



Article

A Simple Method of Coupled Merging and Downscaling for Multi-Source Daily Precipitation Data

Na Zhao ^{1,2,3,*} and Kainan Chen ^{1,4}

¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; chenkn@fio.org.cn

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100101, China

³ Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing 210023, China

⁴ First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China

* Correspondence: zhaon@reis.ac.cn

Abstract: High accuracy and a high spatiotemporal resolution of precipitation are essential for the hydrological, ecological, and environmental fields. However, the existing daily gridded precipitation datasets, such as remote sensing products, are limited both by the coarse resolution and the low accuracy. Despite considerable efforts having been invested in downscaling or merging, a method of coupled and simultaneously downscaling and merging multiple datasets is currently lacking, which limits the wide application of individual popular satellite precipitation products. For the first time, in this study, we propose a simple coupled merging and downscaling (CMD) method for simultaneously obtaining multiple high-resolution and high-accuracy daily precipitation datasets. A pixel-repeated decomposition method was first proposed, and the random forest (RF) method was then applied to merge multiple daily precipitation datasets. The individual downscaled dataset was obtained by multiplying the result of merging by an explanatory rate obtained by RF. The results showed that the CMD method exhibited significantly better performance compared with the original datasets, with the mean absolute error (MAE) improving by up to 50%, the majority of the values of bias ranging between -1 mm and 1 mm, and the majority of the Kling–Gupta efficiency (KGE) values being greater than 0.7. CMD was more accurate than the widely used dataset, Multi-Source Weighted-Ensemble Precipitation (MSWEP), with a 43% reduction in the MAE and a 245% improvement in the KGE. In addition, the long-term estimation suggested that the proposed method exhibits stable good performance over time.

Keywords: precipitation; downscaling and merging; daily; China



Citation: Zhao, N.; Chen, K. A Simple Method of Coupled Merging and Downscaling for Multi-Source Daily Precipitation Data. *Remote Sens.* **2023**, *15*, 4377. <https://doi.org/10.3390/rs15184377>

Academic Editor: Seon Ki Park

Received: 18 August 2023

Revised: 4 September 2023

Accepted: 4 September 2023

Published: 6 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate precipitation information with a high spatial resolution is critical for understanding climate change and various impact studies in the ecological, hydrological, and agricultural fields [1–4]. However, due to its strong spatial and temporal heterogeneity, precipitation is one of the most difficult climatic variables to estimate, and it is challenging to derive highly accurate and high-resolution daily precipitation fields. Remote sensing and numerical models can provide continuous precipitation fields with varying spatial scales. These different types of precipitation products have proven to be useful in various research fields, including the study of changes in climatic means and extremes, as well as the monitoring of droughts and floods [5–7]. However, they remain limited to rather coarse spatial resolutions such as 0.5° – 0.25° , which is coarser than the scale of many environmental and ecological processes and the associated data requirements for multiple scientific and operational applications.

Spatial interpolation methods are the traditional way to generate uniformly gridded precipitation products by using irregular point observations, such as inverse distance

weighting, kriging, regression analysis, and the high-accuracy surface modeling method (HASM), which have been widely used in hydrological and meteorological studies [8–10]. However, the existing techniques face significant uncertainties in many areas characterized by the inadequate and uneven distribution of instruments measuring rainfall, particularly in regions without rain gauges [11]. Until now, we have not been able to acquire accurate spatial estimates of precipitation through the current networks of weather radars and rainfall-measuring instruments [12].

Currently, the mainstream approach for obtaining high-resolution precipitation data involves downscaling or fusion of the data of multi-source coarse-resolution remote sensing precipitation products and climate model outputs. There are various downscaling methods available, such as the geographically weighted regression method, machine learning, and regional climate models [13–16], each with its own advantages and disadvantages. Despite a growing number of studies focusing on downscaling techniques, previous studies have not simultaneously downscaled multiple popular remote sensing precipitation products. As different remote sensing or reanalysis data have their own strengths, including but not limited to their spatial coverage, time spans, varied performance, and scopes of application [17], it would be beneficial to downscale multiple popular datasets simultaneously, thereby increasing the selection and adaptability of data for researchers. In addition, downscaling and fusion are conceptually compatible with each other. In the downscaling process, multiple data sources are fused to obtain more accurate precipitation data. In data fusion, it is necessary to perform scale conversion on the data to fuse multiple data sources. However, the existing data fusion methods usually ignore the issue of scale transformation, and most of the previous merging approaches have tended to apply simple resampling techniques first before merging [18,19]. Additionally, previous studies usually combined station observations and only a single satellite precipitation dataset during data fusion, and have not fully and effectively utilized multiple data sources [13,20]. Although there are weighting methods available to combine multiple precipitation datasets [18], the significant challenge posed by the different weights when obtaining an appropriate weight is crucial for the accuracy of merged precipitation estimates, but such a weight is difficult to acquire. In light of the advantages and wide applications of the different currently available satellite precipitation products and the reanalysis datasets [5,19,21], it is imperative to simultaneously enhance the resolution and accuracy of the existing satellite precipitation products as well as to generate high-quality precipitation fields by developing a new method of data fusion and downscaling.

Motivated by these concerns, this study proposes a novel and simple method of coupled merging and downscaling for multi-source daily precipitation datasets that integrates several precipitation products in a unified framework. Our approach involves the following steps. First, to avoid the errors introduced by traditional interpolation-based resampling, we used a pixel-repeated decomposition method to unify different remote sensing precipitation data to a spatial resolution of 1 km. Secondly, we constructed regression models for multiple remote sensing data by combining meteorological observations using the random forest (RF) method and obtained the weights of the contribution of the satellite precipitation data to the true precipitation. Third, we used the co-kriging method to correct the residuals combined with the local geographic environmental variables. Finally, we added the results of random forest to those of co-kriging to obtain the final fusion data. In addition, to obtain the downscaled results of multiple remote sensing precipitation sources simultaneously, we used the weight contribution coefficients obtained above to decompose the fusion results to yield the downscaled result for each set of remote sensing precipitation data. The accuracy of our results was assessed using observation data and a widely used precipitation dataset: Multi-Source Weighted-Ensemble Precipitation (MSWEP) [22].

The rest of the article is organized as follows. Section 2 lists the materials, including the study area, the datasets, the proposed methodology, and the evaluation metrics. Section 3 focuses on the results. Discussions and the conclusions are given in Sections 4 and 5, respectively.

2. Materials

2.1. Study Area

Our research focused primarily on Mainland China ($3^{\circ}51'–53^{\circ}33'N$, $73^{\circ}40'–135^{\circ}05'E$) (Figure 1). Due to the impact of monsoons and topographical features, precipitation in China experiences significant spatial and temporal heterogeneity, as well as obvious seasonality, making it an area that is sensitive to climate change. The climate is generally affected by the eastern monsoon, resulting in higher precipitation in the southeast and lower precipitation in the northwest. Research has shown that every $1^{\circ}C$ increase in surface temperature in China corresponds to approximately increases of 10% and 23% in precipitation and extreme precipitation, respectively [23]. Under continued global warming, simulation of the fine spatial distribution of precipitation has always been a hot and difficult research topic [24,25].

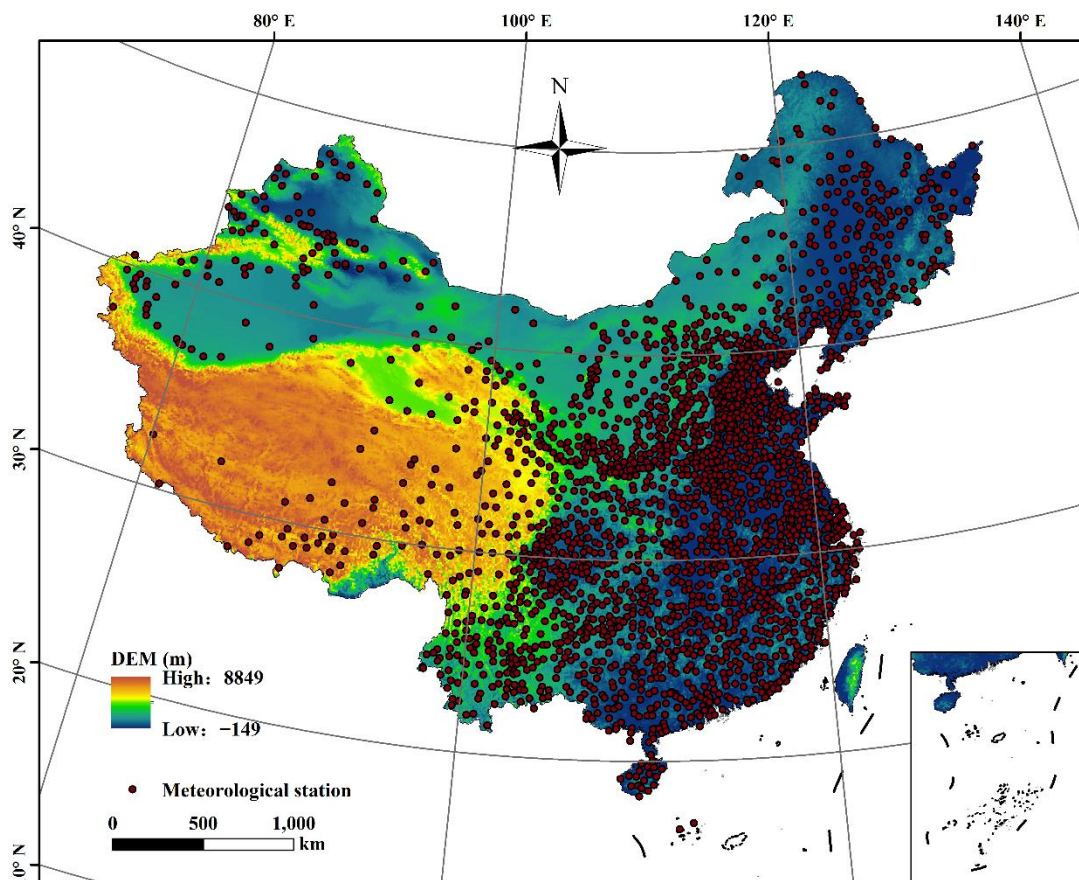


Figure 1. Study area and the distribution of meteorological stations across China.

