although this incurs greater computational demand than using a 100-km radius.

To address this issue, an error-informed localisation strategy is adopted within the statistical interpolation framework. Specifically, a highly efficient implementation is introduced in which only the nearest radar grid cell within a 200 km search radius, identified as the grid cell with the minimum Euclidean distance to the target grid cell, is retained as a supplementary observation. If no valid radar grid cell is available within this radius, no radar observation is included for that target grid cell. The 200 km radius defines the maximum search range, while the effective influence of the selected radar observation is further controlled by the distance-dependent error correlation function.

Radar bias-correction radius

The radius used for radar bias correction has a different role from the localisation radius used in OI. It defines the spatial extent for selecting co-located radar–gauge pairs used to estimate the local radar error, rather than the distance over which radar observations influence the subsequent OI analysis. Briefly, radar values are extracted at gauge locations, and the resulting radar–gauge pairs within the specified radius are used to estimate the local radar error and correct the radar field before blending. R0 uses a radius of 150 km, whereas R1 increases this radius to 300 km. The larger radius allows more co-located radar–gauge pairs to be included in the local error estimation, particularly where gauges are sparse. Therefore, the increase from 150 to 300 km is intended to provide sufficient radar–gauge pairs for robust local error estimation, rather than to increase the interpolation or influence distance of radar observations.

We have revised the manuscript to clarify the purpose of the radar bias-correction radius and its distinction from the OI localisation radius. We have also clarified why a larger radius is used in R1, where more co-located radar–gauge pairs are required to support local radar error estimation under sparser gauge availability.

3.3 Local bias correction of satellite and radar precipitation estimations

The calibrated radar estimates (R1) used in this study differ from the operational Rainfields product (R0) in both temporal and spatial configuration. R0 applies bias correction using gauges within a 150-km radius and a 15-min window. The spatial radius limits the number of nearby co-located radar–gauge pairs available for local bias estimation, while the temporal window is constrained by gauge reporting latency and the short operational update interval. In contrast, R1 uses gauges within a 300-km radius and a 1-hour window, increasing the number of radar–gauge pairs available for estimating the local radar bias. No predefined limit is imposed on the number of pairs within this radius. All available pairs are considered, with their point-scale bias estimates subsequently interpolated using ordinary kriging to construct a spatially continuous bias field. The larger spatial radius allows more spatially distributed gauge observations to be included, while the longer temporal window accommodates additional observations associated with gauge reporting latency. This approach also uses the same available gauge observations for satellite and radar bias correction, subject to the availability of radar observations at each gauge location. However, this configuration is tailored to the Australian radar system and can be adjusted for other radar systems according to their operational configuration and production objectives.

R2-MC5

5. Because the distribution of rain gauges and radar coverage in Australia is denser in the coastal areas, I suggest separating the interpolation and validation for regions with high and low rain-gauge density, and targeting different interpolation weights for gauged data in each region (i.e., put higher weights on gauge measurements in the areas with dense rain gauges). For example, Figures 7a and 7b show an issue of mismatch between the BR-SRG and AGCD datasets, in the area with dense daily rain gauges. This may reflect artifacts in the radar and/or satellite inputs, which are not sufficiently discussed in the manuscript. In this case, I believe the AGCD dataset is more reliable because of the dense distribution of rain gauges in the area to detect rainfall on the ground.

Response: Thank you for this helpful suggestion. We agree that gauge density is an important factor affecting rainfall analysis and validation. We have considered this issue from three aspects, as described below.

Theoretical influence of gauge density on the blending

We understand your suggestion that gauge observations should receive higher weights in regions with dense rain-gauge networks. This behaviour is already inherent in the statistical interpolation framework used in BRAIN. The observation weights are not prescribed according to predefined high- or low-density regions; instead, they are determined automatically from the estimated error covariance structure and the spatial configuration of the available observations. Consequently, when more gauges are available around a target grid point, information from the gauge network is more strongly represented in the local OI analysis. We therefore do not prescribe separate interpolation weights for predefined high- and low-density regions.

Validation under different gauge densities

Following your suggestion, we have added evaluations stratified by local gauge density (see Fig.8 below as an example). To isolate the effect of gauge density from that of radar availability, this analysis is restricted to stations within radar coverage. Gauge density is defined as the mean number of available blending gauges within 200 km of each validation station. Stations at or below the 25th percentile and at or above the 75th percentile are classified as low- and high-density groups, respectively, while the middle 50% are excluded to provide a clearer contrast.

We have added the gauge-density-stratified evaluation to the revised manuscript and expanded the discussion to explicitly assess the influence of gauge density on the validation results.

4.3.3 Performance under different gauge-density conditions

Gauge-density effects were examined using the hourly-gauge 20-fold cross-validation (Fig. 8). The analysis was restricted to areas within radar coverage because radar observations are unavailable outside these areas and the gauge network is generally too sparse for a meaningful comparison between gauge-density conditions. First, gauge density affected the magnitude of performance but did not change the overall relative performance among the rainfall estimates. BR-SRG consistently achieved the best overall performance under both low- and high-density conditions. Second, the influence of gauge density was most evident in correlation, which generally increased under high-density conditions across the gauge-dependent estimation processes. Changes in RMSE and mean bias were less consistent, likely reflecting differences in rainfall characteristics between the low- and high-density regions in addition to gauge-density effects. This density dependence was particularly pronounced for gauge-only IDW interpolation, which showed substantially lower and more variable correlations under low-density conditions but much higher and more consistent correlations under high-density conditions. Notably, however, its mean bias remained close to zero under both low- and high-density conditions.

The relative performance of the rainfall estimates remained consistent across all three evaluation datasets. BR-SRG maintained the best overall performance under both low- and high-density conditions in the independent daily-gauge (Fig. S4) and independent hourly-gauge (Fig. S18) evaluations, confirming that the overall ranking was robust to variations in gauge density. The improvement from low- to high-density conditions was more evident in these independent evaluations, further highlighting the importance of gauge availability for precipitation bias correction, multi-source blending, and gauge-based interpolation.

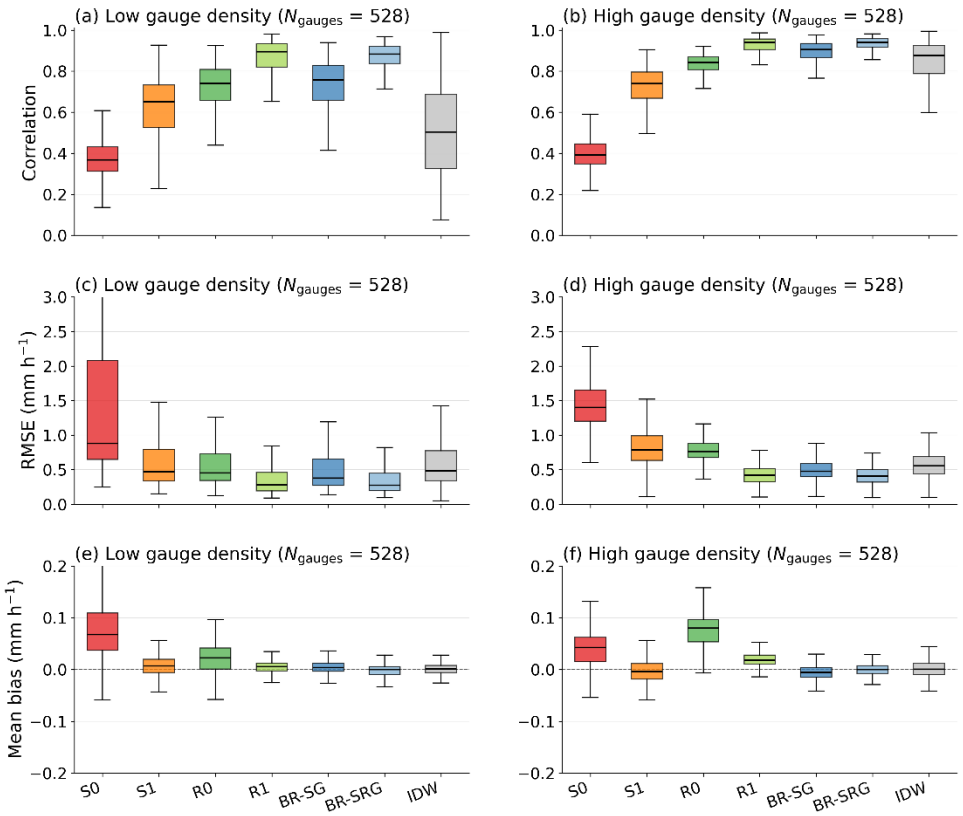
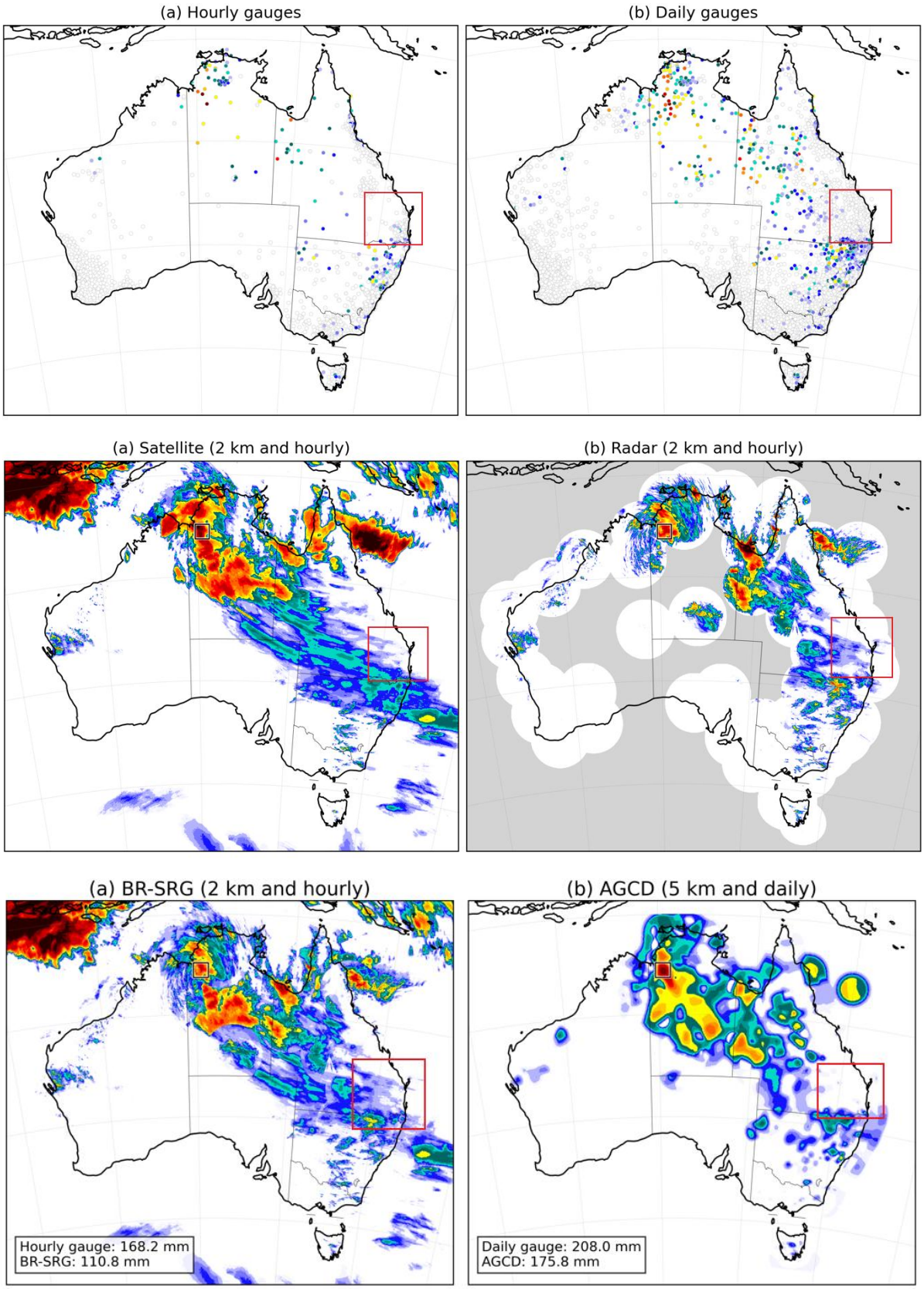


Figure 8. Performance of different rainfall estimates under low- and high-gauge-density conditions in the hourly-gauge 20-fold cross-validation, considering evaluation gauges within radar coverage only. Low- and high-density conditions correspond to the lower (≤25th percentile) and upper (≥75th percentile) tails of the gauge-density distribution, respectively, with gauge density defined as the number of available gauges within a 200-km radius. Panels show (a, b) correlation, (c, d) RMSE, and (e, f) mean bias. Each box represents the distribution of station-based metrics across 528 gauges.

