https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Open Access Earth System Science Data Discussions

TPHiPr: A long-term high-accuracy precipitation dataset for the Third Pole region based on high-resolution atmospheric modeling and dense observations

Yaozhi Jiang[1], Kun Yang[1,2*], Youcun Qi[3], Xu Zhou[2], Jie He[2], Hui Lu[1], Xin Li[2], Yingying Chen[2], Xiaodong Li[4], Bingrong Zhou[5], Ali Mamtimin[6], Changkun Shao[1], Xiaogang Ma[1], Jiaxin Tian[1], Jianhong Zhou[1]

[1] Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies, Tsinghua University, Beijing, China.

[2] National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China.

[3] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[4] State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China

[5] Qinghai Institute of Meteorology Science, Xining, China

[6] Institute of Desert Meteorology/Taklimakan Desert Meteorology Field Experiment Station, China Meteorological Administration, Urumq, China

Correspondence to: Kun Yang (yangk@tsinghua.edu.cn)

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Abstract: Reliable precipitation data are highly necessary for geoscience research in the Third Pole (TP) region but still lacking, due to the complex terrain and high spatial variability of precipitation here. Accordingly, this study produces a long-term (1979-2020) high-resolution (1/30°) precipitation dataset (TPHiPr) for the TP by merging the atmospheric simulation-based ERA5_CNN with gauge observations from more than 9000 rain gauges, using the Climatology Aided Interpolation and Random Forest methods. Validation shows that the TPHiPr is generally unbiased and has a root mean square error of 4.5 mm day^{-1}, a correlation of 0.84 and a critical success index of 0.67 with respect to all independent rain gauges in the TP, demonstrating that this dataset is remarkably better than the widely-used global/quasi-global datasets, including the fifth-generation atmospheric reanalysis of the European Centre for Medium-Range Weather Forecasts (ERA5), the final run version 6 of the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) and the Multi-Source Weighted-Ensemble Precipitation version 2 (MSWEP V2). Moreover, the TPHiPr can better detect precipitation extremes compared with the three widely-used datasets. Overall, this study provides a new precipitation dataset with high accuracy for the TP, which may have broad applications in meteorological, hydrological and ecological studies. The produced dataset can be accessed via https://doi.org/10.11888/Atmos.tpdc.272763 (Yang and Jiang, 2022).

Keywords: Third Pole region, Precipitation, High-density rain gauges, Atmospheric simulation, Merging

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

1. Introduction

The Third Pole (TP) region is one of the most complex-terrain regions with high elevations and heterogeneous land surfaces, and strong water and energy exchanges between land surface and atmosphere exists in this region (Chen et al., 2021). Moreover, it is the source of many large Asian rivers, providing abundant water resources and hydropower within and beyond this region (Yao et al., 2022). Meanwhile, the TP suffers from frequent natural hazards (e.g. flash floods, debris flows, landslides), especially in the periphery of the TP (Cui and Jia, 2015). Reliable gridded precipitation data is essential for understanding of hydrological processes, planning of water resources and prevention of natural hazards in the TP (Gao et al., 2021; Wang et al., 2018).

At present, quasi-global and regional precipitation datasets, including gauge-based products, satellite-based products and reanalysis products, have played an important role over the TP. These datasets include the Asian Precipitation-Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE; Yatagai et al., 2012), the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG; Huffman et al., 2019), the TRMM Multisatellite Precipitation Analysis (TMPA; Huffman et al., 2007), the China Meteorological Forcing Dataset (CMFD; He et al., 2020), the fifth generation ECMWF atmospheric reanalysis (ERA5; Hersbach et al., 2020), the High Asia Refined analysis (HAR; Maussion et al., 2014) and its version 2 (HAR V2; X. Wang et al., 2020), et al. Among these products, gauge-based products may have large errors in the TP, since they are mostly interpolated based on sparse gauge observations. Satellite or satellite-gauge combined products are most widely used in the TP. However, they are proved to misrepresent solid precipitation and orographic precipitation, and show large uncertainties in winter and in the western and southeastern TP (Gao et al., 2020; Lu and Yong, 2018; Xu et al., 2017). Atmospheric simulation with fine spatial resolution can give reasonable atmospheric water transport and precipitation spatial variability in complex terrain (Curio et al., 2015; Maussion et al., 2014; Norris et al., 2017; Ouyang et al., 2021; Sugimoto et al., 2021; Wang et al., 2020b; Zhou et al., 2021), moreover, it is skillful in estimating solid precipitation (Lundquist et al., 2019; Maussion et al., 2014). However, currently atmospheric simulation-based datasets consistently overestimate precipitation amount in the TP (Gao et al., 2015; Wang et al., 2020b; Zhou et al., 2021). As a result, substantial differences exist among these datasets in the TP in terms of both amount and spatial

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

variability of precipitation (D. Li et al., 2020; Lu and Yong, 2018; Tan et al., 2020; Wang and Zeng, 2012; You et al., 2012). In addition, these datasets typically have a horizontal resolution coarser than 10 km, which is insufficient to represent the fine-scale precipitation variability and cannot be applied locally.

Errors in precipitation products hinder the correct understanding of water cycle processes in the TP. For example, Immerzeel et al. (2015) found that the simulated runoff in the upper Indus using APHRODITE is much smaller than the observations and further confirmed that APHRODITE severely underestimates precipitation amount in this region. Savéan et al. (2015) pointed out that precipitation from rain gauges with poor spatial representativeness leads to irrational runoff component simulations in the central Himalaya. Jiang et al. (2022) demonstrated that currently widely-used satellite-based precipitation products cannot close the basin-scale water budget in the eastern edge of the TP. Some other studies also demonstrated the high uncertainties in current precipitation products for simulations of snow cover (Gao et al., 2020), soil moisture (Yang et al., 2020) and river discharge (Alazzy et al., 2017).

Merging multiple precipitation products is widely conducted to mitigate precipitation uncertainties (Hong et al., 2021; Ma et al., 2022; Shen et al., 2014). Ma et al. (2018) used a dynamic Bayesian model to merge multiple satellite precipitation products in the TP and showed that the merged precipitation has higher accuracy than the raw satellite data; Li et al. (2021) produced a high-accuracy precipitation dataset for the southern TP by merging three satellite-based precipitation datasets with high-density rain gauge data. Wang et al. (2020a) developed a long-term precipitation dataset for the Yarlung Tsangpo River basin by merging data from satellite, reanalysis and rain gauges. Although encouraging progresses have been made, there are still some limitations. First, these works either corrected gridded precipitation with data from sparse rain gauge networks or were conducted in sub-regions of the TP. Second, most works have merged satellite products with rain gauge data, while both the two sources of precipitation perform poorly in reflecting heterogeneous precipitation in the complex-terrain TP. Therefore, substantial improvements are still needed for producing high-accuracy precipitation data in the TP.

Therefore, the main goal of this study is to produce a long-term high-resolution precipitation dataset with high accuracy for the TP, by merging unprecedented high-density rain gauge data with high-resolution atmospheric simulation-based precipitation. The atmospheric simulation-based precipitation is selected as the background field, mainly due to its advanced skill in giving the spatial variability of precipitation

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

in complex terrain and estimating solid precipitation, which is especially important in high mountains and the western TP.

2. Data

2.1 Rain gauge data

Rain gauge data used in this study are obtained from several sources, including the China Meteorological Administration (CMA), the Ministry of Water Resources of China (MWR), the Department of Hydrology and Meteorology of Nepal (DHM), the Global Historical Climatology Network (GHCN; Menne et al., 2012), and some other field observation networks (Chen et al., 2014, 2015; Luo, 2018; Wei and Wang, 2019; Wang, 2021; Yang, 2018; Yang et al., 2017; Zhang, 2018; Zhao, 2018; Zhao et al., 2017). These networks provide either daily or sub-daily records. In addition, our group has set up more than 80 rain gauges over the TP since 2017, deployed in the Yadong Valley, the south slope of Gangdise Range, the eastern edge of the TP, the surroundings of the Namco Lake and the Inner TP. Observations from this network are also used in this study.

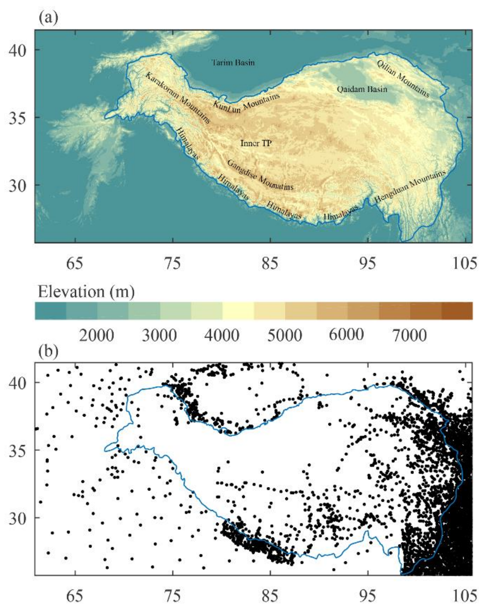

Figure 1: (a) Topography of the Third Pole region. (b) Spatial distribution of rain gauges used in this study. The blue line denotes the 2500 m contour of elevation, which is obtained from Zhang (2019).