2.2. Datasets

In this study, we proposed a novel coupled data merging and downscaling approach, which considers the explanatory power of each data source for actual precipitation using the RF method, and considers the relationship with local geographic environmental factors using the co-kriging method. Four prominent precipitation products, including the Climate Prediction Center morphing technique (CMORPH), the Global Satellite Mapping of Precipitation (GSMAP), the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement Mission (IMERG), the new fifth-generation atmospheric reanalysis of the European Centre for Medium Range Weather Forecasts (ERA5), and high-density station observations during the last five years, were utilized as the input data sources to train and validate proposed algorithm. These datasets, particularly in China and the

northwest region, have been validated to have a relatively high level of accuracy and performance [17,26,27], and have been widely applied in multiple fields [28–31].

Daily precipitation observations from more than 2400 meteorological stations were obtained from the China Meteorological Administration (CMA) (Figure 1). Although these observation sites provide a more comprehensive and detailed understanding of China's climatic characteristics compared with the 800+ publicly available national standard meteorological stations, in some regions, especially in western areas, observations are still sparse. This dataset has been subject to ongoing quality control, including extreme value checks, regional limit value checks, and spatiotemporal consistency checks using RHtests software V2 [32,33]. The spatial consistency check compares the time series of precipitation at the target station with those from nearby stations, while the internal consistency check is designed to identify erroneous reports caused by incorrect units, readings, or coding. Finally, considering the availability of the station observations, 2417 stations in total from 2017 to 2021 were used in this study.

This study utilized the global CMORPH V1.0 precipitation estimate that was developed by the Climate Prediction Center of the National Oceanic and Atmospheric Administration (NOAA) [34]. This estimate combines highly detailed cloud information from the infrared (IR) spectrum with relatively accurate precipitation data retrieved from the passive microwave (PMW) range, allowing it to cover the area between 60°S and 60°N with a 30 min temporal resolution and on a $0.25^\circ \times 0.25^\circ$ grid. Since January 1998, the estimation has undergone reprocessing and bias correction. The validation, conducted using gauge-based precipitation analysis, demonstrated that the CMORPH product effectively captures the precipitation distribution compared with six alternative satellite-derived precipitation estimates in China [35]. We used the suggested pixel repetition technique to obtain the 250×250 0.01° grid boxes for the designated $0.25^\circ \times 0.25^\circ$ grid cells.

A popular high-resolution global precipitation dataset, GSMAP, was used in this study. It includes three distinct products: the near-real-time product (GSMAP_NRT), the microwave-IR combined product (GSMAP_MVK), and the gauge-calibrated rainfall product (GSMAP_Gauge). Of these products, GSMAP_Gauge V7, with a spatial resolution of 0.1° , was used and was calibrated using global daily gauge data [36]. The 10×10 0.01° grid boxes at the target $0.1^\circ \times 0.1^\circ$ grid cells from 2017 to 2021 were obtained by using the pixel repetition method.

The global precipitation measurement project's produce IMERG is a frequently used multi-satellite merged precipitation retrieval product [37]. It contains IMERG-E, IMERG-L, and IMERG-F, each of which is obtained using distinct algorithms. IMERG-F has been found to be the most effective in most cases. This study used the daily calibrated products of IMERG-F at a spatial resolution of 0.1° from the latest IMERG V06B product. The 10×10 0.01° grid boxes at the target $0.1^\circ \times 0.1^\circ$ grid cells were obtained by using the pixel repetition approach.

ERA5 is the latest atmospheric reanalysis created by the European Centre for Medium Weather Forecasts. It replaced the production of ERA-Interim reanalysis, which concluded on 31 August 2019 [38]. ERA5 became publicly accessible in early February 2019, featuring a spatial resolution of 0.25° . Like ERA-Interim, ERA5 spans from 1979 onwards and has been extensively utilized in various applications and evaluations [39,40]. The 250×250 0.01° grid boxes at the target $0.25^\circ \times 0.25^\circ$ grid cells during 2017–2021 were obtained before merging and downscaling.

MSWEP is a recently launched precipitation dataset that offers global coverage from 1979 to the present, providing precipitation estimates every 3 h [41]. This dataset combines information from gauge observations, satellite data, and reanalysis datasets, incorporating two gauge observation datasets, two reanalysis datasets, and three satellite products. With its integrated data sources, MSWEP aims to deliver reliable precipitation estimates on a global scale. Despite its wide utilization for various purposes in both global and regional applications [42–44], it is essential to consider its limitations and potential uncertainties when using it for research or applications. The MSWEP dataset, similar to any other

precipitation dataset, is subject to uncertainties and potential biases, and has a coarse spatial resolution, typically ranging from 0.1 to 0.25 degrees. This level of resolution may not capture small-scale variations in precipitation accurately, especially in regions with a complex topography or strong spatial heterogeneity [41,45].

2.3. Methods

2.3.1. A Simple Coupled Merging and Downscaling (CMD) Method

The four-step flowchart displayed in Figure 2 presents the processes used to merge the gauge observations, three satellite precipitation products, and one reanalysis dataset, as well as simultaneously downscaling the four coarse satellite and reanalysis precipitation datasets. The first step involves time-matching the gauge observations and the satellite and reanalysis products, and pixel decomposition using the pixel repetition method. It should be noted that the daily precipitation of the remote sensing precipitation products is usually the sum of precipitation from 00:00 to 24:00 in the Central Time Zone. Considering that the statistical period for daily data of precipitation stations in China is from 20:00 the previous day to 20:00 on the current day in the Eastern Eight Time Zone, we recalculated the daily satellite and reanalysis precipitation corresponding to the statistical period of the precipitation stations using data with temporal resolutions higher than the daily scale for each precipitation product. The second step is to initially merge the satellite and reanalysis precipitation datasets and the gauge observations at a spatial resolution of $0.01^\circ \times 0.01^\circ$ using the RF method, which is an ensemble learning algorithm used for regression, classification, and selection ranking [46,47]. In addition, the explanatory rate of each data source was calculated in this step using the variance analysis method. Third, the final data fusion result was obtained by summing the result of the RF method and the residuals corrected using co-kriging together with the local explanatory variables, including altitude, latitude, longitude, slope, and some atmospheric environmental variables selected from the ERA5 datasets using the RF method. The individual downscaled precipitation dataset with a spatial resolution of $0.01^\circ \times 0.01^\circ$ was obtained by multiplying the merged results by the explanatory power of the corresponding data.

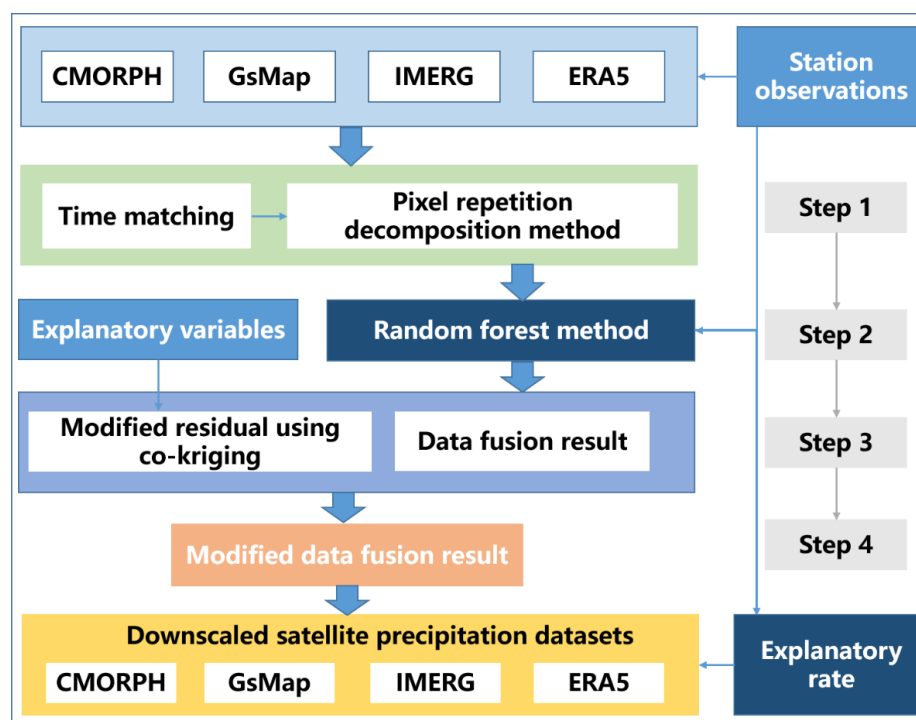


Figure 2. The framework of the proposed coupled merging and downscaling method.