Difference between BRAIN-SRG and AGCD in Figure 7

We also examined the specific mismatch between BRAIN-SRG and AGCD that identified in Figure 7. First, we note that the gauge markers shown in the figure are enlarged for visualisation and therefore should not be interpreted as representing the actual spatial coverage of individual gauges. Second, the two analyses use different gauge networks and operate at different temporal scales:

- AGCD is a daily rainfall analysis and can incorporate the substantially denser daily gauge network.
- BRAIN produces hourly rainfall analyses and uses only hourly gauges, which are considerably less dense.



More importantly, examination of the underlying observations shows that several gauges in the highlighted area reported light rainfall during this period, although this rainfall is not represented in the corresponding AGCD field. Some light rainfall signals are also present in both the satellite and radar observations. Therefore, the light rainfall represented by BRAIN-SRG in this area is not supported by a single isolated data source, but by information from

multiple observational sources. Through the blending process, these consistent signals from gauges, radar, and satellite contribute to the light rainfall represented in the final BRAIN-SRG analysis.

We therefore consider the light rainfall in BRAIN-SRG to be supported by multiple lines of observational evidence, although its exact magnitude and spatial distribution remain uncertain. This example also illustrates an important motivation for multi-source rainfall blending: when individual rainfall products or observing systems provide different information, combining complementary observations can provide a more robust basis for rainfall analysis than relying on any single source alone.

We have revised the introduction of AGCD and the caption of Figure 7 (now Fig. 10) to better clarify the differences between AGCD and BRAIN in terms of the gauge networks used and their temporal resolutions.

3.5.3 IDW implementation and comparison with AGCD

…

At the daily scale, BRAIN was compared with the AGCD product, an existing national operational daily gridded rainfall product. AGCD is a 5-km gauge-based gridded analysis derived from daily gauge observations using an anomaly-based approach combining Barnes successive-correction analysis with a climatological background (Jones et al., 2009). The daily gauge network used for AGCD is substantially denser than the sub-daily gauge network available to BRAIN. Hourly BRAIN estimates were accumulated to daily totals for comparison with AGCD. The BRAIN daily accumulation accounts for the UTC-to-local-time conversion, although exact alignment with the daily accumulation period at every location is not possible because Australia spans multiple time zones, which may introduce minor temporal mismatches. These differences in input gauge networks, temporal resolution, accumulation periods, and product design should therefore be considered when interpreting the comparison between BRAIN and AGCD.

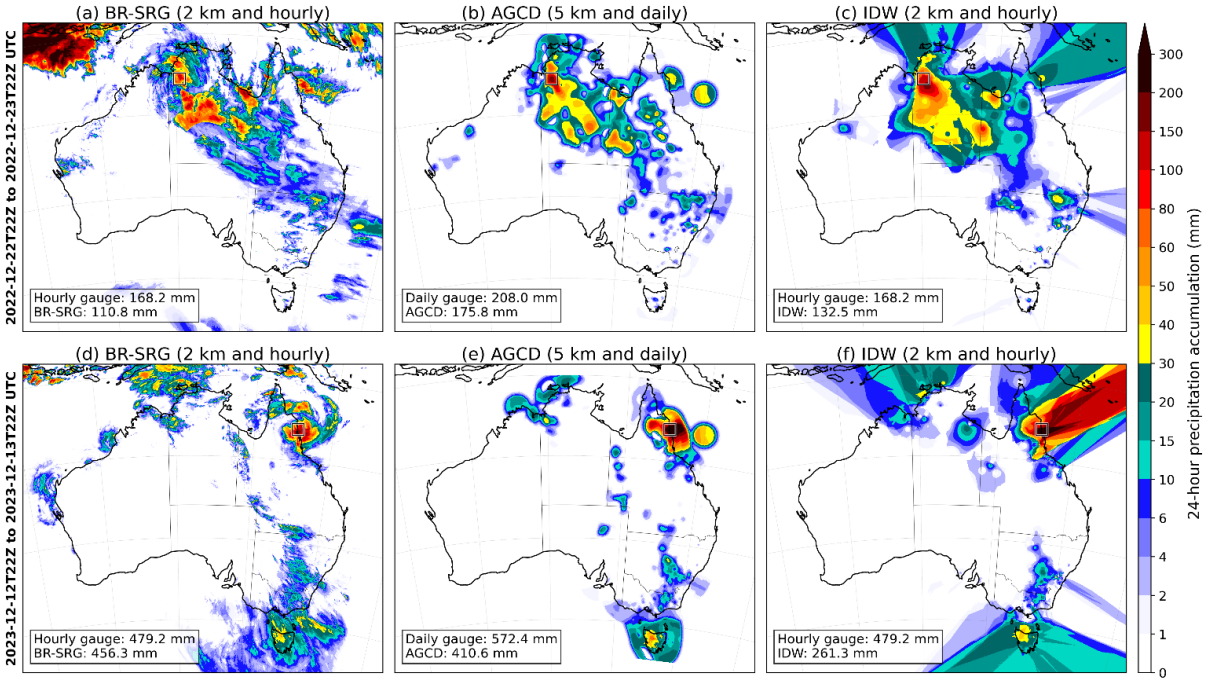


Figure 10. Example rainfall analyses from the three-source blend (BR-SRG; 2-km and hourly), AGCD (5-km and daily), and gauge-based IDW interpolation (2-km and hourly) for two selected 24-h periods during extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. Note that AGCD incorporates the denser daily-gauge network (Fig. S6b, d), whereas BR-SRG and IDW use the comparatively sparser hourly-gauge network (Fig. S6a, c). BR-SRG and IDW are accumulated over the indicated UTC periods, whereas AGCD represents the standard daily gauge accumulation period of approximately 9 am to

9 am local time, ending on 23 December 2022 and 14 December 2023, respectively, with the exact UTC period varying across Australian time zones. The red square indicates the location of the gauge with the highest observed 24-hour precipitation for each event. Further details of the satellite and radar estimates and gauge observations are provided in Figs. S5–S6 and Tables S1–S2.

R2-MC6

6. The hourly rainfall measurements contain many zeros, so even though the RMSE values seems to be low, the dataset can have an issue of significantly underestimating high rainfall values. Figure7 shows that the BR-SRG can underestimate the maximum rainfall substantially compared to the hourly and daily observed data (as noted in the figure caption). This is a critical limitation, as accurate representation of extreme rainfall is essential for applications such as storm and flood analysis. I recommend extending the validation to explicitly assess performance across different rainfall intensities (e.g., low, moderate, and high thresholds). In addition, presenting time series comparisons (or RMSE time series) at selected key locations would provide a more robust and informative evaluation of the dataset's performance, particularly during high-impact events.

Response: Thank you for this important comment. We agree that evaluation across different rainfall intensities and during high-impact events provides important additional information beyond the overall performance metrics. Following your suggestions, we have extended the evaluation in both aspects.

Performance across different rainfall intensities

We have extended the evaluation by separately assessing low-, moderate-, and high-intensity rainfall. We performed this analysis for all three validation settings used in the study: 20-fold hourly gauge cross-validation, independent daily gauge validation, and independent hourly gauge validation. For conciseness, only the independent hourly gauge validation results are shown below as a representative example.

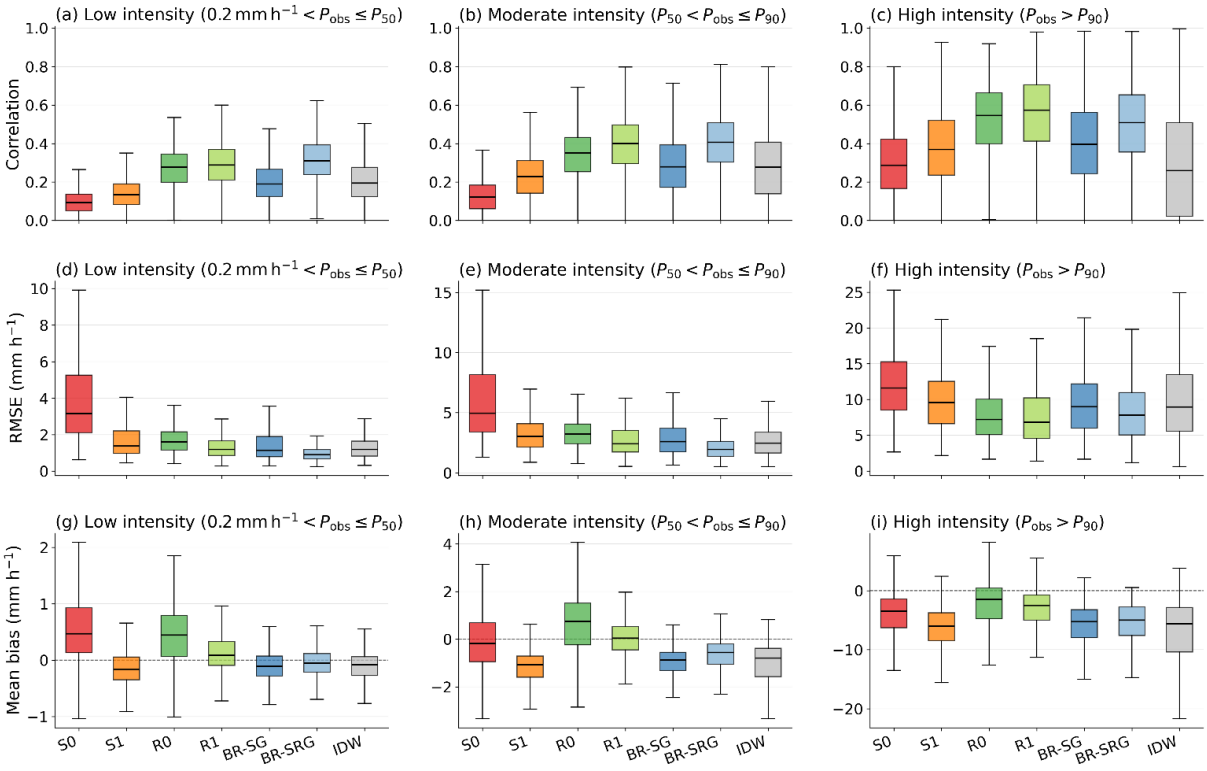


Figure S17. Performance of different rainfall estimates across rainfall-intensity categories in the independent hourly-gauge evaluation. At each gauge, observed rainfall was divided into three categories: low (0.2 mm h^{-1} <

Pobs ≤ P50), moderate (P50 < Pobs ≤ P90), and high (Pobs > P90), where P50 and P90 are the 50th and 90th percentiles of observed rainfall above 0.2 mm h^{-1} at each gauge, respectively. Panels show (a–c) correlation, (d–f) RMSE, and (g–i) mean bias. Each box represents the distribution of station-based metrics across gauges. As the rainfall-intensity categories are defined based on observations, the mean bias for these categories should be interpreted with caution, particularly for high-intensity rainfall, where observation-based stratification can contribute to a tendency towards negative conditional bias.

