Review of egusphere-2026-666: "Design and trial implementation of a continental-scale, kilometre-resolution hourly precipitation analysis for Australia using satellite, radar and gauges**"** by Zhang et al.

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.

Specific Comments:

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

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

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

Line 318 - 324: The "mean bias correction" in the title of this section is confusing. The description (line 319-324) 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 et al. 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.

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).

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

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? $1/d^2$), 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.

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.

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?

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.

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 Line 318-324.

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.

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.

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.

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 200km 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 in 1)
3) Would this same argument work for satellite QPE? If not, please explain why.

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”?

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

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

Line 535: “...framework over short distances” - what are the short distances? Also see comment on Line 285.

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?