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

A series of quality control procedures are applied to the rain gauge data following the method of Hamada et al. (2011), including outlier check, repetition check, and spatial consistency check. Detailed judgment criteria for each check can refer to Hamada et al. (2011). In addition, for each rain gauge, data records for a certain year less than 60 days are removed since they are likely to suffer from a technical broken. After the quality control, data from 9798 rain gauges are eventually selected for precipitation merging and these data have temporal coverages ranging from a few months to more than 40 years. Figure 1 shows the spatial distribution of these rain gauges.

Rain gauge observations usually suffer from measurement errors, including wind-induced undercatch, wet loss and evaporation loss. This especially happens in the TP where the wind is strong and solid precipitation accounts for a large proportion of the total precipitation. Therefore, the measurement errors are corrected in this study. For gauges where observed wind speed and air temperature are provided, the empirical relationships provided by Ye et al. (2007) and Ma et al. (2015) are used to correct the measurements. For gauges without wind speed and air temperature observations, the Random Forest (RF) model is used to correct precipitation. This is achieved with the following steps: first, the RF model is trained at above-corrected gauges, using wind speed and air temperature from ERA5 and original observed daily precipitation as model input and the corrected precipitation as the target; then, the trained model is applied to gauges without wind speed and air temperature observations to estimate corrected precipitation, using wind speed and air temperature from ERA5.

2.2 Gridded precipitation dataset

The background precipitation dataset used in this study is called ERA5_CNN, which was produced by the downscaling method presented in our previous work (Jiang et al., 2021). This dataset is an atmospheric simulation-based dataset, derived from combing a short-term high-resolution WRF simulation (Zhou et al., 2021) with ERA5 reanalysis. More specifically, a two-year high-resolution WRF simulation is firstly obtained and used for training a convolutional neural network (CNN)-based downscaling model. Then, the trained model is used to downscale the long-term ERA5 precipitation to generate the ERA5_CNN. The ERA5_CNN has a high horizontal resolution of 1/30° and daily temporal resolution, covering the period from 1979 to 2020. Our previous evaluations showed that the ERA5_CNN can give fine-scale spatial variability of precipitation in the complex-terrain TP and has high

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

spatial correlations with rain gauge data. However, the ERA5_CNN generally overestimates precipitation in the TP, which is inherited from atmospheric simulation (Jiang et al., 2021). Therefore, its accuracy needs to be further improved by merging it with high-density gauge observations.

For comparison, three typically widely-used precipitation datasets, including ERA5 reanalysis, satellite-based IMERG and the Multi-Source Weighted-Ensemble Precipitation version 2 (MSWEP V2; Beck et al., 2019), are also utilized in this study. The ERA5 is the latest generation reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides 0.25° precipitation data at 1-hour intervals. IMERG is a satellite precipitation dataset retrieving from the combination of both microwave and infrared observations and is currently the most widely-used in the world, with a horizontal resolution of 0.1° and the highest temporal resolution of half an hour. The IMERG Final Run V6 (hereafter IMERG), which has been corrected with monthly rain gauge data, is used in this study. The MSWEP V2 with a horizontal resolution of 0.1° is a merged dataset that has combined multiple satellite, gauge, and reanalysis precipitation datasets. Moreover, it is corrected with observed discharge from many catchments worldwide.

3. Methods

3.1 Merging algorithm

3.1.1 General flowchart

This study merges the ERA5_CNN precipitation with high-density rain gauge data based on the idea of the Climatology Aided Interpolation (CAI; Willmott and Robeson, 1995), in which the anomalies/ratios of meteorological variables are interpolated and then added/multiplied to the climatology, instead of directly interpolating the meteorological variables. The CAI method has been widely applied for gridding precipitation and shown good performance (Contractor et al., 2020; Schamm et al., 2014; Xie et al., 2007). Figure 2 shows the flowchart for merging ERA5_CNN and rain gauge data. Three main parts are involved in the merging procedure, including the construction of monthly precipitation climatology, monthly precipitation and daily precipitation. Details are listed below.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

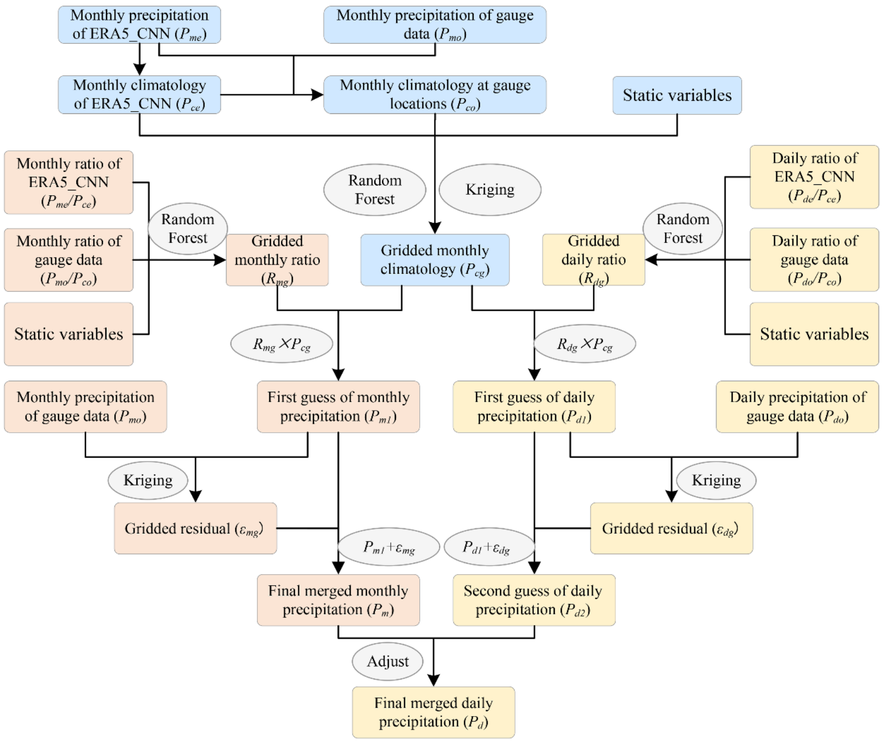

Figure 2: General flowchart of the merging algorithm. The static variables include the elevation, the standard deviation of elevation and the identifier of the clusters with different precipitation characteristics. The subscript '*o*' represents 'observation', '*e*' represents 'ERA5_CNN', '*g*' represents 'gridded', '*c*' represents 'climatology', '*m*' represents 'monthly' and '*d*' represents 'daily'.

(1) Construction of monthly precipitation climatology.

Since the length of the data records varies from gauge to gauge, it is undesirable to obtain monthly climatology fields via directly interpolating the observed multi-year average monthly precipitation. Therefore, we first construct monthly precipitation climatology at gauge locations based on the monthly precipitation climatology of ERA5_CNN, using the following formula:

$$P_{co} = P_{ce} \times \frac{P_{co1}}{P_{ce1}}, \tag{1}$$

where P_{co} is the constructed monthly precipitation climatology at gauge locations, P_{ce} is the monthly precipitation climatology of ERA5_CNN averaged over 1979-2020, P_{co1} is the monthly precipitation of

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

rain gauge averaged over the observing period, which varies from gauge to gauge, and P_{ce1} is the monthly precipitation of ERA5_CNN averaged over the same observing period at the collocated grids.

The precipitation climatology fields for the 12 months are then constructed by interpolating the monthly climatology at gauge locations using a Random Forest (RF; Breiman, 2001) and Kriging-based method, in which the climatology of ERA5_CNN is taken as an auxiliary and will be introduced in section 3.1.2.

(2) Construction of gridded monthly precipitation

In this study, the ratios of monthly precipitation to its climatology are adopted for constructing monthly precipitation fields. There are four steps for constructing monthly precipitation fields. First, the ratios of observed monthly precipitation (P_{mo}) to the precipitation climatology (P_{co}) are calculated at gauge locations; second, the ratios (P_{mo}/P_{co}) are gridded using the RF method; third, the gridded ratios (R_{mg}) are multiplied by the gridded monthly precipitation climatology (P_{cg}) obtained in step (1) to construct the first guess of gridded monthly precipitation fields (P_{m1}); finally, the residuals of the first guess against gauge observations are gridded using the Kriging method and added to the first guess to construct the final monthly precipitation fields (P_m).

(3) Construction of gridded daily precipitation

The procedures for constructing daily precipitation fields are similar to monthly precipitation, with only two differences. First, the ratios are daily precipitation to monthly climatology (P_{do}/P_{co} and P_{de}/P_{ce}) in this part. Second, the daily precipitation fields after residual correction (P_{d2}) are further adjusted to ensure that the sum of the daily precipitation amount in a month is equal to the corresponding monthly precipitation amount obtained in step (2), given that monthly precipitation fields are more reliable due to their less spatial variability than daily fields (He et al., 2020).

In the above procedures, gridding multiple variables, including the monthly climatology, the ratios of monthly/daily precipitation to monthly climatology and the monthly/daily residuals, is achieved based on the RF and Ordinary Kriging, which will be introduced in section 3.1.2.