The intensity-based evaluation provides additional insight into the performance of the different rainfall estimates. BRAIN-SRG maintains relatively high correlation across all three intensity classes. Its RMSE also remains comparatively low, although R1 performs slightly better than BRAIN-SRG for high-intensity rainfall. The mean-bias results show increasingly negative values from moderate- to high-intensity rainfall for BRAIN-SRG. However, these intensity-stratified bias results should be interpreted with caution because the rainfall-intensity classes are defined based on observations, which can contribute to a tendency towards negative conditional bias, particularly at high intensities.

Nevertheless, the relative differences among the rainfall estimates suggest that the satellite input may contribute to the negative bias in BRAIN-SRG. As shown in the figure, S1 exhibits a clear negative bias for moderate- and high-intensity rainfall, whereas R1 has a much smaller bias. Because BRAIN-SRG combines information from satellite, radar, and gauges according to their estimated error characteristics, the satellite still retains a non-negligible contribution to the blended estimate. Consequently, its tendency to underestimate heavier rainfall may be partially propagated to BRAIN-SRG, leading to the negative bias observed for moderate- and high-intensity rainfall. However, the inclusion of radar information reduces this negative bias, with BRAIN-SRG showing smaller underestimation than BRAIN-SG, particularly for moderate- and high-intensity rainfall. This demonstrates the benefit of incorporating complementary radar information, although the underestimation inherited partly from the satellite input is not completely eliminated. Further improvements in satellite rainfall retrieval and bias correction, particularly for heavier rainfall, could therefore further improve the representation of high-intensity rainfall in BRAIN.

We have added this rainfall-intensity-stratified evaluation to the revised manuscript and expanded the discussion to explicitly describe the intensity-dependent performance and the propagation of errors from the input rainfall products into the blended analysis.

For example,

Comparison with the radar estimates. The bias-corrected radar product (R1) showed strong overall performance within the radar coverage. In the hourly cross-validation, R1 had slightly higher correlation and lower RMSE than BR-SRG, whereas BR-SRG showed a smaller absolute bias (Fig. 5). In contrast, BR-SRG generally showed higher correlation, lower RMSE, and smaller absolute bias than R1 in both the independent daily-gauge (Fig. 6) and independent hourly-gauge (Fig. S16) evaluations. The rainfall-intensity analysis provides further insight into these differences. BR-SRG generally performed better than R1 in terms of correlation and RMSE at low and moderate rainfall intensities, whereas R1 performed better at high rainfall intensities. For bias, however, R1 generally showed smaller absolute values than BR-SRG across the rainfall-intensity categories, with the difference becoming more evident at high intensities (Fig. 7, Figs. S3 and S17). However, the intensity-stratified bias results should be interpreted with caution, as the rainfall-intensity categories are defined based on observations, which can contribute to a tendency towards negative conditional bias, particularly at high intensities.

Time-series evaluation during high-impact events

Following your suggestion, we have also added time-series comparisons at selected gauge locations during two high-impact rainfall events, Cyclone Jasper in December 2023 and Cyclone Ellie in December 2022.

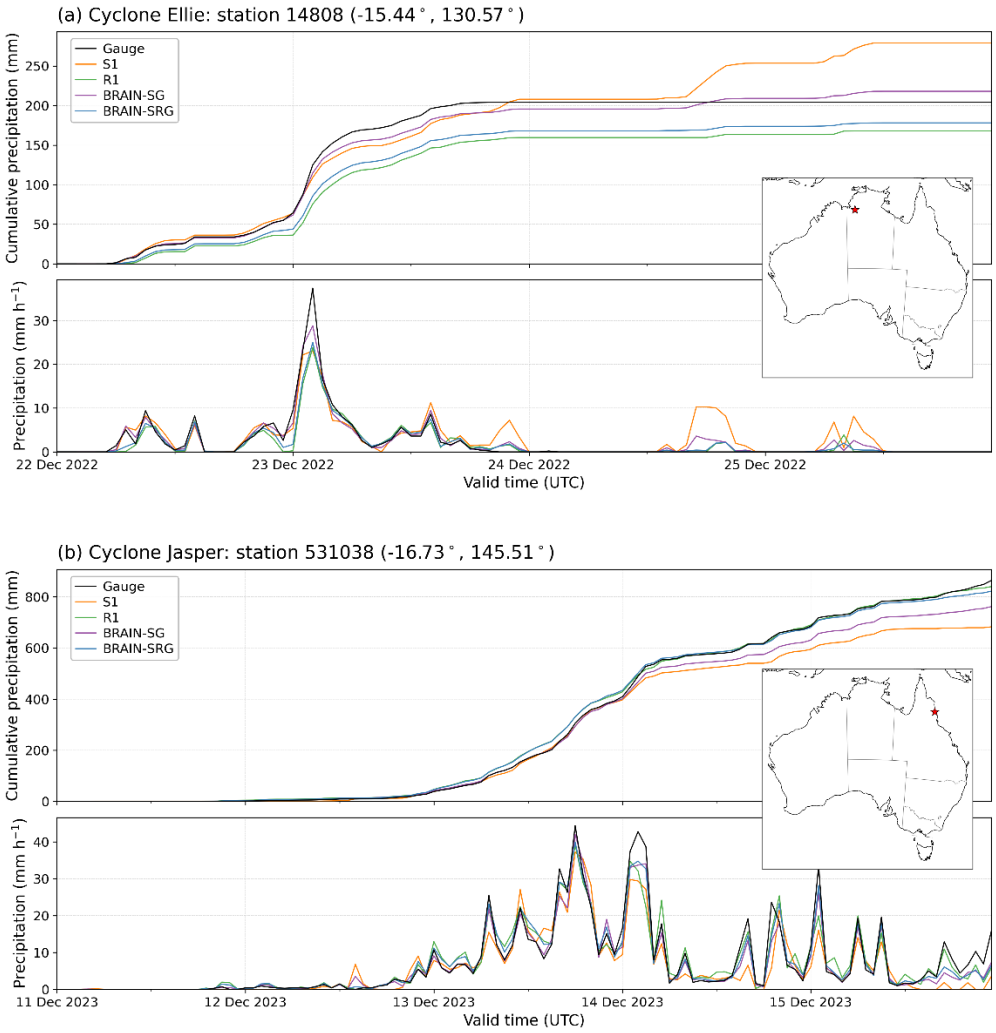


Figure 14. Temporal evolution of precipitation at selected gauge locations during (a) Cyclone Ellie and (b) Cyclone Jasper. Each panel shows cumulative precipitation (top) and hourly precipitation (bottom) from gauge observations and different rainfall estimates. Insets indicate the locations of the selected gauges.

One limitation of the current framework is the propagation of residual errors from input observations into the blended analysis. Residual errors that are not identified through quality control, particularly for gauge observations, or are not fully removed by bias correction, particularly for satellite and radar estimates, can propagate into the blended analysis. Although the statistical interpolation weighting accounts for the intensity-dependent error characteristics of each source, an individual source generally does not receive complete weight, as this would require the error of the other sources to become overwhelmingly large in comparison. Consequently, residual biases in sources retaining non-zero weights can propagate into the final analysis.

For example, during Cyclone Jasper, underestimation in the satellite estimates contributed to underestimation in BR-SRG even when radar closely represented the observed accumulation (Fig. 14a). Conversely, during Cyclone Ellie, radar substantially underestimated the observed accumulation, while the satellite estimates were higher and partly compensated for the radar underestimation in the blended analysis (Fig. 14b). These contrasting cases demonstrate the value of retaining complementary information from multiple data sources, rather than simply increasing the contribution of a single source.

Together with the relative differences among S1, R1, BR-SG, and BR-SRG in the rainfall-intensity analysis (Fig. 7), these cases suggest that residual underestimation in the satellite estimates may contribute to the negative bias observed in BR-SRG, although the intensity-stratified bias is also affected by observation-based stratification.

Further improvements should therefore focus on bias correction, particularly in better accounting for systematic biases across different rainfall intensities.

As noted above, these comparisons are also subject to spatial representativeness differences between point-scale gauge observations and gridded rainfall estimates, which can be particularly important for localised intense rainfall. Nevertheless, the two event-based analyses provide additional evidence that BRAIN can reproduce the temporal evolution of major rainfall events while also highlighting its limitations in representing some local rainfall extremes.

We have added these event-based time-series comparisons to the revised manuscript and expanded the discussion of BRAIN performance during high-impact rainfall events.

Detailed Comments:

R2-DC1

Lines 85-88: Please add references for stating that IDW was used in flood modelling and flood forecasting. From my experience in flood modelling, the rainfall dataset generated using the IDW interpolation is very poor and should not be used as an input in flood models.

Response: Thank you for this comment. We agree that our original wording did not sufficiently distinguish the operational use of IDW for estimating basin/sub-basin rainfall from its use for generating gridded rainfall products.

IDW is the current official methodology used by the Bureau for estimating basin and/or sub-basin rainfall from sub-daily gauge observations for operational flood forecasting and hydrological modelling, rather than for generating a regional or national gridded rainfall product. We have revised the relevant text to make this distinction explicit and added references to Malone (1999) and Malone and Carroll (2022) to support this statement.

Please refer to our detailed response to R2-MC2 for further discussion of the role of IDW in this study and the corresponding revisions.

R2-DC2

Lines 125-127: This validation method is not new, and it has been applied in other studies (Yu et al. 2020; Xia et al. 2025), so I don't think it is a contribution of the paper.

Response: Thank you for pointing this out. We agree that this validation approach has been applied in previous studies and should not be presented as a methodological contribution of this study. Our intention was to highlight the comprehensive evaluation of BRAIN under different validation settings rather than to claim novelty in the validation method itself.

We have revised relevant statements in the manuscript to remove this validation approach as a methodological contribution. We have also added the references (Yu et al., 2020; Xia et al., 2025) where appropriate to acknowledge its previous applications.

The study makes several key contributions:

(i) It systematically examines the role of the latest Himawari CRRPH satellite rainfall product as a high-resolution background field and investigates its interaction with radar and sub-daily gauge observations within the blended analysis, building on recent advances by the Bureau in satellite-based rainfall estimation.

(ii) It presents a continental-scale trial implementation for Australia that integrates all currently available and operationally manageable satellite, radar, and gauge observations within a unified framework, extending previous multi-source efforts towards a potentially operational national product.

(iii) It applies and tailors statistical interpolation theory to Australian observational and computational constraints, including the adoption of an error-informed localisation strategy and a simplified nearest-point representation for radar observations, thereby enabling scalable, grid-cell-wise and time-step-wise blending at kilometre and hourly resolutions with minimal computational overhead.

R2-DC3

Section 2.1: Please add more details about the quality control of gauge data (what values are considered extremely high and why).

Response: Thank you for this comment. We agree that the criteria used to identify unusually high gauge observations require further explanation.

