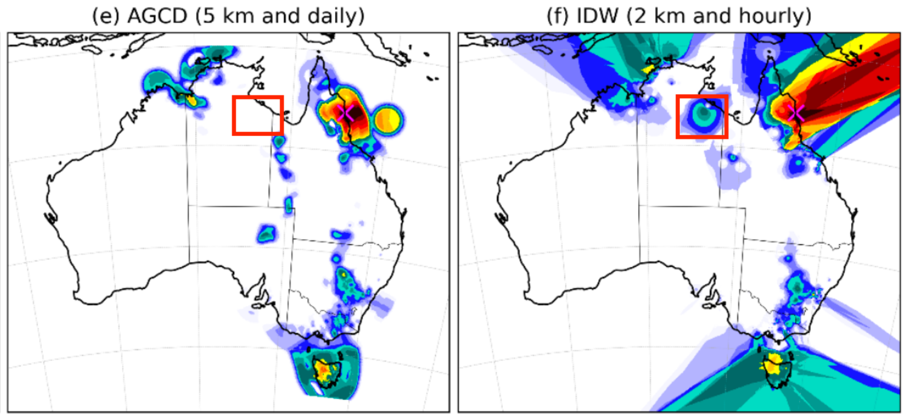
The manuscript presents an hourly, 2 km resolution blended rainfall dataset (BRAIN) for Australia, derived from satellite, radar, and gauge observations. Given the current lack of a national hourly gridded rainfall product, this dataset addresses a clear need. The manuscript is generally well written,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

I attached detailed comments and suggestions to the PDF file.

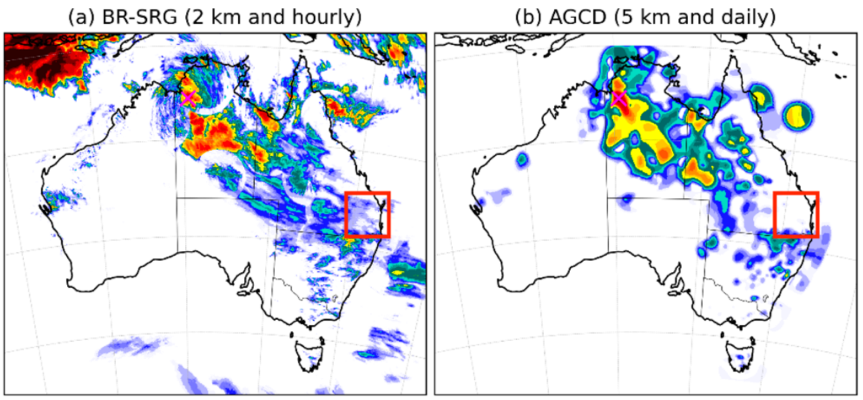
Major comments

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 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).
2. IWD should not be used as 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. This framing may mislead readers, particularly for sensitive applications such as flood modelling, where reliable rainfall estimates are critical.
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.



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.
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.



6. The hourly rainfall measurements contain many zeros, so even though the RMSE values seem to be low, the dataset can have an issue of significantly underestimating high rainfall values. Figure 7 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.

Detailed comments

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.

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.

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

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?

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).

Section 3.2: The method mentioned in this section is undercited. 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.

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.

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.