Random forest is a popular machine learning algorithm that combines the concepts of ensemble learning and decision trees [47]. It is widely used for both regression-based prediction and variable selection tasks in various domains, including the socioeconomic and eco-environmental fields [48–51]. In RF, multiple decision trees are created using random subsets of the training data and random subsets of the input features. Each tree independently learns from the data and makes predictions. The final prediction in the regression tasks is obtained by averaging the predictions of all the individual trees. In regression-based prediction, random forests are effective because they can handle a large number of input features and automatically capture the complex nonlinear relationships between the features and the target variable. The algorithm is relatively robust against overfitting and can handle missing values and outliers in the data. Additionally, by measuring how much the performance of the model decreases when a particular feature is excluded, the algorithm determines the relative importance of each feature, allowing for more interpretable and efficient models. In summary, its ability to handle complex data, handle missing values, and provide the importance of the features makes it a valuable tool for regression-based prediction and variable selection tasks.

In this research, by considering gauge precipitation as the ground truth and disregarding the potential scaling discrepancies between the gauge precipitation and satellite precipitation, the fusion process was conducted using a functional relationship between gauge precipitation and multiple satellite precipitation datasets established through the RF method. Specifically, to achieve the best performance of RF in this study (i.e., the highest R^2), the model was built with 100 trees, and the ‘random_state’ parameter was set to 42. The constructed RF model for fusion can be expressed as the following equation:

$$P_{i,d}^{Gauge} = f_{RF}(X_{i,d}, \beta_{i,d}) + \varepsilon_{i,d} \quad (1)$$

where $P_{i,d}^{Gauge}$ is the gauge precipitation at station i on day d , f_{RF} denotes the relationships between the features $X_{i,d}$ and the target variable $P_{i,d}^{Gauge}$, $X_{i,d} = (P_{i,d}^{0.01^\circ, CMORPH}, P_{i,d}^{0.01^\circ, GSMaP}, P_{i,d}^{0.01^\circ, IMERG}, P_{i,d}^{0.01^\circ, ERA5})$ is a vector of the precipitation datasets, $\beta_{i,d} = (\beta_{i,d}^0, \beta_{i,d}^1, \beta_{i,d}^2, \beta_{i,d}^3, \beta_{i,d}^4)$ is the regression coefficient obtained using RF at location i for day d , and $\varepsilon_{i,d}$ is the residual. The regression coefficients in Equation (1) at the 0.01° grid cells can be obtained from observations and the four daily precipitation products: $P_{i,d}^{0.01^\circ, CMORPH}$, $P_{i,d}^{0.01^\circ, GSMaP}$, $P_{i,d}^{0.01^\circ, IMERG}$, and $P_{i,d}^{0.01^\circ, ERA5}$.

Co-kriging is a well-known geostatistical interpolation method that combines the information from two or more correlated variables to estimate the values at unsampled locations, and has been widely applied in various fields, such as hydrology and water resources, environmental monitoring, and agriculture, and predictions of crop yield [52,53]. It is an extension of the traditional kriging method, which is used for spatial interpolation and prediction. The main advantage of co-kriging is that it leverages the relationship between the primary and secondary variables to improve the accuracy of the interpolation. By incorporating additional information from the secondary variable, co-kriging can reduce the estimation error and provide more reliable predictions. This makes co-kriging particularly useful in situations where only sparse data for the primary variable are available [53]. Combined with the local explanatory variables of the residual $\varepsilon_{i,d}$ selected using RF, co-kriging was used to interpolate the residual $\varepsilon_{i,d}$ and yield the modified residual fields in this study.

Through application of the established RF function, the obtained regression coefficient $\beta_{i,d}$, and the co-kriging method, the merged results $P_{i,d}^{fusion}$ at location i on day d can be given as:

$$P_{i,d}^{fusion} = f_{RF}(X_{i,d}, \beta_{i,d}) + f_{co-kriging}(\varepsilon_{i,d}) \quad (2)$$

Furthermore, by using RF, we obtained the explanatory rate r_j of the j th precipitation dataset in Equation (1) (i.e., CMORPH, GSMAP, IMERG, and ERA5; $j = 1, \dots, 4$), and finally obtained the individual downscaled result of the j th precipitation dataset $P_{i,d,j}^{downscaling}$:

$$P_{i,d,j}^{downscaling} = P_{i,d}^{fusion} * r_j \quad (3)$$

2.3.2. Validation Method

To evaluate the performance of the proposed CMD approach, we used the 10-fold cross-validation method [54,55]. The dataset was randomly divided into 10 equal subsets using the Subset Features tool in ArcGIS software V10.6. We divided the dataset into 10 spatially random subsets, each containing an equal number of observations. We used nine of these subsets for training the CMD model and reserved the remaining one for testing. This method ensured a comprehensive and robust evaluation of the CMD approach, taking the spatial distribution of the data into account while maintaining an even distribution of the observations across all subsets. The weights applied to each precipitation dataset during CMD training were determined using the RF model. Specifically, the weights were 0.89, 0.91, 0.91, and 0.92 for GSMAP, ERA5, CMORPH, and IMERG, respectively. The performance of the RF model was evaluated using the coefficient of determination (R^2). The R^2 of the testing set was found to be 0.69 in this study.

We conducted a comparison of our estimates, the station observations, and a commonly used merged dataset, MSWEP V2. To align our results with the spatial resolution of the original MSWEP dataset (0.1°), we resampled our data from 0.01° using a straightforward bilinear interpolator. The performance of CMD was evaluated by the calculating mean errors across the 10 cross-validation procedures and quantified using common statistical metrics that include the correlation coefficient (CC), the mean absolute error (MAE), the root mean square error (RMSE), bias, and Kling–Gupta efficiency (KGE) [56]. The equations for these metrics are given in Equations (4)–(8), respectively.

$$CC = \frac{\sum_{i=1}^m (y_i - \bar{y})(y_i^* - \bar{y}^*)}{\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^m (y_i^* - \bar{y}^*)^2}} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^m |y_i - y_i^*|}{m} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (y_i - y_i^*)^2}{m}} \quad (6)$$

$$Bias = \frac{\sum_{i=1}^m |y_i - y_i^*|}{\sum_{i=1}^m y_i^*} \quad (7)$$

$$KGE = 1 - \sqrt{(CC - 1)^2 + \left(\frac{\bar{y}}{\bar{y}^*} - 1\right)^2 + \left(\frac{\sigma_y \bar{y}^*}{\sigma_{y^*} \bar{y}} - 1\right)^2} \quad (8)$$

where m denotes the amount of data; y_i and y_i^* are the estimated and observed precipitation at the i th site, respectively; \bar{y} and \bar{y}^* are the average of y_i and y_i^* ; and σ_y and σ_{y^*} are the standard deviation of y_i and y_i^* , respectively. CC measures the strength and direction of the linear relationship between two variables. It ranges from -1 to 1 , where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no

correlation. MAE calculates the average magnitude of the errors between the predicted and actual values. It provides a measure of how accurate the predictions are and is less sensitive to outliers. MAE ranges from 0 to infinity, with lower values indicating better accuracy. RMSE is similar to MAE but gives more weight to larger errors due to the differences being squared. It provides a measure of the spread of errors. RMSE ranges from 0 to infinity, with lower values indicating better accuracy. Bias quantifies the consistency or systematic deviation between the predicted and actual values. It represents the difference between the average predicted and actual values, and can be positive or negative. A bias of zero indicates no systematic deviation. KGE evaluates the overall performance of the model by assessing three aspects: correlation, variability, and distribution. It ranges from negative infinity to 1, with 1 indicating a perfect match, 0 indicating the same pattern with a different magnitude, and negative values indicating a poor match.

3. Results

Figure 3 displays the spatial pattern of the average error (MAE) of the four original satellite and reanalysis products (Figure 3a–d); the five results for 2017–2021, including the merged result (Figure 3e); and the four downscaled satellite and reanalysis daily precipitation datasets (Figure 3f–i). The value of MAE was obtained by averaging the MAE value of daily precipitation over the past five years. All the datasets exhibited higher accuracy in northern China and showed large values of MAE in southern China. The accuracy of the four original coarse precipitation datasets varied spatially across China, with higher MAE values observed in southeastern China (>4 mm). GSMAP performed better than the other three precipitation products, with most MAE values ranging from 0 to 3.8 mm. The error of the merged and downscaled results showed consistently similar spatial patterns, and most MAE values ranged between 0–3 mm for the five outputs of the proposed CMD method (Figure 3e–i), which are lower than those of the source datasets. In summary, according to the mean error indices during 2017–2021 (Figure 3, and Figures S1–S4 in the Supplementary Materials), CMD not only shows a good ability to merge the four satellite and reanalysis precipitation datasets but also shows high performance in simultaneously downsampling the four coarse datasets.