The gauge quality-control procedure includes several complementary checks rather than applying a single fixed threshold for extreme rainfall. In particular, anomalously high values are assessed against (1) the historical rainfall distribution at the station, with values exceeding two and four times the historical 99.9th percentile flagged at different levels; (2) local Intensity-Frequency-Duration (IFD) statistics, with observations exceeding the 1% Annual Exceedance Probability (AEP), 1-in-1000-year, and twice the 1-in-1000-year rainfall thresholds progressively flagged; and (3) surrounding gauge observations, where unusually high rainfall is flagged when it differs substantially from nearby stations within 50 km. These complementary checks are intended to distinguish potentially erroneous extreme values from genuine heavy rainfall events using temporal climatology, local rainfall extremes, and spatial consistency. The quality-control procedure also includes checks for other potential issues, such as repeated identical non-zero values and suspicious zero-rainfall periods.

We have expanded Section 2.1 to describe these quality-control procedures and provide the criteria used to identify potentially anomalous high rainfall observations.

2.1 In situ observations

…

Both the hourly and daily datasets were subjected to rigorous quality control procedures, including checks for anomalously high values based on historical rainfall distributions and gridded Intensity-Frequency-Duration (IFD) statistics, spatial intercomparisons to detect anomalous high values or prolonged periods of zero values indicative of instrument failure or gauge blockage, intercomparisons between sub-daily and daily observations where both were available, and checks for long sequences of identical non-zero values. Specifically, potential rainfall spikes were identified using thresholds of two and four times the historical 99.9th percentile. Extreme rainfall observations were also checked against local IFD statistics, including the 1% Annual Exceedance Probability (AEP) and 1-in-1000-year rainfall thresholds. Spatial consistency checks further compared unusually high observations with neighbouring gauges within 50 km. These complementary checks use historical, climatological, and spatial information to identify potentially erroneous high rainfall values while retaining plausible extreme rainfall observations.

R2-DC4

Lines 170-172: What is the source and the reliability of the additional independent hourly data? Why are they not included in the interpolation and only used for extra validation?

Response: Thank you for this important question. We agree that the source and reliability of these additional hourly observations, and the rationale for using them only for validation, should be explained more clearly.

These additional hourly gauge observations were obtained from third-party providers, including local councils, water utilities, and other water management organisations, through ongoing data-sharing arrangements. The database has continued to expand throughout the project. Importantly, these observations were subjected to the

same quality-control procedures as the primary hourly gauge dataset. This ensures a consistent level of quality control between the hourly gauges used for blending and those used for independent validation.

At the time when the bias correction and blending analyses were performed, these additional observations were not yet available and therefore could not be incorporated into the product generation. They became available and were subsequently collated after these analyses had been completed. We therefore used these newly acquired observations as an independent validation dataset rather than retrospectively incorporating them into the existing analyses. This also provides a valuable independent assessment because none of these observations were used in the bias correction or blending processes.

To address this comment, we have revised the manuscript to:

- provide a clearer description of the sources, characteristics, and quality control of the additional hourly gauge observations;
- explain that these observations became available after the bias correction and blending analyses had been completed and were therefore used as an independent validation dataset.

2.1 In situ observations

…

An additional independent hourly dataset (approximately 2000 stations) was obtained from third-party providers, including local councils, water utilities, and other water management organisations through ongoing data-sharing arrangements. As this dataset has continued to expand throughout the project, these observations became available after completion of the bias correction and blending analyses and were therefore used for supplementary independent validation. These observations were subjected to the same quality-control procedures as the primary hourly gauge dataset. The spatial distribution of these hourly gauge stations is shown in Fig. S15.

R2-DC5

Section 2.2: Since when has radar rainfall data been available in Australia? Please justify whether increasing the calibration distance from 150 km to 300 km to include more gauges makes sense, or if the improvement is mainly from increasing the time window (15-min to 1-hour).

Response: Thank you for these insightful questions. We address the two issues separately below.

Availability of radar rainfall data in Australia

Weather radar observations have been available in Australia since the late 1990s. The network initially covered mainly major population centres, such as Sydney and Melbourne, and has since been progressively expanded. The radar network (2022-2023) used in this study provides coverage of approximately 72% of the Australian land area.

We have revised Section 2.2 to provide this background information and clarify the temporal availability and spatial coverage of radar observations in Australia.

2.2 Radar-based observations

Radar rainfall estimates (Bureau of Meteorology, 2022) were obtained from the Bureau's operational Rainfields system, which provides real-time data at a 1-km/5-min resolution. Weather radar observations have been available in Australia since the late 1990s, with the network progressively expanding from coverage concentrated around major population centres to its current broader coverage. The radar network (2022-2023) used in this study covers approximately 72% of the Australian land area. Further information on Australian weather radar data availability and archiving is provided through the Australian Unified Radar Archive (https://opus.nci.org.au/spaces/NDP/pages/399803502/Australian+Unified+Radar+Archive).

Radar calibration radius and temporal window

The operational radar calibration currently used by the Bureau typically considers gauges within a 150 km radius and a 15-minute temporal window. In this study, we increased the spatial radius to 300 km and the temporal window to 1 hour. Both modifications were intended to increase the number of co-located radar–gauge pairs available for estimating the local radar error, but they address different limitations.

Increasing the spatial radius from 150 to 300 km allows more co-located radar–gauge pairs to be included in the local error estimation, particularly in areas with sparse gauge coverage. Extending the temporal window from 15 minutes to 1 hour addresses a different issue by allowing more gauge observations to become available given their reporting latency. Therefore, the larger spatial radius addresses the spatial availability of gauges, whereas the longer temporal window addresses their temporal availability.

We have revised Section 2.2 and Section 3.3 to clarify the different purposes of the spatial and temporal windows and the rationale for their use in the radar calibration.

2.2 Radar-based observations

…

In this study, we used hourly radar accumulations derived from 5-min calibrated radar rain rates as a blending input for the target hourly resolution. However, we recalibrated the radar data using gauges within a 300-km radius over a 1-hour window to increase the availability of co-located radar-gauge pairs for estimating local radar errors. The larger spatial radius allows more co-located radar-gauge pairs to be included, particularly in gauge-sparse regions, while the longer temporal window increase gauge availability by accommodating their reporting latency. These two modifications therefore address the spatial and temporal availability of gauge information, respectively. We refer to the operational calibrated product as R0 and the recalibrated version as R1. In addition, to match the spatial resolution with satellite data, the original 1-km radar grids were aggregated to 2 km. Figure 1c shows an example of the calibrated radar rainfall hourly accumulation (R1) at a continental scale, recorded at 0800 UTC on 1 March 2022.

3.3 Local bias correction of satellite and radar precipitation estimations

Bias correction for satellite and radar data followed a Kalman filtering approach similar to Chumchean et al. (2006), but was adapted from the original mean-field framework to estimate spatially varying local bias fields at the target hourly timescale. Bias estimation was based on co-located satellite/radar–gauge pairs, obtained by extracting satellite or radar estimates at gauge locations and pairing them with the corresponding gauge observations. All available gauges were included in the satellite bias correction, whereas only gauges within radar coverage were included in the radar bias correction because radar estimates are unavailable outside the radar coverage area. Bias at each gauge location was estimated using a Kalman filter with an AR(1) model to capture temporal dependence, and the resulting point estimates were interpolated using ordinary kriging to generate a continuous spatially varying bias field. The corrected product was then obtained by applying the bias field multiplicatively to the raw product.

The calibrated radar estimates (R1) used in this study differ from the operational Rainfields product (R0) in both temporal and spatial configuration. R0 applies bias correction using gauges within a 150-km radius and a 15-min window. The spatial radius limits the number of nearby co-located radar–gauge pairs available for local bias estimation, while the temporal window is constrained by gauge reporting latency and the short operational update interval. In contrast, R1 uses gauges within a 300-km radius and a 1-hour window, increasing the number of radar–gauge pairs available for estimating the local radar bias. No predefined limit is imposed on the number of pairs within this radius. All available pairs are considered, with their point-scale bias estimates subsequently interpolated using ordinary kriging to construct a spatially continuous bias field. The larger spatial radius allows

more spatially distributed gauge observations to be included, while the longer temporal window accommodates additional observations associated with gauge reporting latency. This approach also uses the same available gauge observations for satellite and radar bias correction, subject to the availability of radar observations at each gauge location. However, this configuration is tailored to the Australian radar system; for different radar systems, adjustments can be made based on their operational configuration and production objectives.

R2-DC6

Section 3.2: The method mentioned in this section is under cited. Equations 1-5 were developed in other studies. Please add references to the original paper that developed the OI method (Gardin L., 1965), and other studies had already used the method to interpolate blended rainfall datasets. In this section, the radar-to-radar and radar-gauge correlations are ignored. However, according to Seo et al. (2011), the covariance of radar-gauge can be considerable at large scales due to the significant variability. Please justify.

Response: Thank you for these comments and helpful references. We address the two issues separately below.

References for the OI methodology

We agree that Section 3.2 was under-cited. The OI formulation used in this study is based on established statistical interpolation theory rather than being newly developed here. We have therefore added the original reference to Gandin (1965), together with more recent and accessible studies describing the OI methodology and its application to precipitation analysis and multi-source rainfall blending. These references have been added around Equations (1)–(5) to more clearly acknowledge the methodological basis of the framework.

3.2 Multi-source statistical interpolation

A range of approaches can be selected for producing blended gridded datasets (e.g., Yin et al. 2021; Gavahi et al. 2023; Khedhaouiria et al. 2022; Chappell et al. 2013; Shen et al. 2014; Yu et al. 2020). Statistical interpolation, also commonly referred to as optimal interpolation (OI), was originally developed within the framework of objective meteorological analysis (Gandin, 1965) and has been widely used interpolation method in many meteorological organisations (e.g., Mitchell et al. 1990; Huang et al. 1997; De Mey 1994; Lorenc et al. 1991) for blending or assimilating observations with weather models, particularly before the development of more advanced variational methods (e.g., Gauthier et al. 2007; Rawlins et al. 2007; Rabier et al. 2000; Rabier 2005). More recently, OI and related statistical interpolation approaches have been widely applied to precipitation interpolation and blending (e.g., Bergman, 1978; Chua et al., 2022c; Mahfouf et al., 2007; Xia and Wang, 2025; Xie and Xiong, 2011; Yu et al., 2020).

Statistical interpolation enables the integration of multi-source precipitation data by updating a background estimate—in this case, the satellite-derived first guess—based on observations (gauge/radar fields) to generate an analysis, with adjustments weighted according to the estimated error characteristics of each data source. The formulation adopted in this study follows the established OI framework (Gandin, 1965; Xie and Xiong 2011; Yu et al. 2020; Mahfouf et al. 2007; Zhang et al., 2026), as described by Eqs. (1)-(5) below. The method assumes Gaussian error distributions, which may not strictly hold for precipitation but are commonly used approximations to avoid added complexity from data transformation, while bias correction is applied to both satellite and radar data to help satisfy the assumption of unbiased errors. Statistical interpolation operates as an iterative, grid-based analysis adjusting each target grid cell using observations within a predefined search radius (e.g., 200 km).

Radar–radar and radar–gauge error correlations

In this study, we adopted a simplified error covariance structure to maintain a computationally tractable framework for national-scale, near-real-time precipitation analysis.

For radar–radar error correlations, BRAIN uses a nearest-radar approximation, whereby only the nearest radar grid cell is selected as the radar observation for each target grid cell. Consequently, each local OI analysis contains at most one radar observation, and no radar–radar covariance terms arise within the local observation-error covariance matrix. The suitability of this nearest-grid-point approximation and its limited impact on the resulting analysis were evaluated in our previous study (Zhang et al., 2026).

