https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

Open Access Earth System Science Data Discussions

A new daily gridded precipitation dataset based on gauge observations across mainland China

Jingya Han, Chiyuan Miao*, Jiaojiao Gou, Haiyan Zheng, Qi Zhang, Xiaoying Guo

State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

* Corresponding author: Chiyuan Miao (miaocy@bnu.edu.cn)

Abstract

Freely accessible long-term precipitation estimates with fine spatiotemporal resolution and high-quality play essential roles in hydrologic, climatic, and numerical modeling applications. However, the existing daily gridded precipitation datasets over China either are constructed with insufficient gauge observations or neglect topographic effects and boundary effects on interpolation. Using daily observations from 2,839 gauges across China and nearby regions from 1961 to the present, this study compared eight different interpolation schemes that adjust the climatology based on monthly precipitation constraint and topographic characteristic correction, using an algorithm that combines the daily climatology field with a precipitation ratio field. Results from these eight interpolation schemes are cross validated using 45,992 high-density daily gauge observations from 2015 to 2019 over China. Of these eight schemes, the one with the best performance merges the Parameter-elevation Regression on Independent Slopes Model (PRISM) in the daily climatology field and interpolates station observations into the ratio field using an inverse distance weighting method. This scheme has a correlation coefficient of 0.78, a root-mean-square deviation of 8.8 mm/d, and a Kling-Gupta efficiency of 0.69 for comparisons between the 45,992 high-density gauge observations and the best interpolation scheme for the 0.1° latitude × longitude grid cells from 2015 to 2019. Therefore, this scheme is used to construct a new long-term gauge-based gridded precipitation dataset across mainland China (called

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

CHM_PR, as a member of the China Hydro-Meteorology dataset) with spatial resolutions of 0.5°/0.25°/0.1°. This precipitation dataset is expected to facilitate the advancement of drought monitoring, flood forecasting, and hydrological modeling. Free access to the dataset can be found at https://doi.org/10.6084/m9.figshare.21432123.v2 (Han and Miao, 2022).

1 Introduction

As one of the key components of the hydrological cycle, precipitation can influence the distribution of water resources (Rodell et al., 2018), sustain agriculture (Beck et al., 2020), replenish aquifers (Fischer and Knutti, 2016; Kucera et al., 2013), and enable economic prosperity (Trenberth et al., 2003; Kirschbaum et al., 2017). Each of the last three decades has been successively warmer at Earth's surface than any preceding decade since 1850 (IPCC, 2021). With an ever-warming climate, the Earth's water cycle has been amplified, resulting in frequent severe extreme precipitation events (Fischer and Knutti, 2016; Myhre et al., 2018). The intensification of water transport and exchanges between the atmosphere and surface is having profound impacts on the redistribution of water resources by moisture flux, which exaggerates the contrast between wet and dry meteorological regimes, seasons, and events (Allen and Ingram, 2002; Allan et al., 2020). Constructing a high-quality, long-term daily precipitation dataset is essential to current hydrometeorology research (Sun et al., 2018; Beck et al., 2019). However, due to the spatial heterogeneity and temporal variability of daily precipitation, it is challenging to derive accurate spatiotemporal patterns of daily precipitation.

The measurement of precipitation relies mainly on direct measurement using rain gauges, disdrometers, and radar and on indirect estimation using satellite systems (Shen et al., 2014; Beck et al., 2019; Sun et al., 2018). Among these approaches, rain-gauge observations are the most reliable and widely used tool for directly measuring precipitation. However, gauge observations reflect only point precipitation, and the

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

uneven distribution of gauges increases the limitations of gauge applications over a region. It is vital to interpolate these spatially irregular gauge observations to areal averages, since multiple scientific and operational applications (e.g., estimating local climate variables in data-sparse regions, monitoring climate change at the regional or global scale, and validating climate models with observations) require good-quality, high-spatiotemporal-resolution precipitation datasets (Haylock et al., 2008; Xie et al., 2007; Harris et al., 2020). Spatial interpolation methods are usually applied to convert irregular point observations to regional measurements (Ahrens, 2006), thus generating evenly gridded precipitation products that are widely used in hydrology and meteorology studies (Schamm et al., 2014; Golian et al., 2019).

In China, the original gauge observations used as the benchmark of various precipitation datasets mainly come from two suites of gauge observations provided by the China Meteorological Administration (CMA): ~700 benchmark stations and ~2,400 gauges comprising other ordinary national automatic weather stations (Shen and Xiong, 2016). Using observations from the former (the ~700 stations), a monthly precipitation dataset has been established over China that covers the period of 1901–2017 (Peng et al., 2019). To reach a higher temporal resolution, a daily gridded precipitation dataset has been built across China using the same raw precipitation data with a time span from 1961 to 2019 (Qin et al., 2022). Through a fusion of remote sensing products and reanalysis datasets into *in situ* station data, the China Meteorological Forcing Dataset (CMFD) has been produced to serve as a high-resolution (three hours, 0.1° × 0.1°) input forcing dataset for hydrological and ecosystem models beginning in 1979 (He et al., 2020). Generally, the quantitative accuracy of a gauge-based dataset can be improved by enhancing the density of gauge observations (Merino et al., 2021; Hofstra and New, 2009). Xie et al. (2007) developed a widely used gauge-based analysis of daily precipitation over East Asia (EA05) with a collection of daily precipitation observations from over 700 stations from CMA and about 1,000 hydrological station observations from the Chinese Yellow River Conservation Commission. Using observations from approximately 2,400 gauges from 1961 to the present, Wu and Gao (2013) created a

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

daily gridded dataset with a resolution of 0.25°×0.25° over China (CN05.1), and Zhao et al. (2014) constructed the second version of 0.5° × 0.5° gridded daily precipitation dataset over China (CMA V2.0). Further accounting for topographic effects including elevation, slope, proximity to coastlines, and the locations of temperature inversions, Shen et al. (2010) developed the China Gauge-based Daily Precipitation Analysis (CGDPA) with spatial resolutions of 0.5° × 0.5° and 0.25° × 0.25° using a topographic correction algorithm. The aforementioned datasets only involve gauges inside China's boundaries, except for EA05, which covers the East Asia domain. This limitation can lead to boundary effects such that grid cells near the boundaries suffer worse positioning accuracy relative to interior grid cells (Ahrens, 2006). In addition, different interpolation algorithms can produce different results even with the same inputs. Comparing the performance of different interpolation techniques is crucial to determining the optimal interpolation method.

Given these limitations and the important role these datasets play in many applications, it is of great urgency to establish long-term, continuously updated daily precipitation series with multiple spatial resolutions that are free to use. Therefore, this study aims to construct a long-term (from 1961 to the present) daily precipitation dataset with different spatial resolutions (0.5° × 0.5°, 0.25° × 0.25°, and 0.1° × 0.1°) based on 2,839 gauge observations in and around China (2,419 gauges over mainland China and 420 gauges from nearby regions). Eight interpolation schemes are proposed and evaluated by cross validation using 45,992 gauge observations for the period of 2015–2019 over China. Finally, a new gridded precipitation dataset across mainland China (a member of the China Hydro-Meteorology datasets, hereinafter called CHM_PR) covering the period 1961–2021 with spatial resolutions of 0.5°/0.25°/0.1° is provided publicly for applications and will be updated yearly.

2 Data

2.1 Raw gauge data used for interpolation

Daily rain gauge datasets (from 1961 to the present) from 2,419 stations across

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

mainland China and 420 stations surrounding China's boundaries are collected from China Meteorological Administration (CMA, http://data.cma.cn) and Global Historical Climatology Network-Daily Version 3 (GHCND, https://www.ncei.noaa.gov), respectively (Figure 1a). In this study, the CMA gauge dataset is available to use for mainland China (data in Hong Kong, Macao, and Taiwan are currently not accessible for use). The coverage of stations is relatively sparse over northwestern China and the Tibetan Plateau compared with eastern and southern China. The daily precipitation is the accumulated precipitation amount between 20:00 and 20:00 (local time in Beijing). This dataset has been subjected to strict quality controls, including (1) extreme values check, (2) internal consistency check of daily values, (3) spatial and temporal consistency check, and (4) manual verification (Zhang et al., 2020). The gauge observations from neighboring countries come from the GHCND dataset, which contains records from over 80,000 stations in 180 countries and territories. Quality controls are routinely applied to assure the basic consistency of the dataset (Menne et al., 2012). Stations with less than 5% missing calendar days in an individual year are used for interpolation. Changes in the number of stations over time are shown in Figure 1c.

2.2 High-density gauge observations used for cross validation

High-density daily observations from nearly 68,000 automatic weather stations with a range of 2015–2019 across China are provided by the National Meteorological Information Center of CMA (Li et al., 2018). After removing stations with a missing rate of over 20% for the period of 2015–2019, 45,992 good-quality stations are left (Figure 1b).

