Response to the Reviewers

General response: Thank you for the reviewers' comments and these comments are critical for improving this work. All comments are considered and addressed carefully. The comments are given in black typeface and the authors' responses are given in blue typeface. Line numbers refer to those in the marked-up version of the revised manuscript.

Response to Reviewer 1

"TPHiPr: A long-term high-accuracy precipitation dataset for the Third Pole region based on high-resolution atmospheric modelling and dense observations" by Jiang et al.

Overall summary:

Long-term high-accuracy precipitation dataset for the Third Pole is greatly needed, which is also a great challenge for researchers due to complex and bad weather systems over TP. This study seems to provide a novel strategy for potentially catering this challenge. While after carefully reading this manuscript, there are still various aspects confusing me. The largest issues mainly include: (1) the authors generated the TPHiPr based on ERA5_CNN and gauge observations, why not the ERA5_Land? (2) The robustness of the algorithm needs to be furtherly demonstrated and revealed, especially the RF as a black box without a detailed description on the parametrization strategy; (3) the current validations of TPHiPr, compared with other gridded precipitation datasets, are not convinced enough; and (4) the overall language and structures are relatively poor, of which still needs to be greatly improved. Therefore, I recommend a "Major" at this stage.

Response: Thanks for the reviewer's helpful comments! We have thoroughly considered the concerns and revised the manuscript accordingly, mainly including the following aspects:

(1) We added more details about the ERA5_CNN and further clarify the reasons for selecting the ERA5_CNN as the background field.

(2) We further clarified the underlying logic for using such an algorithm and provided more details about the methods.

(3) We added the evaluation results of a regional precipitation dataset (AERA5_Asia) in the revised manuscript to extend the comparison with other datasets.

(4) We have carefully edited the language of the manuscript.

Major concerns:

1. The TPHiPr is generated based on ERA5_CNN and gauge-based observations. Why did not the authors use the ERA5_Land with the high resolutions of 0.1 deg and hourly. Even though the authors should give some detailed descriptions on the ERA5_CNN. Recently, various investigations found that the ERA5_Land has many advantages, compared with satellite-based precipitation estimates. For instance, "Do ERA5 and ERA5-Land Precipitation Estimates Outperform Satellite-based Precipitation Products? A Comprehensive Comparison between State-of-the-art Model-based and Satellite-based Precipitation Products over Mainland China" .

Response: There are two reasons for selecting ERA5_CNN rather than ERA5_land in our work: first, the ERA5_CNN (1/30°) has a higher spatial resolution than ERA5_land (0.1°), which can provide finer spatial variability of precipitation in complex terrain (as shown in Figure R1); second, as shown in Figure R1 and reported in the reviewer's suggested article (Xu et al., 2022), although ERA5_land has a higher horizontal resolution than ERA5, they only have slight differences in the Third Pole (TP) region. However, our previous work showed that the ERA5_CNN has smaller wet biases than the ERA5 in the TP (Figure

R2; Jiang et al., 2021) and is more skillful in representing the spatial variability of precipitation (Figure R3; Jiang et al., 2022). We have further clarified the advantages of using the ERA5_CNN as a background dataset. Please refer to Line 168-174 in the marked-up version of the revised manuscript.

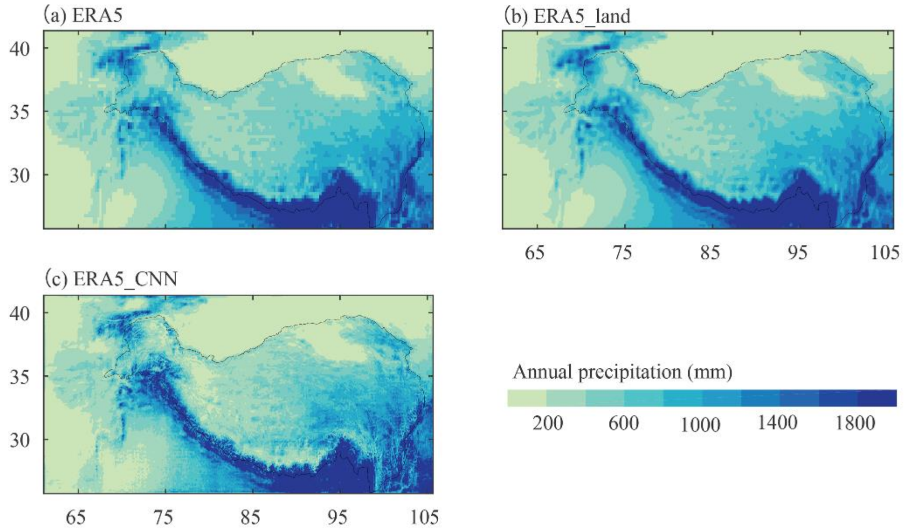

Figure R1: Spatial patterns of annual precipitation from (a) ERA5, (b) ERA5_land and (c) ERA5_CNN. The precipitation is averaged over 2008-2020.

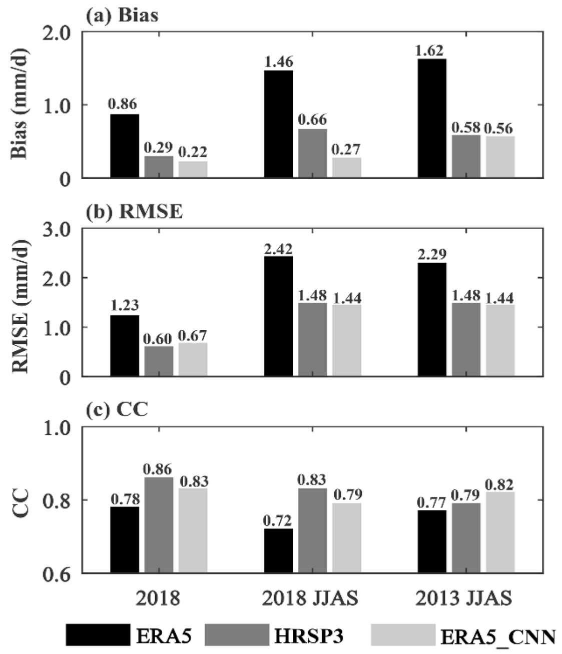

Figure R2: Spatial error metrics for ERA5, high-resolution WRF simulation (HRSP3) and ERA5_CNN with respect to three different periods, namely, whole 2018, June to September of 2018, and June to September of 2013. The error metrics were calculated based on the time-averaged precipitation of 96 CMA stations (Jiang et al., 2021).