For radar–gauge error correlations, these are assumed to be zero in the current implementation. Explicitly representing this term requires the errors associated with both radar and gauge observations to be independently characterised. In particular, robust estimation of gauge error requires a sufficiently dense gauge network to estimate the areal rainfall reference independently of an individual gauge, which is difficult to achieve consistently across Australia at the national scale. Seo et al. (2011), for example, investigated radar–gauge error covariance using exceptionally dense gauge networks specifically designed to provide reliable estimates of areal rainfall. Importantly, although they showed that radar–gauge error covariance can exhibit considerable variability, particularly at larger spatial scales, the mean radar–gauge error correlations remained very close to zero across all spatial and temporal scales examined (see their Table 2).

Therefore, we consider the zero radar–gauge error cross-correlation assumption to be a reasonable simplification for the current national-scale implementation, while recognising that cross-source error correlations may exist and could be further investigated where sufficiently dense reference observations are available.

We have revised Section 3.2 to explicitly state and justify these assumptions.

3.2 Multi-source statistical interpolation

…

In this study, both gauge and radar are incorporated as observations. Radar–gauge observational error cross-correlations are assumed to be zero. Explicit estimation of these cross-correlations requires independent characterisation of both radar and gauge errors. In particular, robust estimation of gauge representativeness errors generally requires a sufficiently dense gauge network to provide an independent estimate of areal rainfall, which is difficult to achieve consistently across Australia. Seo et al. (2011), using high-density gauge networks, showed that radar–gauge error correlations can vary across individual samples and scales, but their mean correlations remained close to zero across the spatial and temporal scales examined. We therefore adopt zero radar–gauge error cross-correlation as a practical simplifying assumption for the national-scale implementation. Furthermore, gauges at different locations are also assumed to be independent. Accordingly, radar–gauge and gauge–gauge observational error correlations are set to zero. However, a full representation of radar-to-radar error correlations would require retaining potentially hundreds of radar grid cells within the influence radius (e.g., 200 km) for each target grid cell, substantially increasing the size of the observation-error covariance matrix (i.e., in Eq. (5)), computational cost, and potential for matrix rank deficiency.

To address this issue, an error-informed localisation strategy is adopted within the statistical interpolation framework. Specifically, a highly efficient implementation is introduced in which only the nearest radar grid cell within a 200 km search radius, identified as the grid cell with the minimum Euclidean distance to the target grid cell, is retained as a supplementary observation. If no valid radar grid cell is available within this radius, no radar observation is included for that target grid cell. The 200 km radius defines the maximum search range, while the effective influence of the selected radar observation is further controlled by the distance-dependent error correlation function. In cases where multiple grid cells have the same minimum distance, the grid cell with the lower error variance is retained. This nearest-point approximation reflects a deliberate selection strategy that prioritises the most locally relevant radar information while avoiding dilution by redundant contributions. Consequently, each local analysis contains at most one radar observation, and no radar–radar error covariance terms arise in the resulting observation-error covariance matrix. The suitability of this nearest-point approximation has been evaluated in our previous study (Zhang et al., 2026). As a result, this approach enables scalable, grid-cell-wise blending suitable for continental-scale precipitation analysis at kilometre spatial resolution and hourly temporal resolution under practical computational constraints.

References

Seo, B.-C. and Krajewski, W. F.: Investigation of the scale-dependent variability of radar-rainfall and rain gauge error covariance, Advances in Water Resources, 34, 152–163, https://doi.org/10.1016/j.advwatres.2010.10.006, 2011.

Zhang, Y., Wang, Q. J., Frost, A. J., Pudashine, J., Trewin, B., Velasco-Forero, C., Su, C.-H., and Villani, V.: A localisation strategy for scalable and efficient multi-source precipitation blending at high resolutions over large domains, Journal of Hydrology, 677, 135831, https://doi.org/10.1016/j.jhydrol.2026.135831, 2026.

R2-DC7

Figure 7: The "x" symbol is in red, and it is really hard to see in the figure. Please change the shape and colour of the symbol.

Response: Thank you for this helpful suggestion. We agree that the red "×" symbols in the original figure were difficult to distinguish against the rainfall colour scale. We have revised Figure 7 by using more visible symbols and contrasting colours to clearly identify the intended locations.

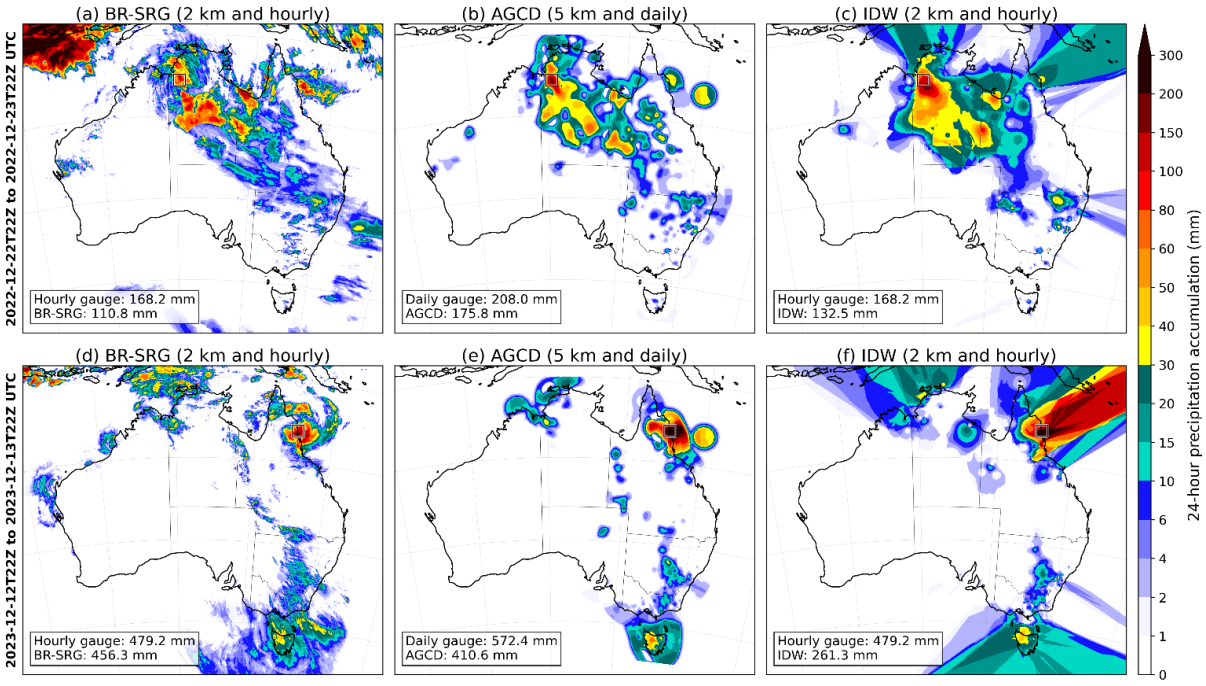


Figure 10. Example rainfall analyses from the three-source blend (BR-SRG; 2-km and hourly), AGCD (5-km and daily), and gauge-based IDW interpolation (2-km and hourly) for two selected 24-h periods during extreme precipitation events: (a–c) Cyclone Ellie, from 22:00 UTC on 22 December to 22:00 UTC on 23 December 2022, and (d–f) Cyclone Jasper, from 22:00 UTC on 12 December to 22:00 UTC on 13 December 2023. Note that AGCD incorporates the denser daily-gauge network (Fig. S6b, d), whereas BR-SRG and IDW use the comparatively sparser hourly-gauge network (Fig. S6a, c). BR-SRG and IDW are accumulated over the indicated UTC periods, whereas AGCD represents the standard daily gauge accumulation period of approximately 9 am to 9 am local time, ending on 23 December 2022 and 14 December 2023, respectively, with the exact UTC period varying across Australian time zones. The red square indicates the location of the gauge with the highest observed 24-hour precipitation for each event. Further details of the satellite and radar estimates and gauge observations are provided in Figs. S5–S6 and Tables S1–S2.

References

Nguyen, C., Vaze, J., Mateo, C. M. R., Hutchinson, M. F., & Teng, J. (2026). Elevation-dependent spatial interpolation of hourly rainfall for accurate flood inundation modelling. Hydrology and Earth System Sciences, 30(1), 45-66.

Ochoa-Rodriguez, S., Wang, L. P., Willems, P., & Onof, C. (2019). A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. Water Resources Research, 55(8), 6356-6391.

Seo, B. C., & Krajewski, W. F. (2011). Investigation of the scale-dependent variability of radar-rainfall and rain gauge error covariance. Advances in Water Resources, 34(1), 152-163.

Xia, H., & Wang, K. (2025). Hourly, kilometer-scale precipitation merged from rain gauge, ground-based radar and satellites over East Asia: methods, evaluation and applications. Journal of Hydrology, 134148.

Yu, J., Li, X. F., Lewis, E., Blenkinsop, S., & Fowler, H. J. (2020). UKGrsHP: A UK high-resolution gauge–radar–satellite merged hourly precipitation analysis dataset. Climate Dynamics, 54(5), 2919-2940.

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 careful review and constructive comments. We appreciate your positive assessment of the overall presentation of the data, methodology, and results. We also appreciate your comments regarding the description of the IDW gauge-based rainfall estimates and the interpretation of the hourly and daily validation results.

In response, we have revised the manuscript to provide a clearer and more complete description of the IDW implementation and its role in the study. We have also clarified the interpretation of the validation results, expanded the evaluation and discussion where necessary, and provided further explanation of the limitations and potential artefacts identified in the blended rainfall estimates.

Detailed responses to your major and specific comments are provided below.

Specific Comments:

R3-SC1

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 relevant reference. We agree that the Multi-Radar Multi-Sensor (MRMS) system provides an important example of an operational national-scale multi-sensor QPE framework and should be acknowledged in our discussion.

We have revised the Introduction to include the MRMS multi-sensor QPE approach and added the suggested reference (Martinaitis et al., 2020) to provide broader context for the operational development of multi-source precipitation products.

Introduction

… Physically based approaches have also been developed, in which dynamic weighting rules are assigned according to observational quality and environmental conditions, including the operational national-scale Multi-Radar Multi-Sensor (MRMS) system in the United States (e.g., Martinaitis et al., 2020). …

R3-SC2

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 question. 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 valid radar grid cell is available within this distance, no radar observation is included in the interpolation for that target grid cell.

We note that 200 km represents the maximum search radius rather than the effective influence distance. The actual contribution of the selected radar observation is controlled by the distance-dependent error correlation function, which decreases rapidly with distance (Fig. 4b). Therefore, distant radar observations receive relatively small weights and have limited influence on the analysis.

We have revised the manuscript to explicitly state the 200 km maximum search distance and clarify its role in the nearest-radar selection.

To address this issue, an error-informed localisation strategy is adopted within the statistical interpolation framework. Specifically, a highly efficient implementation is introduced in which only the nearest radar grid cell within a 200 km search radius, identified as the grid cell with the minimum Euclidean distance to the target grid cell, is retained as a supplementary observation. If no valid radar grid cell is available within this radius, no radar observation is included for that target grid cell. The 200 km radius defines the maximum search range, while the effective influence of the selected radar observation is further controlled by the distance-dependent error correlation function. In cases where multiple grid cells have the same minimum distance, the grid cell with the lower error variance is retained. This nearest-point approximation reflects a deliberate selection strategy that prioritises the most locally relevant radar information while avoiding dilution by redundant contributions. Consequently, each local analysis contains at most one radar observation, and no radar–radar error covariance terms arise in the resulting observation-error covariance matrix. The suitability of this nearest-point approximation has been evaluated in our previous study (Zhang et al., 2026). As a result, this approach enables scalable, grid-cell-wise blending suitable for continental-scale precipitation analysis at kilometre spatial resolution and hourly temporal resolution under practical computational constraints.

