

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2022-299-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-299**

Anonymous Referee #2

Referee comment on "TPHiPr: a long-term (1979–2020) high-accuracy precipitation dataset (1/30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations" by Yaozhi Jiang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-299-RC2>, 2022

General comments:

High-resolution precipitation over the Tibetan Plateau (TP) region is important in climate science and other related fields. Climate models can simulate high spatial-temporal resolution precipitation datasets but generally overestimate the precipitation amount. The gauge-based rainfall observations are relatively accurate but only short-period, sparse-distribution records. This manuscript tries to take the advantage of both two and generates a high-resolution 1/30° long-term (1979–2020) precipitation dataset (TPHiPr) over the TP. A high-resolution pre-derived precipitation dataset (ERA5-CNN) and a dense gauge-based dataset are used. The manuscript describes the merged procedure and then intercompared the TPHiPr with independent station observations and several global datasets. The TPHiPr will benefit the researchers who are working on the climate or related works. However, before the manuscript was published in the journal, the below comments should be answered or clarified.

Major comments:

- From the data construction procedure (flow chart) and description in section 3, the RF and Kriging were repeatedly used to convert data between grid cells and gauge stations. However, the manuscript does not provide the reasons and also does not describe the methods in detail. Machine Learning has been used in climate sciences for decades and it includes many different algorithms. The RF is only one of them. Similarly, ordinary Kriging is also one of the interpolation methods. There should be specific reasons to choose those two approaches. It is necessary to provide them clearly in the manuscript.
- L193–196. "the daily precipitation fields after residual correction (Pd2) are further adjusted to ensure that the sum of the daily precipitation amount in a month..." At a certain station/grid cell in the TP, the non-raining day in a month should be very

common. Let's take an assumption. When the above monthly precipitation is greater than "the sum of the daily precipitation amount in a month", how do you perform the "adjust" on both rainy days and non-raining days? If you only add the differences in the amount on rainy days, this would enhance daily extreme. Otherwise, it will increase the frequency of rainfall if both rainy or non-raining days are "adjusted". A detailed "adjust" process is needed.

- Figure 2 and section 3 present the data construction procedure based on the ERA5_CNN and observations at gauged stations. Over regions without observation (e.g., northwest TP in Figure 1b), is the TPHiPr directly from ERA5_CNN or another approach? Compared to Figure 3 and Figure 1b, it seems that regions without stations also show non-zero differences between TPHiPr and ERA5_CNN.

Minor comments:

- The latitude and longitude labels on both the x-axis and y-axis are needed for all figures with the map.
- L124 To correct the biases of gauged precipitation, wind speed and air temperature from ERA 5 are used. Why do you use both variables from ERA5? Do you have any justification?
- What interpolated methods are used to convert the TPHiPr from grid cell to station location when they are intercompared?
- L266-268 it is necessary to explicitly the station location in Figure 1 or in an additional figure. Also, the temporal range/resolution of those rain gauge-based precipitation should be given.
- Figure 7 shows the mean seasonal precipitation amounts from different databases. The spatial patterns of those datasets are very similar and cannot be distinguished by eye. I suggest plotting the differences between the three reference datasets and the TPHiPr.