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

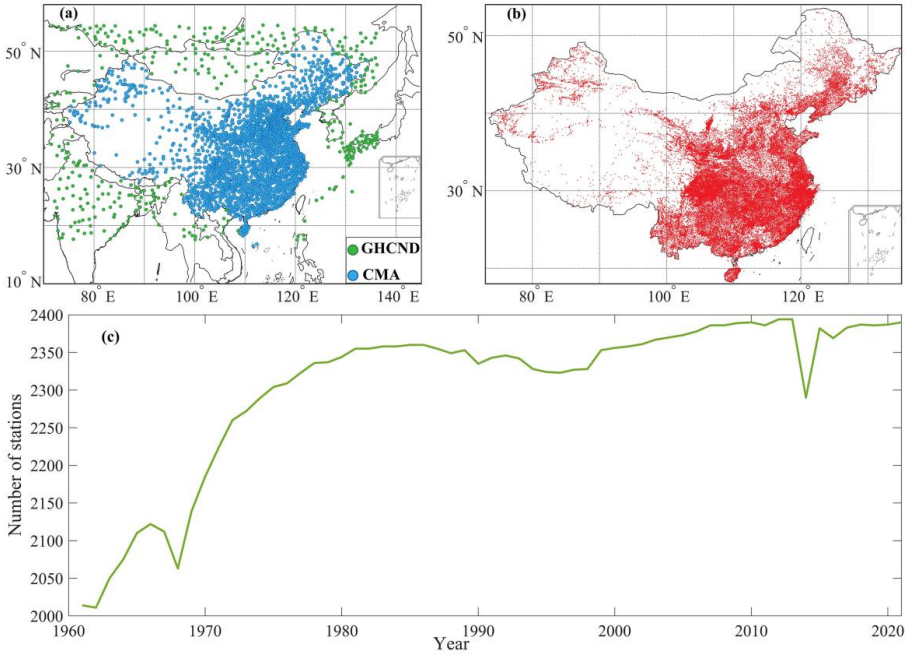

Figure 1. (a) Distribution of 2,839 stations used in interpolation and (b) 45,992 stations used for cross validation. (c) Quality-controlled number of stations for interpolation over time.

2.3 SRTM-DEM

The 3 arc second (90 m resolution) digital elevation model (DEM) applied in this study is collected from the Shuttle Radar Topography Mission (SRTM) (https://urs.earthdata.nasa.gov/). SRTM uses dual radar antennas to acquire interferometric radar data and processes digital topographic data (Farr et al., 2007). We resampled the SRTM-DEM into 0.05° × 0.05° grid cells.

2.4 PRISM

The monthly climatology generated by Parameter-elevation Regression on Independent Slopes Model (PRISM) is used for the monthly precipitation constraint and topographic characteristic correction of the daily climatology field (https://prism.oregonstate.edu/). PRISM incorporates local climate-elevation relationships, topographic facets,

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

proximity to coastlines, and several measures of terrain complexity, and it is the most widely used climatology dataset in the world (Daly et al., 1994; Daly et al., 2002). The original spatial resolution is 0.04° × 0.04° for the monthly climatology of PRISM between 1961 and 1990; we regridded the spatial resolution into 0.05° × 0.05° grid cells for the following climatology adjustment.

3 Methodology

3.1 Interpolation scheme

Due to the high spatial variability of precipitation relative to other climate variables, directly interpolating the daily rain-gauge observations into grid cells could produce a dataset with misleading daily precipitation characteristics (Xie et al., 2007; Chen et al., 2002; Shen et al., 2010). To eliminate such interpolation errors, the overall strategy for establishing a daily gridded precipitation dataset is to construct a relatively continuous daily climatology field (Shen et al., 2010). Then, we would build an intermediate field of the interpolated variable based on this daily climatology field, such as a daily precipitation anomalies field or a field of the ratio between daily precipitation and daily climatology. Previous studies have demonstrated that interpolating the ratio (between daily precipitation and daily climatology) yields a better performance than interpolating anomalies for constructing daily gridded precipitation (Xie et al., 2007; Yatagai et al., 2012; Di Luzio et al., 2008). Therefore, "daily climatology field (*Cd*) × field of the ratio between daily precipitation and daily climatology (*P/Cd*)" is employed as the interpolation scheme for constructing the new gridded precipitation dataset in this study (Figure 2), as developed by Xie et al. (2007). Figure 3 shows a flowchart of the gridding analysis system.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

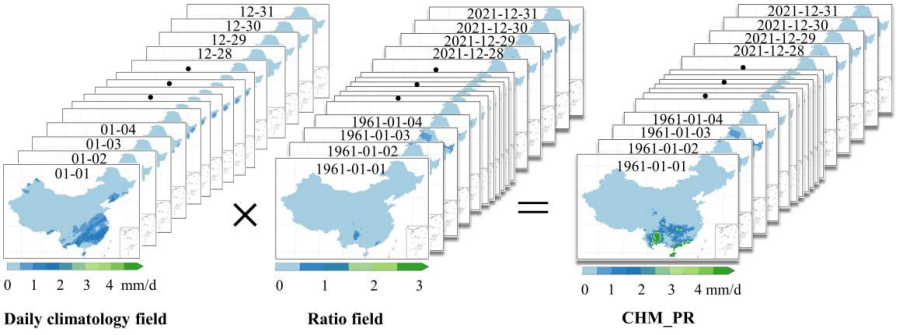

Figure 2. Interpolation strategy for generating daily gridded precipitation dataset (CHM_PR) in this study.

190

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

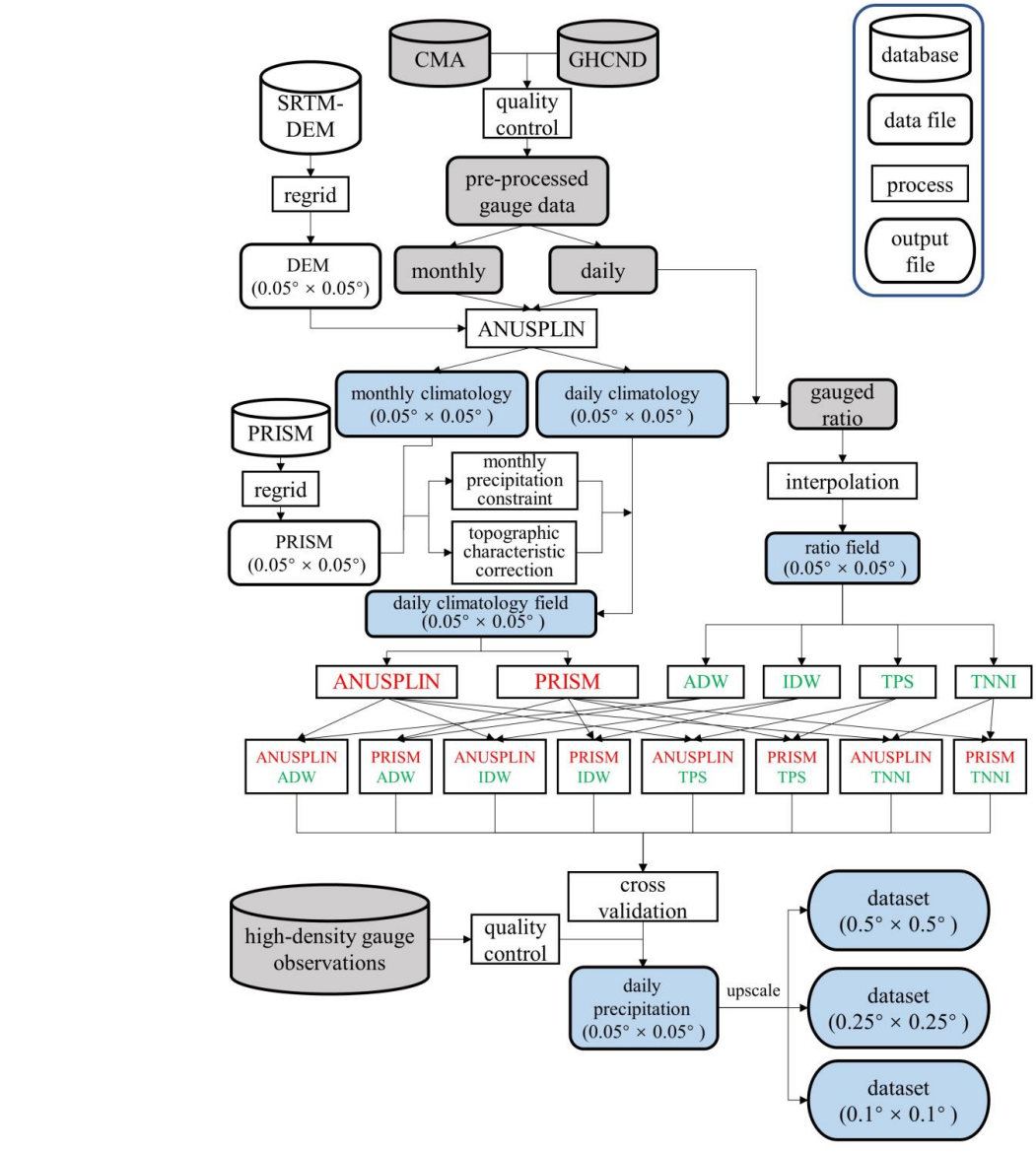

Figure 3. Flowchart of the gridding analysis system. Gray shading represents gauge data; blue shading represents gridded data. Approaches (using ANUSPLIN software or using PRISM data) used for generating daily climatology fields are marked in red, and interpolation methods (angular-distance weighting (ADW), inverse distance weighting (IDW), thin plate spline (TPS), and triangulation-based nearest neighbor interpolation (TNNI)) used for producing ratio fields are marked in green.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