R3-SC3

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. In our implementation, the radar/satellite bias is estimated at gauge locations, interpolated to the corresponding grids to construct a spatially varying bias field, and then applied locally to each grid cell. Therefore, “local bias correction” provides a clearer description than the previously used term “mean bias correction”.

To address this comment, we have revised the section title by replacing “mean bias correction” with “local bias correction” and revised the corresponding methodological description to clarify how the spatially varying bias field is estimated and applied.

3.3 Local bias correction of satellite and radar precipitation estimations

Bias correction for satellite and radar data followed a Kalman filtering approach similar to Chumchean et al. (2006), but was adapted from the original mean-field framework to estimate spatially varying local bias fields at the target hourly timescale. Bias estimation was based on co-located satellite/radar–gauge pairs, obtained by extracting satellite or radar estimates at gauge locations and pairing them with the corresponding gauge observations. All available gauges were included in the satellite bias correction, whereas only gauges within radar coverage were included in the radar bias correction because radar estimates are unavailable outside the radar coverage area. Bias at each gauge location was estimated using a Kalman filter with an AR(1) model to capture temporal dependence, and the resulting point estimates were interpolated using ordinary kriging to generate a continuous spatially varying bias field. The corrected product was then obtained by applying the bias field multiplicatively to the raw product.

The calibrated radar estimates (R1) used in this study differ from the operational Rainfields product (R0) in both temporal and spatial configuration. R0 applies bias correction using gauges within a 150-km radius and a 15-min window. The spatial radius limits the number of nearby co-located radar–gauge pairs available for local bias estimation, while the temporal window is constrained by gauge reporting latency and the short operational update interval. In contrast, R1 uses gauges within a 300-km radius and a 1-hour window, increasing the number of radar–gauge pairs available for estimating the local radar bias. No predefined limit is imposed on the number of pairs within this radius. All available pairs are considered, with their point-scale bias estimates subsequently interpolated using ordinary kriging to construct a spatially continuous bias field. The larger spatial radius allows more spatially distributed gauge observations to be included, while the longer temporal window accommodates additional observations associated with gauge reporting latency. This approach also uses the same available gauge observations for satellite and radar bias correction, subject to the availability of radar observations at each gauge location. However, this configuration is tailored to the Australian radar system; for different radar systems, adjustments can be made based on their operational configuration and production objectives.

R3-SC4

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. No predefined limit is imposed on the number of gauges used for radar bias estimation. All available co-located radar–gauge pairs within the 300-km radius are considered. Therefore, a larger number of pairs can be available in regions with dense gauge networks, such as the south-

eastern coast. However, these gauge pairs are not simply averaged to obtain a single bias estimate over the 300-km radius. Bias is first estimated at individual gauge locations from the co-located radar–gauge pairs. These point-scale bias estimates are then spatially interpolated using ordinary kriging to construct the continuous bias field applied to the radar estimates. Thus, the spatial bias correction is determined by both the available point-scale bias estimates and their spatial configuration through the kriging interpolation, rather than simply by the total number of gauges within the 300-km radius.

3.3 Local bias correction of satellite and radar precipitation estimations

…

The calibrated radar estimates (R1) used in this study differ from the operational Rainfields product (R0) in both temporal and spatial configuration. R0 applies bias correction using gauges within a 150-km radius and a 15-min window. The spatial radius limits the number of nearby co-located radar–gauge pairs available for local bias estimation, while the temporal window is constrained by gauge reporting latency and the short operational update interval. In contrast, R1 uses gauges within a 300-km radius and a 1-hour window, increasing the number of radar–gauge pairs available for estimating the local radar bias. No predefined limit is imposed on the number of pairs within this radius. All available pairs are considered, with their point-scale bias estimates subsequently interpolated using ordinary kriging to construct a spatially continuous bias field. The larger spatial radius allows more spatially distributed gauge observations to be included, while the longer temporal window accommodates additional observations associated with gauge reporting latency. This approach also uses the same available gauge observations for satellite and radar bias correction, subject to the availability of radar observations at each gauge location. However, this configuration is tailored to the Australian radar system; for different radar systems, adjustments can be made based on their operational configuration and production objectives.

R3-SC5

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 the suggestion. We have revised Figure 1 to mark the tropical, subtropical, and temperate subregions.

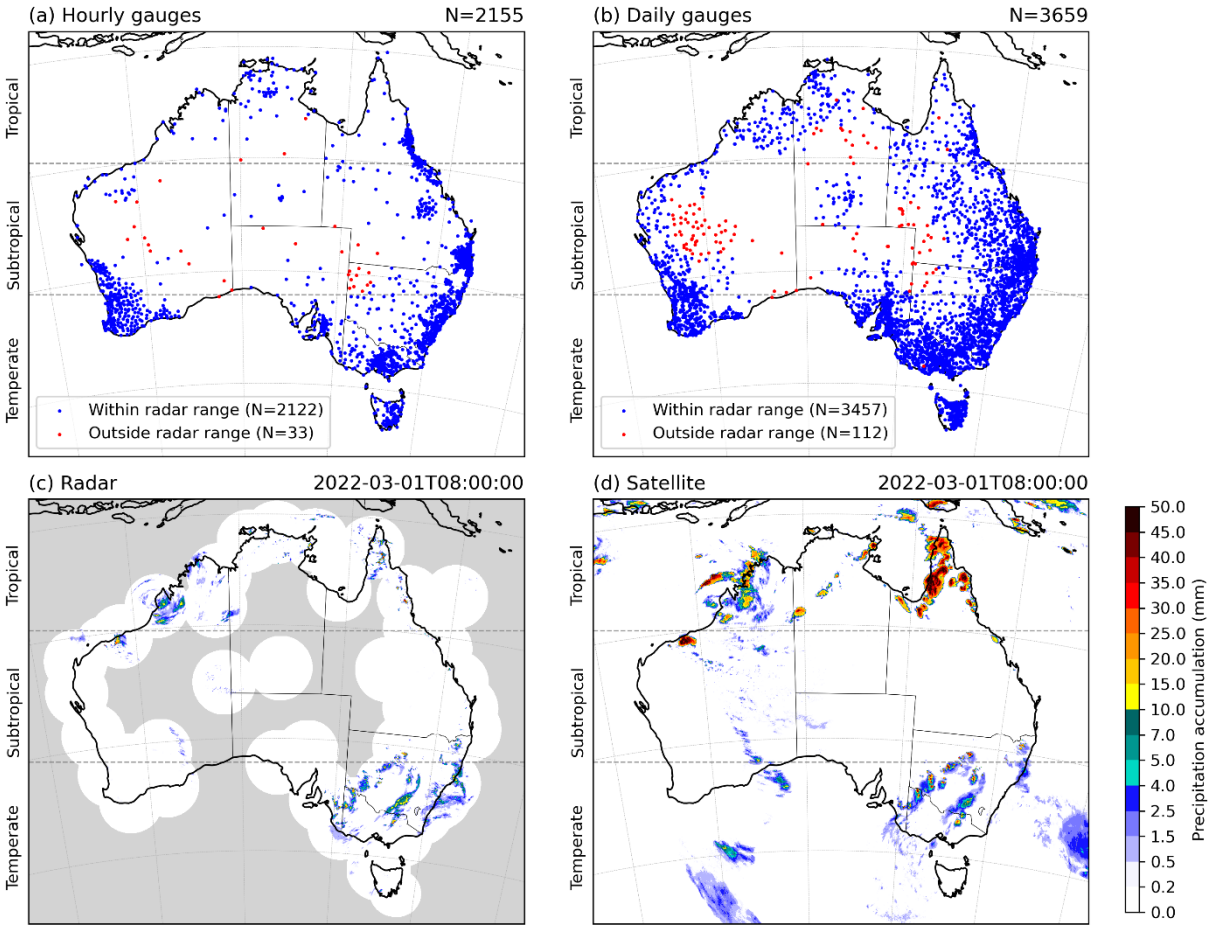


Figure 1. Spatial distribution and coverage of the data sources used in this study: (a) hourly gauges (aggregated from sub-daily) (N=2,155) categorised as within (N=2,122, blue) or outside (N=33, red) the maximum radar range, (b) daily gauges (N=3,659) similarly categorised as within (N=3,457, blue) or outside (N=112, red) the maximum radar range, (c) radar-based hourly precipitation accumulation at 0800 UTC on 1 March 2022, with a maximum radar range of approximately 300 km and a spatial resolution of 1 km, and (d) satellite-based precipitation (CRRPH) for the same time, with seamless spatial coverage at a 2-km resolution. The tropical, subtropical, and temperate subregions used in subsequent analyses are indicated by the horizontal dashed lines.

R3-SC6

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 comment. We have revised the main manuscript to provide a clearer description of the IDW interpolation and its configuration. Specifically, the IDW estimate is calculated using the five nearest available gauges based on Euclidean distance, with weights proportional to the inverse square of distance ($1/d^2$) and no maximum search radius. We have also clarified the role of IDW as a simple gauge-only reference for comparison with the proposed multi-source blending framework.

3.5 Evaluation strategies

The evaluation was designed primarily to assess the added value of integrating satellite, radar, and gauge observations within the BRAIN framework. The three-source satellite–radar–gauge analysis was therefore

evaluated against individual-source fields and reduced-source blending configurations (i.e., satellite-gauge blending) to determine how the incorporation of additional observational information affects the analysis performance.

BRAIN was also compared with rainfall products and approaches currently used operationally by the Bureau to provide additional operational context. Gauge-based IDW interpolation was included as a simple gauge-only reference because it is currently used in the Bureau's operational flood forecasting system to estimate catchment- and/or sub-catchment-scale rainfall from sub-daily gauge observations for hydrological modelling (Malone, 1999; Malone and Carroll, 2022), rather than to generate a national sub-daily gridded rainfall product. AGCD was included as an existing national operational daily gridded rainfall product (Jones et al., 2009).

3.5.3 IDW implementation and comparison with AGCD

Gauge-based IDW interpolation was included as a simple gauge-only reference. The primary IDW configuration used the five nearest available gauges, selected according to Euclidean distance, with an inverse-distance power of 2 and no maximum search distance. This configuration is referred to as IDW-K5 and is shortened to IDW in the main evaluation. To examine the sensitivity of the comparison to the IDW configuration, two additional configurations were considered. IDW-K5-L retained the five-gauge configuration of IDW-K5 but imposed a maximum search distance of 200 km, allowing the effect of the distance constraint to be examined. IDW-K20-L used up to 20 gauges within a 200-km radius, consistent with the gauge-selection configuration used in BRAIN. The sensitivity of the results to these IDW configurations is examined in Section 5.2.

R3-SC7

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 suggestion and recommendation. We have revised the text to explicitly include bias as one of the three primary performance metrics, together with correlation and RMSE.

3.5.2 Performance metrics and stratified evaluation

Performance was evaluated using three metrics: correlation, root mean square error (RMSE), and mean bias. Correlation measures the strength of the linear relationship between estimated and observed rainfall, with values closer to 1 indicating stronger agreement. RMSE measures the overall magnitude of the errors, with lower values indicating better accuracy. Bias measures systematic overestimation or underestimation, with values closer to zero indicating smaller systematic bias. These metrics were calculated separately for each gauge to characterise spatial variations in performance.

