A DETAILED LIST OF RESPONSES

TO REVIEWER #1

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.

Response: We would like to thank you for your constructive comments on our manuscript. Your insightful review has enhanced our paper considerably. We have updated the dataset to 2022 (https://doi.org/10.6084/m9.figshare.21432123.v3). Below is a point-by-point response to your comments.

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).

Response: Thanks for your comments. In consideration of users' research demands for various spatial resolutions, CHM_PR provides daily precipitation series with multiple spatial resolutions so that users can easily find the best matches. Meanwhile, as our strategy for creating products, the daily precipitation is first produced on a 0.05-degree latitude-longitude grid. Computing the analyzed values (here, daily precipitation) at a finer resolution enables improved correction of the orographic effects, which exhibit rapid changes with elevation (Daly et al. 1994; Xie et al. 2007). Also, this makes it convenient to generate analyzed values at various resolutions for different applications.

The main reason for not including a spatial resolution of 0.05-degree in the dataset is due to the prevailing demand for daily precipitation data in China and the limitations of gauge density. In China's hydrometeorology research, daily precipitation with spatial resolutions of 0.1-degree, 0.25-degree, and 0.5-degree are the most widely used. Additionally, when considering gauge density, only 0.29% of all 0.05-degree grid cells contain at least one station out of the 2,839 available gauges. This may not provide sufficient support for estimating daily precipitation at a high resolution of 0.05 degrees.

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.

Response: Thank you for your insightful suggestion. The gridded data use the areal average precipitation over the grid cell. We have added this clarification to section 4.1.

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

Response: Thanks for pointing this out. Due to a high rate of missing daily observations from 1961 to 1980, up to 15% of stations could not reach the threshold for quality control (i.e. a rate of missing daily precipitation not more than 5%) and were removed. With the vigorous development of hydrometeorological observation in China since the 1980s, precipitation data quality has been getting better and better, and the missing rate has been getting lower and lower, which caused a jump in the number of stations that met this quality control requirement beginning in 1981 (Shen et al., 2014). If we had kept the number of gauges used for interpolation steady over the full 62 year-span, about 300 gauges available for the period 1981–2022 would have been excluded, which would have been a great loss of real observed precipitation information over the last 40 years. Therefore, the strategy adopted in this study was that all observational data that met the quality control conditions were used for data interpolation, which led to some differences in the number of sites used every year. These slight differences could be partly compensated for by using correlation decay distance (CDD1 and CDD2), which confirmed there were at least three stations involved in the interpolation for each grid cell so that there would not be a sharp change in the number of stations used for interpolation for each grid cell.

We have added discussion about this in Section 4.1 of the manuscript as follows:
"Due to a high rate of missing daily observations from 1961 to 1980, up to 15% of stations could not reach the threshold for quality control (i.e., a rate of missing daily precipitation not more than 5%) and were removed. With the vigorous development of hydrometeorological observation in China since the 1980s, precipitation data quality has been improving, which caused a jump in the number of stations that met quality control requirements beginning in 1981 (Shen et al., 2014). If we had kept the number of gauges used for interpolation steady over the full 62 year-span, about 300 gauges available for the period 1981–2022 would have been excluded, which would have been a great loss of real observed precipitation information. Therefore, the strategy adopted in this study was that all observational data that met the quality control conditions were used for data interpolation, which led to some differences in the number of sites used every year. These slight differences could be partly compensated for by using correlation decay distance (CDD1 and CDD2), which confirmed there were at least three stations involved in the interpolation for each grid cell so that there would not be a sharp change in the number of stations used for interpolation for each grid cell."

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.