3.2 Building the daily climatology field (*Cd*)

The raw daily climatology is first calculated for all stations. The definition of raw daily climatology is the Fourier-truncated 30-year mean daily precipitation series during the period of 1971–2000 for the 365 calendar days (Figure 4). We use Fourier truncation to remove the high-frequency noise of the 30-year mean daily precipitation series across stations and retain the accumulation of the first six harmonic components as the raw daily climatology (Xie et al., 2007). The raw 0.05° × 0.05° gridded daily climatology field (Cd_0) is then interpolated from stations with SRTM-DEM as a covariate using ANUSPLIN software (Hutchinson and Xu, 2004). To eliminate the systematic bias of the raw daily climatology field on the monthly climatology field (Cm), the monthly accumulation of the daily climatology field is then constrained by the monthly climatology field. We compare two types of monthly climatology fields to determine which adjusts better for the systematic bias: 1) an ANUSPLIN-type monthly climatology field or 2) a PRISM-type monthly climatology field. The ANUSPLIN-type monthly climatology field is produced by interpolating monthly precipitation climatology (1971–2000) of stations to the 0.05° latitude × longitude grids with a covariate of SRTM-DEM using ANUSPLIN software. The 0.05° × 0.05° regridded monthly climatology of PRISM is used as the PRISM-type monthly climatology field.

The climatology adjustment steps for one grid cell are as follows:

1) Calculate Cd_{0_m} $(m = 1, 2, 3, \dots, 12)$, which is the monthly total of the raw gridded daily climatology field for the 12 calendar months.
2) Match the monthly total series derived using the raw gridded daily climatology field to the monthly climatology field month by month.
3) Compute the scaling factor for the individual calendar day (SF_j $(j = 1, 2, 3, \dots, 365)$) of the raw daily climatology field to the monthly climatology field:

$$SF_j = \frac{Cm}{w_{m-1} \times Cd_{0\_m-1} + w_m \times Cd_{0\_m} + w_{m+1} \times Cd_{0\_m+1}} \quad (1)$$

$$(m = 1, 2, 3, \dots, 11, 12;\ j = 1, 2, 3, \dots, 365;$$
$$m \text{ is the corresponding month for day } j)$$

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

where Cm is the monthly climatology field for month m; Cd_{0_m-1}, Cd_{0_m}, and Cd_{0_m+1} are the monthly total of month $m-1$, m, and $m+1$ respectively, which are calculated from the raw gridded daily climatology field; w_{m-1}, w_m, and w_{m+1} are the corresponding weights for month $m-1$, m, and $m+1$, respectively, which are inversely proportional to the interval between the calendar day *j* and the center of the month (Xie et al., 2007). Note that the weight w_{m-1} is zero when *m* = 1, and so is the weight w_{m+1} when *m* = 12.

4) The adjusted daily climatology field (Cd) is defined as

$$Cd = Cd_0 \times SF \quad (2)$$

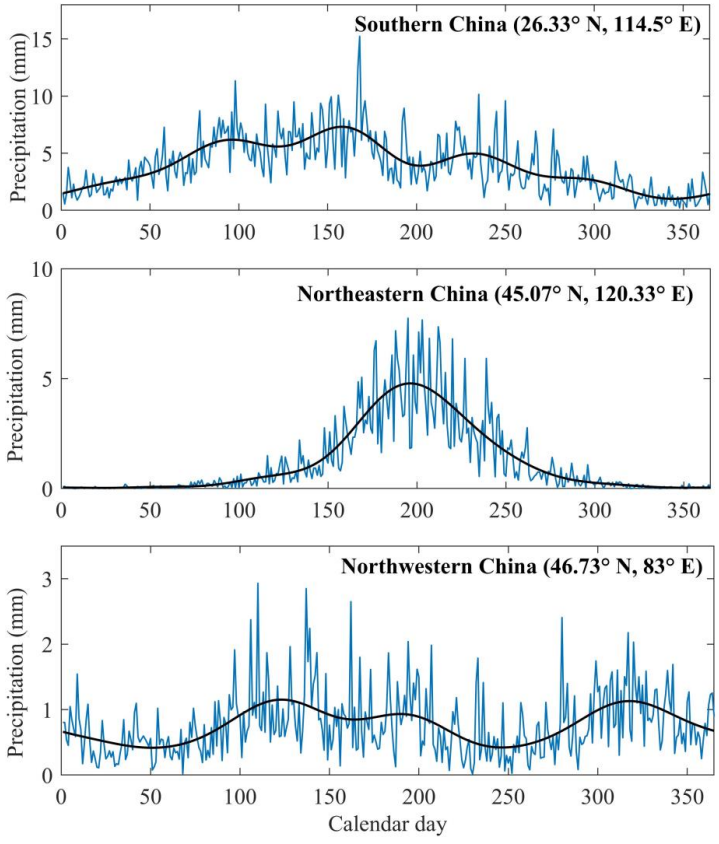

Figure 4. Time series of 30-year (1971–2000) mean daily precipitation (blue lines) and daily climatology derived by Fourier truncation (black lines) across three randomly selected stations in Southern China (top), Northeastern China (middle), and Northwestern China (bottom).

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

3.3 Constructing a field of the ratio between daily precipitation and daily climatology (*P/Cd*)

The ratio of daily precipitation to the spatiotemporally corresponding daily climatology field is calculated across stations. To compare the performances of different interpolation methods, four widely used interpolation approaches for precipitation are adopted to generate the 0.05° × 0.05° gridded ratio field from station data. The four interpolation methods used in this study are angular-distance weighting (ADW) (Shepard, 1968; Caesar et al., 2006), inverse distance weighting (IDW) (Shepard, 1984; Eischeid et al., 2000), thin plate spline (TPS) (Hutchinson, 1995; Camera et al., 2014), and triangulation-based nearest neighbor interpolation (TNNI) (Thiessen, 1911; Sibson, 1978). A brief overview of the main characteristics of the four methods is given below.

3.3.1 ADW

The ADW interpolation method we used in this study is the modified Shepard's algorithm, which introduces the concept of correlation decay distance (CDD), also called correlation length scale or decorrelation length (Shepard, 1984; Dunn et al., 2020). The CDD is defined as the distance at which the correlation between one station and all other stations decays below 1/*e*, approximately corresponding to the significance level of 0.05 for the correlation within large samples (Jones et al., 1997; Harris et al., 2020). The number of stations for interpolating the target grid cell is well constrained by the CDD, thus improving the interpolation precision (New et al., 2000; Mitchell and Jones, 2005; Hofstra and New, 2009).

For every station, correlations (r) and distances (x) with the other 2,838 stations are shown in Figure 5, and the ordinary least squares method is used to fit an exponential decay function:

$$r = e^{-x/CDD} \tag{3}$$

The estimated CDD is 244.7 km (95% confidence interval: 244.5–244.8 km) at the 0.05 significance level.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

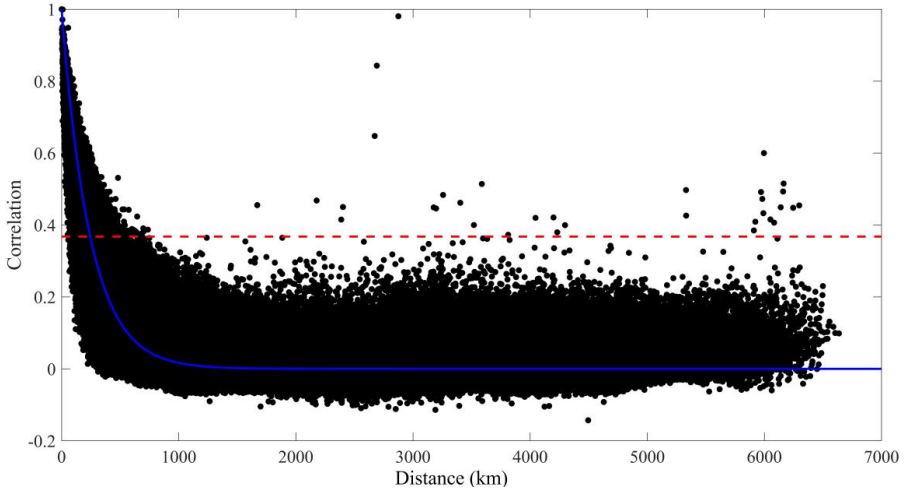

Figure 5. Estimation of correlation decay distance (CDD) for daily precipitation series across stations in the interpolated domain. Black points show the pair of distance-correlation for each station. The blue line is the exponential curve fitted to the data by ordinary least squares. The red dashed line marks where correlation equals 1/*e*.