R3-SC8

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 have revised the manuscript to clarify that nearest-neighbour interpolation, rather than bilinear interpolation, was used to establish the spatial correspondence between gridded precipitation products and gauge observations. Specifically, for each gauge, we extracted the precipitation value from the single grid cell with the minimum Euclidean distance to the gauge location. We also clarified that this approach avoids additional spatial smoothing of the gridded precipitation fields that could be introduced by bilinear interpolation.

3.5.2 Performance metrics and stratified evaluation

For all gauge-based evaluations, gridded estimates were matched to gauge locations using nearest-neighbour interpolation, whereby the value from the grid cell with the minimum Euclidean distance to each gauge location was selected. This approach preserves the values of the original gridded analyses and avoids additional spatial smoothing that may be introduced by interpolation methods such as bilinear interpolation. Nevertheless, comparisons between point-based gauge observations and grid-cell rainfall estimates are subject to inherent point-to-area representativeness differences, which should be considered when interpreting the evaluation results.

R3-SC9

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 for recognising the scope of the current study. We agree that comparison with existing global precipitation products would be scientifically valuable for further assessing the benefits of incorporating local gauge and radar observations. We have therefore revised the manuscript to clarify why such comparisons are not included in the current study, particularly given differences in spatial and temporal resolution, operational objectives, and observational inputs. We have also added discussion identifying systematic comparison with global precipitation products as an important direction for future work once a long-term BRAIN dataset becomes available.

There are several existing global precipitation products (e.g., Beck et al., 2019; Ma et al., 2025) that could potentially be used for comparison with BRAIN. However, these products differ from BRAIN in their spatial and temporal resolutions, operational objectives, and observational inputs. In particular, BRAIN is designed as a high-resolution operational product for Australia with an update latency of less than one hour and incorporates a large number of hourly gauge observations and regional radar data that are not generally used in global precipitation products. These differences need to be carefully considered when conducting direct product comparisons. The current study therefore focuses on the development and evaluation of the BRAIN framework, including its performance relative to its constituent data sources and different blending configurations. We are currently working towards the development of a long-term BRAIN dataset, which will enable more systematic comparisons with existing global precipitation products. Such comparisons will also provide an important opportunity to assess the potential benefits of incorporating regional observational datasets.

R3-SC10

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. Yes, the substantial improvement in the error distribution from S0 to S1 is largely attributable to the introduction of local bias correction.

To clarify, S0 does not apply any bias correction and directly uses the original CRRPH satellite precipitation product. In contrast, S1 applies the local bias correction described in Section 3.2. This correction substantially reduces systematic errors in the satellite estimates, contributing to the narrower and more symmetric error distribution shown in Figure 3.

We have revised the manuscript to clarify the distinction between S0 and S1 and to explicitly explain the role of local bias correction in improving the satellite error distribution.

2.3 Satellite-based observations

…

As an infrared satellite precipitation product, CRRPH infers surface rainfall indirectly from cloud properties observed by the Himawari AHI. Retrieval errors may therefore occur when the observed cloud characteristics do not adequately represent the occurrence or intensity of surface precipitation (Karagiannidis et al., 2023; So and Shin, 2018). In addition, the calibrated lookup tables used to convert retrieved cloud properties into rainfall rates represent empirical relationships and may not fully capture variations in the cloud–rainfall relationship under different precipitation conditions. To reduce systematic discrepancies with gauge observations, a spatially varying local bias correction was applied as described in Section 3.3. S0 denotes the original CRRPH product with any bias correction, while S1 denotes the bias-corrected product used as the background field for blending. However, the local gauge-based bias correction reduces systematic discrepancies between CRRPH and gauge observations but cannot fully correct errors in rainfall occurrence, intensity, or spatial structure. Similar to the radar correction, the estimated bias field depends on the availability and representativeness of nearby gauges and may therefore retain residual biases or introduce additional uncertainty, particularly in gauge-sparse regions.

R3-SC11

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 this comment. The improvement in the radar error distribution from R0 to R1 is associated with the additional local bias correction applied in this study.

R0 represents the operational Rainfields calibrated radar product, whereas R1 applies an additional local bias correction using a larger spatial radius (300 km compared with 150 km) and a longer temporal window (1 hour compared with 15 min). These modifications increase the number of co-located radar–gauge pairs available for estimating the local radar bias, contributing to the narrower and more symmetric error distribution centred around zero.

We have revised the manuscript to clarify the distinction between R0 and R1 and to explain how the larger spatial radius and longer temporal window contribute to the improved radar error distribution.

2.2 Radar-based observations

…

In this study, we used hourly radar accumulations derived from 5-min calibrated radar rain rates as a blending input for the target hourly resolution. However, we recalibrated the radar data using gauges within a 300-km radius over a 1-hour window to increase the availability of co-located radar-gauge pairs for estimating local radar errors. The larger spatial radius allows more co-located radar-gauge pairs to be included, particularly in gauge-sparse regions, while the longer temporal window increase gauge availability by accommodating their reporting latency. These two modifications therefore address the spatial and temporal availability of gauge information, respectively. We refer to the operational calibrated product as R0 and the recalibrated version as R1.

3.3 Local bias correction of satellite and radar precipitation estimations

…

The calibrated radar estimates (R1) used in this study differ from the operational Rainfields product (R0) in both temporal and spatial configuration. R0 applies bias correction using gauges within a 150-km radius and a 15-min window. The spatial radius limits the number of nearby co-located radar–gauge pairs available for local bias estimation, while the temporal window is constrained by gauge reporting latency and the short operational update

interval. In contrast, R1 uses gauges within a 300-km radius and a 1-hour window, increasing the number of radar–gauge pairs available for estimating the local radar bias. No predefined limit is imposed on the number of pairs within this radius. All available pairs are considered, with their point-scale bias estimates subsequently interpolated using ordinary kriging to construct a spatially continuous bias field. The larger spatial radius allows more spatially distributed gauge observations to be included, while the longer temporal window accommodates additional observations associated with gauge reporting latency.

R3-SC12

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 agree that our original statement was confusing and could incorrectly imply that R1 and BR-SRG differ primarily because R1 uses more gauges. We have therefore removed this statement and revised our interpretation of this result.

The evaluation results show that the relative performance of R1 and BR-SRG in terms of RMSE is not consistent across the hourly cross-validation and independent hourly and daily evaluations. However, the rainfall-intensity-stratified evaluation shows a more consistent pattern. R1 generally performs better for high-intensity precipitation, particularly with a smaller negative bias, whereas BR-SRG shows increasing negative bias with rainfall intensity. The satellite estimates exhibit an even greater negative bias at high rainfall intensities, indicating that residual underestimation in the satellite background can propagate into the blended estimates. Within radar coverage, radar already receives a relatively large contribution when its estimated error is smaller than that of the satellite background. However, as long as the satellite estimate remains informative according to the estimated error characteristics, it retains some contribution to the analysis. Conversely, there are also cases where the satellite background provides useful information relative to radar, as illustrated in Fig. 14a. Therefore, simply increasing the radar contribution would not necessarily improve the blended estimates under all rainfall conditions.

We now recognise the underestimation of high-intensity precipitation as a potential limitation of the current implementation and have expanded the discussion in Section 5.3. Further improvements could focus on reducing intensity-dependent biases in the input precipitation estimates, particularly the satellite background, and refining the error modelling to better account for variations in data-source performance with rainfall intensity and different precipitation event types. These developments could reduce the propagation of source-specific errors into the blended analysis while retaining the complementary information provided by multiple data sources.

5.3 Limitations and future directions

One limitation of the current framework is the propagation of residual errors from input observations into the blended analysis. Residual errors that are not identified through quality control, particularly for gauge observations, or are not fully removed by bias correction, particularly for satellite and radar estimates, can propagate into the blended analysis. Although the statistical interpolation weighting accounts for the intensity-dependent error characteristics of each source, an individual source generally does not receive complete weight, as this would require the error of the other sources to become overwhelmingly large in comparison. Consequently, residual biases in sources retaining non-zero weights can propagate into the final analysis.

For example, during Cyclone Jasper, underestimation in the satellite estimates contributed to underestimation in BR-SRG even when radar closely represented the observed accumulation (Fig. 14a). Conversely, during Cyclone Ellie, radar substantially underestimated the observed accumulation, while the satellite estimates were higher and partly compensated for the radar underestimation in the blended analysis (Fig. 14b). These contrasting cases

demonstrate the value of retaining complementary information from multiple data sources, rather than simply increasing the contribution of a single source.

Together with the relative differences among S1, R1, BR-SG, and BR-SRG in the rainfall-intensity analysis (Fig. 7), these cases suggest that residual underestimation in the satellite estimates may contribute to the negative bias observed in BR-SRG, although the intensity-stratified bias is also affected by observation-based stratification. Further improvements should therefore focus on bias correction, particularly in better accounting for systematic biases across different rainfall intensities.

R3-SC13

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 comment. Yes, the lower RMSE of BR-SRG compared with BR-SG outside radar coverage can be partly attributed to the contribution of nearby radar observations. However, we would like to clarify that this contribution differs from horizontal extrapolation of the radar precipitation field.

In BRAIN, nearby radar observations can contribute beyond the radar coverage and are blended with the local satellite background and available gauge observations according to their estimated errors and spatial correlations. In some cases, the nearby radar observation may provide better precipitation information than the satellite estimate. In other cases, the radar estimate itself may not be more accurate than the satellite estimate, but it can still provide complementary information through the blending. After blending the different data sources, the resulting BR-SRG estimate can therefore perform better than BR-SG.

We have revised the manuscript to clarify this point and have further discussed how nearby radar information can improve the blended estimates beyond the radar coverage.

5 Discussion

5.1 Methodological contributions of the BRAIN framework

…

A further distinctive feature is that radar information is not restricted by a hard coverage mask. For each target location, BRAIN searches for the nearest available radar observation within a predefined maximum distance (200 km in this study). Consequently, radar information can still contribute to the blended estimate when the target location lies outside radar coverage, provided that a sufficiently close and reliable radar observation is available. Its effective contribution is determined by both the distance-dependent error correlation (Fig. 4b) and the estimated error variance, with greater weight generally assigned to observations that are closer to the target location and have lower estimated errors. Even when the nearby radar estimate is not more accurate than the local satellite estimate on its own, it can provide complementary information that can improve the resulting estimate after blending.

This formulation allows radar information to contribute both within and near radar-covered regions, providing a gradual transition across radar-coverage boundaries while retaining the satellite estimate as the spatially continuous background. This differs from simple radar extrapolation, in which rainfall information within the radar domain is spatially extended into areas without radar observations. In BRAIN, the radar field itself is not extrapolated. The satellite estimate remains the background at every target location, while available radar and gauge observations contribute through the statistical interpolation according to their estimated error characteristics and spatial correlations.

R3-SC14

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 these thoughtful comments and questions. We agree that our original explanation included several possible contributing factors and was not sufficiently clear. We have revised this section and clarified the IDW configuration and our interpretation of the results.

For points 1 and 2, BR-SRG uses a maximum search distance of 200 km for incorporating observations, whereas the primary IDW configuration uses the five nearest available gauges without a maximum search distance. Therefore, IDW can use gauges located more than 200 km away in gauge-sparse regions. However, we agree that the difference in search distance alone does not explain the better daily RMSE of IDW outside radar coverage, nor does it demonstrate that gauges more than 200 km away provide better precipitation information than the local satellite estimate. We have therefore removed this interpretation from the revised manuscript.