3.1.2 Gridding method

Gridding monthly precipitation climatology and precipitation ratio is the key for merging ERA5_CNN and rain gauge data. The main gridding method used in this study is the RF model, which is an ensemble

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

machine learning model based on the decision tree algorithm and can learn the complex non-linear relationships between multiple covariates and the target variable. The RF is easy to implement and has robust prediction accuracy, thus making it a widely-used method for the correction and downscaling of meteorological variables (Baez-Villanueva et al., 2020; He et al., 2016; Sekulić et al., 2021; Zhang et al., 2021). The general formulation for griding precipitation at multiple timescales with the RF can be expressed as follow:

$$\begin{cases} P_{cg} = f_1\left(x_{c,1}, x_{c,2}, \dots, x_{c,n}\right) + \varepsilon_1, & for\ monthly\ precipitation\ climatology \\ P = P_{cg} \times f_2(x_1, x_2, \dots, x_n) + \varepsilon_2, & for\ monthly\ and\ daily\ precipitation \end{cases}, \quad (2)$$

where P_{cg} is the monthly precipitation climatology, P is the monthly or daily precipitation, $f_1(\bullet)$ and $f_2(\bullet)$ are the non-linear regressive relationship built with the RF model, $x_{c,i}$ and x_i are the covariates used to predict the precipitation climatology or the ratio of monthly/daily precipitation to the climatology, and ε_1 and ε_2 are the residuals of the estimated precipitation.

Multiple covariates are used to build the RF model. For gridding monthly precipitation climatology, the target for training the RF model is the monthly precipitation climatology at the gauge locations, and the inputs are monthly precipitation climatology from ERA5_CNN at nine grids around the target location, longitude, latitude, elevation and standard deviation of elevation around the target location. In addition, the study area is divided into 25 clusters according to the monthly variation of precipitation and the identifier for the cluster is also input into the RF model. For griding the ratio of monthly/daily precipitation to monthly climatology, the training target is the observed ratio of monthly/daily precipitation to monthly climatology, and the inputs are the same as those for griding precipitation climatology except that the ratios of monthly/daily precipitation to monthly climatology are input to the model rather than monthly climatology. Model training performs for each month, i.e. samples from all gauges and all years in a month are gathered together and used for model training.

As shown in Eq. (2), there are residuals (ε_1 and ε_2) between the precipitation estimates from the RF model and the gauge observations. Therefore, we first calculate the differences between the gauge observations and the precipitation estimates from RF at each gauge. Then, the Ordinary Kriging is used to interpolate the differences. Finally, the difference field is added to the precipitation estimates from RF to obtain the final estimates of precipitation.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Open Access Earth System Science Data Discussions

3.2 Evaluation metrics

Several metrics are used for validating the merged precipitation, including relative bias (Rbias), root mean square error (RMSE), correlation coefficient (CC), probability of detection (POD), false alarm ratio (FAR) and critical success index (CSI). The formulas and perfect values for these metrics are listed in Table 1.

Table 1 The error metrics used in this study

Metrics	Formula	Perfect value
Relative bias	$Rbias = \frac{\sum_{i=1}^{n}(M_i - O_i)}{\sum_{i=1}^{n} O_i}$	0
Root mean square error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2}$	0
Correlation coefficient	$CC = \frac{\sum_{i=1}^{n}(M_i - \overline{M})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{n}(M_i - \overline{M})^2}\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}}$	1
Probability of detection	$POD = \frac{H}{H + MM}$	1
False alarm ratio	$FAR = \frac{F}{H + F}$	0
Critical success index	$CSI = \frac{1}{POD^{-1} + (1 - FAR)^{-1} - 1}$	1

where n is the number of days, M_i and O_i are the merged and observed precipitation at a specific day, respectively, $\overline{M}$ and $\overline{O}$ are the mean values of merged and observed precipitation, respectively. H is the days when both merged data and observation have precipitation. MM is the days when only observation has detected precipitation. F is the days when only merged data has detected precipitation. For calculating POD, FAR and CSI, a threshold of 0.1mm day^{-1} is adopted for distinguishing precipitation and non-precipitation day.

4. Results

4.1 Validation of the merging algorithm

4.1.1 Merging effect on precipitation amount and spatial pattern

The spatial patterns of average annual precipitation from ERA5_CNN and the merged data (TPHiPr) during 1979-2020 are shown in Fig. 3a and b. It can be found that ERA5_CNN and TPHiPr have similar

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

spatial patterns of precipitation in the TP. Both have large precipitation amounts in the southeast of the TP and along the Himalayas, while having small precipitation amounts in the Qaidam Basin, the Tarim Basin and the Inner TP. The similar spatial patterns of ERA5_CNN and TPHiPr demonstrate that the merging algorithm generally retains the spatial characteristics of precipitation from ERA5_CNN.

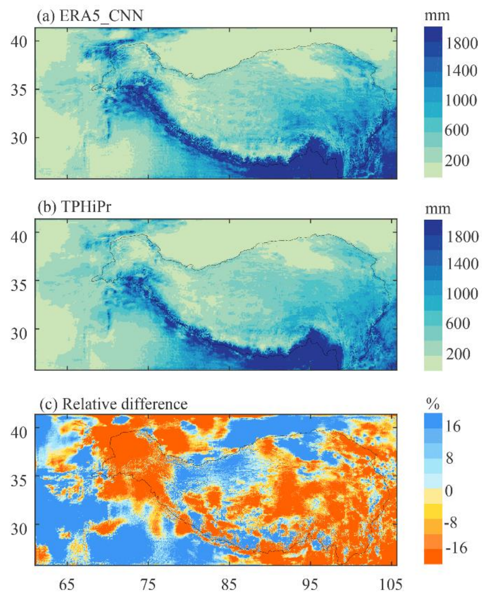

Figure 3: Spatial patterns of the annual average precipitation from (a) ERA5_CNN and (b) the merged data (TPHiPr), as well as (c) the relative difference between them. The precipitation is averaged over the period from 1979 to 2020. The relative difference is calculated by subtracting ERA5_CNN from TPHiPr, and then dividing by ERA5_CNN.

The relative difference between ERA5_CNN and TPHiPr is also calculated and shown in Fig. 3c. Generally, by merged with rain gauge data, the precipitation amount is reduced in the TP. The precipitation amount averaged over the study area decreases from 696.4 mm year^{-1} of ERA5_CNN to 600.9 mm year^{-1} of TPHiPr. This corresponds to previous works that have demonstrated the overestimation in the atmospheric simulation-based precipitation datasets (Gao et al., 2015; Jiang et al., 2021; Wang et al., 2020b; Zhou et al., 2021). Spatially, the precipitation decrease is evident (up to 20%) in the central and eastern TP, the western Himalayas, the Karakoram and the Tarim Basin, while precipitation amount increases in the Qaidam Basin and its north, the southwest of the TP and the eastern

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Kunlun.

4.1.2 Validation with independent gauge data

In this study, about 10% of the total rain gauges are randomly excluded for independent validation of TPHiPr, and several metrics against rain gauge data are calculated for ERA5_CNN and TPHiPr at these rain gauges based on daily precipitation.

Figure 4 compares the boxplot of these metrics for ERA5_CNN and TPHiPr. TPHiPr has remarkably better performance than the ERA5_CNN. In terms of the Rbias, ERA5_CNN generally overestimates precipitation in the TP, with the median Rbias value for all these rain gauges of 16.6%. In comparison, the overestimation is largely reduced in TPHiPr, which has a median value of 0.5%. Also, TPHiPr shows smaller RMSE values (with a median value of 4.5 mm day^{-1}) than the ERA5_CNN (with a median value of 8.6 mm day^{-1}). Regarding CC, ERA5_CNN has values between 0.40 and 0.60 at most rain gauges (the median value is 0.53), while they are generally larger than 0.70 for TPHiPr with a median value of 0.84, indicating that precipitation from the TPHiPr has highly consistent temporal variations with rain gauge data. In addition, it can be seen that the Rbias (Fig. 4a) and RMSE (Fig. 4b) for TPHiPr are less divergent than those for ERA5_CNN, implying that TPHiPr has more spatially homogeneous accuracy than ERA5_CNN.