The ADW method accounts for the importance of both distance and the isolation of stations in interpolation (New et al., 2000). Only stations within the range of the CDD for the center of the target grid cell (L) are involved in the interpolation. The weight for each involved station (i) is a function of the distance weight (D_i) and angular weight (A_i):

$$D_i = \left(e^{-x/CDD}\right)^n \tag{4}$$

where x is the distance between station i and the center of target grid cell L; n is a constant and usually set to 4, in accordance with previous studies (Harris et al., 2020; Dunn et al., 2020; Efthymiadis et al., 2006).

$$A_i = 1 + \frac{\sum_k D_k[1-cos(\theta_k-\theta_i)]}{\sum_k D_k} \left(i \neq k\right) \tag{5}$$

where k represents the surrounding stations relative to station i; D_k is the distance weight for the surrounding stations k; θ_i and θ_k are the angles relative to the north of the center of target grid cell L for station i and the surrounding stations k. Finally, the

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

weights for all contributing stations are standardized to sum to 1.0.

The angular-distance weight (W_i) equals

$$W_i = D_i \times A_i \tag{6}$$

295

3.3.2 IDW

The concept of CDD is introduced to the IDW method for interpolation. We set CDD1 = 244.7 km, which represents the boundary at which the search radius has a good correlation between stations and the target grid cell. The minimum distance satisfying
300 the condition that at least three stations are included in the search radius is around 1,336 km. Therefore, CDD2 = 1,336 km is used as a second choice of search radius if there are not at least three stations located within the range of CDD1. We employ S_0 representing the target grid cell, i representing the surrounding stations, $y(s_i)$ denoting the station observations, and d_{0i} denoting the distance between S_0 and i. The estimation
305 of daily precipitation in grid cell S_0 is $\hat{y}(s_0)$:

$$\hat{y}(s_0) = \sum_{i=1}^{n} \lambda_i y(s_i) \tag{7}$$

$$\lambda_i = d_{0_i}^{-\alpha} / \sum_{i}^{n} d_{0i}^{-\alpha} \tag{8}$$

$$\sum_{i}^{n} \lambda_i = 1 \tag{9}$$

where λ_i is the distance weight for the interpolated stations; n is the number of stations
310 involved in the interpolation; and the parameter α is the geometric form of weight. A high magnitude of α represents a strong correlation decay per unit of distance, thus the stations that are close to the target grid cell would be assigned a greater weight (Lu and Wong, 2008). In this study, we use $\alpha = 2$, which is widely used for the IDW method to show the Euclidean distance between the centers of grid cells and the interpolated
315 stations (Ly et al., 2013; Ahrens, 2006).

3.3.3 TPS

Splines are developed with the use of spatial covariate functions (Wahba and Wendelberger, 1980; Camera et al., 2014). TPS regards the spatial distribution as simply

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

320 a function of observations, and there is no need to first estimate a covariate function (Hutchinson, 1995). Thus, the interpolation precision is improved. The TPS function offers a trade-off between data fidelity and smoothness of fit (Tait et al., 2006; Haylock et al., 2008). The degree of smoothing is determined by minimizing generalized cross validation (Hutchinson, 1998).

325

3.3.4 TNNI

The Delaunay triangulation net is built by location of vertices of triangulations with rainfall amount as the third dimension (Delaunay, 1934). Within finite point sets, the Delaunay triangulation is demonstrated to be the only optimal method (Sibson, 1978).
330 This uniqueness guarantees the stability of interpolation. TNNI estimates the value of the target grid cell as a value of the nearest sample, and this can reflect the characteristic of regional precipitation (Vivoni Enrique et al., 2004).

3.4 Eight combination schemes for the daily climatology field and ratio field

335 The two types of daily climatology fields and four types of ratio fields constitute eight combination schemes of interpolation strategies (Table 1). We compared the performances of the eight combination schemes to choose the optimal scheme to construct the dataset.

340 Table 1 An overview of eight combination schemes for daily climatology field and ratio field

No.	Scheme name	Interpolation method for raw daily climatology field	Monthly climatology field type	Interpolation method used for ratio field
1	ANUSPLIN + ADW	ANUSPLIN	ANUSPLIN	ADW
2	PRISM + ADW	ANUSPLIN	PRISM	ADW
3	ANUSPLIN + IDW	ANUSPLIN	ANUSPLIN	IDW

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

4	PRISM + IDW	ANUSPLIN	PRISM	IDW
5	ANUSPLIN + TPS	ANUSPLIN	ANUSPLIN	TPS
6	PRISM + TPS	ANUSPLIN	PRISM	TPS
7	ANUSPLIN + TNNI	ANUSPLIN	ANUSPLIN	TNNI
8	PRISM + TNNI	ANUSPLIN	PRISM	TNNI

3.5 Cross validation

To improve the cross-validation efficiency, 45,992 high-density gauge observations over China were used to evaluate the eight interpolation schemes with a spatial resolution of 0.1° × 0.1°. The cross-validation steps are as follows:

1) Remove the 2,839 gauge stations used for interpolation from the 45,992 stations; the observations of the remaining stations are employed as the "true values".

2) Distribute the remaining stations into the corresponding 0.1° × 0.1° grid cells according to their longitudes and latitudes.

3) Apply the cross validation only for the grid cells where stations are located. The average of station observations is calculated as the validation value for the 0.1° × 0.1° grid cells where multiple stations are located. The correlation coefficient (CC), root-mean-square error (RMSE), and Kling-Gupta efficiency (KGE) between validation value (V_n) and estimation value (Y_n) are used as evaluation indicators for cross validation:

$$\mathrm{CC} = \frac{\frac{1}{N}\sum_{n=1}^{N}(V_n-\bar{V})(Y_n-\bar{Y})}{\sigma_V \sigma_Y} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(V_n - Y_n)^2} \tag{11}$$

$$\mathrm{KGE} = 1 - \sqrt{(CC-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{12}$$

$$\alpha = \frac{\sigma_Y}{\sigma_V}, \beta = \frac{\bar{Y}}{\bar{V}} \tag{13}$$

where N is the length of the daily precipitation series for 2015–2019; σ_V and $\bar{V}$ are the standard deviation and mean value for the validated daily precipitation series, respectively; σ_Y and $\bar{Y}$ are the standard deviation and mean value for the estimated daily precipitation series, respectively; α is a variability bias term, and β is a measure of mean bias (Gupta et al., 2009).

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

4 Results and discussion

4.1 Optimal scheme derived by cross validation

Generally, the performances of CC, RMSE, and KGE in southeast China are all better than those in northwest China and the Tibetan Plateau, where the density of stations is relatively sparse (Figures 6, 7, 8). The interpolated grid values are closer to gauge observations in simple terrain (e.g., the North China Plain) compared with complex terrain (e.g., the Yungui Plateau, the Loess Plateau, and the Tibetan Plateau). In terms of CC, the performances of scheme 3 (ANUSPLIN + IDW) and scheme 4 (PRISM + IDW) are best among the eight schemes, both with CC values of 0.78. The CC values for scheme 7 (ANUSPLIN + TNNI; 0.76) and scheme 8 (PRISM + TNNI; 0.76) are a little smaller than those for schemes 3 and 4. Scheme 1 (ANUSPLIN + ADW) and scheme 2 (PRISM + ADW) share the same CC value of 0.71. The lowest CC value (0.63) is calculated for scheme 5 (ANUSPLIN + TPS) and scheme 6 (PRISM + TPS). For the RMSE values, the order of the eight interpolation schemes is as follows: scheme 4 (8.8 mm/d) < scheme 3 (8.83 mm/d) < scheme 1 (10.09 mm/d) = scheme 2 (10.09 mm/d) < scheme 7 (10.14 mm/d) < scheme 8 (10.23 mm/d) < scheme 5 (11.16 mm/d) < scheme 6 (11.17 mm/d). The results for the comprehensive index, KGE, combining the characteristics of correlation, variability bias, and mean bias show that using scheme 4 would obtain the best interpolation results, at a KGE value of 0.69 across China. Schemes 3, 7, and 8 perform slightly worse than scheme 4, all with the same value of 0.68. The KGE values for scheme 1 (0.56) and scheme 2 (0.57) are a little worse than those listed above. Using scheme 5 and scheme 6 results in the worst performance based on KGE values, which are about 0.49 and 0.5, respectively. Overall, using scheme 4, which applies PRISM monthly climatology to adjust the daily climatology field, combined with an IDW-interpolated ratio field, provides the optimal performance among these validation indices. Therefore, the optimal scheme (PRISM + IDW) is used to construct the new 61-yr CHM_PR dataset across mainland China.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