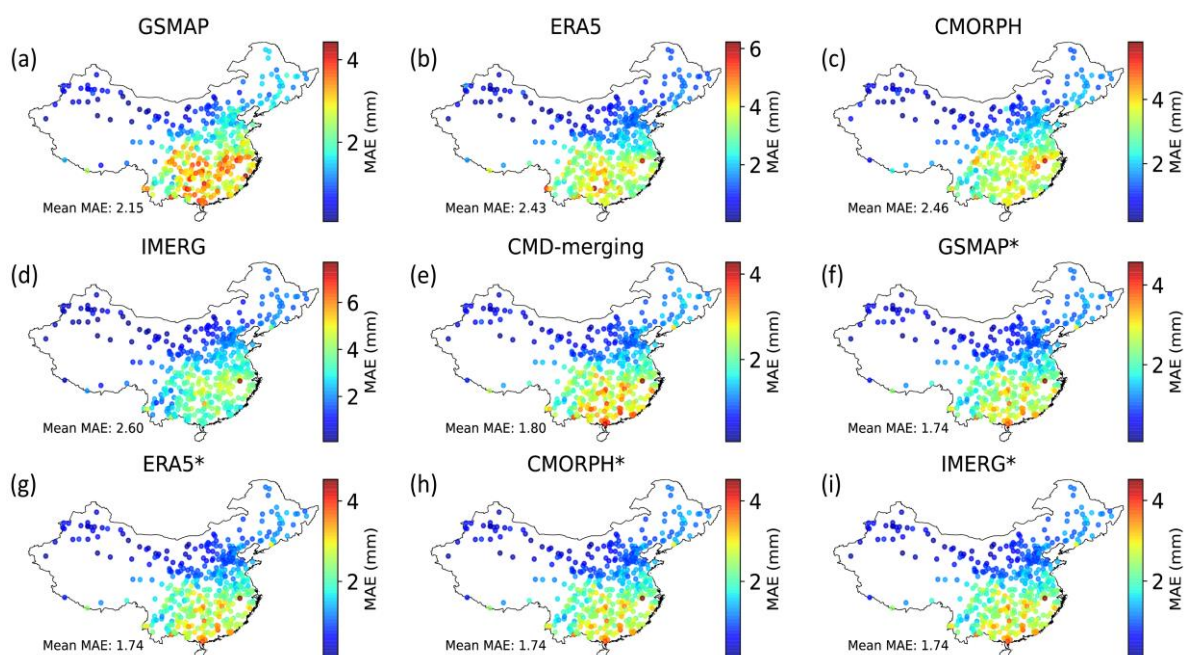


Figure 3. Spatial patterns of errors (MAE) of different precipitation datasets in mainland China. * denotes the individual downscaled precipitation products.

Figure 4 shows the regionally averaged error of different daily precipitation datasets, randomly selected from four seasons in 2021 (Day 1: January 16; Day 2: April 20; Day 3: July 20; Day 4: October 16) as examples. According to the MAE, RMSE, and KGE, the merged results and the downscaled results are more accurate than the individual original satellite and reanalysis datasets. On January 16, the MAE and RMSE of the merged result of the proposed CMD method improved by 37.2% and 36.7%, respectively. Earlier research indicated that while the downscaled outcomes consistently demonstrated enhanced precision compared with the initial satellite-based precipitation products, no notable advancements were observed before and after downscaling [57]. In this study, the downscaled results of GSMAP, ERA5, CMORPH, and IMERG using CMD showed significant improvements, with the MAE improving by 21.1%, 37.85%, 49.6%, and 50%, respectively, compared with the original datasets. CMD performed well in terms of merging together with downscaling, with a KGE value greater than 0.8. On April 20, compared with the original satellite and reanalysis datasets, the merged result of CMD improved by 42.3% and 42.2% on average, in the MAE and RMSE, while the individual downscaled results improved by 41.8% and 37%, 46.1% and 40.5%, 50.7% and 50.9%, 42.7% and 49%, respectively, when compared with the original dataset. The KGE values (>0.8) of the merged and downscaled results of CMD are significantly larger than those of the original datasets (<0.5). On July 20, the MAE of the merged result of CMD improved by 24.4%, 33.7%, 41.3%, and 47.8% compared with GSMAP, ERA5, CMORPH, and IMERG, respectively, and the MAE of the downscaled results of these coarse datasets improved by 23.4%, 33.1%, 40.8%, and 47.3%, respectively. CMD performed well in terms of merging and downscaling, with lower RMSE and larger KGE values compared with the original precipitation datasets. On October 16, the MAE (RMSE) value of the merged results of CMD improved by 26.5–63.1% (24.4–63.1%) compared with the four original precipitation datasets, and the downscaled results of these coarse datasets using CMD improved by 27.5–63.7% (26.5–64.2% for RMSE). In terms of the KGE, all five CMD results showed higher accuracy than the four original datasets, with the KGE approaching 0.9. In addition, we found that by using CMD, the CMORPH precipitation products improved the most in almost all cases, followed by IMERG. The five outputs of CMD, including one merged result and four downscaled results, generally had comparable levels of accuracy.

We presented and compared the spatial distribution of different precipitation datasets in Figure 5, taking two days (January 6 and July 20) in the dry and wet seasons of 2021 as examples due to space limitations. The spatial patterns of daily precipitation varied among the input data sources. On January 6 (Figure 5a–j), GSMAP and ERA5 exhibited similar spatial patterns, while CMORPH and IMERG exhibited significantly different patterns compared with the others. The merged and downscaled results of CMD showed consistent spatial patterns and had a similar spatial distribution to GSMAP and ERA5 but with significant differences in specific regions. In a comparison of some local observations, IMERG and CMORPH performed the worst, and GSMAP performed the best. The five outputs of the proposed CMD method performed better than the individual satellite and reanalysis datasets. On July 20 (Figure 5k–t), the original input precipitation data, including GSMAP, ERA5, CMORPH, and IMERG, still showed different spatial patterns, with CMORPH and IMERG exhibiting the largest differences. The five results of the proposed CMD method exhibited spatial patterns similar to GSMAP and ERA5 but with local differences. Compared with the station observations, IMERG performed the worst, followed by CMORPH, and ERA5 performed better than GSMAP. The downscaled results showed large improvements, especially for CMORPH and IMERG, indicating CMD's good ability to retrieve local details. The results of the CMD method showed the best performance in terms of both the spatial patterns and local accuracy. In addition, combined with the results of error shown in Figures 3 and 4, the difference in accuracy between the merged results and the four downscaled results did not exceed 0.1 mm according to the MAE, meaning that, to a large extent, these five outputs of the proposed CMD method are close to the true precipitation.

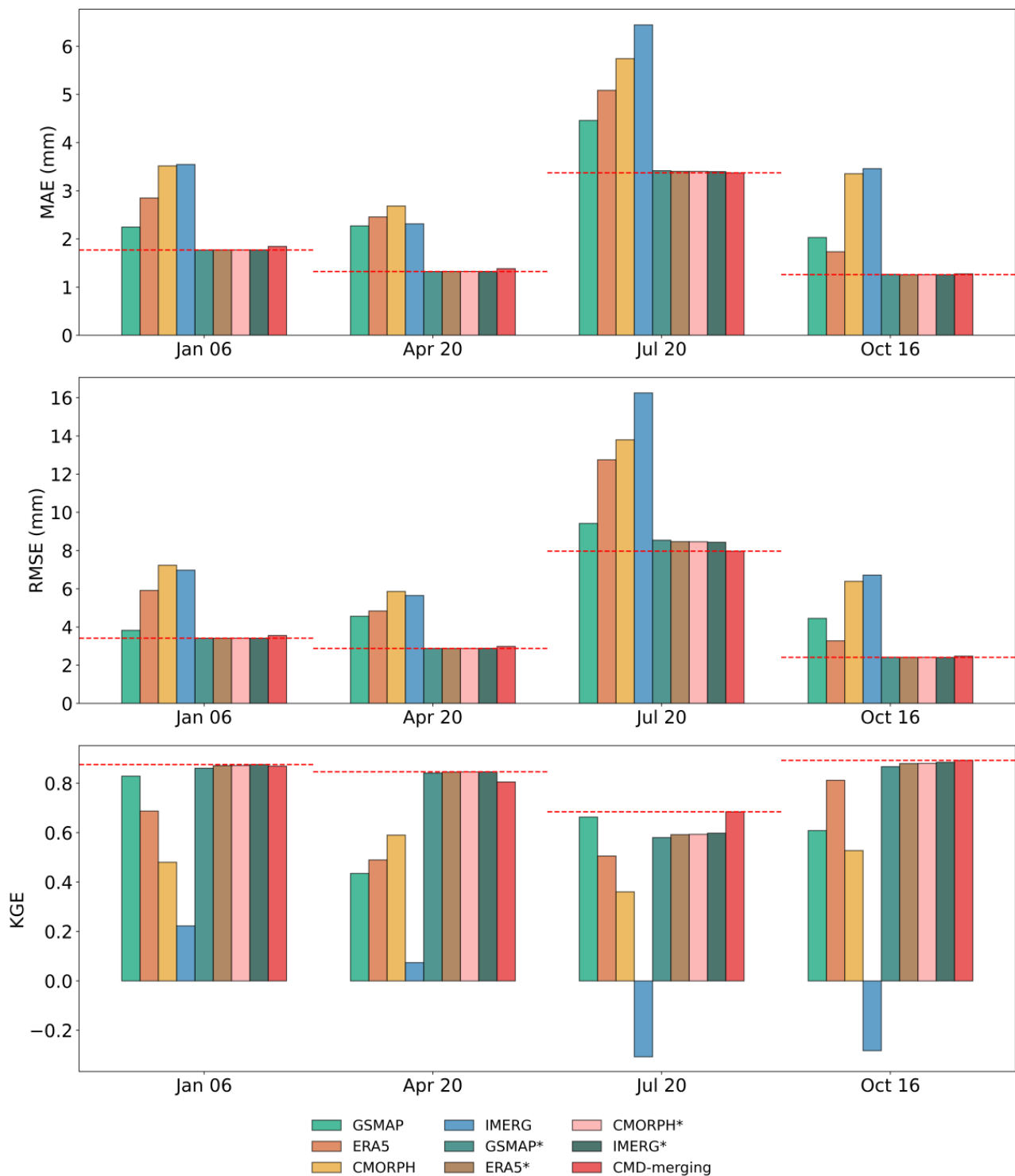


Figure 4. Regionally averaged error of different precipitation datasets over four days as examples. * denotes the downsampled result.