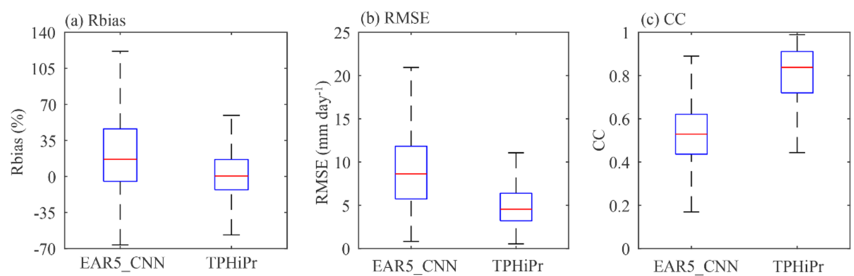

Figure 4: Comparison of error metrics for ERA5_CNN and TPHiPr at 966 independent rain gauges. The box represents the distribution of the metrics for all the independent rain gauges in the TP.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

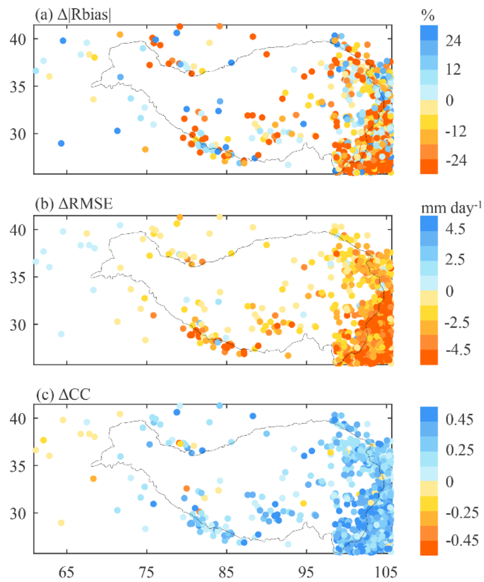

Figure 5: Spatial distribution of error metrics differences between ERA5_CNN and TPHiPr. The differences are calculated by subtracting the metrics of ERA5_CNN from those of TPHiPr.

Figure 5 shows the differences in the three metrics between ERA5_CNN and the TPHiPr at each rain gauge. After the merging, the rain gauges with better Rbias, RMSE and CC account for 68%, 97% and 96% of the total validation rain gauges, respectively. More than 50% of the rain gauges have RMSE reductions larger than 3.0 mm day^{-1} and about 67% of the rain gauges have CC improved by more than 0.2. Moreover, obvious improvements can be found at many east rain gauges. In the western region, improvements can also be found at many rain gauges in the high elevations, while the metrics change little at some rain gauges outside the 2500 m contour.

In summary, by merged with rain gauge data, the accuracy of ERA5_CNN is well improved in the TP, especially in regions where high-density rain gauges are located.

4.2 Comparison with other datasets

We also compare the merged precipitation data with other widely-used precipitation products. The comparison focuses mainly on three aspects: the amount and spatial patterns of precipitation, the error metrics against rain gauge data and the ability to reproduce precipitation extremes.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

4.2.1 Precipitation amount and spatial patterns

Figure 6 compares the average annual precipitation amount from multiple datasets in the Third Pole region (above 2500 m contour) for 2008-2020. Among the four datasets, ERA5 has the largest precipitation amount of 810.8 mm year^{-1}, followed by TPHiPr (640.1 mm year^{-1}) and MSWEP V2 (501.5 mm year^{-1}), and IMERG has the smallest precipitation amount of 424.7 mm year^{-1}.

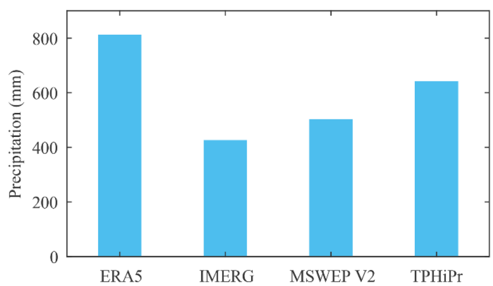

Figure 6: Average annual precipitation of the four datasets for the TP (above 2500 m contour) during 2008-2017.

Figure 7 shows spatial patterns of the average annual and seasonal precipitation during 2008-2020 from the four precipitation datasets. Generally, the average annual precipitation (Fig. 7a-7d) from all the four datasets decreases from the southeast to the northwest because the monsoon has brought abundant water vapor to the southeastern region of the study area while its impact is reduced in the northwest. In addition, high mountains along the Himalayas block the northward moisture and result in large precipitation amounts in this region, which is revealed by all these datasets. As shown in Fig. 7, precipitation from IMERG and MSWEP V2 varies more smoothly in space than that from TPHiPr and ERA5. Moreover, compared with ERA5, TPHiPr presents more details related to local topography. For example, the dry belt in the northern slope of the central Himalayas (around 90°E, 29°N), which was proved in the results of Wang et al. (2019), is more evident in TPHiPr than in ERA5. Besides, TPHiPr shows greater spatial variability of precipitation than ERA5 in the Hengduan Mountains where the topography is much complex with many large mountain ranges and valleys.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

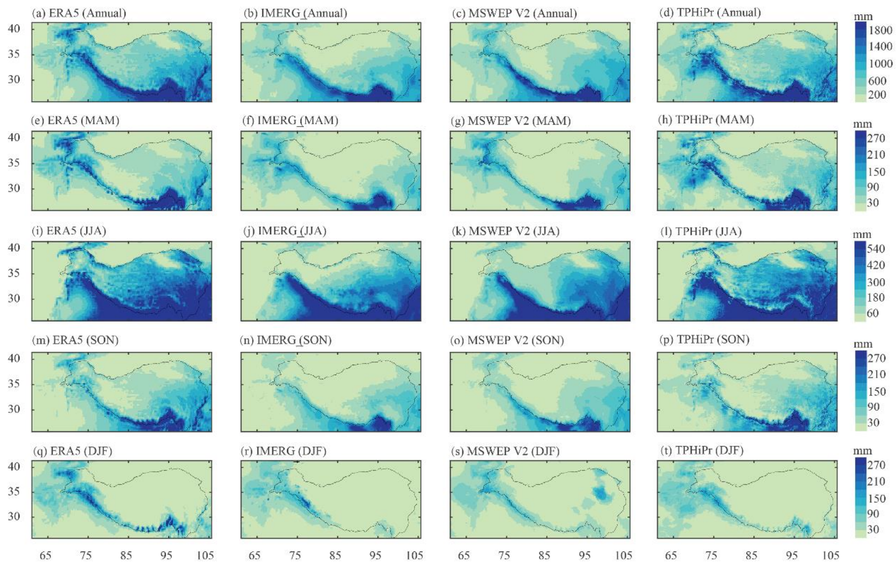

Figure 7: Spatial patterns of average (a-d) annual and (e-t) seasonal precipitation from ERA5 (first column), IMERG (second column), MSWEP V2 (third column) and TPHiPr (fourth column). The precipitation is averaged over the period from 2008 to 2020. MAM: March to May; JJA: June to August; SON: September to November; DJF: December to February.

With respect to the seasonal variations of precipitation, affected by the monsoon climate, most parts of the TP have large precipitation in summer but small precipitation in winter. In the westerly-dominant western TP, the precipitation is large in spring and winter but small in summer. All these datasets can generally capture the seasonal cycles of precipitation in the TP. In summer (Fig. 7i-l), the differences between these datasets mainly occur in the Inner TP, where TPHiPr and ERA5 show larger precipitation than the IMERG and MSWEP V2. In spring (Fig. 7e-h) and winter (Fig. 7q-t), apparent differences between these datasets are shown in the Karakoram and the western Himalayas. TPHiPr and ERA5 yield large precipitation amounts in these regions, while the precipitation amount from IMERG and MSWEP V2 is relatively small. This is likely because solid precipitation accounts for a large part of the total precipitation in these regions and the model-based ERA5 and TPHiPr are more skillful in estimating solid precipitation than the IMERG and MSWEP V2, which has also been pointed out in the work of D. Li et al. (2020).

4.2.2 Comparison of error metrics

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

The performance of the three widely-used global/quasi-global datasets is evaluated using the rain gauge data used for independent validation in section 4.1.2 and compared with that of TPHiPr in this study. Note that the evaluation in this section span a shorter period from 2008 to 2020 considering the availability of the IMERG data.

Figure 8 compares the boxplots of the Rbias, RMSE and CC of the four datasets. In terms of the Rbias (Fig. 8a and the first column in Fig. 9), ERA5 overestimates precipitation at most rain gauges in the TP with a median value of 16.9%. The other three datasets generally have small relative biases and the median values for IMERG, MSWEP V2 and TPHiPr are -0.7%, -0.4% and -0.2%, respectively. For RMSE (Fig. 8b and the second column in Fig. 9), the three global/quasi-global datasets have similar RMSE in the TP, with the median value of 7.4 mm day^{-1} for ERA5, 7.1 mm day^{-1} for IMERG and 6.4 mm day^{-1} for MSWEP V2, while the RMSE for TPHiPr has a median value of 4.5 mm day^{-1}, which is remarkably smaller than those of the other three datasets. Particularly, the correlations between the precipitation from TPHiPr and rain gauge data are remarkably larger than those of the other three datasets (Fig. 8c and the third column in Fig. 9). The values of CC for ERA5 are between 0.30 and 0.60 at most gauges, with a median value of 0.55. The IMERG and MSWEP V2 have higher correlations with rain gauge data and both of them have a median value of 0.64. By contrast, PHiPr has a CC value larger than 0.70 at about 80% of the total rain gauges, resulting in a median value for all gauges of 0.84.