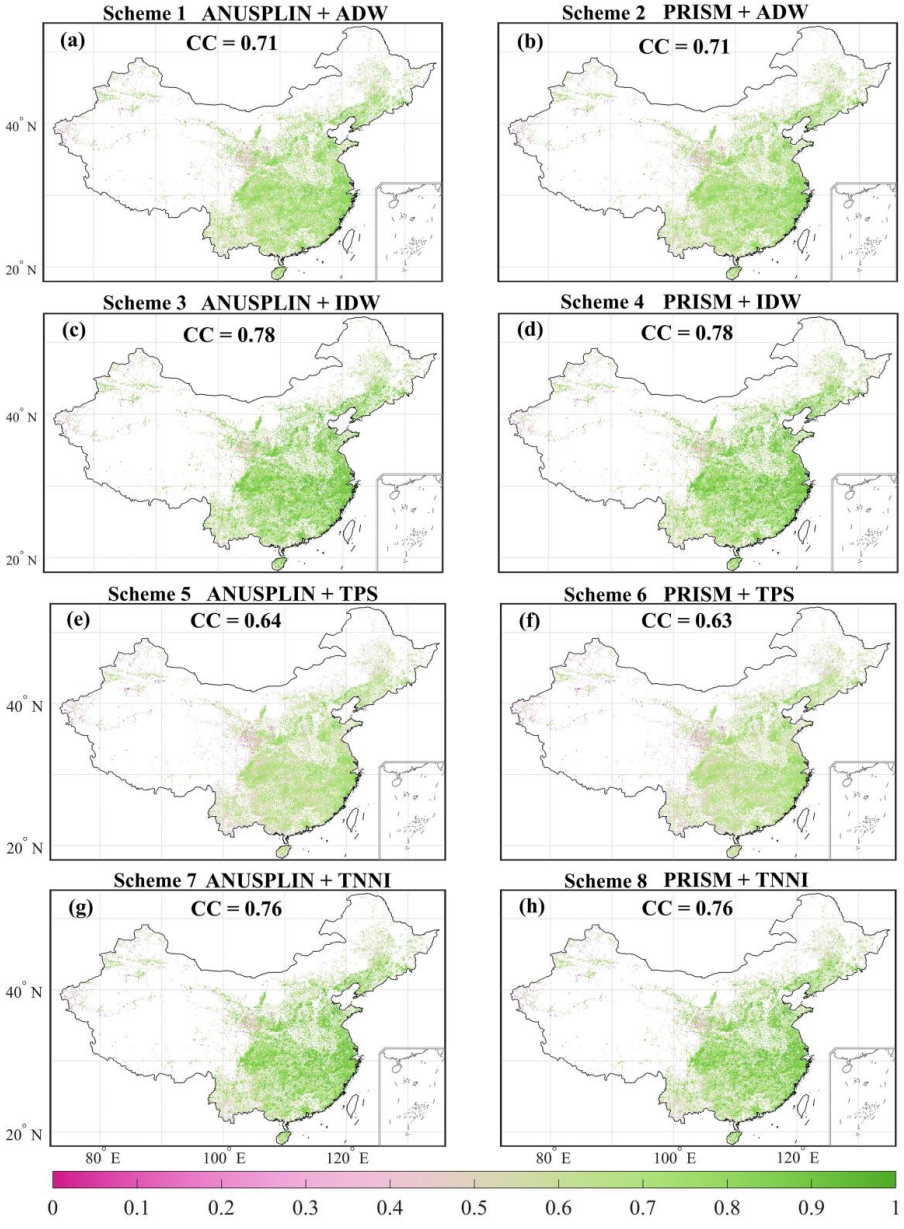

Figure 6. (a–h) Spatial pattern of the correlation coefficient (CC) for eight combination schemes. Numbers in each subplot represent the median of CC values across all grid cells involved in cross validation.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

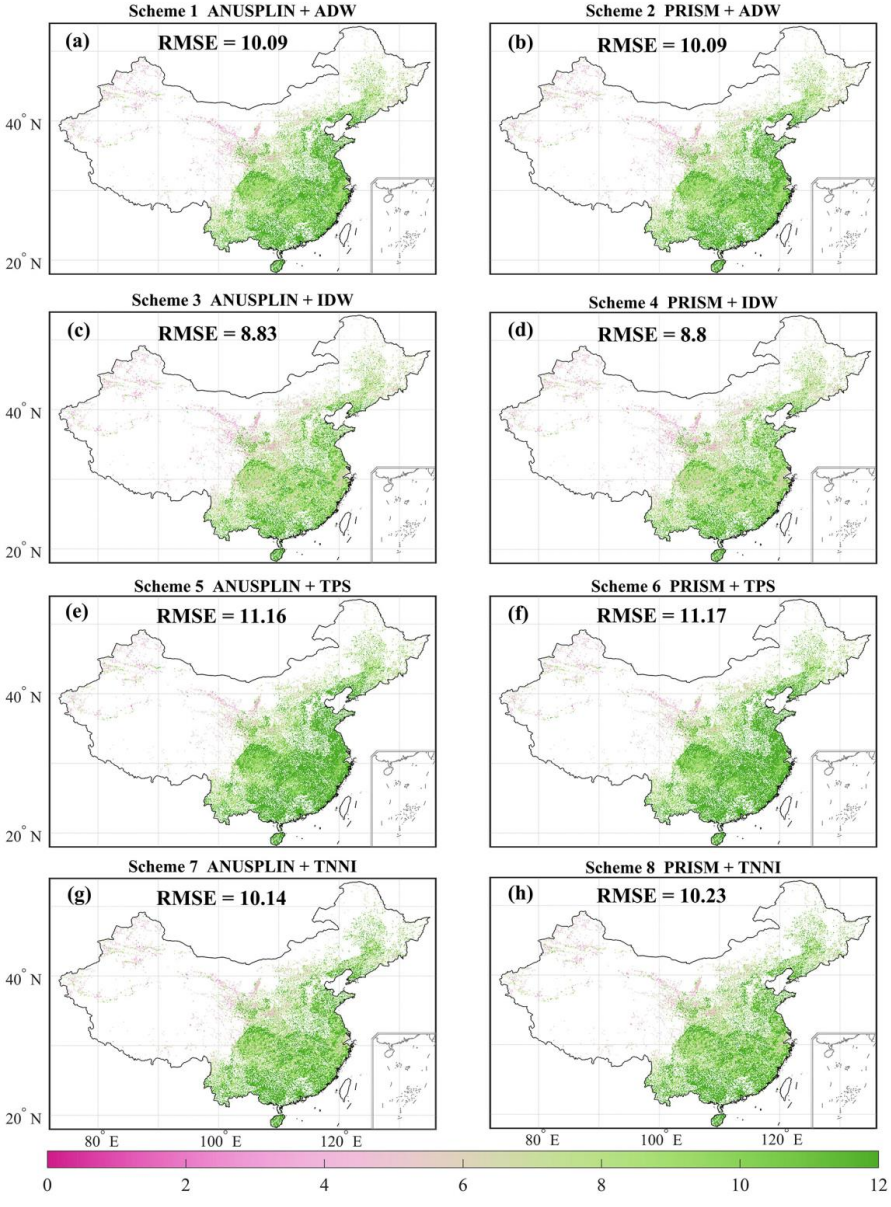

Figure 7. The same as Figure 6, but for root-mean-square error (RMSE). The unit is mm/d.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

