Revision notes

The revision is much improved. The added calibration of DEM smoothing scale and elevation scaling, plus the automated false-zero occurrence analysis, strengthen the methodological contribution and the dataset's reliability.

The CHRain product shows strong correspondence with observations (e.g., r≈0.949 for hourly comparisons) and clear advantages over BARRA-SY and radar during the 2017 event.

Thank you for your evaluation. We addressed the comments from the reviewer in detail below.

Major comments:
1. Reproducibility & reusability (scripts/data):
Given the practical value of the pipeline (hourly disaggregation → quality control/false-zero removal → ANUSPLIN interpolation → evaluation), please add a Code & Data Availability section and share:
o the disaggregation scripts implementing your three criteria and 9am–8am alignment,
o the occurrence-analysis code (including the 0.5 threshold and σ=0.25 setting),
o ANUSPLIN .cmt files and any DEM smoothing scripts (e.g., focal mean windows), and
o a small worked example (one storm day) so users can reproduce Figure/Tables for their area.
These elements are all described textually but are not yet available as artifacts; making them public would maximise impact and reuse.

We provided Python scripts with detailed comments to generate rainfall splines using the ANUSPLIN program, including the processes:

- Disaggregating daily data to hourly data from 9 am to 8 am the next day, and applying the quality control to the combined data (Disagg_d2h.py)
- Generating input files (.dat) from the hourly rainfall dataset to run the ANUSPLIN (Prepare_hourly_input.py)
- Generating the ANUSPLIN command file (.cmd) associated with the input files
- A small sample of one day of rainfall data (24 hours) at 2 rain gauge stations, and the DEM data, so that the user can test the provided scripts
- The 1 km DEM in the analysis was smoothed using the focal mean method in ArcGIS (https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-focal-statistics-works.htm)

Users need to contact the Fenner School of Environment & Society (https://fennerschool.anu.edu.au/research/products/anusplin-version-4-4) to get the ANUSPLIN program and the program to run the false zeros analysis and removal. We cannot share the source code of the ANUSPLIN program due to licensing constraints.

We included Python scripts and a sample as suggested by the reviewer in the Code and Data Availability in the revised manuscript (https://doi.org/10.5281/zenodo.17686121).

2. False-zero detection—report a small audit:
Great addition. You report ~0.26% removed; consider adding (i) a short manual audit of a random subset to estimate precision/recall, and (ii) a sensitivity note on the 0.5 threshold. Including a simple confusion matrix for the 30 March 2017 case (Appendix B) would be ideal.

The false zero removal is a two-step process, including an initial ANUSPLIN occurrence analysis followed by a standalone FORTRAN program (occflg1) that converts the data and fitted occurrence values into standard bad data flags for input into the main ANUSPLIN interpolation. Users need to contact the Fenner School of Environment & Society (https://fennerschool.anu.edu.au/research/products/anusplin-version-4-4) to get the ANUSPLIN program and the program to run the false zeros analysis and removal.

We added the analysis of occurrence-based corrections with hourly radar rainfall data provided by BoM in Section 2.3 in the revised manuscript. Since there are artifacts in the radar rainfall data on 30/3/2017, we apply the analysis on 8 successive days during the two flood events in 2022.

We also conducted a sensitivity analysis on the threshold from 0.4 to 0.8, and they gave similar results to the chosen value of 0.5.