Response: Thanks for your constructive suggestions. The definition of "daily climatology" is the mean value for each day over a specified time range (https://iridl.ldeo.columbia.edu/dochelp/StatTutorial/Climatologies/index.html). In this study, it does refer to the mean daily precipitation amount. Because this term is frequently employed in atmospheric sciences, we would like to keep the term "daily climatology" in the manuscript. We have revised the phrases related to the term "daily climatology" according to your suggestions and to better clarify the use of gauges versus grid cells. The revised phrases we used in this study are divided into three categories:

1) Gauge-based climatology of daily precipitation, which is defined as the Fourier-truncated 30-year mean daily precipitation series produced from gauge observations for the period of 1971–2000 for each of the 365 calendar days. We revised the phrase "raw daily climatology" in the original version into "gauge-based climatology of daily precipitation" to better convey that the process is based on gauge observations instead of grid cells. To produce the gauge-based climatology of daily precipitation, the simple mean daily precipitation amounts (mm/d) for the period (1971–2000) are first calculated for each gauge, and then high-frequency fluctuations are removed, with only the first few harmonics retained by Fourier truncation. The revised expressions for this phrase in the manuscript are listed as follows:

"First, the gauge-based climatology of daily precipitation was calculated using gauge observations. The definition of gauge-based climatology of daily precipitation is the Fourier-truncated 30-year mean daily precipitation series produced from gauge observations for the period of 1971–2000 for each of the 365 calendar days (Figure 4). We used Fourier truncation to remove the high-frequency noise of the 30-year mean daily precipitation series for each station and retained the accumulation of the first six harmonic components as the gauge-based climatology of daily precipitation (Xie et al., 2007). After Fourier truncation, approximately 75% of all stations preserve a variation of 40% to 75% in the truncated mean daily precipitation series relative to the total variation in the mean daily precipitation."

2) The unadjusted 0.05° × 0.05° gridded daily climatology field, which is interpolated from the gauge-based climatology of daily precipitation with SRTM-DEM as a covariate using ANUSPLIN software. We have changed the phrase "raw 0.05° × 0.05° gridded daily climatology field" in the previous manuscript into the phrase "unadjusted 0.05° × 0.05° gridded daily climatology field" to emphasize this is the original 0.05° × 0.05° gridded daily climatology field without any adjustment processes applied, such as monthly precipitation constraint or topographic characteristic correction. We have altered the corresponding explanation for this phrase in the manuscript as follows:

"The unadjusted 0.05° × 0.05° gridded daily climatology field (Cd_0) was then interpolated from the gauge-based climatology of daily precipitation with SRTM-DEM as a covariate using the ANUSPLIN software (Hutchinson and Xu, 2004)."

3) The adjusted gridded daily climatology field, which was adjusted by using the monthly climatology field to consider a monthly precipitation constraint and topographic characteristic correction. We have added this expression ("the adjusted gridded daily climatology field") into the manuscript as follows:

"To minimize systematic bias from the unadjusted 0.05° × 0.05° gridded daily climatology field on the monthly climatology field (Cm), the monthly accumulation of the unadjusted 0.05° × 0.05° gridded daily climatology field was then constrained by the monthly climatology field. This produced an adjusted gridded daily climatology field that uses a monthly precipitation constraint and topographic characteristic correction."

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.

Response: Many thanks for your advice. After Fourier truncation, approximately 75% of all stations preserve a variation of 40% to 75% in the truncated mean daily precipitation series relative to the total variation in the mean daily precipitation. We have added this expression to Section 3.2.

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.

Response: Thank you for your suggestion. We have carefully checked and revised the tense.

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

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

Response: Done.

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

Response: Done.

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

Response: Done.

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.

Response: Done.

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

Response: Done.

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).

Response: Done.

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 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.)

Response: Done.

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

Response: Done.

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

Response: Since the phrase "cross validation" is not appropriate here, we have revised the sentence as follows:

"Eight interpolation schemes were considered and validated using 45,992 gauge observations for the period of 2015–2019 over China."

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

Response: Done.

Line 121 are collected -> were collected.

Response: Done.

Line 124 available to use for -> available for

Response: Done.

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

Response: Done.

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

Response: Done.

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).

Response: Done. Also, since the term "cross validation" is not appropriate here, we have revised it to "validation".

Line 152: is -> was

Response: Done.

Line 165:

for the following climatology adjustment. -> for adjustment based on climatology

Response: Done.

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?)

Response: The monthly total of the unadjusted $0.05° \times 0.05°$ gridded daily climatology field is derived by taking the sum of the unadjusted 0.05° × 0.05° gridded daily climatology field for the month. This process uses the gridded daily climatology field, which is interpolated from the gauge-based climatology of daily precipitation that has experienced Fourier truncation.