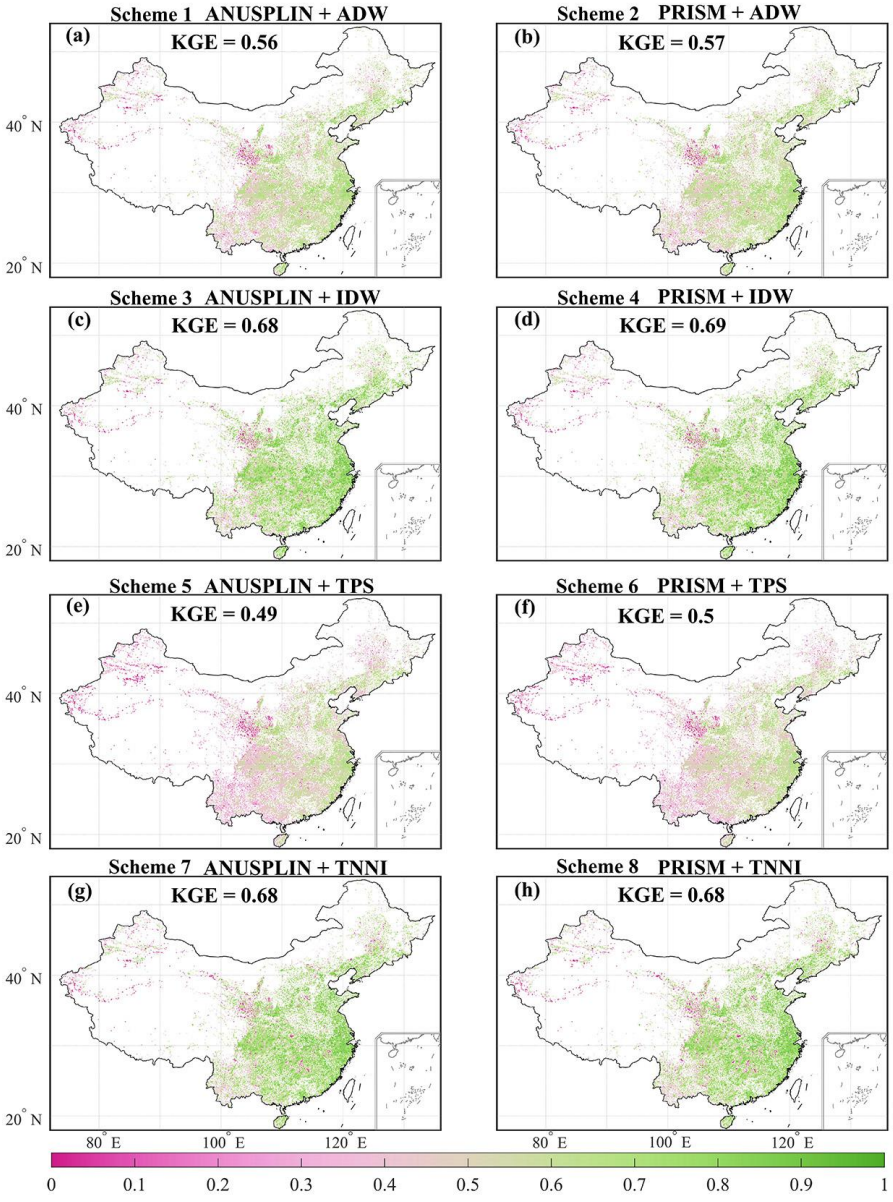

Figure 8. The same as Figure 6, but for Kling-Gupta efficiency (KGE).

4.2 Comparison with other gauge-based datasets

We compared the performance of the CHM_PR dataset with other three datasets—CGDPA (Shen et al., 2010), CN05.1 (Wu and Gao, 2013), and CMA V2.0 (Zhao et al., 2014) (Figures 9–12). To derive a uniform time span among the different datasets,

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

monthly precipitation series for the period of January 1, 2008, to December 31, 2015, were calculated in the CHM_PR, CGDPA, CN05.1, and CMA V2.0 datasets. Results show that the temporal pattern of the monthly precipitation series is generally consistent among different datasets, with a maximum bias of 5 mm/month for the dry (December-January-February) and wet (June-July-August) seasons (Figure 9). The average annual wet-day (>1 mm/day) precipitation amount and frequency between 2008 and 2015 for different datasets share similar spatial patterns, with a general decrease from southeastern China to northwestern China (Figures 10, 11). The median differences in the multi-year annual wet-day precipitation amount across grid cells for CHM_PR − CGDPA, CHM_PR − CN05.1, and CHM_PR − CMA V2.0 are −24.79 mm/yr, −7.43 mm/yr, and 13.87 mm/yr, respectively. The multi-year annual wet-day frequency is higher in CGDPA and CMA V2.0 than in the CHM_PR dataset, with median differences of 6.38 days/yr and 3.25 days/yr across grid cells, respectively. The mean annual wet-day frequency in CN05.1 is 9.63 days/yr less than that in the CHM_PR dataset. The broad features of mean annual maximum 1-day precipitation amount (Rx1day) are comparable among the four datasets from 2008 to 2015 (Figure 12). The 8-year average Rx1day values across China are 48.19 mm/day, 35.29 mm/day, 39.72 mm/day, and 40.19 mm/day for the CGDPA, CN05.1, CMA V2.0, and CHM_PR datasets, respectively. Generally, the difference of spatial pattern among different datasets is a combined effect of the gauge density involved and whether or not orographic effects are considered in the interpolation algorithm. The spatial patterns of mean precipitation (Figures 10, 11) and extreme precipitation (Figure 12) agree well among the different datasets in eastern and southern China, where gauge density is relatively high. This indicates the density of input gauges could be a dominant factor affecting interpolation output (Morrissey et al., 1995). Agreement in spatial patterns of mean and extreme precipitation is poorer in northwestern China and the Tibetan Plateau, which is driven by interpolation algorithms. In particular, the heavy precipitation in the southern Tibetan Plateau is well captured by the CGDPA and CHM_PR datasets, but the CN05.1 and CMA V2.0 datasets failed to capture it. This suggests the importance of orographic effects and boundary effects in interpolation processes, because heavier

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

rainfall appears over mountainous regions than nearby plains.

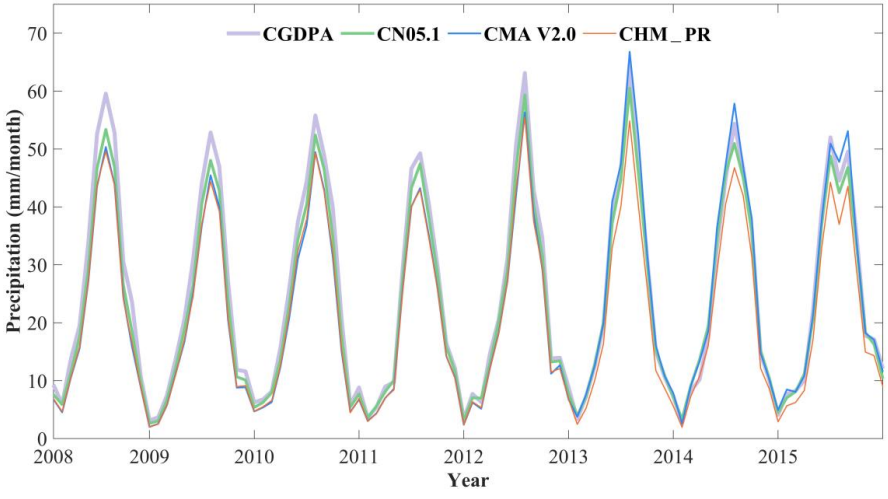

Figure 9. Monthly precipitation series from January 1, 2008, to December 31, 2015, for the CGDPA, CN05.1, CMA V2.0, and CHM_PR datasets.

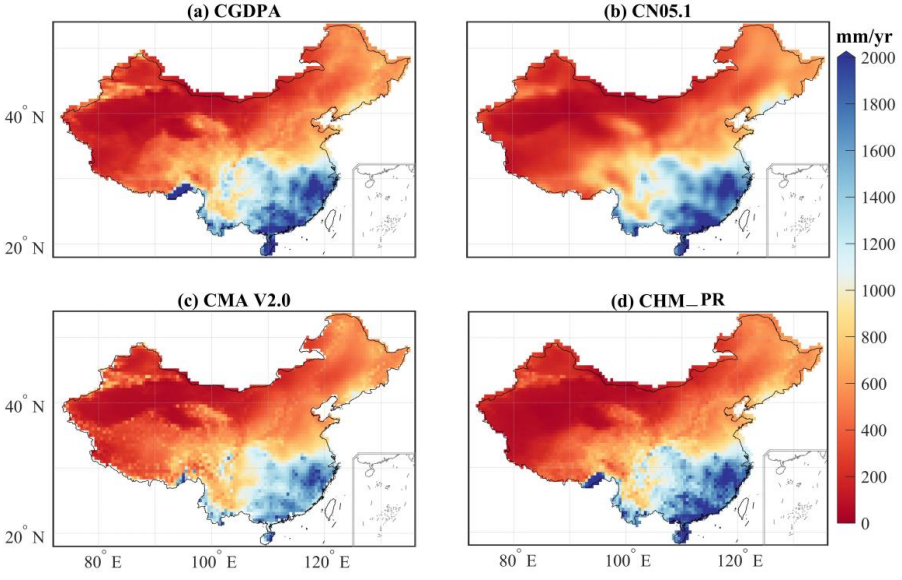

Figure 10. Spatial pattern of average annual wet-day (>1 mm/day) precipitation amount during the period of 2008 to 2015 for the (a) CGDPA, (b) CN05.1, (c) CMA V2.0, and (d) CHM_PR datasets.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

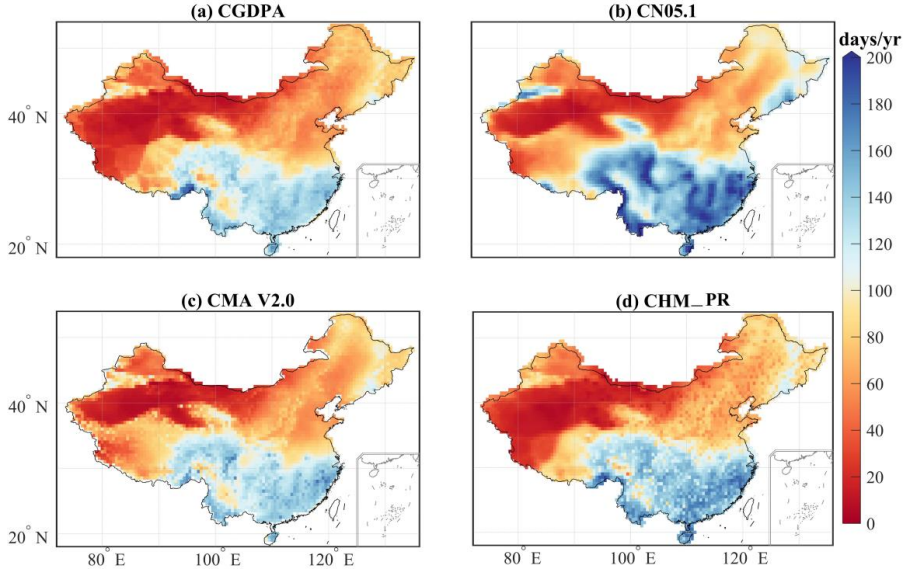

Figure 11. Spatial pattern of mean annual wet-day (>1 mm/day) frequency from 2008 to 2015 for the (a) CGDPA, (b) CN05.1, (c) CMA V2.0, and (d) CHM_PR datasets.

