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 will revise the manuscript accordingly, mainly including the following aspects:

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

(2) We will further clarify the underlying logic for using such an algorithm and provide more details about the methods.

(3) We will include some regional precipitation datasets in our evaluation results and extend the comparison with other datasets.

(4) We will edit the language of the manuscript carefully.

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 attached in the supplement); second, as shown in Figure R1 and reported in the reviewer's suggested article, the ERA5_land and ERA5 only have slight differences in the Third Pole (TP) region. However, our previous work showed that the ERA5_CNN is more skillful in representing the spatial variability of precipitation and has smaller wet biases than the ERA5 in the TP (Figure R2 attached in the supplement; Jiang et al., 2021). In the revised manuscript, we will further clarify the advantages of using the ERA5_CNN as a background dataset.
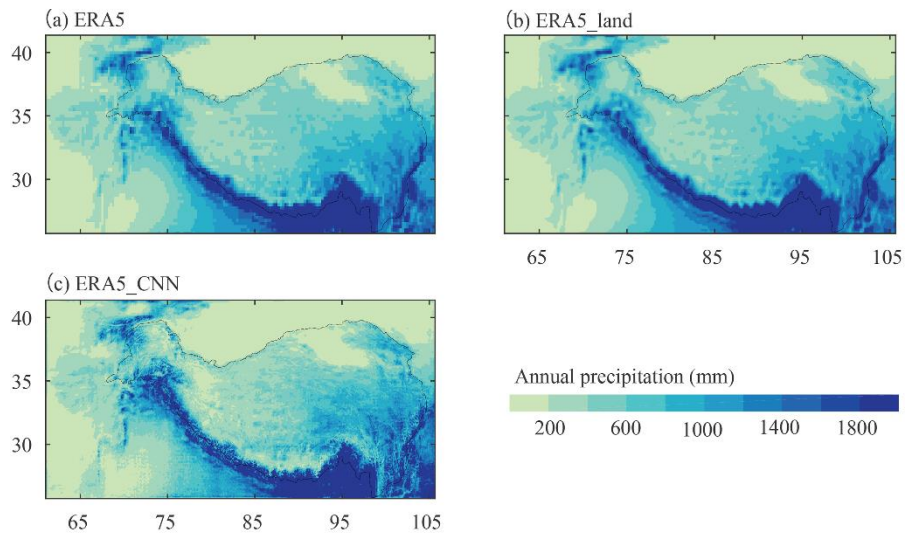
Figure R1 Spatial patterns of annual precipitation from (a) ERA5, (b) ERA5_land and (c) ERA5_CNN. The precipitation is averaged over 2008-2020.
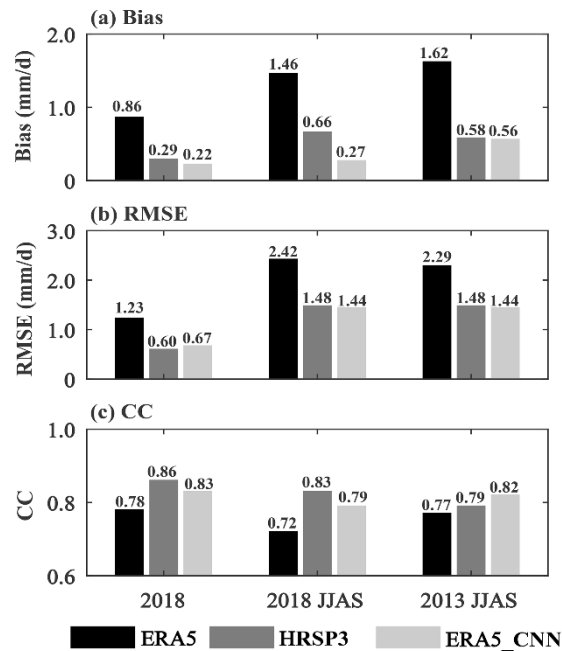
Figure R2. Spatial error metrics for ERA5, high-resolution WRF simulation (HRSP3) and ERA5_CNN with respect to three different periods, namely, whole 2018, June to September of 2018, and June to September of 2013. The error metrics were calculated based on the time-averaged precipitation of 96 CMA stations (Jiang et al., 2021).

Jiang, Y., Yang, K., Shao, C., Zhou, X., Zhao, L., Chen, Y., 2021. 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

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 will conduct a more comprehensive literature review about the merging algorithms of precipitation in 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 will refine it in the revised manuscript.

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 Trees and aggregates estimates from multiple Decision Tree, therefore, it is less likely to suffer from overfitting and has good generalization capability. 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 R3 (attached in the supplement) shows that the RF generally performs better than the other three methods.

In the revised manuscript, we will further clarify the underlying logic of the merging algorithm and introduce more about the RF and Kriging.
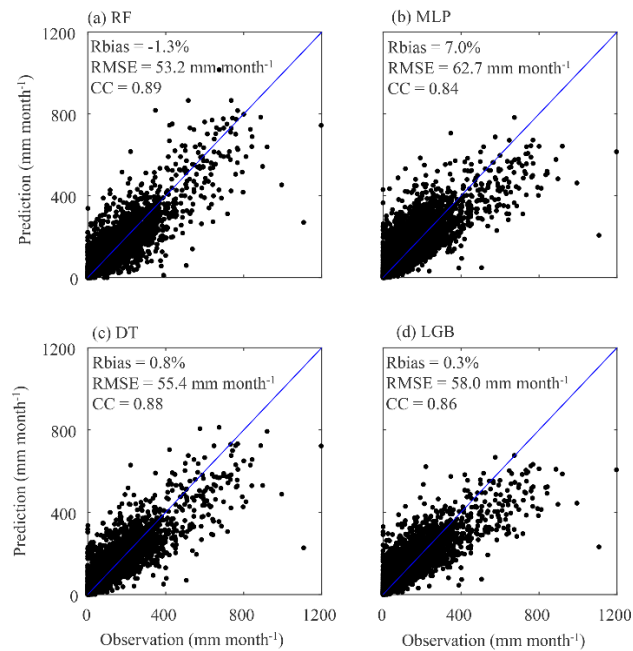
Figure R3 Comparison between the monthly precipitation in 2018 estimated by four machine learning models and the observed monthly precipitation. RF: Random Forest; MLP: Multi-layer Perceptron; DT: Decision Trees; LGB: LightGBM.

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 a higher spatial resolution than ERA5, we will add the evaluation results of ERA5_land in section 4.2 in the revised manuscript.

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 will compare the performance of some regional precipitation datasets (such as AERA5-Asia) with our produced product in the revised manuscript.

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 will conduct an evaluation of the AERA5-Asia and compare its performance with our produced product.

8. The language still need to be greatly improved.

Response: We will carefully edit 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 will make these details clearer in the abstract in 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. We will give more details about the ERA5_CNN in 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. 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. (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 will further clarify the objective of this work in 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 103 in the manuscript. We have aggregated the sub-daily records to daily precipitation. We will carefully check these points and make them clearer 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 will add the evaluation results of ERA5_land in section 4.2.

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 will remove this figure.