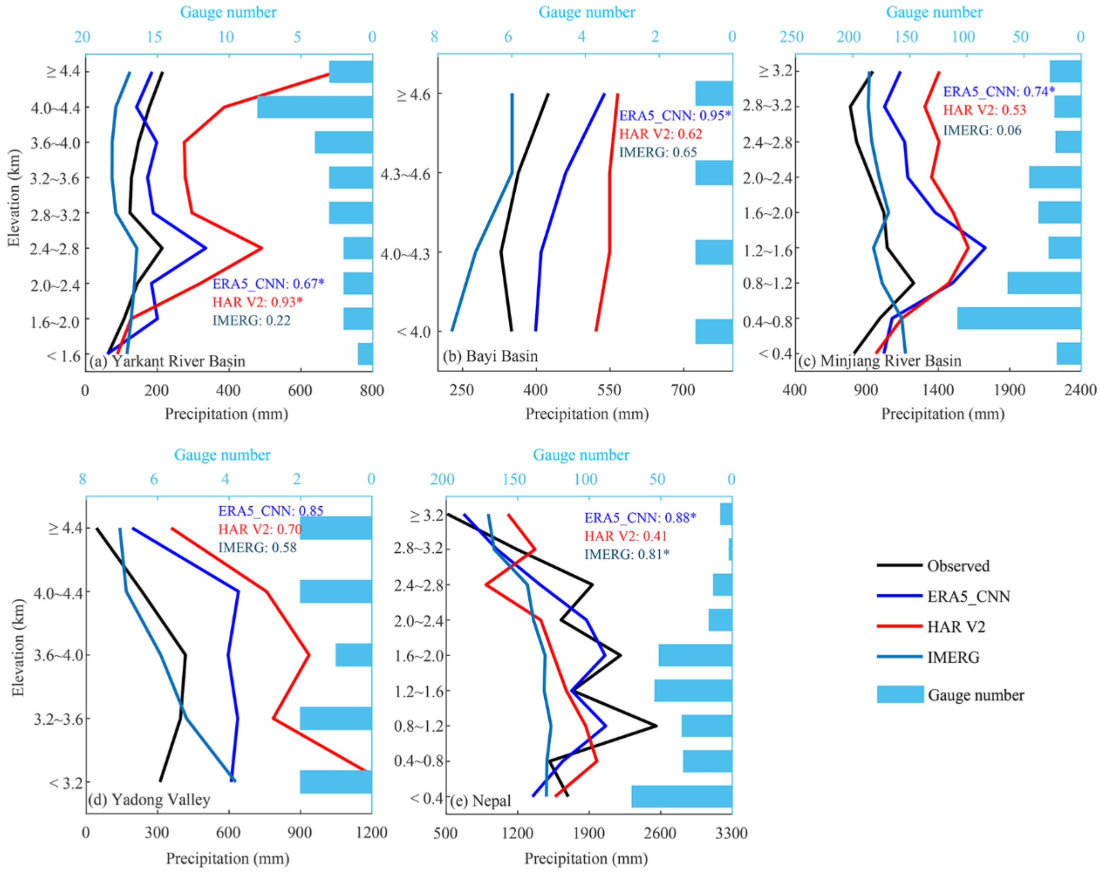

Figure R3: Comparison between the altitude dependence of precipitation from ERA5_CNN, IMERG and HAR V2 and that from rain gauge data in five networks. The lines show the average precipitation amount in each altitude zone and the bars denote the number of rain gauges in each zone. The numbers in the figures give the spatial correlations of the precipitation amount between the rain gauge data and precipitation products. The "∗" represents the correlation that is significant at the 95% confidence level (Jiang et al., 2022).

Jiang, Y., Yang, K., Shao, C., Zhou, X., Zhao, L., Chen, Y.: A downscaling approach for constructing high-resolution precipitation dataset over the Tibetan Plateau from ERA5 reanalysis, Atmos. Res., 256, 105574. https://doi.org/10.1016/j.atmosres.2021.105574, 2021.

Jiang, Y., Yang, K., Yang, H., Lu, H., Chen, Y., Zhou, X., Sun, J., Yang, Y., Wang, Y.: Characterizing basin-scale precipitation gradients in the Third Pole region using a high-resolution atmospheric simulation-based dataset, Hydrol. Earth Syst. Sci., 26, 4587–4601, https://doi.org/10.5194/hess-26-4587-2022, 2022.

Xu, J., Ma, Z., Yan, S., Peng, J.: Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China, J. Hydrol., 605, 127353, https://doi.org/10.1016/j.jhydrol.2021.127353, 2022.

2. The organization of Introduction is relatively poor and some very related research is just only simply mentioned, for example, in line 79. The authors need to pay great attentions to give a comprehensive review of the merging algorithms to meet the standard of the big journal, ESSD. For instance, some

recently representative merging algorithms, "A Morphology-based Adaptively Spatio-Temporal Merging Algorithm (MASTMA) for optimally combining multi-source gridded precipitation products with various resolutions", as well as AERA5_Asia and AIMERG.

Response: Thanks for the comment. We conducted a more comprehensive literature review about the merging algorithms of precipitation in the revised manuscript. Please refer to Line 83-93 in the marked-up version of the revised manuscript.

3. As for merging algorithm, at least two major issues should be concerned: (1) the description of the flowchart is not very clear and readable; (2) the robustness of the algorithm should be further manifested, for instance, why RF is used in this study? And it is also like a black box without any introduction of the parametrizations.

Response: Thanks for the comments. The flowchart in the previous manuscript is indeed very complex and we have refined it. Figure R4 (i.e. Figure 2 in the revised manuscript) gives the refined flowchart.