Figure 6 shows the time series of the errors for different precipitation datasets for the years 2020 and 2021. The results showed that, overall, the five results (the result of CMD fusion and the four individual downsampled results of the CMORPH, GSMAP, IMERG, and ERA5 datasets) of the proposed CMD method performed better than the original satellite and reanalysis precipitation datasets, with lower MAE and BIAS values and larger KGE values. The time evolution of MAE indicated that IMERG, CMORPH, and GSMAP performed relatively worse than the others. By using the proposed CMD

method (Figure 6c,d), the result of merged and the individual downscaled results showed comparable performance, and the downscaled results performed relatively better over time, with 80% of days within the two years having a MAE less than 3 mm and 45% of the days in 2020–2021 having an MAE less than 1 mm. Overvaluation and undervaluation were observed in the nine datasets (Figure 6e,f): 95% of the days in 2020 and 2021 were underestimated by IMERG, and most cases were overestimated by the other datasets. Compared with the original precipitation datasets, the outputs of the CMD method showed good performance in terms of bias, with more than 97% of cases having a bias of less than 1 mm, while the probability of a bias less than 1 ranged between 43% and 82% for the original coarse precipitation datasets. For the five outputs of CMD, the downscaled result of IMERG performed better than the others, with 98% of the days in the two years having a bias of less than 1, and the downscaled result of GSMAP performed the worst, with 96% of the days within the two years having a bias of less than 1, followed by the merged result (Figure 6g,h). Before downscaling, 41% of cases in 2020 and 2021 had a bias of less than -2 mm. However, the probability of a bias of less than 1 was 98% after downscaling using the CMD method. The likelihood of having a KGE greater than 0.5 ranged from 0 to 57% for the four coarse precipitation datasets, and from 0 to 2% for when the KGE was greater than 0.8. By using the proposed CMD method, the probability of a KGE greater than 0.5 reached 80% (Figure 6i–l). In addition, a comparison of the five output results based on KGE, which evaluates the overall performance of the model by assessing three aspects including correlation, variability, and distribution [56], revealed that the fused result performed best, followed by the downscaled IMERG data, while the downscaled GSMAP data performed worst.

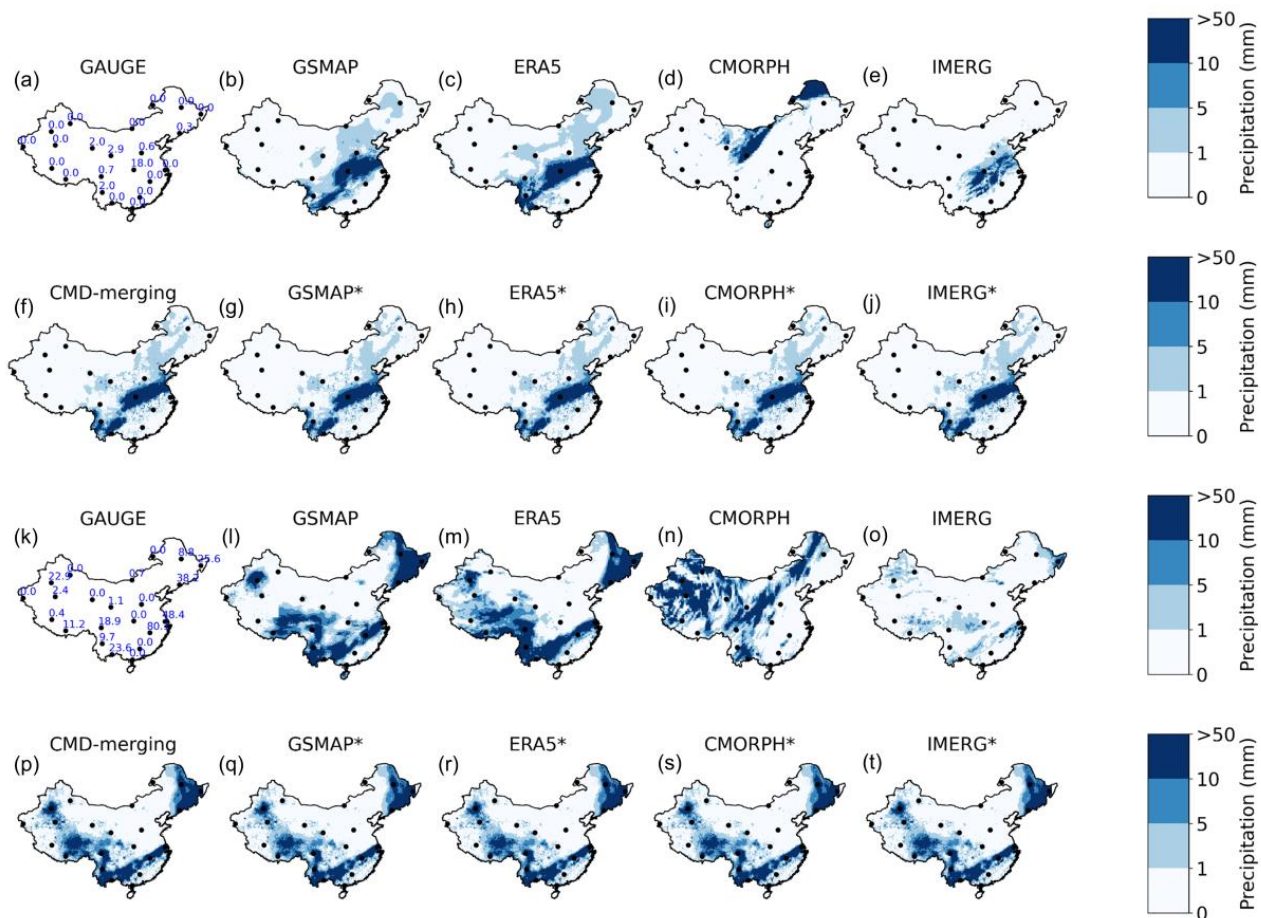


Figure 5. Spatial distributions of different precipitation datasets: (a–j) January 6; (k–t) July 20. * denotes the downscaled result.

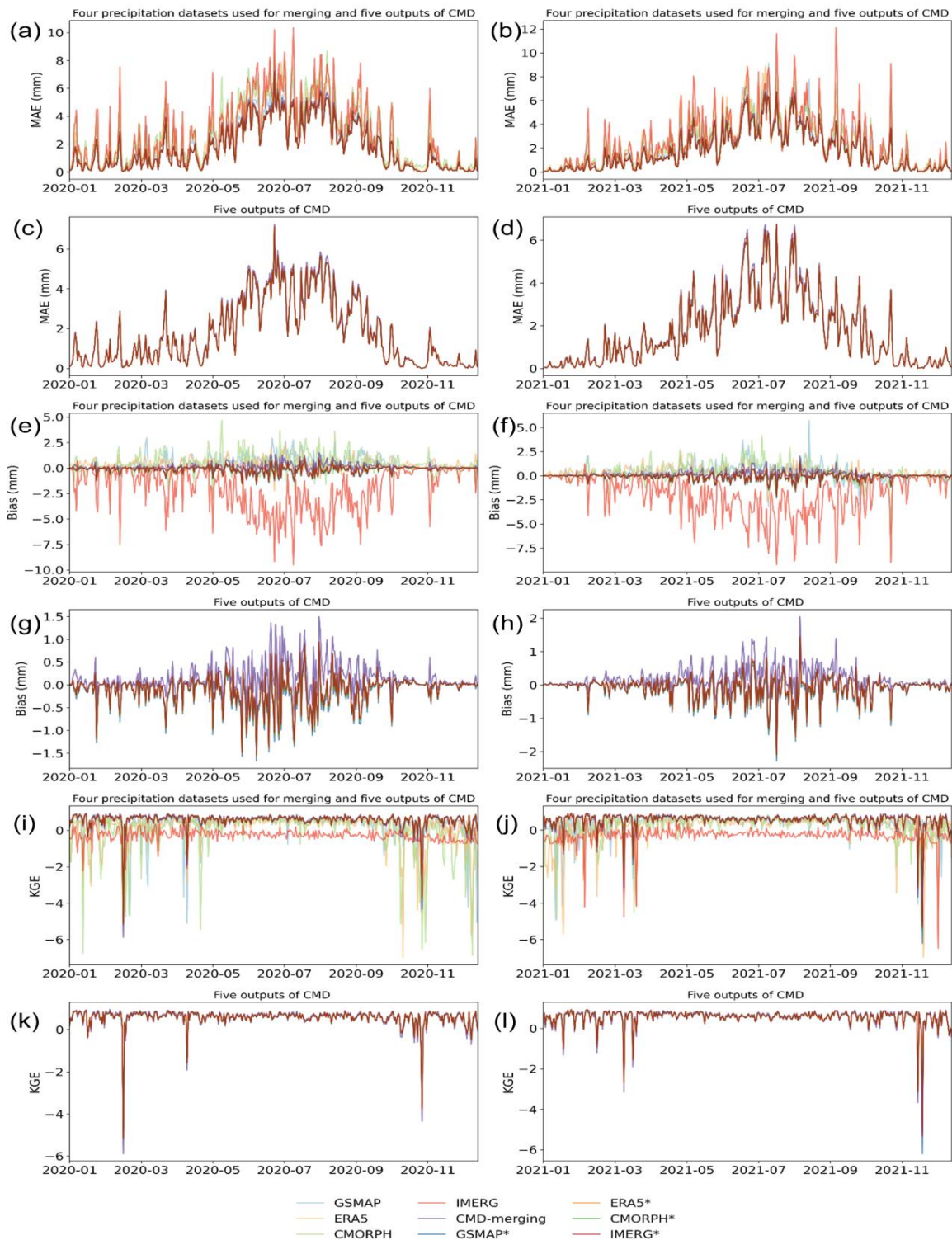


Figure 6. Time series of the errors of different daily datasets for the two years used as examples ((a,c,e,g,i,k) for errors in the year 2020; (b,d,f,h,j,l) for errors in the year 2021; * denotes downscaled data).