For point 3, our further analysis indicates that temporal error cancellation is an important factor. At the hourly scale (Fig. 11c, f), IDW shows a nearly balanced contribution of positive and negative errors, particularly outside radar coverage. This results in a very small mean bias, although the RMSE (and/or MAE) remains relatively large, indicating that the small bias does not necessarily reflect small hourly errors. When hourly precipitation is accumulated to daily totals (Fig. S20c, f), these opposing errors can partly cancel over time, reducing the accumulated error and thereby contributing to the lower MAE and RMSE of IDW in the daily evaluation while maintaining a small bias.

Temporal error cancellation can also occur for satellite and blended estimates. However, our analysis shows that their positive and negative errors are less balanced than those of IDW, particularly outside radar coverage. This makes the cancellation effect more evident for IDW and helps explain why its relative performance improves substantially from the hourly to the daily evaluation. We have added this analysis to the revised manuscript and included a dedicated subsection to discuss these results and their implications in detail.

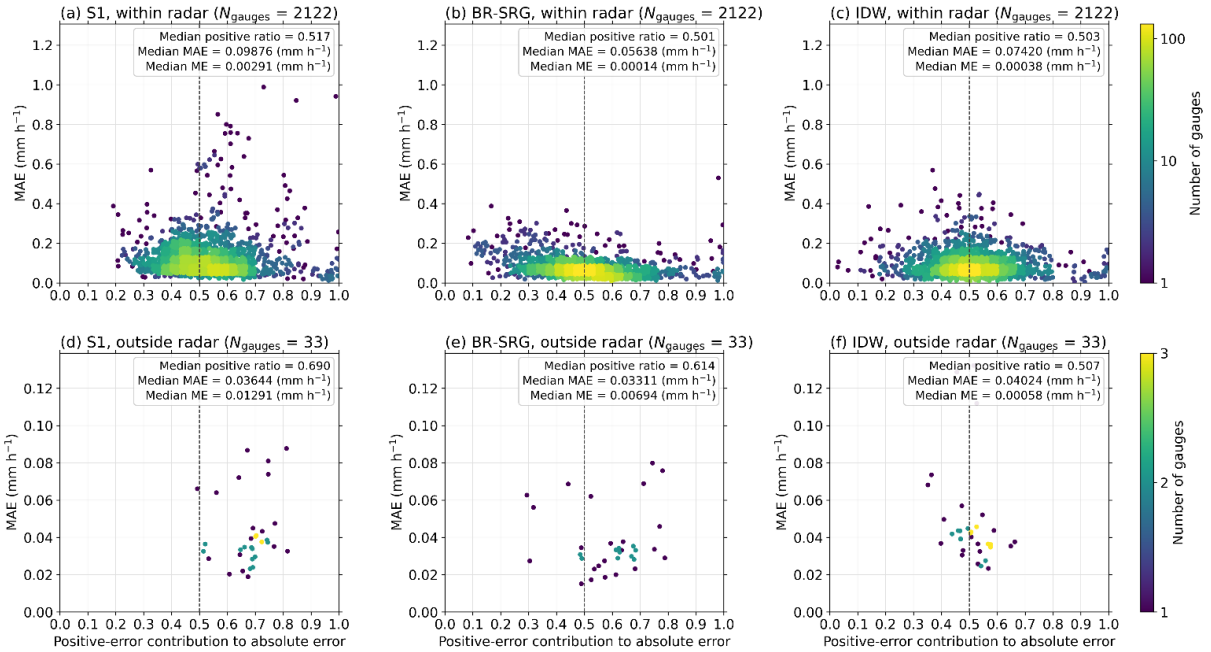


Figure 11. Relationship between positive-error contribution and mean absolute error (MAE) across evaluation gauges in the hourly-gauge 20-fold cross-validation for S1, BR-SRG, and IDW, evaluated (a–c) within and (d–f) outside radar coverage. Each point represents an individual gauge. Positive-error contribution is defined as the magnitude of positive errors relative to the total absolute error, such that a value of 0.5 indicates equal contributions from positive and negative errors. The vertical dashed line marks this balance point.

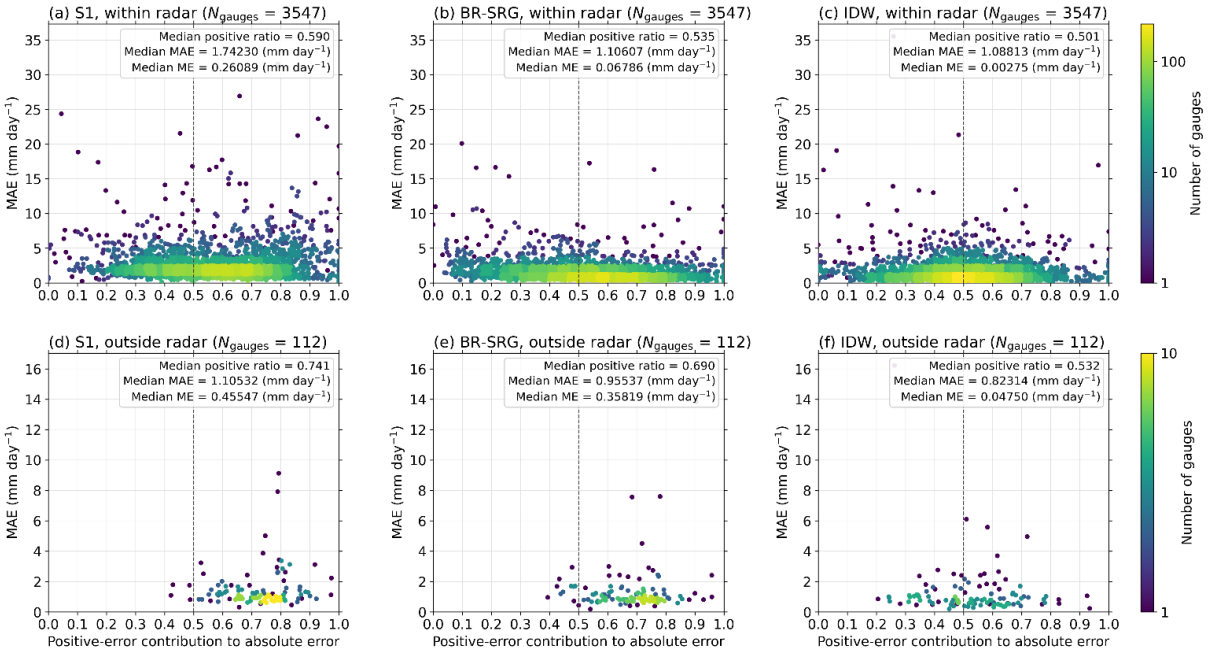


Figure S20. Relationship between positive-error contribution and mean absolute error (MAE) across evaluation gauges in the independent daily-gauge evaluation for S1, BR-SRG, and IDW, evaluated (a–c) within and (d–f) outside radar coverage. Each point represents an individual gauge. Positive-error contribution is defined as the magnitude of positive errors relative to the total absolute error, such that a value of 0.5 indicates equal contributions from positive and negative errors. The vertical dashed line marks this balance point.

5.2 Behaviour and sensitivity of gauge-based interpolation

The evaluation revealed two notable behaviours of IDW: its ME remained consistently close to zero despite relatively large RMSE and lower correlation (Fig. 5 and Fig. S16), while its relative performance improved from hourly to daily evaluation, achieving a lower RMSE than BR-SRG at the daily timescale (Fig. 6), a pattern consistently observed across different subregions outside radar coverage (Fig. S2). To examine these behaviours, we compared the station-level contributions of positive and negative errors (Fig. 11, Fig. S20, and Fig. S21). The median positive-error contribution of IDW remained close to 0.5 across all three evaluations despite relatively large MAE, indicating nearly equal contributions from positive and negative errors. At the hourly scale, this balance results in a near-zero ME despite relatively large individual errors, as reflected by the MAE and RMSE. The near-zero ME of IDW therefore primarily reflects error cancellation rather than uniformly small estimation errors. When hourly precipitation is accumulated to daily totals, these opposing errors can further offset over time, reducing the accumulated daily errors and contributing to the improved relative performance of IDW at the daily timescale (Fig. 6 and Fig. S20).

R3-SC15

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, and the original reference should have been to Text S2.2. We have thoroughly reorganised the Supplementary Material and updated the corresponding cross-references throughout the manuscript.

R3-SC16

Line 509: "(see Fig. 1a, Fig. S2, and Text S2a)". I couldn't find Text S2a.

Response: Thank you for pointing this out. This was a typo, and the original reference should have been to Text S2.2. We have thoroughly reorganised the Supplementary Material and updated the corresponding cross-references throughout the manuscript.

R3-SC17

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 R3-SC12.

R3-SC18

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

Response: Thank you for pointing this out. Please see our detailed response to R3-SC2 regarding the definition and interpretation of the distance considered in the framework. We have also revised the corresponding text in the manuscript to clarify this point.

5.1 Methodological contributions of the BRAIN framework

….

A further distinctive feature is that radar information is not restricted by a hard coverage mask. For each target location, BRAIN searches for the nearest available radar observation within a predefined maximum distance (200 km in this study). Consequently, radar information can still contribute to the blended estimate when the target location lies outside radar coverage, provided that a sufficiently close and reliable radar observation is available.

Its effective contribution is determined by both the distance-dependent error correlation (Fig. 4b) and the estimated error variance, with greater weight generally assigned to observations that are closer to the target location and have lower estimated errors. Even when the nearby radar estimate is not more accurate than the local satellite estimate on its own, it can provide complementary information that can improve the resulting estimate after blending.

This formulation allows radar information to contribute both within and near radar-covered regions, providing a gradual transition across radar-coverage boundaries while retaining the satellite estimate as the spatially continuous background. This differs from simple radar extrapolation, in which rainfall information within the radar domain is spatially extended into areas without radar observations. In BRAIN, the radar field itself is not extrapolated. The satellite estimate remains the background at every target location, while available radar and gauge observations contribute through the statistical interpolation according to their estimated error characteristics and spatial correlations.

R3-SC19

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.

We have added a discussion of this issue and its potential causes and implications in the revised manuscript.

5.3 Limitations and future directions

…

A second limitation concerns the spatial continuity of the blended precipitation field, particularly across radar coverage boundaries. Outside radar coverage, the analysis is primarily constrained by the satellite background and gauge observations, whereas within radar coverage, radar observations provide an additional constraint and can strongly influence the analysis when their estimated errors are smaller. When satellite and radar estimates are broadly consistent, the transition is generally smooth. However, large discrepancies between the two sources can produce visible spatial transitions near radar coverage boundaries. This occurs because the analysis within radar coverage is adjusted towards radar when its estimated errors are much smaller than those of the satellite, while areas outside radar coverage remain primarily constrained by the satellite background (Fig. 10). Importantly, radar information in the present framework is not restricted by a hard coverage boundary. Through error-informed localisation, surrounding radar observations can influence nearby areas beyond the nominal radar coverage, with their influence gradually decreasing with distance. This provides a smoother transition between radar-dominated and satellite-dominated regions than approaches relying solely on co-located observations, substantially reducing abrupt discontinuities at radar coverage boundaries. However, it cannot fully eliminate discontinuities when the input precipitation fields differ substantially. Improving consistency among input sources through better retrieval algorithm, quality control, and bias correction, together with incorporating additional independent data sources, could further improve spatial continuity.