We have revised the sentence as follows: "Calculate $Cd_{0\_(m,j)}$ (m = 1, 2, 3, …, 12; j = 1, 2, 3, …, 365; m is the corresponding month for day j), which is the monthly total of the unadjusted 0.05° × 0.05° gridded daily climatology field, derived by taking the sum of the unadjusted 0.05° × 0.05° gridded daily climatology field 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?)

Response: Thanks for your keen observation. We have checked and revised equations (1) and (2) like this:

"3) Compute the scaling factor $SF_{(m,j)}$ for the individual calendar day of the unadjusted 0.05° × 0.05° daily climatology field to the gridded monthly climatology field:

$$SF_{(m,j)} = \frac{C_{(m,j)}}{w_{(m-1,j)}\, Cd_{0\_(m-1,j)} + w_{(m,j)}\, Cd_{0\_(m,j)} + w_{(m+1,j)}\, Cd_{0\_(m+1,j)}} \quad (1)$$

$$(m = 1, 2, 3, \ldots, 11, 12;\ j = 1, 2, 3, \ldots, 365;$$
$$m \text{ is the corresponding month for day } j)$$

where $C_{(m,j)}$ is the gridded monthly climatology field for the corresponding month m of day j; $Cd_{0\_(m-1,j)}$, $Cd_{0\_(m,j)}$ and $Cd_{0\_(m+1,j)}$ are the monthly total of months $m-1$, m, and $m+1$, respectively, which are calculated from the unadjusted 0.05° × 0.05° gridded daily climatology field; $w_{(m-1,j)}$, $w_{(m,j)}$, and $w_{(m+1,j)}$ are the corresponding weights for months $m-1$, m, and $m+1$, respectively, which are inversely proportional to the interval between the calendar day j and the center of the month (Xie et al., 2007). Note that the weight $w_{(m-1,j)}$ is zero when m = 1, and so is the weight $w_{(m+1,j)}$ when m = 12.

4) The adjusted gridded daily climatology field ($Cd_{(m,j)}$) is defined as

$$Cd_{(m,j)} = Cd_{0\_(m,j)}\, SF_{(m,j)} \quad (2)$$

$(m = 1, 2, 3, \ldots, 11, 12;\ j = 1, 2, 3, \ldots, 365;$

m is the corresponding month for day j)"

Line 255: we used in -> used for

Response: Done.

Equation (4), Should be expressed as e^(-nx_i/GDD)

Response: Thanks for the suggestion. We have added the subscript (x_i) to represent the distance between station i and the center of target grid cell L. To make it easier for readers to understand, we keep the power n outside of the bracket. The expression for equation (4) has been revised as follows:

"$$D_i = \left(e^{-x_i/CDD}\right)^n \quad (4)$$

where x_i is the distance between station i and the center of target grid cell L; n is a constant and usually set to 4, in accordance with previous studies (Harris et al., 2020; Dunn et al., 2020; Efthymiadis et al., 2006)."

Equation (5), What is the definition of surrounding stations?

Response: "Surrounding stations" are all of the stations except for station i. We have added this definition in the manuscript: 'Here, "surrounding stations" refers to all other stations except for station i.'

Equation (6), Normally we do not use 'X' in equations to indicate multiplication.

Response: Many thanks for the advice. We have revised the equation (6) into the following format: "$W_i = D_i A_i$"

For IDW, again what are the surrounding stations?

Response: For IDW, the surrounding stations are defined as the stations that fall in the search radius of the target grid cell.

We have added the definition in the manuscript as follows: "We employed S_0 representing the target grid cell, i representing the surrounding stations that fall in the search radius of the target grid cell, $y(s_i)$ denoting the station observations, and d_{0i} denoting the distance between S_0 and i."

Line 247 – repetition from data section above.

Response: Done.

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.

References:

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, Journal of Applied Meteorology, 33, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

Shen, Y., Zhao, P., Pan, Y., and Yu, J.: A high spatiotemporal gauge-satellite merged precipitation analysis over China, Journal of Geophysical Research: Atmospheres, 119, 3063–3075, https://doi.org/10.1002/2013JD020686, 2014.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A gauge-based analysis of daily precipitation over east Asia, Journal of Hydrometeorology, 8, 607–626, https://doi.org/10.1175/JHM583.1, 2007.