In addition to comparing these results with on-site observations, we also compared them with a widely used fusion dataset, Multi-Source Weighted-Ensemble Precipitation (MSWEP). This dataset, generated by merging gauge, satellite, and reanalysis data, has been widely used in several scientific and practical applications [22,43,58]. The results of CMD

were resampled to 0.1° to match the spatial resolution of MSWEP. Figure 7 shows the spatial patterns of the average errors of different datasets during 2020–2021 (Figure 7a–f) and the time series of the errors for the two exemplary years (Figure 7g–l). The results showed that, compared with the merged and downscaled results of CMD, MSWEP had higher MAE values regarding site observations, and most of the MAE values ranged from 1.5 mm to 6.7 mm, with the majority being above 3 mm. By using CMD, the merged result was greatly improved, with 91% of the MAE values being less than 3 mm. The accuracy of the downscaled results was slightly higher than that of the results of fusion, with 93.5% of the MAE values being below 3 mm. With respect to the station observations, it can be seen that the results of CMD, including the merged and downscaled results, performed better than MSWEP in terms of the MAE, bias, and KGE. Both overestimations and underestimations were observed for the comparison datasets, and the magnitudes of both the overestimations and underestimations of MSWEP were significantly larger than those of all outputs of the CMD method. Moreover, 77% of the CMD's KGE values were greater than 0.5, and 20% of the KGE values were larger than 0.8, yet the majority of the KGE values for MSWEP were below 0.5.

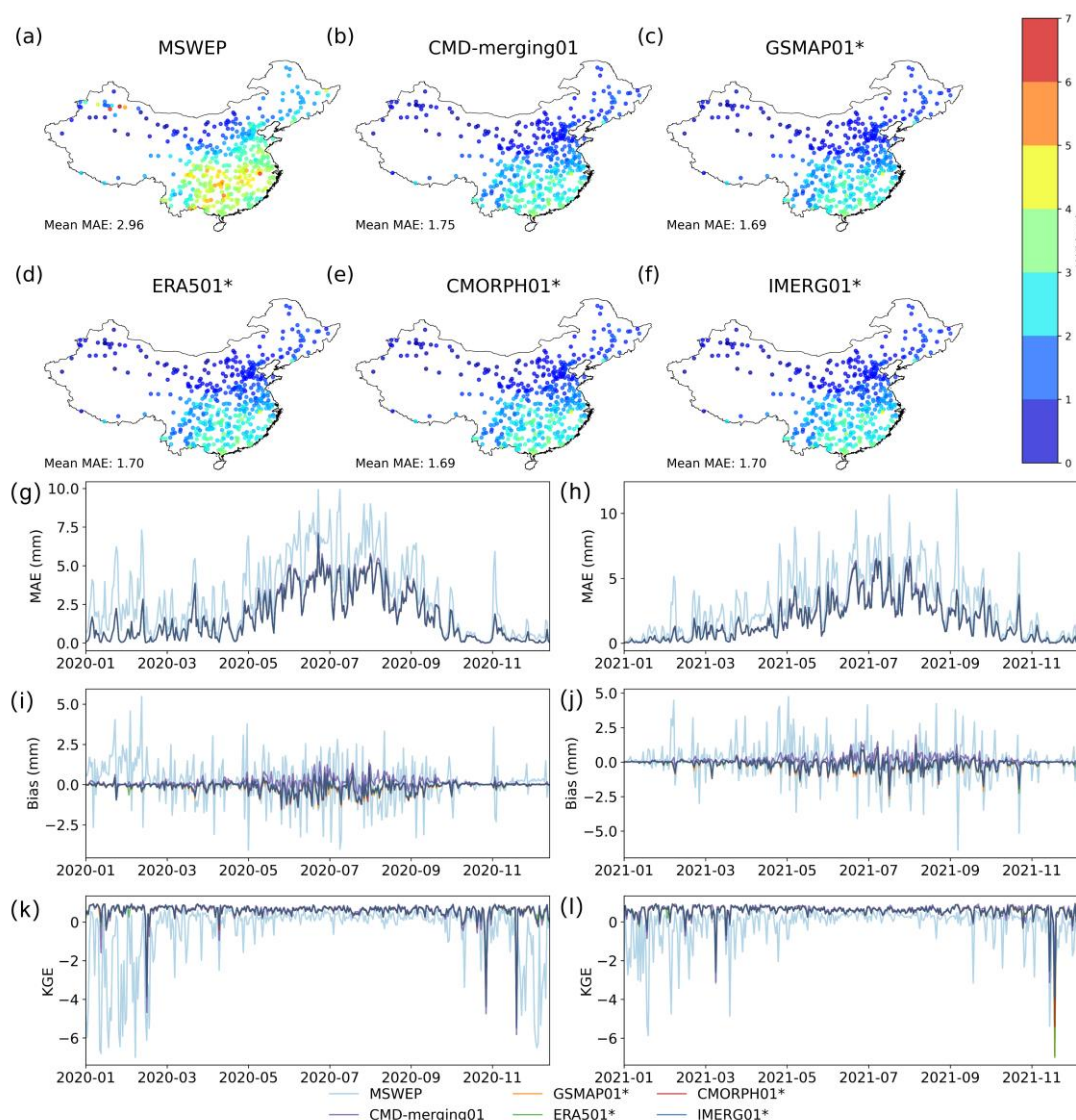


Figure 7. Comparison with the widely used data fusion dataset, MSWEP ((a–f) denote the mean MAE values of different products; (g–l) give the time series of errors in the years 2020 and 2021; 01 means a 0.1° spatial resolution; * indicates downscaling).

4. Discussion

Precipitation is one of the most difficult climate variables to estimate, and the high accuracy and high resolution of precipitation plays a vital role in the natural ecosystem and human society. Remote sensing precipitation data and reanalysis data are often used in agricultural, ecological, hydrological, and other research fields due to their ability to provide spatially continuous information on precipitation. However, their coarse resolution and relatively low accuracy have limited their further widespread use. Moreover, different data products, such as GSMAP, IMERG, and CMORPH, have different spatial coverage and time spans and exhibit significant regional differences in accuracy, which has led to their different scientific and practical applications [6,17]. To provide multiple high-quality precipitation datasets with potential for different applications, a new method is therefore necessary to simultaneously downscale multiple popular remote sensing and reanalysis precipitation products.

For the first time, this study proposed a new simple coupled merging and downscaling method to simultaneously acquire multiple high-quality fine-scaled daily precipitation datasets. First, unlike traditional approaches, we proposed a pixel replication method to decompose data with a coarse resolution, avoiding additional errors introduced at the pixels' boundaries by resampling. Secondly, we made full use of the advantages of random forest to fuse multiple precipitation data sources while providing the rates of contribution of different precipitation datasets to true precipitation. Based on these contribution rates, we further decomposed the results of fusion to obtain downscaled results for individual precipitation sources. It should be noted that having more data sources enabled greater accuracy in the results of fusion using the random forest method, which led to improved accuracy in the downscaled results for individual precipitation datasets. However, as the focus of this study was to propose a new methodological framework, we selected some commonly used precipitation products for this purpose, including three prominent daily satellite precipitation datasets and a popular reanalysis dataset. The cross-validation results showed that CMD's five output results were more accurate than the input datasets (Figures 3, 4 and 6). In terms of the spatial distribution, the proposed method combined the advantages of having multiple sources of input data while also accurately reflecting the local details better than the individual data sources (Figure 5). In addition, a comparison of the results of CMD with the widely used third-party data of MSWEP based on station observations during 2020–2021 demonstrated that the proposed method outperformed MSWEP in both accuracy and long-term time series simulation (Figure 7).

The five output results of CMD, including the fused results and the downscaled results, showed considerable accuracy, with the MAE not exceeding 1 mm. The advantage of this method lies in its ability to significantly improve the accuracy of the original precipitation data that had a high level of error, while also substantially enhancing the spatial distribution. Furthermore, a distinctive feature of the proposed approach in this study is its simultaneous generation of multiple high-resolution and high-accuracy precipitation datasets to meet diverse application needs.