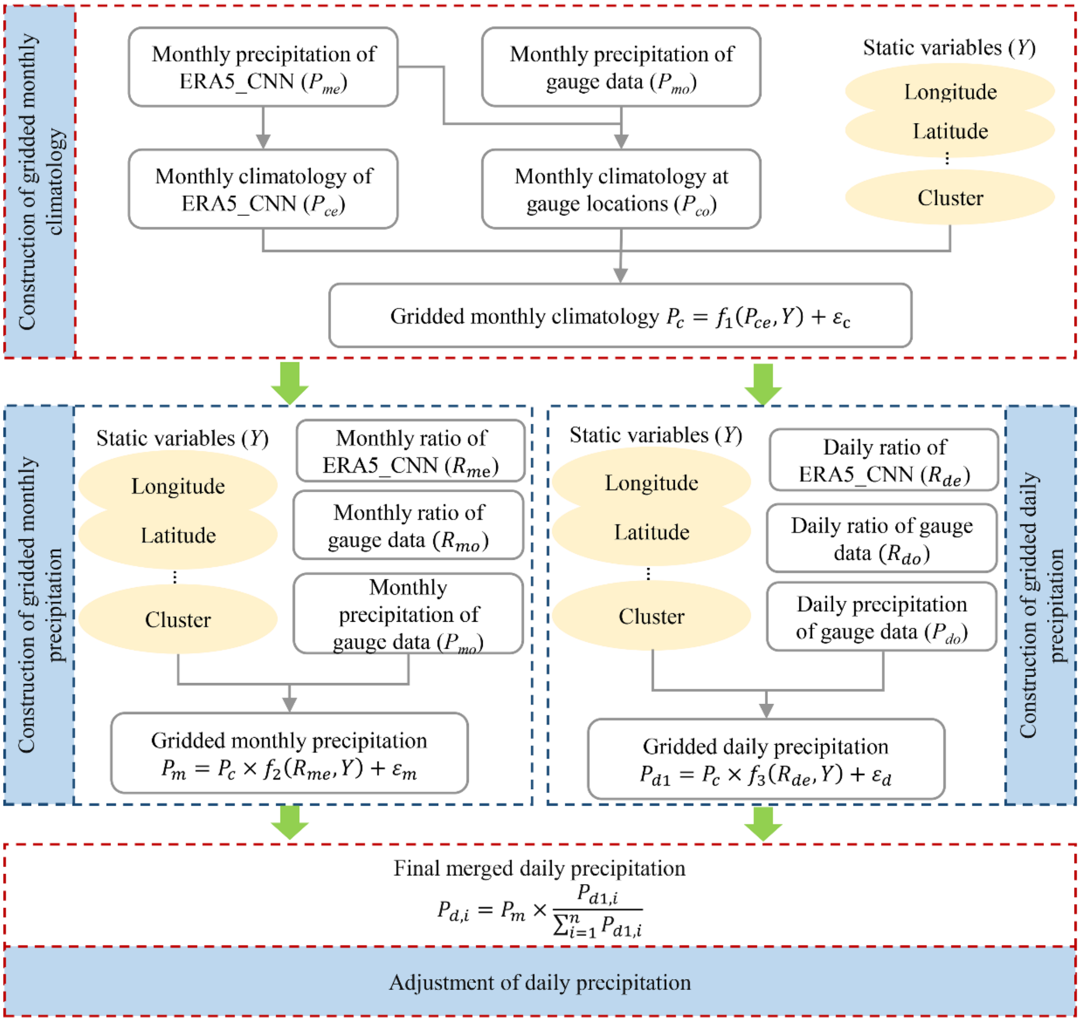

Figure R4: General flowchart of the merging algorithm. The static variables include the longitude, latitude, elevation, standard deviation of elevation and the identifier of the clusters with different precipitation characteristics. The subscript '*o*' represents 'observation', '*e*' represents 'ERA5_CNN', '*c*' represents 'climatology', '*m*' represents 'monthly', '*d*' represents 'daily', '*n*' represents the number of days in a month and '*i*' represents the *i*th day in a month. f_1 (•), f_2(•) and f_3(•) denote the regression models based on Random Forest. ε_c, ε_m and ε_d represent the residuals of estimations from RF, which are interpolated using the Kriging method.

The interpolation algorithm used in our study is based on the idea of Regression Kriging, in which the interpolated target is assigned to the spatial trend (deterministic) and the stochastic component (residual). A regression model is applied to predict the spatial trend and the Ordinary Kriging is used to estimate the stochastic component that is expected to be Gaussian distribution. In this method, multiple regression methods can be combined with Kriging. Machine learning-based regression models combined with Kriging were widely applied in earth science and proved to have good performance, as reported in many previous works (Araki et al., 2015; Cellura et al., 2008; Demyanov et al., 1998).

In terms of different machine learning methods, the Random Forest (RF) is an ensemble method based on Decision Trees. It randomly selects samples for training each Decision Tree and aggregates estimates from multiple Decision Trees. Compared to other machine learning methods, the RF is less sensitive to hyperparameters, less likely to suffer from overfitting and has good generalization capability. Moreover, The RF is easy to implement and has a robust prediction accuracy. Many works have applied the RF in earth science and demonstrated its good performance (Baez-Villanueva et al., 2020; He et al., 2016; Zhang et al., 2021). To demonstrate the reliability of the RF, we compared the performance of four widely-used machine learning methods for estimating the monthly precipitation in 2018. Figure R5 shows that the RF generally performs better than the other three methods. Because the comparison of different method is not in the scope of our study, this figure is not included in the revised manuscript.

In the revised manuscript, we further clarified the underlying logic of the merging algorithm and introduce more about the method. Please refer to Line 258-266 and Line 268-271 in the marked-up version of the revised manuscript.

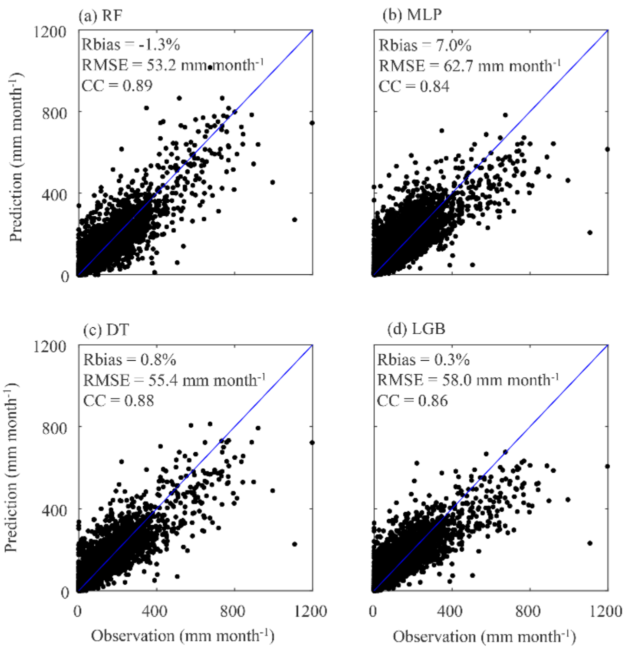

Figure R5: Comparison between the monthly precipitation in 2018 estimated by four machine learning models and the observed monthly precipitation. RF: Random Forest; MLP: Multi-layer Perceptron; DT: Decision Trees; LGB: LightGBM.