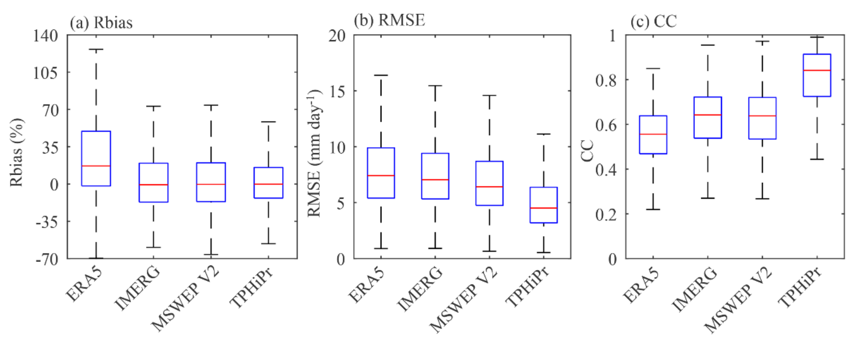

Figure 8: Comparison of (a) Rbias, (b) RMSE and (c) CC for ERA5, IMERG, MSWEP V2 and TPHiPr. The box represents the distribution of the metrics for all the independent rain gauges in the TP.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Open Access Earth System Science Data Discussions

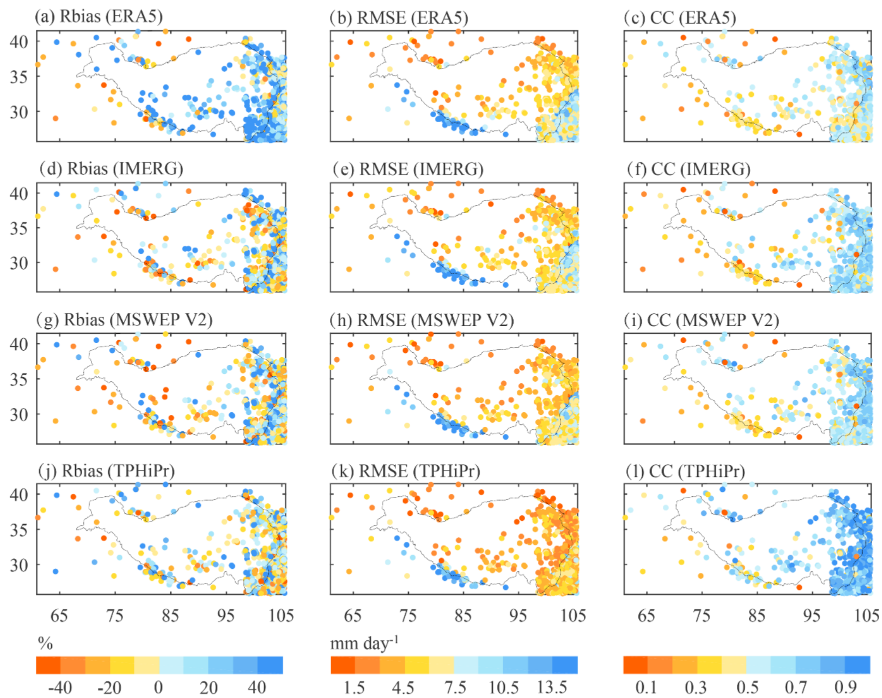

Figure 9: Spatial distribution of Rbias (first column), RMSE (second column) and CC (third column) for (a-c) ERA5, (d-f) IMERG, (g-i) MSWEP V2 and (j-l) TPHiPr. The metrics are calculated at daily scale.

This study also calculates the POD, FAR and CSI for the four datasets to compare their performance in detecting precipitation occurrence. In this section, a threshold of 0.1 mm day^{-1} is used to distinguish rain and no-rain days. Figure 10 compares the boxplots of these metrics for ERA5, IMERG, MSWEP V2 and TPHiPr, and the spatial distributions for these metrics are shown in Fig. 11. Among the four datasets, the ERA5 and MSWEP V2 have high values of POD (both have a median value of 0.97). However, it can be seen from Fig. 10b and Fig. 11 that they also have large FAR values. This is mainly because both ERA5 and MSWEP V2 have data sources from atmospheric reanalysis, which tends to overestimate precipitation frequency in the TP (Hu and Yuan, 2021). In contrast, IMERG, mainly based on satellite estimates, has lower values of POD and FAR. With respect to TPHiPr, Fig. 10 shows that it has relatively high POD values (the median value is 0.93) and the lowest FAR (the median value is 0.29). As a result, TPHiPr gains the highest CSI values among the four datasets, with a median value of 0.67, while all the other datasets have a median CSI value of about 0.55.

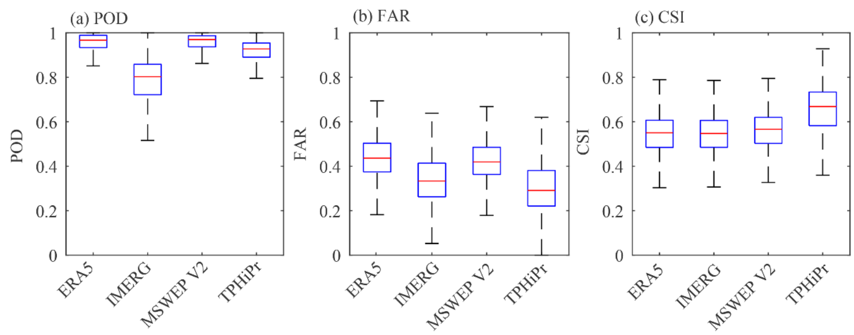

Figure 10: Similar to Fig. 8 but for (a) POD, (b) FAR and (c) CSI. These metrics are calculated using a threshold of 0.1 mm day^{-1}.

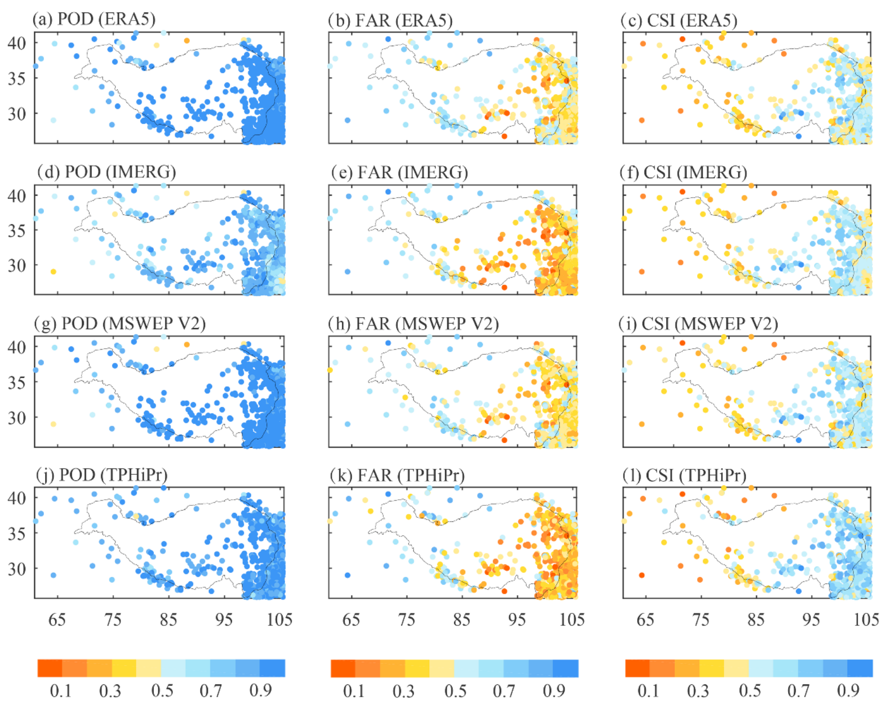

Figure 11: Similar to Fig. 9 but for POD (first column), FAR (second column) and CSI (third column).

In summary, the comparison of these error metrics shows that TPHiPr generally has better performance than the widely-used reanalysis data (ERA5), satellite-based data (IMERG) and multiple-sources merged data (MSWEP V2). In addition, it should be noted that some validation data from CMA, DHM and GHCN have been used to produce the IMERG and MSWEP V2. Therefore, if these data are removed

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

from the validation, more evident superiority of TPHiPr is expected compared with IMERG and MSWEP V2.

4.2.3 Comparison of precipitation extremes

Extreme precipitation is the leading cause of many water-related disasters. Therefore, this study also evaluates the performance of TPHiPr to reproduce extreme precipitation. Following some previous works (Katsanos et al., 2016; Li et al., 2022; Lockhoff et al., 2014), the 90th percentile of daily precipitation on wet days is set as the threshold for extreme precipitation in this study. Due to discontinuous temporal coverages of gauge observations, this study only evaluates the extreme precipitation of these datasets at 136 rain gauges with at least 2-year precipitation records and covering a complete seasonal cycle.

Figure 12 compares the detection skill of these precipitation datasets for extreme precipitation. Compared with the detection skill for all precipitation events (Fig. 10), the detection skill of all the four datasets for extreme precipitation is obviously reduced, with lower POD and CSI but higher FAR. Nevertheless, TPHiPr performs the best among these datasets. The median values of POD, FAR and CSI for TPHiPr are 0.39, 0.42 and 0.28, respectively, which is better than those of the other three datasets.