The detailed analysis was added in lines 188-204 and in Table 2:

"The reliability of the occurrence-based corrections was assessed by comparing the analyses with hourly radar rainfall data over eight successive days during the two flood events in 2022 (with the peak around 28/02/2022 and 29/03/2022). Summary statistics are presented in Table \ref{table2}. As noted above, the radar rainfall is not always reliable. The percent occurrence agreement of the raw hourly rainfall data with the radar data ranged between 56% and 72% for six of the eight days, while there were strong occurrence agreements of 92% and 81% on the two high rainfall days on 27/02/2022 and 28/03/2022. This indicates there were major deficiencies in the radar data, except on the heavy rainfall days when significant rainfall was widespread over the data network. Comparing the occurrence corrections with the radar occurrence data showed strong agreement with the original data occurrence agreements, ranging from 43% to 92% on six of the days and 98% and 83% on 27/02/2022 and 28/03/2022. If the corrections were all correct, comparison with the radar data could be expected to assess them as having an accuracy similar to the initial overall agreements between the rainfall data and the radar data. The strong agreement between column 3 and column 5 in Table 2 is consistent with the corrections being in fact highly reliable, with a <u>true</u> accuracy up to around 98% on <u>all</u> eight days. The true reliability maybe somewhat lower on days with less widespread rainfall and less spatially coherent rainfall occurrence patterns. The occurrence corrections were sufficient to improve the overall occurrence agreement with the radar data on the fourth day in the first event and on all three days in the second event. The overall agreement was unchanged for the other days. A range of occurrence thresholds was tested by assessing the overall occurrence agreement of the corrected data with respect to the occurrence threshold. A range of thresholds from 0.4 to 0.8 gave similar results to the chosen value of 0.5. Further refinements are limited by the overall unreliability of the radar rainfall data."

Table 2. Summary statistics of occurrence corrections with respect to hourly radar data over five successive days, during the flood event in 2022. Occurrence agreement is calculated as the percentage of the number of agreements in occurrence (both zero or both non-zero) divided by the total number of hourly data values.

Date	Average Rain (mm/h)	Percent occurrence agreement of raw data with radar data	Number of estimated false zeros	Percent agreement of corrections with radar data	Percent occurrence agreement of corrected data with radar data
24/02/2022	3.3	57	75	57	57
25/02/2022	0.8	57	36	53	57
26/02/2022	2.3	72	54	87	72
27/02/2022	8.6	92	102	98	96
28/02/2022	10.5	56	85	46	56
28/03/2022	4.2	81	90	83	83
29/03/2022	4.2	65	94	92	67
30/02/2022	3.0	57	83	43	58

3. Clarify what "outperforms" means in spatial analysis:
Where you state "outperforms ANUClimate/AGCD," specify the exact metrics and thresholds (e.g., Bias, Hit Rate, CSI for ≥95th/≥99th percentiles). A compact figure comparing CSI across thresholds would help.

From the results of the spatial analysis, we confirmed that the CHRain dataset outperformed the ANUClimate and AGCD datasets compared at the daily timestep because it can reproduce more details about rainfall variation and capture high rainfall values better. The bias value of the CHRain dataset compared with the ANUClimate dataset is 0.916.

We added the information in line 487 in the revised manuscript as:

"The spatial evaluation indicated that the CHRain outperforms the ANUClimate and AGCD datasets, which are the most commonly used reliable rainfall datasets in Australia, in representing the 5 km sub-grid rainfall distribution at the Richmond River catchment. The 24-hour total CHRain dataset can also capture high rainfall values better than the ANUClimate (Bias = 0.916)."

Minor comments
Abstract wording: Consider changing "correlation coefficient of 0.949 shows that…" to "During the 2017 event, CHRain achieved r=0.949 against hourly gauges," to anchor the statistic to its context.
Daily-alignment note: You already aggregate from 9:00 am–8:00 am to match BoM daily totals; make this alignment explicit the first time daily metrics are mentioned, not just in Methods.

We adapted all the minor comments in the revised manuscript.

Line 13: “During the 2017 event, CHRain achieved the correlation coefficient of 0.949 against hourly gauges, showing that the dataset can adequately reproduce the patterns of hourly rainfall measurements.”

Line 91: “In the areas with sparse distribution of hourly rainfall stations, the daily measurements are disaggregated to hourly data from 9:00 am the previous day to 8:00 am the current day, using patterns from nearby hourly rainfall stations, to match with the daily data provided by the Australian Bureau of Meteorology (BoM).”