Araki, S., Yamamoto, K., Kondo, A., 2015. Application of regression kriging to air pollutant concentrations in Japan with high spatial resolution. Aerosol Air Qual. Res. 15, 234–241.

https://doi.org/10.4209/aaqr.2014.01.0011
Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Xuan Thinh, N., 2020. RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606. https://doi.org/10.1016/j.rse.2019.111606
Cellura, M., Cirrincione, G., Marvuglia, A., Miraoui, A., 2008. Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. Renew. Energy 33, 1251–1266. https://doi.org/10.1016/j.renene.2007.08.013
Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E., Timonin, V., 1998. Neural Network Residual Kriging Application for Climatic Data 2, 215–232.
He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. Water Resour. Res. 52, 8217–8237. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x
Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., Ge, Y., 2021. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. J. Hydrol. 594, 125969. https://doi.org/10.1016/j.jhydrol.2021.125969

4. The precipitation detection index (POD, FRA, and CSI) is mainly applied for evaluating the precipitation estimates at hourly or sub-daily scales. One issue is greatly confusing me is that the detection abilities of TPHiPr has significant improvements from Figs.10–12, so the question is that what are the advantageous parts of the merging algorithm for these contributions?

Response: The merging algorithm itself has no special consideration for improving the detection skills. The significant improvements in detection skills mainly benefit from the high-density gauge observations.

5. The author seems to be aiming at improving ERA5_CNN, however, they compared the qualities of TPHiPr with those of ERA5 in Figs. 6–14, while not ERA5_CNN or ERA5_Land, which is greatly strange. Additionally, the spatial resolution of ERA5 (0.25 deg) is much coarser than that of ERA5_Land (0.1 deg).

Response: The manuscript has already compared the performance of ERA5_CNN and the TPHiPr. Please refer to section 4.1. Nevertheless, given that ERA5_land has similar precipitation amount and spatial pattern to ERA5 (Figure R1) but has a higher spatial resolution, we replaced the evaluation results of ERA5 with those of ERA5_land in the revised manuscript. The results shows that ERA5_land performs similar to ERA5 and our produced TPHiPr also shows superiority to ERA5_land, as shown in Figure R6 and Figure R7 (i.e. Figure 9 and Figure 11 in the revised manuscript). Please refer to section 4.2 and section 5 in the revised manuscript to find the results.

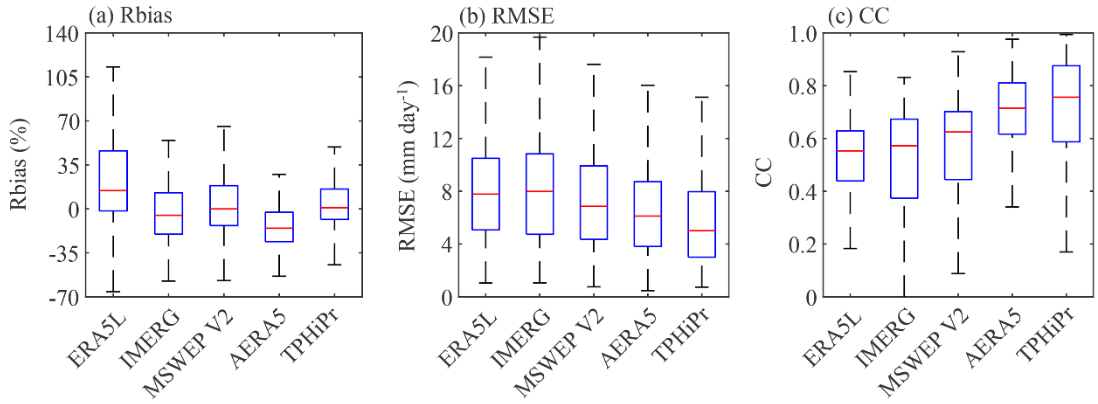

Figure R6: Comparison of (a) Rbias, (b) RMSE and (c) CC for ERA5L, IMERG, MSWEP V2, AERA5 and TPHiPr. The box represents the distribution of the metrics for all the 197 independent rain gauges in the TP.

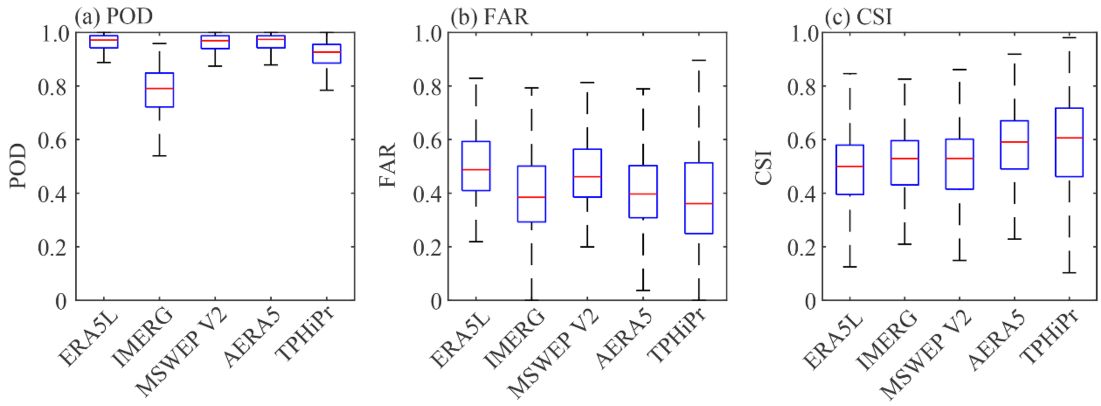

Figure R7: Similar to Figure R6 but for (a) POD, (b) FAR and (c) CSI. These metrics are calculated using a threshold of 0.1 mm day^{-1}.

6. The evaluation section seem not to be robust and comprehensive, which needs to be greatly redesigned and extended.

Response: To make our manuscript more comprehensive, we have added the comparison of a regional precipitation dataset (AERA5-Asia) with our produced dataset in the revised manuscript. The results (Figure R6 and R7) showed that AERA5-Asia performs better than the widely-used global/ quasi-global precipitation datasets (ERA5_land, IMERG and MSWEP V2) over the TP, nevertheless, our produced TPHiPr still exhibits better accuracy than the AERA5-Asia in terms of most error metrics, mainly because our study has used more rain gauge data. Please refer to section 4.2 and section 5 to find the results.

7. As to further demonstrating the quality of TPHiPr and the robustness of the merging algorithm, I recommend the authors to add a discussion paragraph for comparing the characteristics and/or the qualities of AERA5_Asia (AIMERG is optional) with that of TPHiPr: AERA5-Asia “A long-term Asian precipitation dataset (0.1°, 1 hourly, 1951–2015, Asia) anchoring the ERA5-Land under the total volume control by APHRODITE” and AIMERG “a new Asian precipitation dataset (0.1°/half-hourly, 2000–2015)

by calibrating GPM IMERG at daily scale using APHRODITE”.