Although the CMD method provides enhanced estimates, there are still uncertainties associated with the final results. These uncertainties arise from factors such as the density and locations of the observation stations, the inherent uncertainty in the remote sensing and reanalysis precipitation products, and the relationship between precipitation and the predictor's variables. RF models tend to have a higher risk of overfitting, especially when the number of trees in the forest is large and the model's complexity is high, and it is sensitive to noisy or erroneous data. Overfitting occurs when the model captures noise or irrelevant patterns in the training data, leading to poor performance in terms of generalization with unseen data. The effectiveness of co-kriging relies on the spatial correlation between the primary variable of interest and the auxiliary variables. If suitable auxiliary variables are not available or if there is a low correlation between the primary and auxiliary variables, co-kriging may not provide better results. In the study, we trained the method based on the available research datasets to optimize its performance and obtain the

corresponding parameters. In addition, residual correction is a commonly used approach, and its effectiveness has been demonstrated by several studies [57,59,60]. However, due to space limitations, this study mainly focused on the framework of how to simultaneously obtain multiple high-quality datasets and did not provide a separate explanation for the effectiveness of residual correction. Local explanatory variables may influence the degree to which the final results are improved. Future research will investigate the extent of this impact further.

The limitation of this study is that it uses regionally averaged values to assess the contribution rates. However, due to the large area of China, precipitation products in different regions exhibit significant variations in accuracy. It is important to consider different rates of contribution for precipitation in different regions, and future research should focus on specific areas. Furthermore, we evaluated the effectiveness of the proposed method by comparing it with station-based precipitation observations from the past five years. The results clearly indicate that the method achieved a high level of accuracy and consistent performance in simulating precipitation. These findings suggest that the method has the potential to deliver improved results when applied to simulations at different time scales. However, it is essential to acknowledge the inherent spatial and temporal heterogeneity of precipitation. Therefore, it is necessary to further validate the simulation uncertainty for longer periods, such as inter-decadal variations, in future research.

5. Conclusions

In this study, a simple coupled downscaling and merging method was proposed to simultaneously obtain multiple high-quality precipitation datasets using RF. The proposed method, named CMD, was applied to daily data from Mainland China. The results showed that CMD outperformed the original popular satellite and reanalysis datasets of daily precipitation, and the performance of CMD was stable over time in terms of MAE, bias, RMSE, and KGE. The results of CMD improved by 27% on average in terms of the MAE. CMD successfully reduced the magnitude of overestimation and underestimation, with over 97% of the estimated cases having a bias of less than 1 mm, while the probability of a bias of less than 1 ranged from 43% to 82% for the original coarse precipitation datasets. The KGE value changed from less than 0.5 in most cases to more than 0.5 in 80% of cases. When compared with the MSWEP dataset, the majority of MAE values of the MSWEP were greater than 3 mm, while more than 94% of CMD's MAEs were less than 3 mm with respect to the station observations. In addition, CMD performed better in capturing the time series of daily precipitation, while the KGE value of MSWEP was below 0.5 in most cases.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs15184377/s1>. Figure S1: Spatial patterns of the mean CC values during 2017–2021 of different precipitation datasets from Mainland China; * denotes the individual downscaled precipitation products. Figure S2: Spatial patterns of the mean RMSE values during 2017–2021 for different precipitation datasets from Mainland China; * denotes the individual downscaled precipitation products. Figure S3: Spatial patterns of the mean bias values during 2017–2021 of the different precipitation datasets from Mainland China; * denotes the individual downscaled precipitation products. Figure S4: Spatial patterns of the mean KGE values during 2017–2021 of the different precipitation datasets from mainland China; * denotes the individual downscaled precipitation products.

Author Contributions: Conceptualization, N.Z.; methodology, N.Z. and K.C.; software, K.C.; validation, N.Z. and K.C.; formal analysis, N.Z.; investigation, N.Z.; resources, N.Z. and K.C.; data curation, K.C.; writing—original draft preparation, N.Z.; writing—review and editing, N.Z. and K.C.; visualization, K.C.; supervision, N.Z.; project administration, N.Z.; funding acquisition, N.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Major Program of the National Natural Science Foundation of China (No. 42293270), the National Program of National Natural Science Foundation of China (No. 42071374), and the Key Project of Innovation LREIS (KPI001).

Data Availability Statement: Data and additional information can be obtained by directly contacting the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pfister, L.; Brönnimann, S.; Schwander, M.; Isotta, F.A.; Horton, P.; Rohr, C. Statistical reconstruction of daily precipitation and temperature fields in Switzerland back to 1864. *Clim. Past* **2020**, *16*, 663–678. [[CrossRef](#)]
2. Rodell, M.; Famiglietti, J.S.; Wiese, D.N.; Reager, J.T.; Beaudoin, H.K.; Landerer, F.W.; Lo, M.H. Emerging trends in global freshwater availability. *Nature* **2019**, *565*, E7, Correction to *Nature*, **2018** *557*, 651–659. [[CrossRef](#)] [[PubMed](#)]
3. Li, Y.; Guan, K.; Schnitkey, G.D.; DeLucia, E.; Peng, B. Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. *Glob. Chang. Biol.* **2019**, *25*, 2325–2337. [[CrossRef](#)] [[PubMed](#)]
4. Iizumi, T.; Furuya, J.; Shen, Z.H.; Kim, W.; Okada, M.; Fujimori, S.; Hasegawa, T.; Nishimori, M. Responses of crop yield growth to global temperature and socioeconomic changes. *Sci. Rep.* **2017**, *7*, 7800. [[CrossRef](#)]
5. Zhu, S.; Wei, J.A.; Zhang, H.R.; Xu, Y.; Qin, H. Spatiotemporal deep learning rainfall-runoff forecasting combined with remote sensing precipitation products in large scale basins. *J. Hydrol.* **2023**, *616*, 128727. [[CrossRef](#)]
6. Gummadi, S.; Dinku, T.; Shirsath, P.B.; Kadiyala, M.D.M. Evaluation of multiple satellite precipitation products for rainfed maize production systems over Vietnam. *Sci. Rep.* **2022**, *12*, 485. [[CrossRef](#)] [[PubMed](#)]
7. Shi, J.Y.; Wang, B.; Wang, G.Q.; Yuan, F.; Shi, C.X.; Zhou, X.; Zhang, L.M.; Zhao, C.X. Are the Latest GSMaP Satellite Precipitation Products Feasible for Daily and Hourly Discharge Simulations in the Yellow River Source Region? *Remote Sens.* **2021**, *13*, 4199. [[CrossRef](#)]
8. Nie, S.P.; Luo, Y.; Wu, T.W.; Shi, X.L.; Wang, Z.Z. A merging scheme for constructing daily precipitation analyses based on objective bias-correction and error estimation techniques. *J. Geophys. Res. Atmos.* **2015**, *120*, 8671–8692. [[CrossRef](#)]
9. Shen, Y.; Xiong, A. Validation and comparison of a new gauge-based precipitation analysis over mainland China. *Int. J. Climatol.* **2016**, *36*, 252–265. [[CrossRef](#)]
10. Zhao, N.; Yue, T.X.; Li, H.; Zhang, L.L.; Yin, X.Z.; Liu, Y. Spatio-temporal changes in precipitation over Beijing-Tianjin-Hebei region, China. *Atmos. Res.* **2018**, *202*, 156–168. [[CrossRef](#)]
11. Ouyang, L.; Lu, H.; Yang, K.; Leung, L.R.; Wang, Y.; Zhao, L.; Zhou, X.; Zhu, L.; Chen, Y.; Jiang, Y.; et al. Characterizing Uncertainties in Ground “Truth” of Precipitation Over Complex Terrain Through High-Resolution Numerical Modeling. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091950. [[CrossRef](#)]
12. Kidd, C.; Becker, A.; Huffman, G.J.; Muller, C.L.; Joe, P.; Skofronick-Jackson, G.; Kirschbaum, D.B. So, How Much of The Earth’s Surface Is Covered by Rain Gauges? *Bull. Am. Meteorol. Soc.* **2017**, *98*, 69–78. [[CrossRef](#)]
13. Arshad, A.; Zhang, W.C.; Zhang, Z.J.; Wang, S.H.; Zhang, B.; Cheema, M.J.M.; Shalamzari, M.J. Reconstructing high-resolution gridded precipitation data using an improved downscaling approach over the high altitude mountain regions of Upper Indus Basin (UIB). *Sci. Total Environ.* **2021**, *784*, 147140. [[CrossRef](#)] [[PubMed](#)]
14. Chen, S.L.; Xiong, L.H.; Ma, Q.M.; Kim, J.S.; Chen, J.; Xu, C.Y. Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method. *J. Hydrol.* **2020**, *589*, 125156. [[CrossRef](#)]
15. Ge, J.; Qiu, B.; Wu, R.Q.; Cao, Y.P.; Zhou, W.D.; Guo, W.D.; Tang, J.P. Does Dynamic Downscaling Modify the Projected Impacts of Stabilized 1.5 degrees C and 2 degrees C warming on Hot Extremes Over China? *Geophys. Res. Lett.* **2021**, *48*, e2021GL092792. [[CrossRef](#)]
16. Yan, X.; Chen, H.; Tian, B.; Sheng, S.; Wang, J.; Kim, J.-S. A Downscaling-Merging Scheme for Improving Daily Spatial Precipitation Estimates Based on Random Forest and Cokriging. *Remote Sens.* **2021**, *13*, 2040. [[CrossRef](#)]
17. Wu, X.; Zhao, N. Evaluation and Comparison of Six High-Resolution Daily Precipitation Products in Mainland China. *Remote Sens.* **2023**, *15*, 223. [[CrossRef](#)]
18. Hu, L.; Peng, D.; Zhang, M.; Qiu, L. Spatial Interpolation of Meteorological Variables in Yarlung Zangbo River Basin. *J. Beijing Norm. Univ. Nat. Sci.* **2012**, *48*, 449–452.
19. Sakata, S.; Ashida, F.; Zako, M. Hybrid approximation algorithm with Kriging and quadratic polynomial-based approach for approximate optimization. *Int. J. Numer. Methods Eng.* **2007**, *70*, 631–654. [[CrossRef](#)]
20. Xiao, Y.; Xie, G.; An, K. Comparison of interpolation methods for content of soil available phosphor. *Chin. J. Eco-Agric.* **2003**, *11*, 56–58.
21. Haarhoff, S.J.; Kotze, T.N.; Swanepoel, P.A. A prospectus for sustainability of rainfed maize production systems in South Africa. *Crop Sci.* **2020**, *60*, 14–28. [[CrossRef](#)]
22. Beck, H.E.; van Dijk, A.I.J.M.; Levizzani, V.; Schellekens, J.; Miralles, D.G.; Martens, B.; de Roo, A. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 589–615. [[CrossRef](#)]
23. Sun, J.; Ao, J. Changes in precipitation and extreme precipitation in a warming environment in China. *Chin. Sci. Bull.* **2013**, *58*, 1395–1401. [[CrossRef](#)]