The 90th percentile (R90p) of daily precipitation on wet days, the average intensity (R90p_INT) and the frequency (R90p_FRQ) of precipitation greater than R90p are also calculated for each dataset and compared with those of rain gauge data. Figure 13 shows that all these datasets underestimate the intensity but overestimate the frequency of extreme precipitation. TPHiPr has worse performance than IMERG, however, it performs better than the ERA5 and MSWEP V2.

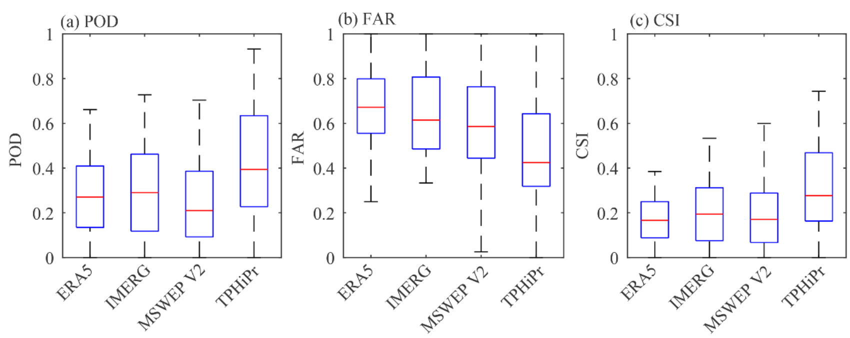

Figure 12: Similar to Fig. 10, but for extreme precipitation. The 90th percentile of observed daily precipitation at each rain gauge is taken as the threshold for calculating these metrics.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

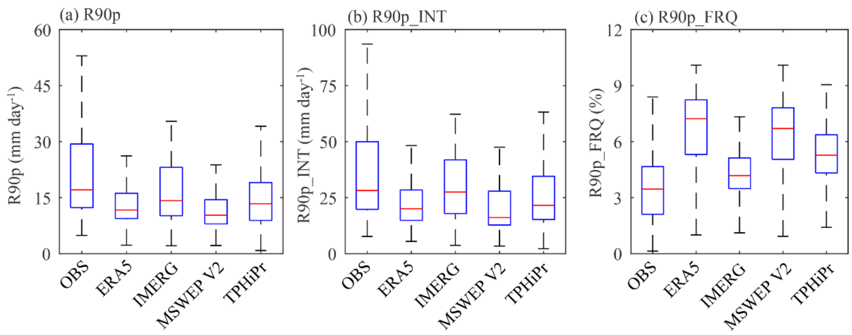

Figure 13: Comparison of (a) R90p, (b) R90p_INT and (c) R90p_FRQ for rain gauge data (OBS), ERA5, IMERG, MSWEP V2 and TPHiPr. R90p represents the 90th percentile of daily precipitation for each dataset. R90p_INT represents the average precipitation intensity of daily precipitation larger than R90p. R90p_FRQ represents the frequency of daily precipitation larger than R90p.

In summary, although the TPHiPr underestimates the intensity but overestimates the frequency of extreme precipitation, it has better performance than the other three datasets in detecting the occurrence of extreme precipitation.

5. Limitations

The above analysis shows that the TPHiPr produced in this study generally has high accuracy in the TP and is superior to the most widely-used global/quasi-global precipitation datasets. However, there are still some limitations in TPHiPr that need to be clarified.

As shown in Fig. 5, by merged with the rain gauge data, the accuracy of the gridded data is generally improved, but the improvements vary greatly in space. In the eastern TP, the improvement is evident, however, the accuracy at some western rain gauges outside the 2500 m contour changes little and even gets worse. This highlights the importance of high-density rain gauge data for precipitation merging, as demonstrated in many previous works that rain gauge density greatly impacts the accuracy of the produced dataset (Berndt et al., 2014; Girons et al., 2015; Xie et al., 2007). Therefore, the TPHiPr may still have large uncertainties in the west of the TP and regions where rain gauges are sparse.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

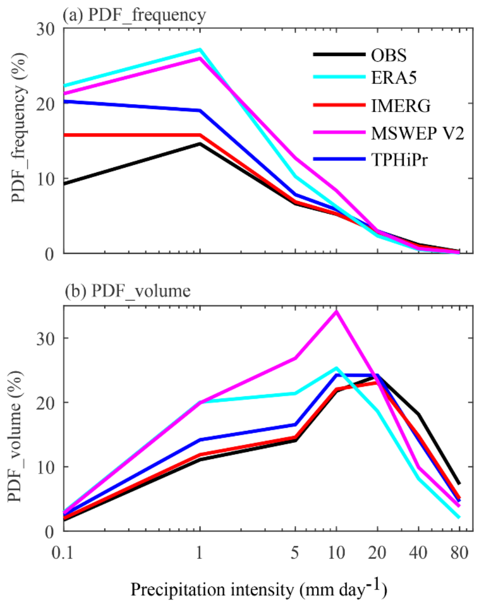

Figure 14: Comparison of the probability density function by (a) precipitation frequency and (b) amount for rain gauge data and the four datasets. The x axis is in log space.

Besides, previous studies have reported that the atmospheric simulation-based datasets generally overestimate the precipitation frequency (Hu and Yuan, 2021; P. Li et al., 2020). Therefore, we investigate the probability distribution function (PDF) of both precipitation frequency and amount in TPHiPr with respect to different precipitation intensities. As shown in Fig. 14, the TPHiPr largely overestimates the frequency of light precipitation (less than 5 mm day^{-1}), but the overestimation is smaller than that in ERA5 and MSWEP V2. In addition, we can find from Fig. 14b that the TPHiPr overestimates the amount of light to moderate precipitation but underestimates the amount of heavy precipitation, and the same is also found in ERA5 and MSWEP V2.

6. Conclusion

This study collects more than 9000 rain gauges over and around the Third Pole (TP) region from multiple sources. Then, the following steps are applied for merging the high-density gauge observations and the atmospheric simulation-based ERA5_CNN: first, the monthly precipitation climatology at gauge locations is obtained by correcting the climatology of ERA5_CNN with rain gauge data and the monthly

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

climatology at gauge locations is interpolated using a Random Forest based method; second, the ratios of observed monthly/daily precipitation to the climatology at gauge locations is interpolated for each month/day using the RF-based method; third, the monthly/daily precipitation fields are obtained by multiplying the interpolated monthly climatology by the interpolated monthly/daily ratios; finally, the daily precipitation fields are further adjusted using the monthly precipitation. Eventually, a long-term (1979-2020) high-resolution (1/30°) precipitation dataset (TPHiPr) is produced for the TP.

We compare the performance of the merged TPHiPr with the original ERA5_CNN data and three widely-used precipitation datasets, including the atmospheric simulation-based ERA5, the satellite-based IMERG and the MSWEP V2 merged from multiple sources. Results show that the TPHiPr retains the general spatial patterns of precipitation from ERA5_CNN but has a reduced wet bias in the TP, resulting in better error metrics than ERA5_CNN at most validation gauges. Meanwhile, the TPHiPr performs better than the three widely-used precipitation datasets in the TP, with respect to errors in both precipitation amount and detection skill. Validation with independent gauges shows that the TPHiPr has a negligible bias, low RMSE (4.5 mm day-1), high correlation (0.84) and high detection skill (CSI=0.67). In addition, the TPHiPr is more skillful than the three datasets in detecting extreme precipitation events, although it overestimates the frequency but underestimates the intensity of extreme precipitation.

In summary, a new high-accuracy precipitation dataset is produced for the data-sparse TP, which can be used for land surface modeling, water resource management, water-related disasters assessment, climate change research, et al. This dataset is expected to deepen our understanding of land surface processes and water cycles in the TP. Nevertheless, further efforts (e.g. setting up more rain gauges in remote regions and developing more skillful merging methods) are still needed for obtaining higher-accuracy precipitation datasets for the TP, as clarified in section 5, the produced data may still have large uncertainties in data-sparse regions and cannot reproduce the observed frequency and intensity of precipitation well.

Data and code availability

The TPHiPr precipitation dataset in NETCDF format is available at the National Tibetan Plateau Data Center, which can be accessed by https://doi.org/10.11888/Atmos.tpdc.272763 (Yang and Jiang, 2022). The codes used for producing this dataset are available upon request to the authors.

Author contributions: **Yaozhi Jiang**: Conceptualization, Investigation, Formal analysis, Methodology,

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Software, Visualization, Writing – original draft preparation; **Kun Yang**: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing; **Youcun Qi**: Data curation, Validation, Writing – review & editing; **Xu Zhou** and **Jie He**: Methodology, Writing – review & editing; **Hui Lu** and **Xin Li**: Supervision, Writing – review & editing; **Yingying Chen**: Data curation, Writing – review & editing; **Xiaodong Li**: Data curation; **Bingrong Zhou**, **Ali Mamtimin**, **Changkun Shao**, **Xiaogang Ma**, **Jiaxin Tian**, and **Jianhong Zhou**: Writing – review & editing.

Acknowledgements: This work is supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (Grant No. 2019QZKK0206), Basic Science Center for Tibetan Plateau Earth System of National Science Foundation of China (Grant No. 41988101), and the National Key Research and Development Program of China (Grant No. 2018YFC1507505).