Response: According to the reviewer’s suggestion. We have conducted an evaluation of AERA5-Asia and compare its performance with our produced product. The results (Figure R6 and R7) showed that the AERA5-Asia performs better than the widely-used global/quasi-global precipitation datasets (ERA5_land, IMERG and MSWEP V2) over the TP, nevertheless, our produced TPHiPr still exhibits better accuracy than the AERA5-Asia in terms of most error metrics, mainly because our study has used more rain gauge data. Please refer to section 4.2 and section 5 to find the results.

8. The language still need to be greatly improved.

Response: We have carefully edited the language of the manuscript.

Specific comments:

1. the spatiotemporal resolutions, temporal span, and extent should be noted in the Title and Abstract following the TPHiPr, making it more clear for readers to know the characteristics of this dataset.

Response: Thanks for the suggestion. We have revised the title to “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”. In addition, we also gave these details in the abstract. Please refer to Line 22-23 in the marked-up version of the revised manuscript.

2. what’s the relationship between the ERA5 and ERA5_CNN in this manuscript, which is fully mixed and confused.

Response: The ERA5_CNN was produced by downscaling the ERA5 using a convolutional neural network (CNN)-based method, which was trained with short-term high-resolution (1/30°) WRF simulation. The ERA5_CNN has a higher spatial resolution(1/30°) than ERA5 and our previous works showed that it is more skillful in representing the spatial variability of precipitation and has smaller wet biases in the TP than the ERA5. These details were given in Line 162-177 in the marked-up version of the revised manuscript.

3. The Introduction still needs to be further improved, especially the aims of this study. For instance, are the authors sure the rain gauge data is unprecedented?

Response: Given that the Third pole region is the hotspot of hydrological, meteorological and ecological studies but precipitation in this region shows large uncertainties due to the complex terrain here, the aim of this study is to produce a long-term, high-resolution and high-accuracy precipitation dataset for the Third Pole region. Our produced dataset has three distinguishing features: (1) our dataset is produced based on a high-resolution atmospheric simulation, which is skillful in modeling solid precipitation and representing spatial variability of precipitation in complex terrain. (2) our dataset has merged rain gauge data from more than 9000 rain gauges, including the rain gauges in the central and western TP set up by our research group. However, most previous works in this region have only merged data from sparse rain gauge networks (generally no more than 1000 gauges) that are mainly parts of the CMA or MWR stations and usually located in the eastern TP. (3) our dataset has a relatively high spatial resolution of 1/30°, while the spatial resolutions of most existing datasets in this region are coarser than 10 km. We have further clarified the objective of this work in the revised manuscript. Please refer to Line 105-116 in the marked-up version of the revised manuscript.

4. What does the blue line mean in line 109? Is it the extent of the TP?

Response: Yes, the blue line denotes the 2500 m contour of elevation, which is considered to be the boundary of the TP.

5. what's the temporal resolution of gauge observations? There are many such points that are not clearly described.

Response: The gauge observations have daily or sub-daily records, which was given in Line 124 and Line 127 in the marked-version of the revised manuscript. We have aggregated the sub-daily records into daily precipitation so that they can be merged with gridded data at a daily scale. These information was added to the marked-up revised manuscript (Line 128-129). In addition, we have checked the unclear points and clarified them in the revised manuscript.

6. the authors seem to generate TPHiPr using ERA5_CNN and the gauge observations, while they compared the quality of TPHiPr with that of ERA5. And the ERA5_Land with high spatial resolution of 0.1 deg is not even mentioned in this study. The logicality needs to be redesigned.

Response: According to the reviewer's suggestion, we have replaced the evaluation results of ERA5 with those of ERA5_land in the revised manuscript. According to the results, ERA5_land performs similar to ERA5 and our produced TPHiPr also shows superiority to ERA5_land. Please refer to section 4.2 and section 5 in the revised manuscript.

7. RF and Kriging are used in many times in Fig. 2, how did you concern the uncertainties and errors from these methods?

Response: We have to acknowledge that these methods indeed contain some uncertainties. In fact, even the most advanced merging algorithms contain uncertainties and we just use methods that are widely accepted and used in earth science. The main contribution of our study is that we use dense rain gauge data and use a background field derived from high-resolution WRF simulation which is skillful in representing precipitation variability in complex terrain, rather than proposing an advanced merging algorithm. Moreover, the results in section 4.1 show that the merging algorithm indeed can improve the quality of the precipitation dataset.

8. Fig. 6 presented very limited information.

Response: Thanks for the comment. We removed this figure in the revised manuscript.

Response to Reviewer 2

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.

Response: Thanks for the reviewer's comments and we believe that these comments are beneficial for improving our work. We have carefully considered these comments and a point-by-point response is given as follows. Line numbers refer to those in the marked-up version of the revised manuscript.

Major comments:

1. 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 algorisms. 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.

Response: Thanks for the comments. The interpolation algorithm used in our study is based on the idea of Regression Kriging, in which the interpolated variable is assigned to the spatial trend (deterministic) and the stochastic component (residual). A regression model is applied to predict the spatial trend and the Ordinary Kriging is used to estimate the stochastic component that is expected to be Gaussian distribution. In this method, multiple regression methods can be combined with Kriging. Machine learning-based regression models combined with Kriging were widely applied in earth science and proved to have good performance, as reported in many previous works (Araki et al., 2015; Cellura et al., 2008; Demyanov et al., 1998).

In terms of different machine learning methods, the Random Forest (RF) is an ensemble method based on Decision Tree. It randomly selects samples for training each Decision Tree and aggregates estimates from multiple Decision Trees. Compared to other machine learning methods, the RF is less sensitive to hyperparameters, less likely to suffer from overfitting and has good generalization capability. Moreover, The RF is easy to implement and has a robust prediction accuracy. Many works have applied the RF in earth science and demonstrated its good performance (Baez-Villanueva et al., 2020; He et al., 2016; Zhang et al., 2021). To demonstrate the reliability of the RF, we compared the performance of four widely-used machine learning methods for estimating the monthly precipitation in 2018. Figure R8 shows that the RF generally performs better than the other three methods. Because the comparison of different method is not in the scope of our study, this figure is not included in the revised manuscript.

In the revised manuscript, we further clarified the underlying logic of the merging algorithm and introduce more about the method. Please refer to Line 258-266 and Line 268-271 (marked-up version).