455

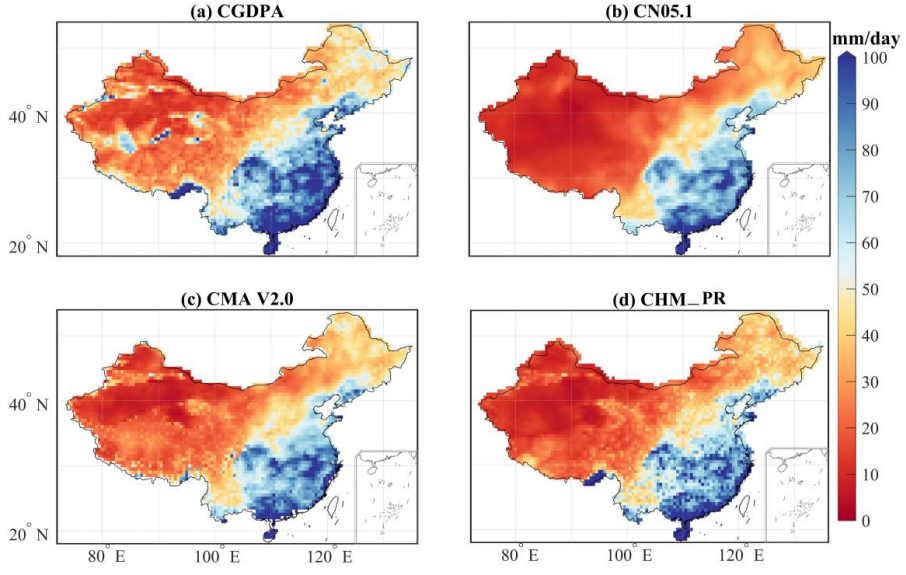

Figure 12. Spatial pattern of mean Rx1day from 2008 to 2015 for the (a) CGDPA, (b) CN05.1, (c) CMA V2.0, and (d) CHM_PR datasets.

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

460 **5 Data availability**

This high-resolution long-term gauge-based daily precipitation dataset covers the period of 1961–2021, and it will continue to be updated annually. It contains data for three spatial resolutions: 0.1° × 0.1° covering the domain of 17.9°N–54.1°N, 71.9°E–136.1°E, 0.25° × 0.25° covering the domain of 17.75°N–54.25°N, 71.75°E–136.25°E,
465 and 0.5° × 0.5° covering the domain of 17.5°N–54.5°N, 71.5°E–136.5°E. The NetCDF-formatted output files of the CHM_PR dataset are freely accessible at https://doi.org/10.6084/m9.figshare.21432123.v2 (Han and Miao, 2022).

6 Conclusions

470 Based on a recent 61-yr time series of daily observations from 2,839 gauges across China and the surrounding areas, this study compares eight different interpolation schemes that use an algorithm combining the daily climatology field with a precipitation ratio field. A cross-validation method is used to evaluate the eight interpolation schemes using 45,992 high-density gauge observations over China. The
475 results indicate that the best-performing scheme is scheme 4, which combines a monthly precipitation constraint and correction for topographic characteristics with the daily climatology field and interpolates station observations of precipitation ratio into grid cells using an inverse distance weighting method. The CC, RMSE, and KGE values for the optimal interpolation scheme in comparison with the high-density gauge
480 observations used for cross validation are 0.78, 8.8 mm/d, and 0.69, respectively. Using the optimal interpolation scheme, we have constructed a new gridded precipitation dataset (CHM_PR) covering mainland China with a daily temporal resolution and at multiple spatial resolutions (0.1° × 0.1°, 0.25° × 0.25°, and 0.5° × 0.5°) for the period of 1961–2021. The CHM_PR dataset shows reliable quality compared with the other
485 available precipitation products.

Author Contributions

JH and CM contributed to designing the research; JH implemented the research and wrote original draft; CM and JG supervised the research; all co-authors revised the

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

manuscript and contributed to the writing.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the State Key Laboratory of Earth Surface Processes and Resource Ecology (2022-ZD-03), the National Natural Science Foundation of China (No. 42041006), and the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (No.2019QZKK0405). We are also grateful to the National Meteorological Information Center of the China Meteorological Administration (NMIC, http://data.cma.cn) for providing the observed climate data.

References

AghaKouchak, A., Mehran, A., Norouzi, H., and Behrangi, A.: Systematic and random error components in satellite precipitation data sets, Geophysical Research Letters, 39, L09406, https://doi.org/10.1029/2012GL051592, 2012.

Ahrens, B.: Distance in spatial interpolation of daily rain gauge data, Hydrology and Earth System Sciences, 10, 197–208, https://doi.org/10.5194/hess-10-197-2006, 2006.

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., Gan, T. Y., Pendergrass, A. G., Rosenfeld, D., Swann, A. L. S., Wilcox, L. J., and Zolina, O.: Advances in understanding large-scale responses of the water cycle to climate change, Annals of the New York Academy of Sciences, 1472, 49–75, https://doi.org/10.1111/nyas.14337, 2020.

Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, Nature, 419, 228–232, https://doi.org/10.1038/nature01092, 2002.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1° precipitation:

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Beck, H. E., Westra, S., Tan, J., Pappenberger, F., Huffman, G. J., McVicar, T. R., Gründemann, G. J., Vergopolan, N., Fowler, H. J., Lewis, E., Verbist, K., and Wood, E. F.: PPDIST, global 0.1° daily and 3-hourly precipitation probability distribution climatologies for 1979–2018, Scientific Data, 7, 1–12, https://doi.org/10.1038/s41597-020-00631-x, 2020.

Caesar, J., Alexander, L., and Vose, R.: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set, Journal of Geophysical Research: Atmospheres, 111, D05101, https://doi.org/10.1029/2005JD006280, 2006.

Camera, C., Bruggeman, A., Hadjinicolaou, P., Pashiardis, S., and Lange, M. A.: Evaluation of interpolation techniques for the creation of gridded daily precipitation (1×1 km^2); Cyprus, 1980–2010, Journal of Geophysical Research: Atmospheres, 119, 693–712, https://doi.org/10.1002/2013JD020611, 2014.

Chen, M., Xie, P., Janowiak, J. E., and Arkin, P. A.: Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations, Journal of Hydrometeorology, 3, 249–266, https://doi.org/10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2, 2002.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, Journal of Applied Meteorology, 33, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

Daly, C., Gibson, W., Taylor, G. H., Johnson, G. L., and Pasteris, P. P.: A knowledge-based approach to the statistical mapping of climate, Climate Research, 22, 99–113, https://doi.org/10.3354/cr022099, 2002.

Delaunay, B.: Sur la sphere vide, Izv. Akad. Nauk SSSR (in French), Otdelenie Matematicheskii i Estestvennyka Nauk, 7, 1–2, 1934.

Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K., and Arnold, J. G.: Constructing

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

retrospective gridded daily precipitation and temperature datasets for the conterminous United States, Journal of Applied Meteorology and Climatology, 47, 475–497, https://doi.org/10.1175/2007JAMC1356.1, 2008.

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., de Guzman, R., Htay, T. M., Wan Ibadullah, W. M., Bin Ibrahim, M. K. I., Khoshkam, M., Kruger, A., Kubota, H., Leng, T. W., Lim, G., Li-Sha, L., Marengo, J., Mbatha, S., McGree, S., Menne, M., de los Milagros Skansi, M., Ngwenya, S., Nkrumah, F., Oonariya, C., Pabon-Caicedo, J. D., Panthou, G., Pham, C., Rahimzadeh, F., Ramos, A., Salgado, E., Salinger, J., Sané, Y., Sopaheluwakan, A., Srivastava, A., Sun, Y., Timbal, B., Trachow, N., Trewin, B., van der Schrier, G., Vazquez-Aguirre, J., Vasquez, R., Villarroel, C., Vincent, L., Vischel, T., Vose, R., and Bin Hj Yussof, M. N. A.: Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3, Journal of Geophysical Research: Atmospheres, 125, e2019JD032263, https://doi.org/10.1029/2019JD032263, 2020.

Efthymiadis, D., Jones, P. D., Briffa, K. R., Auer, I., Böhm, R., Schöner, W., Frei, C., and Schmidli, J.: Construction of a 10-min-gridded precipitation data set for the Greater Alpine Region for 1800–2003, Journal of Geophysical Research: Atmospheres, 111, D01105, https://doi.org/10.1029/2005JD006120, 2006.

Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S., and Lott, N. J.: Creating a serially complete, national daily time series of temperature and precipitation for the western United States, Journal of Applied Meteorology, 39, 1580–1591, https://doi.org/10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2, 2000.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The shuttle radar topography mission, Reviews of Geophysics, 45, RG2004, https://doi.org/10.1029/2005RG000183, 2007.

Fischer, E. M. and Knutti, R.: Observed heavy precipitation increase confirms theory

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

and early models, Nature Climate Change, 6, 986–991, https://doi.org/10.1038/nclimate3110, 2016.

Golian, S., Javadian, M., and Behrangi, A.: On the use of satellite, gauge, and reanalysis precipitation products for drought studies, Environmental Research Letters, 14, 075005, https://doi.org/10.1088/1748-9326/ab2203, 2019.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Han, J.Y. and Miao, C.Y.: A new daily gridded precipitation dataset based on gauge observations across mainland China. https://doi.org/10.6084/m9.figshare.21432123.v2, 2022.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres, 113, D20119, https://doi.org/10.1029/2008JD010201, 2008.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Scientific Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

Hofstra, N. and New, M.: Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe, International Journal of Climatology, 29, 1872–1880, https://doi.org/10.1002/joc.1819, 2009.

Hutchinson, M. F.: Interpolating mean rainfall using thin plate smoothing splines, International Journal of Geographical Information Systems, 9, 385–403, https://doi.org/10.1080/02693799508902045, 1995.

Hutchinson, M. F.: Interpolation of rainfall data with thin plate smoothing splines - part

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

I: two dimensional smoothing of data with short range correlation, Journal of Geographic Information and Decision Analysis, 2, 153–167, 1998.

Hutchinson, M. F. and Xu, T.: Anusplin version 4.2 user guide, Centre for Resource and Environmental Studies. The Australian National University. Canberra, 5, 2004.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press., 2021.

Jones, P. D., Osborn, T. J., and Briffa, K. R.: Estimating sampling errors in large-scale temperature averages, Journal of Climate, 10, 2548–2568, https://doi.org/10.1175/1520-0442(1997)010<2548:ESEILS>2.0.CO;2, 1997.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, Journal of Hydrometeorology, 5, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., McNally, A., Skofronick-Jackson, G., Stocker, E., Wu, H., and Zaitchik, B. F.: NASA's remotely sensed precipitation: a reservoir for applications users, Bulletin of the American Meteorological Society, 98, 1169–1184, https://doi.org/10.1175/BAMS-D-15-00296.1, 2017.

Kucera, P. A., Ebert, E. E., Turk, F. J., Levizzani, V., Kirschbaum, D., Tapiador, F. J., Loew, A., and Borsche, M.: Precipitation from space: advancing earth system science, Bulletin of the American Meteorological Society, 94, 365–375, https://doi.org/10.1175/BAMS-D-11-00171.1, 2013.

Li, R., Wang, K., and Qi, D.: Validating the integrated multisatellite retrievals for global precipitation measurement in terms of diurnal variability with hourly gauge observations collected at 50,000 stations in China, Journal of Geophysical Research: Atmospheres, 123, 10,423– 10,442,

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

https://doi.org/10.1029/2018JD028991, 2018.

Lu, G. Y. and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, Computers & Geosciences, 34, 1044–1055, https://doi.org/10.1016/j.cageo.2007.07.010, 2008.

Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: a review, Biotechnologie, Agronomie, Societe et Environnement, 17, 392–406, https://doi.org/10.6084/M9.FIGSHARE.1225842.V1, 2013.

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

Merino, A., García-Ortega, E., Navarro, A., Fernández-González, S., Tapiador, F. J., and Sánchez, J. L.: Evaluation of gridded rain-gauge-based precipitation datasets: Impact of station density, spatial resolution, altitude gradient and climate, International Journal of Climatology, 41, 3027–3043, https://doi.org/10.1002/joc.7003, 2021.

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, International Journal of Climatology, 25, 693–712, https://doi.org/10.1002/joc.1181, 2005.

Morrissey, M. L., Maliekal, J. A., Greene, J. S., and Wang, J.: The uncertainty of simple spatial averages using rain gauge networks, Water Resources Research, 31, 2011–2017, https://doi.org/10.1029/95WR01232, 1995.

Myhre, G., Samset, B. H., Hodnebrog, Ø., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Forster, P. M., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T. B., Shawki, D., Shindell, D., Shine, K. P., Stjern, C. W., Takemura, T., and Voulgarakis, A.: Sensible heat has significantly affected the global hydrological cycle over the historical period, Nature Communications, 9, 1922, https://doi.org/10.1038/s41467-018-04307-4, 2018.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space–time climate

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

variability. part II: development of 1901–96 monthly grids of terrestrial surface climate, Journal of Climate, 13, 2217–2238, https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2, 2000.

Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth System Science Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, 2019.

Qin, R., Zhao, Z., Xu, J., Ye, J. S., Li, F. M., and Zhang, F.: HRLT: A high-resolution (1 day, 1 km) and long-term (1961–2019) gridded dataset for temperature and precipitation across China, Earth System Science Data Discussions (in review), 1–38, https://essd.copernicus.org/preprints/essd-2022-79/, 2022.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., and Lo, M. H.: Emerging trends in global freshwater availability, Nature, 557, 651–659, https://doi.org/10.1038/s41586-018-0123-1, 2018.

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

Shen, Y., Feng, M., Zhang, H., and Gao, F.: Interpolation Methods of China Daily Precipitation Data (in Chinese), Journal of Applied Meteorological Science, 21, 279–286, 2010.

Shen, Y. and Xiong, A.: Validation and comparison of a new gauge-based precipitation analysis over mainland China, International Journal of Climatology, 36, 252–265, https://doi.org/10.1002/joc.4341, 2016.

Shen, Y., Zhao, P., Pan, Y., and Yu, J.: A high spatiotemporal gauge-satellite merged precipitation analysis over China, Journal of Geophysical Research: Atmospheres, 119, 3063–3075, https://doi.org/10.1002/2013JD020686, 2014.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, Proceedings of the 1968 23rd ACM national conference, 517–524, https://doi.org/10.1145/800186.810616, 1968.

Shepard, D. S.: Computer Mapping: The SYMAP Interpolation Algorithm, in: Spatial

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

Statistics and Models, edited by: Gaile, G. L., and Willmott, C. J., Springer Netherlands, Dordrecht, 133–145, 10.1007/978-94-017-3048-8_7, 1984.

Sibson, R.: Locally equiangular triangulations, The Computer Journal, 21, 243–245, https://doi.org/10.1093/comjnl/21.3.243, 1978.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, Bulletin of the American Meteorological Society, 81, 2035–2046, https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2, 2000.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Reviews of Geophysics, 56, 79–107, https://doi.org/10.1002/2017RG000574, 2018.

Tait, A., Henderson, R., Turner, R., and Zheng, X.: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface, International Journal of Climatology, 26, 2097–2115, https://doi.org/10.1002/joc.1350, 2006.

Thiessen, A. H.: Precipitation averages for large areas, Monthly Weather Review, 39, 1082–1089, https://doi.org/10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2, 1911.

Tian, Y. and Peters-Lidard, C. D.: A global map of uncertainties in satellite-based precipitation measurements, Geophysical Research Letters, 37, L24407, https://doi.org/10.1029/2010GL046008, 2010.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, Bulletin of the American Meteorological Society, 84, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205, 2003.

Vivoni Enrique, R., Ivanov Valeri, Y., Bras Rafael, L., and Entekhabi, D.: Generation of triangulated irregular networks based on hydrological similarity, Journal of Hydrologic Engineering, 9, 288–302, https://doi.org/10.1061/(ASCE)1084-0699(2004)9:4(288), 2004.

Wahba, G. and Wendelberger, J.: Some new mathematical methods for variational

https://doi.org/10.5194/essd-2022-373
Preprint. Discussion started: 12 December 2022

730 objective analysis using splines and cross validation, Monthly Weather Review, 108, 1122–1143, https://doi.org/10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2, 1980.

Wu, J., and Gao, X.: A gridded daily observation dataset over China region and comparison with the other datasets [in Chinese], Chinese Journal of Geophysics, 735 56, 1102–1111, https://doi.org/10.6038/cjg20130406, 2013.

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

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: 740 APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, Bulletin of the American Meteorological Society, 93, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.

Zhang, Y., Ren, Y., Ren, G., and Wang, G.: Precipitation trends over mainland China 745 from 1961–2016 after removal of measurement biases, Journal of Geophysical Research: Atmospheres, 125, e2019JD031728, https://doi.org/10.1029/2019JD031728, 2020.

Zhao, Y., Zhu, J., and Xu, Y.: Establishment and assessment of the grid precipitation datasets in China for recent 50 years (in Chinese), Journal of the Meteorological 750 Sciences , 34, 414–420, https://doi.org/10.3969/2013jms.0008, 2014.