Competing interests: The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

References

Alazzy, A.A., Lü, H., Chen, R., Ali, A.B., Zhu, Y., Su, J.: Evaluation of Satellite Precipitation Products and Their Potential Influence on Hydrological Modeling over the Ganzi River Basin of the Tibetan Plateau, Adv. Meteorol., 2017, 1-23, https://doi.org/10.1155/2017/3695285, 2017.

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.: 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, 2020.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F.: MSWEP v2 Global 3-hourly 0.1° precipitation: Methodology and quantitative assessment, Bull. Am. Meteorol. Soc., 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Berndt, C., Rabiei, E., Haberlandt, U.: Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios, J. Hydrol., 508, 88–101, https://doi.org/10.1016/j.jhydrol.2013.10.028, 2014.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1007/978-3-030-62008-0_35, 2001.

Chen, F., Ding, L., Piao, S., Zhou, T., Xu, B., Yao, T., Li, X.: The Tibetan Plateau as the engine for Asian environmental change: the Tibetan Plateau Earth system research into a new era, Sci. Bull., 66, 1263–1266, https://doi.org/10.1016/j.scib.2021.04.017, 2021.

Chen, R., Song, Y., Kang, E., Han, C., Liu, J., Yang, Y., Qing, W., Liu, Z.: A cryosphere-hydrology observation system in a small alpine watershed in the Qilian mountains of China and its meteorological gradient, Arctic, Antarct. Alp. Res., 46, 505–523, https://doi.org/10.1657/1938-4246-46.2.505, 2014.

Chen, R., Song, Y., Liu, J., Yang, Y., Qing, W., Liu, Z., Han, C.: Evaporation and precipitation dataset in Hulugou outlet in Upstream of Heihe River (2011), Natl. Tibet. Plateau Data Cent. [data set], https://dx.doi.org/10.3972/heihe.110.2013.db, 2015.

Contractor, S., Donat, M.G., Alexander, L. V., Ziese, M., Meyer-Christoffer, A., Schneider, U., Rustemeier, E., Becker, A., Durre, I., Vose, R.S.: Rainfall Estimates on a Gridded Network (REGEN) - A global land-based gridded dataset of daily precipitation from 1950 to 2016, Hydrol.

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Earth Syst. Sci., 24, 919–943, https://doi.org/10.5194/hess-24-919-2020, 2020.

Cui, P., Jia, Y.: Mountain hazards in the Tibetan Plateau: Research status and prospects, Natl. Sci. Rev., 2, 397–399, https://doi.org/10.1093/nsr/nwv061, 2015.

Curio, J., Maussion, F., Scherer, D.: A 12-year high-resolution climatology of atmospheric water transport over the Tibetan Plateau, Earth Syst. Dyn., 6, 109–124, https://doi.org/10.5194/esd-6-109-2015, 2015.

Gao, H., Wang, J., Yang, Y., Pan, X., Ding, Y., Duan, Z.: Permafrost Hydrology of the Qinghai-Tibet Plateau: A Review of Processes and Modeling, Front. Earth Sci., 8, 1–13, https://doi.org/10.3389/feart.2020.576838, 2021.

Gao, Y., Chen, F., Jiang, Y.: Evaluation of a convection-permitting modeling of precipitation over the Tibetan Plateau and its influences on the simulation of snow-cover fraction, J. Hydrometeorol., 21, 1531–1548, https://doi.org/10.1175/JHM-D-19-0277.1, 2020.

Gao, Y., Xu, J., Chen, D.: Evaluation of WRF mesoscale climate simulations over the Tibetan Plateau during 1979-2011, J. Clim., 28, 2823–2841, https://doi.org/10.1175/JCLI-D-14-00300.1, 2015.

Girons, L.M., Wennerström, H., Nordén, L.Å., Seibert, J.: Location and density of rain gauges for the estimation of spatial varying precipitation, Geogr. Ann. Ser. A Phys. Geogr., 97, 167–179, https://doi.org/10.1111/geoa.12094, 2015.

Hamada, A., Arakawa, O., Yatagai, A.: An automated quality control method for daily rain-gauge data, Glob. Environ. Res., 15, 183–192, 2011.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Sci. Data, 7, 1–11, https://doi.org/10.1038/s41597-020-0369-y, 2020.

He, X., Chaney, N.W., Schleiss, M., Sheffield, J.: 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, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan,

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hong, Z., Han, Z., Li, X., Long, D., Tang, G., Wang, J.: Generation of an improved precipitation dataset from multisource information over the tibetan plateau, J. Hydrometeorol., 22, 1275–1295, https://doi.org/10.1175/JHM-D-20-0252.1, 2021.

Hu, X., Yuan, W.: Evaluation of ERA5 precipitation over the eastern periphery of the Tibetan plateau from the perspective of regional rainfall events, Int. J. Climatol., 41, 2625–2637, https://doi.org/10.1002/joc.6980, 2021.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol., 8, 38–55, https://doi.org/10.1175/JHM560.1, 2007.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Kidd, R.J.C., Nelkin, E.J., Sorooshian, S., Tan, J., Xie, P.: Algorithm Theoretical Basis Document (ATBD) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), NASA 29, 2019.

Immerzeel, W.W., Wanders, N., Lutz, A.F., Shea, J.M., Bierkens, M.F.P.: Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff, Hydrol. Earth Syst. Sci., 19, 4673–4687, https://doi.org/10.5194/hess-19-4673-2015, 2015.

Jiang, Y., Yang, K., Li, Xiaodong, Zhang, W., Shen, Y., Chen, Y., Li, Xin: Atmospheric simulation-based precipitation datasets outperform satellite-based products in closing basin-wide water budget in the eastern Tibetan Plateau, Int. J. Climatol., 1–17, https://doi.org/10.1002/joc.7642, 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.

Katsanos, D., Retalis, A., Tymvios, F., Michaelides, S.: Analysis of precipitation extremes based on satellite (CHIRPS) and in situ dataset over Cyprus, Nat. Hazards, 83, 53–63, https://doi.org/10.1007/s11069-016-2335-8, 2016.

Li, D., Yang, K., Tang, W., Li, X., Zhou, X., Guo, D.: Characterizing precipitation in high altitudes of

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

the western Tibetan plateau with a focus on major glacier areas, Int. J. Climatol., 1–14, https://doi.org/10.1002/joc.6509, 2020.

Li, K., Tian, F., Khan, M.Y.A., Xu, R., He, Z., Yang, L., Lu, H., Ma, Y.: A high-accuracy rainfall dataset by merging multiple satellites and dense gauges over the southern Tibetan Plateau for 2014–2019 warm seasons, Earth Syst. Sci. Data, 13, 5455–5467, https://doi.org/10.5194/essd-13-5455-2021, 2021.

Li, P., Furtado, K., Zhou, T., Chen, H., Li, J.: Convection-permitting modelling improves simulated precipitation over the central and eastern Tibetan Plateau, Q. J. R. Meteorol. Soc., 1–22, https://doi.org/10.1002/qj.3921, 2020.

Li, Y., Pang, B., Ren, M., Shi, S., Peng, D., Zhu, Z., Zuo, D.: Evaluation of Performance of Three Satellite-Derived Precipitation Products in Capturing Extreme Precipitation Events over Beijing, China, Remote Sens, 14, 2698, https://doi.org/10.3390/rs14112698, 2022.

Lockhoff, M., Zolina, O., Simmer, C., Schulz, J.: Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations, J. Clim., 27, 607–623, https://doi.org/10.1175/JCLI-D-13-00194.1, 2014.

Lu, D., Yong, B.: Evaluation and hydrological utility of the latest GPM IMERG V5 and GSMaP V7 precipitation products over the Tibetan Plateau, Remote Sens., 10, https://doi.org/10.3390/rs10122022, 2018.

Lundquist, J., HugHes, M., Gutmann, E., KapnicK, S.: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks, Bull. Am. Meteorol. Soc., 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1, 2019.

Luo, L.: Meteorological observation data from the integrated observation and research station of the alpine environment in Southeast Tibet (2007-2017), Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/AtmosphericPhysics.tpe.68.db, 2018.

Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y., Long, D., Li, C., Han, Z., Liu, R.: Performance of Optimally Merged Multisatellite Precipitation Products Using the Dynamic Bayesian Model Averaging Scheme Over the Tibetan Plateau, J. Geophys. Res. Atmos., 123, 814–834, https://doi.org/10.1002/2017JD026648, 2018.

Ma, Y., Zhang, Y., Yang, D., Farhan, S. B.: Precipitation bias variability versus various gauges under

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

different climatic conditions over the Third Pole Environment (TPE) region, Int. J. Climatol., 35, 1201–1211, https://doi.org/10.1002/joc.4045, 2015.

