‘A new daily gridded precipitation dataset…’ by Han et al.

The authors have produced a grid-based precipitation data product for mainland China. The authors used long-term daily precipitation data from 2419 stations in China since 1961 for interpolation, and additional short-term (2015-2019) gauge data for a much larger number of stations (>40,000) for evaluating 8 different spatial interpolation schemes. The analysis was extensive, and rigorous, and the recommended ‘optimal’ scheme is justified and well supported with empirical evidence, and final data products at 3 different spatial resolutions will no doubt be most useful for wider applications.

The manuscript is mostly readable, not difficult to follow. English expressions at times are a bit odd, so is the tense.

Overall, the manuscript along with the data product(s) are publishable, with additional effort to improve the clarity and quality of presentation.

Major comments/suggestions:

About spatial resolution: I wonder about the wisdom of making data products available at all 3 spatial resolutions, namely, 0.1, 0.25, and 0.5 degree. Users of these products surely would be able to resample the dataset at finer resolution to coarse ones. I also wonder why data product at 0.05-degree resolution was not attempted since both gridded daily and monthly climatologies at 0.05 were used for this manuscript. 0.05-degree is commonly for areas of comparable size such as Australia (Jeffrey et al. 2001), and datasets at much finer spatial resolution (0.01-degree) are available for densely gauged areas such as Japan (Hatono et al. 2022).

Can the authors make it absolutely clear whether the gridded data refer to point precipitation at the centre of each grid cell, or to areal average precipitation over the grid cell. This has considerable implications for how these datasets are used and/or resampled, especially when the centres of grid cells of differing resolutions are co-located.

There were fewer stations, up to 15%, over the two two decades (1961-1980) for interpolated precipitation products (Fig. 1). This warrants discussion towards the end of manuscript.

Clarify the notion of ‘daily climatology’ as this term is not widely used and understood. Based on my understanding of the manuscript, there are three daily climatologies:

Raw daily climatology – simply mean daily precipitation amount (mm/d) for the period (1971-2000)

Smoothed daily climatology – high frequency fluctuations are removed with only the first few harmonics retained. (the authored all this the raw daily climatology, but why?)

Adjusted daily climatology – the smoothed daily precipitation was adjusted proportionally so that the monthly precipitation was preserved.

I would recommend use of the mean daily precipitation amount instead of 'daily climatology'. It is much easier to understand.

Discarding high-frequency 'noise' in the mean daily precipitation (line 202-204) would lead to a reduction in total variation in the daily climatology. How much variation preserved relative to the total variation in the mean daily precipitation? Please indicate a range.

There are issues with most of the equations used in the manuscripts see comments below

Check the tense used throughout the manuscript. Use the past tense to describe what you did, and present or present perfect to describe what others have said or done.

Minor comments/edits (the original in black, revised in blue)

Line 43: warmer at Earth's surface -> warmer at the Earth's surface

Line 44: between the atmosphere and surface -> between the atmosphere and land surface

Line 52: dataset is essential to current hydrometeorology research -> dataset is essential for hydrometeorological research

Line 55: The measurement of precipitation relies mainly on direct measurement using rain gauges disdrometers, and radar and on indirect estimation using satellite systems. -> Collection of precipitation data relies mainly on measurements using ground-based rain gauges, and estimates using sensing technologies such as weather radar and satellite.

Line 60: However, gauge observations reflect only point precipitation, and -> However, precipitation data measured with gauges are point observations only, and

Line 67 Spatial interpolation methods are usually applied to convert irregular point observations to regional measurements (Ahrens, 2006), thus generating evenly gridded precipitation products that are widely used in hydrology and 70 meteorology studies (Schamm et al., 2014; Golian et al., 2019).

Spatial interpolation methods are usually applied to irregular point observations to produce evenly distributed precipitation grid (Ahrens, 2006) for application in hydrological and meteorological studies (Schamm et al., 2014; Golian et al., 2019).

Line 78: To reach a higher temporal resolution, a daily gridded precipitation dataset has been built across China using the same raw precipitation data with a time span from 80 1961 to 2019 (Qin et al., 2022).

To achieve a higher temporal resolution, a daily gridded precipitation dataset has been produced for China using the same raw precipitation data for the same period from 1961 to 2019 (Qin et al., 2022).

(NB, 'across China' is used extensively in the manuscript. The word 'across' suggests 'from one side to another', and the word may not be appropriate in all cases. Please review its usage in the manuscript.)

Line 99 boundaries suffer worse positioning accuracy relative - > boundaries suffer positioning inaccuracy

Line 111: Eight interpolation schemes are proposed and evaluated by cross validation -> Eight interpolation schemes were considered and evaluated with cross validation

Line 115 is provided publicly for applications -> is available in the public domain

Line 121 are collected -> were collected.

Line 124 available to use for -> available for

Line 126: The coverage of stations is relatively sparse over northwestern China -> Stations are sparsely distributed in northwestern China

Line 140 stations with a range of 2015–2019 across China -> stations for the period 2015–2019 in China

Line 142
After removing stations with a missing
rate of over 20% for the period of 2015–2019, 45,992 good-quality stations are left (Figure 1b).

Once stations with more than 20% missing data were removed, there were 45,992 good-quality stations available for cross validation (Figure 1b).

Line 152: is -> was

Line 165:

for the following climatology adjustment.

for adjustment based on climatology

Line 217: the monthly total of the raw gridded daily climatology field for

(unclear, is this the sum of the smoothed mean daily precipitation for the month?)

Line 224

Equation (1)

(Something is not quite correct here. There is no index j on the right hand side. Is SF value the same for every day of the month? In other words, are there only 12 distinct SF values for each year?)

Line 255:

we used in -> used for

Equation (4)

Should be expressed as e^(-nx_i/GDD)

Equation (5)

What is the definition of surrounding stations?

Equation (6)

Normally we do not use 'X' in equations to indicate multiplication.

For IDW, again what are the surrounding stations?

Line 247 – repetition from data section above.

References

Hatono et al. 2022. Scientific Data. https://doi.org/10.1038/s41597-022-01548-3

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Modelling and Software 16 (4), 309–330.