24. Ding, Y.H.; Shi, X.L.; Liu, Y.M.; Liu, Y.; Li, Q.Q.; Qian, F.F.; Miao, Q.Q.; Zhai, Q.Q.; Gao, K. Multi-year simulations and experimental seasonal predictions for rainy seasons in China by using a nested regional climate model (RegCM_NCC). part I: Sensitivity study. *Adv. Atmos. Sci.* **2006**, *23*, 323–341. [[CrossRef](#)]
25. Wu, S.Y.; Wu, Y.J.; Wen, J.H. Future changes in precipitation characteristics in China. *Int. J. Climatol.* **2019**, *39*, 3558–3573. [[CrossRef](#)]
26. Zhu, H.; Chen, S.; Li, Z.; Gao, L.; Li, X. Comparison of Satellite Precipitation Products: IMERG and GSMaP with Rain Gauge Observations in Northern China. *Remote Sens.* **2022**, *14*, 4748. [[CrossRef](#)]
27. Wang, Y.; Zhao, N. Evaluation of Eight High-Resolution Gridded Precipitation Products in the Heihe River Basin, Northwest China. *Remote Sens.* **2022**, *14*, 1458. [[CrossRef](#)]
28. Opere, A.O.; Waswa, R.; Mutua, F.M. Assessing the Impacts of Climate Change on Surface Water Resources Using WEAP Model in Narok County, Kenya. *Front. Water* **2022**, *3*, 789340. [[CrossRef](#)]
29. Du, L.; Li, X.; Yang, M.; Sivakumar, B.; Zhu, Y.; Pan, X.; Li, Z.; Sang, Y.-F. Assessment of spatiotemporal variability of precipitation using entropy indexes: A case study of Beijing, China. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 939–953. [[CrossRef](#)]
30. Morales-Acuña, E.; Linero-Cueto, J.R.; Canales, F.A. Assessment of Precipitation Variability and Trends Based on Satellite Estimations for a Heterogeneous Colombian Region. *Hydrology* **2021**, *8*, 128. [[CrossRef](#)]
31. Smith, L.B.; Liang, C.T. Technical solutions in reserve design for habitat conservation planning: A case study of the Sonoran Desert Conservation Plan. *Ecol. Soc. Am. Annu. Meet. Abstr.* **2002**, *87*, 271.
32. Cao, L.J.; Wei, Y.Z. Progress in Research on Homogenization of Climate Data. *Adv. Clim. Chang. Res.* **2012**, *3*, 59–67. [[CrossRef](#)]
33. Wang, X.L.; Wen, Q.H.; Wu, Y. Penalized Maximal t Test for Detecting Undocumented Mean Change in Climate Data Series. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 916–931. [[CrossRef](#)]
34. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P.P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503. [[CrossRef](#)]
35. Shen, Y.; Xiong, A.; Wang, Y.; Xie, P. Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res. Atmos.* **2010**, *115*, D02114. [[CrossRef](#)]
36. Song, X.-P.; Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Tyukavina, A.; Vermote, E.F.; Townshend, J.R. Global land change from 1982 to 2016. *Nature* **2018**, *563*, E26, Correction to *Nature* **2018**, *560*, 639–643. [[CrossRef](#)]
37. Pradhan, R.K.; Markonis, Y.; Godoy, M.R.V.; Villalba-Pradas, A.; Andreadis, K.M.; Nikolopoulos, E.I.; Papalexiou, S.M.; Rahim, A.; Tapiador, F.J.; Hanel, M. Review of GPM IMERG performance: A global perspective. *Remote Sens. Environ.* **2022**, *268*, 112754. [[CrossRef](#)]
38. Lakew, H.B.; Moges, S.A.; Asfaw, D.H. Hydrological Evaluation of Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin: A Case Study of Gilgel Abbay. *Hydrology* **2017**, *4*, 39. [[CrossRef](#)]
39. Hwang, S.-O.; Park, J.; Kim, H.M. Effect of hydrometeor species on very-short-range simulations of precipitation using ERAS. *Atmos. Res.* **2019**, *218*, 245–256. [[CrossRef](#)]
40. Urraca, R.; Huld, T.; Gracia-Amillo, A.; Javier Martinez-de-Pison, F.; Kaspar, F.; Sanz-Garcia, A. Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. *Sol. Energy* **2018**, *164*, 339–354. [[CrossRef](#)]
41. Beck, H.E.; Vergopolan, N.; Pan, M.; Levizzani, V.; van Dijk, A.I.J.M.; Weedon, G.P.; Brocca, L.; Pappenberger, F.; Huffman, G.J.; Wood, E.F. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6201–6217. [[CrossRef](#)]
42. Satgé, F.; Espinoza, R.; Zolá, R.P.; Roig, H.; Timouk, F.; Molina, J.; Garnier, J.; Calmant, S.; Seyler, F.; Bonnet, M.-P. Role of Climate Variability and Human Activity on Poopó Lake Droughts between 1990 and 2015 Assessed Using Remote Sensing Data. *Remote Sens.* **2017**, *9*, 218. [[CrossRef](#)]
43. Chen, L.; Dirmeyer, P.A. Impacts of Land-Use/Land-Cover Change on Afternoon Precipitation over North America. *J. Clim.* **2017**, *30*, 2121–2140. [[CrossRef](#)]
44. Martens, B.; Miralles, D.; Hans, L.; van der Schalie, R.; Jeu, R.; Fernández-Prieto, D.; Beck, H.; Dorigo, W.; Verhoest, N. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev. Discuss.* **2016**, *10*, 1903–1925. [[CrossRef](#)]
45. Anh Nguyet, D.; Kawasaki, A. Integrating biophysical and socio-economic factors for land-use and land-cover change projection in agricultural economic regions. *Ecol. Model.* **2017**, *344*, 29–37. [[CrossRef](#)]
46. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [[CrossRef](#)]
47. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
48. Legasa, M.N.; Manzanar, R.; Calviño, A.; Gutiérrez, J.M. A Posteriori Random Forests for Stochastic Downscaling of Precipitation by Predicting Probability Distributions. *Water Resour. Res.* **2022**, *58*, e2021WR030272. [[CrossRef](#)]
49. King, C.; Strumpf, E. Applying random forest in a health administrative data context: A conceptual guide. *Health Serv. Outcomes Res. Methodol.* **2022**, *22*, 96–117. [[CrossRef](#)]
50. Nicodemus, K.K.; Malley, J.D.; Strobl, C.; Ziegler, A. The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinform.* **2010**, *11*, 110. [[CrossRef](#)]
51. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random Forests for Classification in Ecology. *Ecology* **2007**, *88*, 2783–2792. [[CrossRef](#)] [[PubMed](#)]

52. Xiao, M.; Zhang, G.; Breikopf, P.; Villon, P.; Zhang, W. Extended Co-Kriging interpolation method based on multi-fidelity data. *Appl. Math. Comput.* **2018**, *323*, 120–131. [[CrossRef](#)]
53. Adhikary, S.K.; Muttill, N.; Yilmaz, A.G. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrol. Process.* **2017**, *31*, 2143–2161. [[CrossRef](#)]
54. Ghorbanpour, A.K.; Hessels, T.; Moghim, S.; Afshar, A. Comparison and assessment of spatial downscaling methods for enhancing the accuracy of satellite-based precipitation over Lake Urmia Basin. *J. Hydrol.* **2021**, *596*, 126055. [[CrossRef](#)]
55. Chen, Y.; Huang, J.; Sheng, S.; Mansaray, L.R.; Liu, Z.; Wu, H.; Wang, X. A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. *Remote Sens. Environ.* **2018**, *214*, 154–172. [[CrossRef](#)]
56. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [[CrossRef](#)]
57. Zhao, N. A Method for Merging Multi-Source Daily Satellite Precipitation Datasets and Gauge Observations over Poyang Lake Basin, China. *Remote Sens.* **2023**, *15*, 2407. [[CrossRef](#)]
58. Yang, Y.; Donohue, R.J.; McVicar, T.R. Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resour. Res.* **2016**, *52*, 8260–8276. [[CrossRef](#)]
59. Duan, Z.; Bastiaanssen, W.G.M. First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling-calibration procedure. *Remote Sens. Environ.* **2013**, *131*, 1–13. [[CrossRef](#)]
60. Liu, Z.; Xu, Z.; Charles, S.P.; Fu, G.; Liu, L. Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. *Int. J. Climatol.* **2011**, *31*, 2006–2020. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.