Ma, Z., Xu, J., Ma, Y., Zhu, S., He, K., Zhang, S., Ma, W., Xu, X.: A Long-Term Asian Precipitation Dataset (0.1°, 1-hourly, 1951–2015, Asia) Anchoring the ERA5-Land under the Total Volume Control by APHRODITE, Bull. Am. Meteorol. Soc., 1146–1171, https://doi.org/10.1175/BAMS-D-20-0328.1, 2022.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., Finkelnburg, R.: Precipitation seasonality and variability over the Tibetan Plateau as resolved by the high Asia reanalysis, J. Clim., 27, 1910–1927, https://doi.org/10.1175/JCLI-D-13-00282.1, 2014.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G.: An overview of the global historical climatology network-daily database, J. Atmos. Ocean. Technol., 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1, 2012.

Norris, J., Carvalho, L.M.V., Jones, C., Cannon, F., Bookhagen, B., Palazzi, E., Tahir, A.A.: The spatiotemporal variability of precipitation over the Himalaya: evaluation of one-year WRF model simulation, Clim. Dyn., 49, 2179–2204, https://doi.org/10.1007/s00382-016-3414-y, 2017.

Ouyang, L., Lu, H., Yang, K., Leung, L.R., Wang, Y., Zhao, L., Zhou, X., LaZhu, Chen, Y., Jiang, Y., Yao, X.: Characterizing uncertainties in ground “truth” of precipitation over complex terrain through high-resolution numerical modeling, Geophys. Res. Lett., 1–11, https://doi.org/10.1029/2020gl091950, 2021.

Savéan, M., Delclaux, F., Chevallier, P., Wagnon, P., Gonga-Saholiariliva, N., Sharma, R., Neppel, L., Arnaud, Y.: Water budget on the Dudh Koshi River (Nepal): Uncertainties on precipitation, J. Hydrol., 531, 850–862, https://doi.org/10.1016/j.jhydrol.2015.10.040, 2015.

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., Stender, P.: Global gridded precipitation over land: A description of the new GPCC First Guess Daily product, Earth Syst. Sci. Data, 6, 49–60, https://doi.org/10.5194/essd-6-49-2014, 2014.

Sekulić, A., Kilibarda, M., Protić, D., Bajat, B.: A high-resolution daily gridded meteorological dataset for Serbia made by Random Forest Spatial Interpolation, Sci. Data, 8, 1–12, https://doi.org/10.1038/s41597-021-00901-2, 2021.

Shen, Y., Xiong, A., Hong, Y., Yu, J., Pan, Y., Chen, Z., Saharia, M.: Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau, Int. J. Remote Sens., 35, 6843–6858, https://doi.org/10.1080/01431161.2014.960612, 2014.

Sugimoto, S., Ueno, K., Fujinami, H., Nasuno, T., Sato, T., Takahashi, H.G.: Cloud-Resolving-Model Simulations of Nocturnal Precipitation over the Himalayan Slopes and Foothills, J. Hydrometeorol., 22, 3171–3188, https://doi.org/10.1175/JHM-D-21-0103.1, 2021.

Tan, X., Ma, Z., He, K., Han, X., Ji, Q., He, Y.: Evaluations on gridded precipitation products spanning more than half a century over the Tibetan Plateau and its surroundings, J. Hydrol., 582, https://doi.org/10.1016/j.jhydrol.2019.124455, 2020.

Wang, A., Zeng, X.: Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau, J. Geophys. Res. Atmos., 117, 1–12, https://doi.org/10.1029/2011JD016553, 2012.

Wang, J.: Daily meteorological Data of Nam Co Station China during 2019-2020, Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/Meteoro.tpdc.271782, 2021.

Wang, X., Pang, G., Yang, M.: Precipitation over the tibetan plateau during recent decades: A review based on observations and simulations, Int. J. Climatol., 38, 1116–1131, https://doi.org/10.1002/joc.5246, 2018.

Wang, X., Tolksdorf, V., Otto, M., Scherer, D.: WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis, Int. J. Climatol., 1–20, https://doi.org/10.1002/joc.6686, 2020.

Wang, Y., Wang, L., Li, X., Zhou, J., Hu, Z.: An integration of gauge, satellite, and reanalysis precipitation datasets for the largest river basin of the Tibetan Plateau, Earth Syst. Sci. Data, 12, 1789–1803, https://doi.org/10.5194/essd-12-1789-2020, 2020a.

Wang, Y., Yang, K., Zhou, X., Chen, D., Lu, H., Ouyang, L., Chen, Y., Lazhu, Wang, B.: Synergy of orographic drag parameterization and high resolution greatly reduces biases of WRF-simulated precipitation in central Himalaya, Clim. Dyn., 54, 1729–1740, https://doi.org/10.1007/s00382-019-05080-w, 2020b.

Wang, Y., Yang, K., Zhou, X., Wang, B., Chen, D., Lu, H., Lin, C., Zhang, F.: The Formation of a Dry-Belt in the North Side of Central Himalaya Mountains, Geophys. Res. Lett., 46, 2993–3000, https://doi.org/10.1029/2018GL081061, 2019.

Wei, D., Wang, X.: Meteorological observation data in an alpine steppe site of Shenzha Station (2015-018), Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/Meteoro.tpdc.270117, 2019.

Willmott, C.J., Robeson, S.M.: Climatologically aided interpolation (CAI) of terrestrial air temperature, Int. J. Climatol., 15, 221–229, https://doi.org/10.1002/joc.3370150207, 1995.

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

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., Hou, A.: Ground validation of GPM IMERG and trmm 3B42V7 rainfall products over Southern Tibetan plateau based on a high-density rain gauge network, J. Geophys. Res., 122, 910–924, https://doi.org/10.1002/2016JD025418, 2017.

Yang, K., Jiang, Y.: A long-term (1979-2020) high-resolution (1/30°) precipitation dataset for the Third Polar region (TPHiPr), Natl. Tibet. Plateau Data Cent. [data set], https://doi.org/10.11888/Atmos.tpdc.272763, 2022.

Yang, S., Li, R., Wu, T., Hu, G., Xiao, Y., Du, Y., Zhu, X., Ni, J., Ma, J., Zhang, Y., Shi, J., Qiao, Y.: Evaluation of reanalysis soil temperature and soil moisture products in permafrost regions on the Qinghai-Tibetan Plateau, Geoderma, 377, 114583, https://doi.org/10.1016/j.geoderma.2020.114583, 2020.

Yang, W., Yao, T., Zhu, M., Wang, Y.: Comparison of the meteorology and surface energy fluxes of debris-free and debris-covered glaciers in the southeastern Tibetan Plateau, J. Glaciol., 63, 1090–1104, https://doi.org/10.1017/jog.2017.77, 2017.

Yang, W.: Conventional ice surface meteorological data for Parlung Glacier No. 4 and Debris-covered 24K Glacier in southeast Tibet from June to September (2016), Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/AtmosPhys.tpe.249475.db, 2018.

Yao, T., Bolch, T., Chen, D., Gao, J., Immerzeel, W., Piao, S., Su, F., Thompson, L., Wada, Y., Wang, L., Wang, T., Wu, G., Xu, B., Yang, W., Zhang, G., Zhao, P.: The imbalance of the Asian water tower, Nat. Rev. Earth Environ., https://doi.org/10.1038/s43017-022-00299-4, 2022.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., Kitoh, A.: Aphrodite constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges,

https://doi.org/10.5194/essd-2022-299
Preprint. Discussion started: 15 September 2022

Bull. Am. Meteorol. Soc., 93, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.

Ye, B., Yang, D., Ding, Y., Han, T., Koike, T.: A bias-corrected precipitation climatology for China, J. Hydrometeorol., 62, 3–13 , https://doi.org/10.1175/JHM-366.1, 2007.

You, Q., Fraedrich, K., Ren, G., Ye, B., Meng, X., Kang, S.: Inconsistencies of precipitation in the eastern and central Tibetan Plateau between surface adjusted data and reanalysis, Theor. Appl. Climatol., 109, 485–496, https://doi.org/10.1007/s00704-012-0594-1, 2012.

Zhang, G.: Dataset of river basins map over the TP (2016), Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/BaseGeography.tpe.249465.file, 2019.

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., Ge, Y.: 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, 2021.

Zhang, Y.: Meteorological observation dataset of Shiquan River Source (2012-2015), Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/Meteoro.tpdc.270548, 2018.

Zhao, L., Wu, T., Xie, C., Li, R., Wu, X., Yao, J., Yue, G., Xiao, Y.: Support Geoscience Research, Environmental Management, and Engineering Construction with Investigation and Monitoring on Permafrost in the Qinghai-Tibet Plateau, China. Bull. Chinese Acad. Sci. (Chinese Version), 32, 1159–1168, https://doi.org/10.16418/j.issn.1000-3045.2017.10.015, 2017.

Zhao, L.: Meteorological Datasets of Xidatan station (XDT) on the Tibetan Plateau in 2014-2018, Natl. Tibet. Plateau Data Cent.[data set], https://doi.org/10.11888/Meteoro.tpdc.270084, 2018.

Zhou, X., Yang, K., Ouyang, L., Wang, Y., Jiang, Y., Li, X., Chen, D., Prein, A.: Added value of kilometer-scale modeling over the third pole region: a CORDEX-CPTP pilot study, Clim. Dyn., https://doi.org/10.1007/s00382-021-05653-8, 2021.