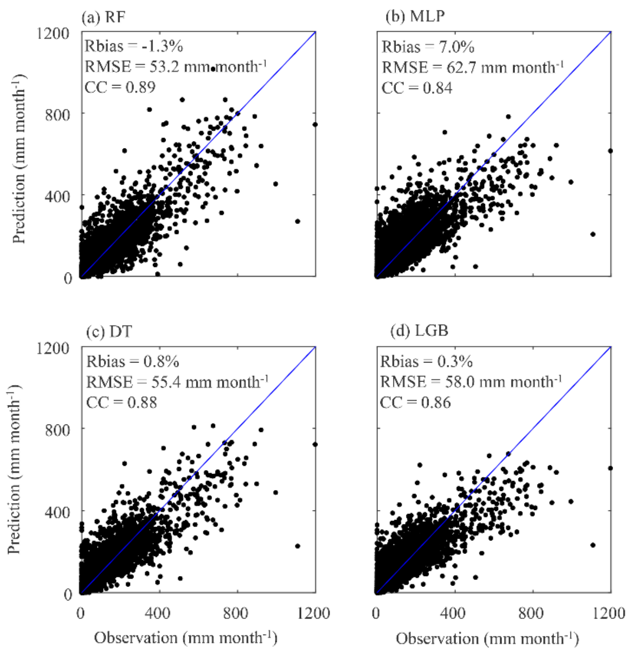

Figure R8 Comparison between the monthly precipitation in 2018 estimated by four machine learning models and the observed monthly precipitation. RF: Random Forest; MLP: Multi-layer Perceptron; DT: Decision Trees; LGB: LightGBM.

Araki, S., Yamamoto, K., Kondo, A., 2015. Application of regression kriging to air pollutant concentrations in Japan with high spatial resolution. Aerosol Air Qual. Res. 15, 234–241. https://doi.org/10.4209/aaqr.2014.01.0011

Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Xuan Thinh, N., 2020. RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606. https://doi.org/10.1016/j.rse.2019.111606

Cellura, M., Cirrincione, G., Marvuglia, A., Miraoui, A., 2008. Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. Renew. Energy 33, 1251–1266. https://doi.org/10.1016/j.renene.2007.08.013

Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E., Timonin, V., 1998. Neural Network Residual Kriging Application for Climatic Data 2, 215–232.

He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. Water Resour. Res. 52, 8217–8237. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., Ge, Y., 2021. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. J. Hydrol. 594, 125969. https://doi.org/10.1016/j.jhydrol.2021.125969

2. 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.

Response: Thanks for the comments. We adjust the daily precipitation as follows:

$$P_{d,i} = P_m \times \frac{P_{d1,i}}{\sum_{i=1}^{n} P_{d1,i}}$$

Where $P_{d,i}$ is the adjusted precipitation for the ith day in a month, $P_{d1,i}$ is the precipitation before adjustment for the ith day, P_m is the monthly precipitation and n is the number of days in that month. When the monthly precipitation (P_m) is no-zero but the sum ($\sum_{i=1}^{n} P_{d1,i}$) of the daily precipitation amount in that month is zero, we will search the nearest grid that has a non-zero $\sum_{i=1}^{n} P_{d1,i}$ and then disaggregate P_m to daily precipitation according to the day-to-day variation of precipitation in the nearest grid. In fact, the differences between P_m and ($\sum_{i=1}^{n} P_{d1,i}$) are small at most case and the adjustment does not increase the daily extreme. These contents were added to the revised manuscript (Please refer to Line 246-252 in the marked-up version of the revised manuscript).

3. 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.

Response: In regions without observation, the correction value is also non-zero. In the merging algorithm, the RF model is trained at gauge locations but the trained model is applied to all grids in the study area, which will result in precipitation changes in ungauged regions. In addition, the Kriging-based residual correction can also change the precipitation amount, although its impact is more evident in regions close to the gauges and less in regions far from the gauges.

Minor comments:

1. The latitude and longitude labels on both the x-axis and y-axis are needed for all figures with the map.

Response: Thanks for the comment. We have added the latitude and longitude labels for all maps in the revised manuscript.

2. 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?

Response: The ERA5 is the latest generation of reanalysis which has assimilated lots of in situ data. Our evaluation based on CMA stations showed that the wind speed and air temperature from ERA5 generally have better performance than those from two other datasets in the Third Pole (Figure R9 and R10). In addition, the results of Huai et al. (2021) also demonstrated the superiority of the near-surface climate from ERA5 to some other reanalysis datasets in the Third Pole region. Moreover, the ERA5 has a long time series, which can be used for correcting the early gauged precipitation. We have given the reasons why we use wind speed and air temperature from ERA5 in the revised manuscript. Please refer to Line 156-160 in the marked-up version of the revised manuscript.

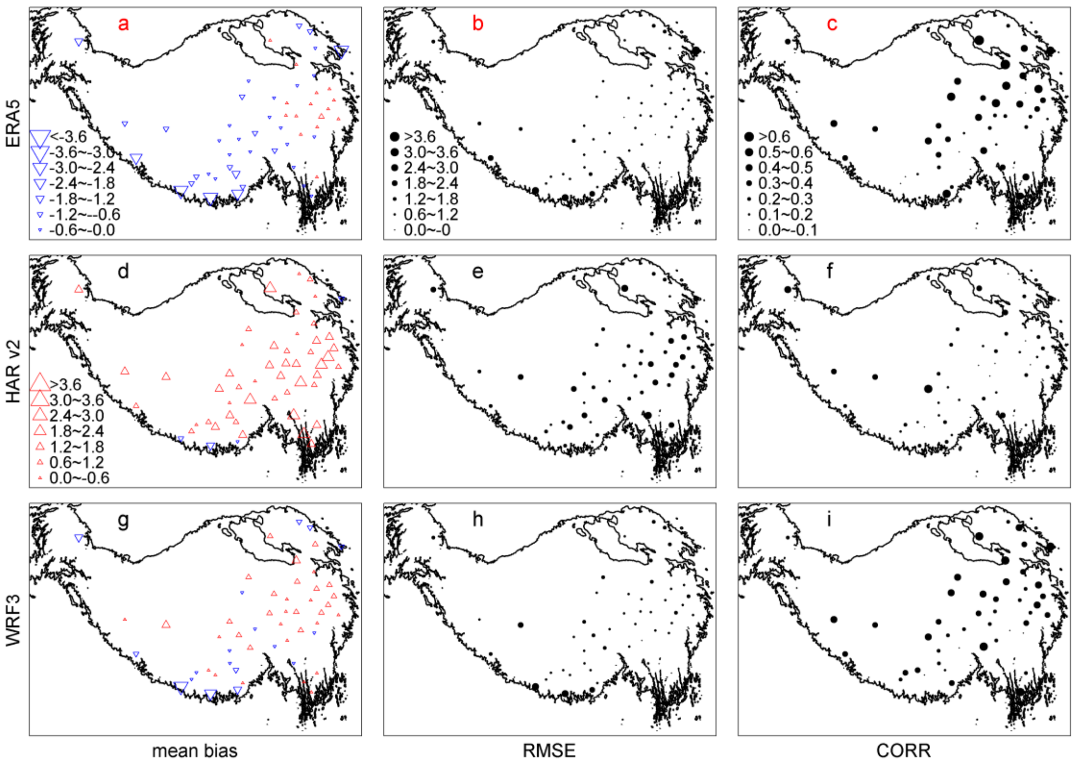

Figure R9 Error metrics at each station based on daily 10-m wind speed derived from (a–c) ERA5, (d–f) HAR v2 and (g–i) WRF3 versus observation for the period from June to September of 2013.

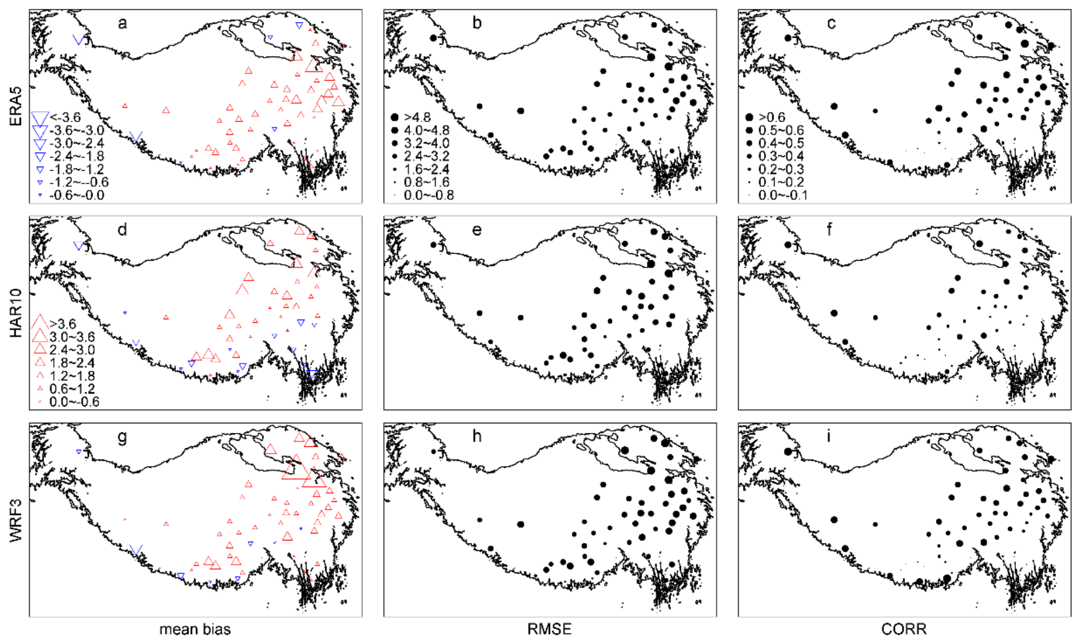

Figure R10 Error metrics at each station based on daily 2-m air temperature derived from (a–c) ERA5, (d–f) HAR v2 and (g–i) WRF3 versus observation for the period from June to September of 2013.

Huai, B., Wang, J., Sun, W., Wang, Y., Zhang, W., 2021. Evaluation of the near-surface climate of the recent global atmospheric reanalysis for Qilian Mountains, Qinghai-Tibet Plateau. Atmos. Res. 250, 105401. https://doi.org/10.1016/j.atmosres.2020.105401

3. What interpolated methods are used to convert the TPHiPr from grid cell to station location when they are intercompared?

Response: We compared the gauge observations with the precipitation from the nearest TPHiPr grid. This information was added to the revised manuscript (see Line 306-307 in the marked-up version). Our dataset has a spatial resolution of 1/30°, the spatial scale of our dataset is more close to gauge observations than other coarse datasets. Nevertheless, we have to acknowledge that a spatial scale mismatch still exists, to some extent, between these two datasets. In the future, very high-resolution datasets are still needed for dealing with this problem,.

4. 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.

Response: Thanks for the comment. We have gave the locations of the independent rain gauges for validation in Figure R11c. The temporal range varies from gauge to gauge and it is hard to define the temporal range for each rain gauge, instead, we counted the length of data records for each rain gauge and shown in Figure R11b. In addition, we counted the number of rain gauges in each year and shown in Figure R11d. These figures have been added to the revised manuscript (i.e. Figure 1 in the revised manuscript).

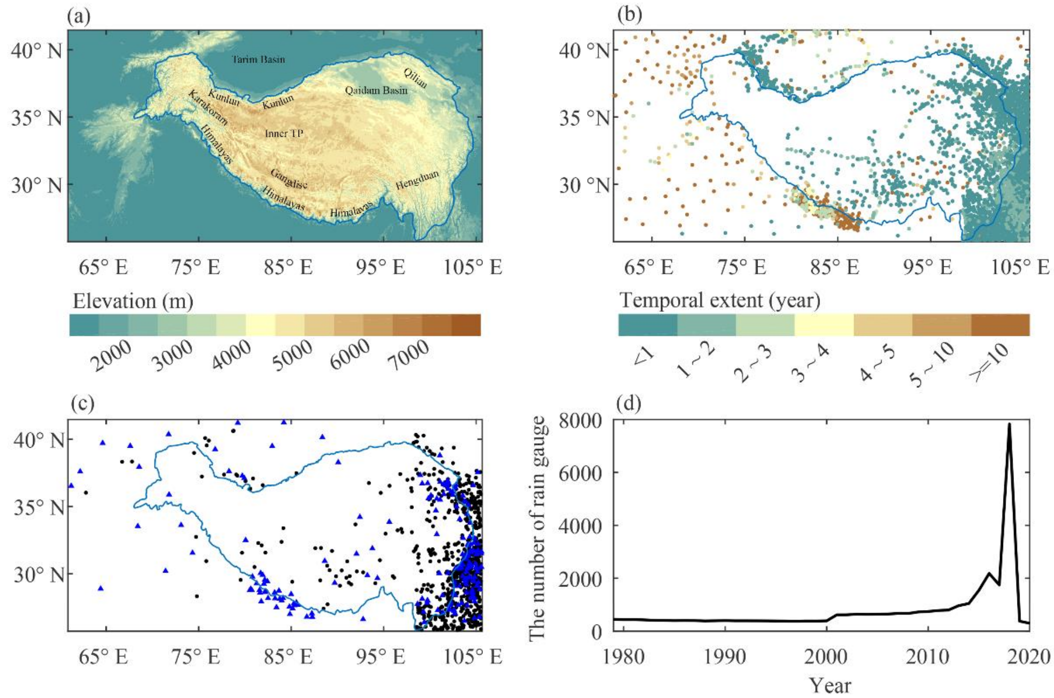

Figure R11: (a) Topography of the Third Pole region. (b) Spatial distribution of rain gauges used in this study and their temporal extent. (c) The independent rain gauges used for validation, in which rain gauges marked by both black dot and blue triangle are used in the analysis period of 1979-2020 (section 4.1.2), and rain gauges marked by blue triangle are used in the analysis period of 2008-2015 (section 4.4.2), considering the availability of gauge data in different periods. (d) The number of available rain gauges in each year.

5. 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.

Response: That is a good suggestion. We gave the differences between these datasets and the TPHiPr in Figure R12 and Figure R13 (i.e. Figure 6 and Figure 8 in the revised manuscript). Besides, description of these figures were given in Line 380-382, Line 393-400 and 412-419 in the marked-up version of the revised manuscript.

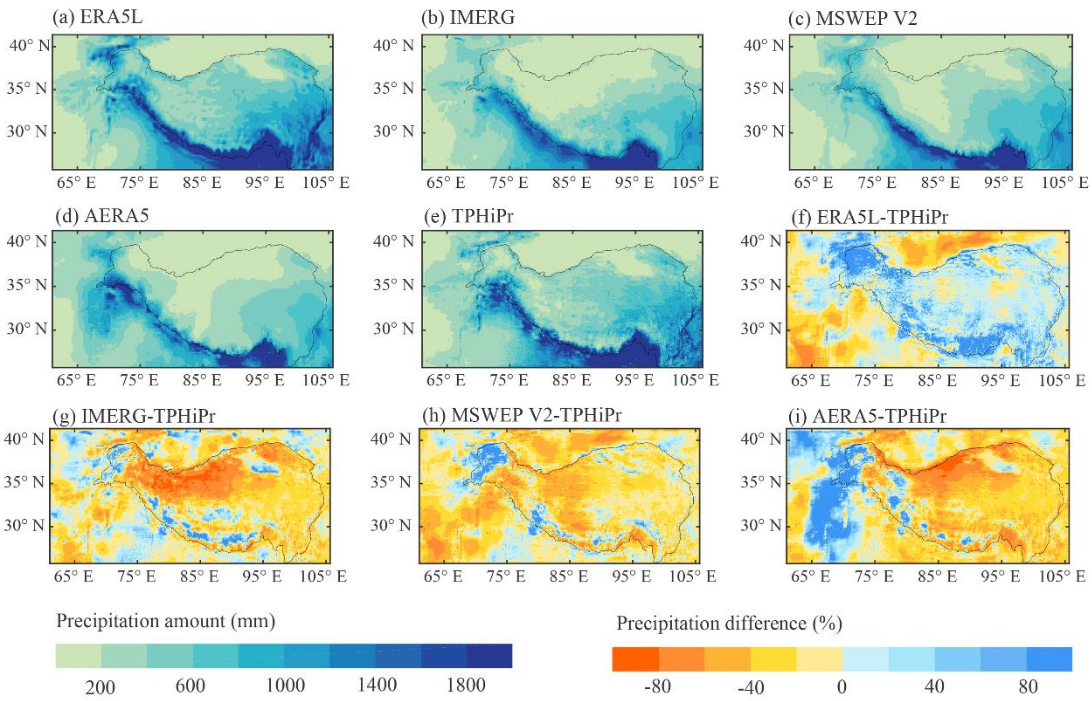

Figure R12: Spatial patterns of (a-e) the average annual precipitation during 2008-2015 from the five datasets and (f-i) the relative differences between TPHiPr and the other four datasets. The differences are calculated by subtracting TPHiPr from and the other four datasets then dividing by TPHiPr.

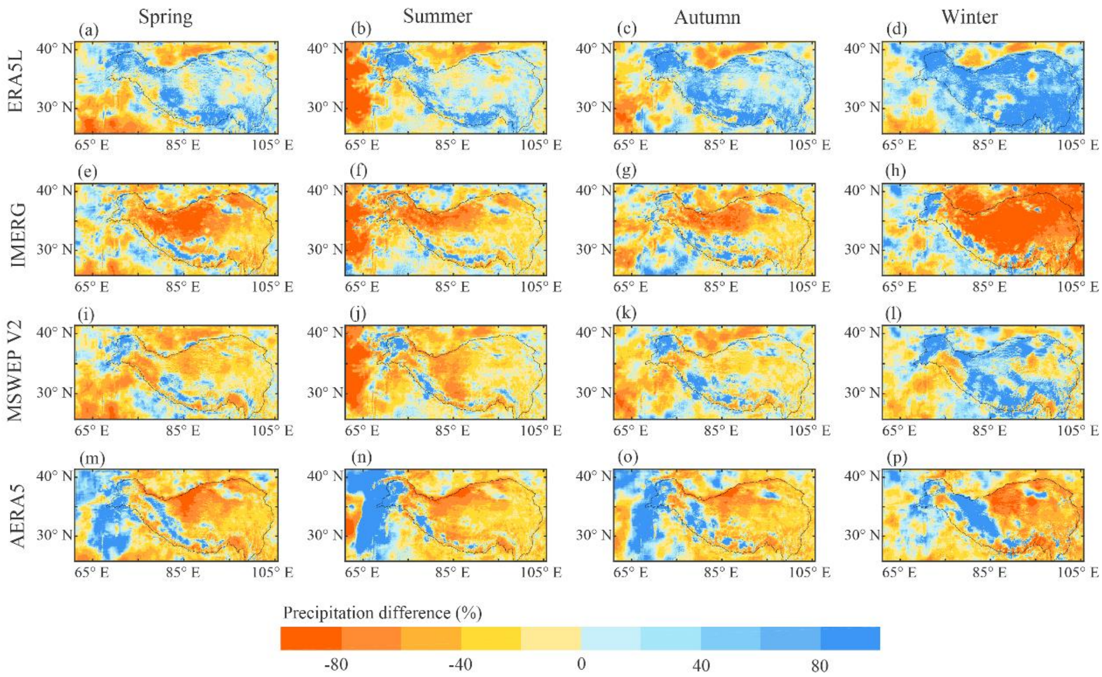

Figure R13: Spatial patterns of the relative differences in average seasonal precipitation between TPHiPr and the other four datasets. The differences are calculated by subtracting TPHiPr from the other four datasets and then dividing by TPHiPr.