Evaluation of Multiple Forcing Data Sets for Precipitation and Shortwave Radiation over Mainland China

Fan Yang[1], Hui Lu[1,2], Kun Yang[1,2,3,4], Jie He[4], Wei Wang[1,5], Chengwei Li[1], Menglei Han[1], Yishan Li[1]

[1]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, 100084, China
[2]The Joint Center for Global Change Studies, Beijing, 100875, China
[3]CAS Center for Excellence in Tibetan Plateau Earth System, Beijing, 100101, China
[4]Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
[5]Changjiang Institute of Survey, Planning, Design and Research, Wuhan, 430010, China

Correspondence to: Hui Lu (luhui@tsinghua.edu.cn)

Abstract. Precipitation and shortwave radiation play important roles in climatic, hydrological and biogeochemical cycles. Currently, several global and regional forcing data sets, such as Global Land Data Assimilation System (GLDAS), China Meteorological Administration (CMA) Land Data Assimilation System (CLDAS) and China Meteorological Forcing Dataset (CMFD), can provide estimates of these two variables for China, while CN05.1, an interpolation product of CMA gauge observation, can provide high resolution precipitation for China. In this study, precipitation and shortwave radiation from CN05.1, CMFD, CLDAS and GLDAS were inter-compared with one another and against independent ground station observations during 2008-2014. The results demonstrate that all the four forcing data sets can capture the spatial distribution characteristics of precipitation over mainland China while the annual mean precipitation of CLDAS is smaller than others in most area. The time series of precipitation anomaly from the forcing data sets also match well with each other expect CMFD after August 2014. All forcing data sets show the temporal variations in dry region are greater than wet region. Compared with the independent precipitation observation data provided by the Ministry of Water Resources (MWR) in the middle and lower reaches of the Yangtze River, CLDAS performs best and the annual mean precipitation of CMFD also match well with the MWR station data. However, GLDAS shows a large dispersion degree and CN05.1 obviously overestimates the precipitation with the highest bias in eight months. As for shortwave radiation, CMFD is consistent with observation data, while CLDAS and GLDAS heavily overestimate shortwave radiation when compared against station data. Spatially, the three forcing data sets have some common distribution features. Compared with CMFD, CLDAS and GLDAS have higher radiation values in most areas of mainland China. However, the metrics we calculated indicate that CLDAS performs better than GLDAS. For temporal variations, CLDAS is closer to CMFD than GLDAS, while their amplitude of anomalies are diverse. Also, the temporal variation difference of shortwave radiation from the three forcing data sets mainly exists the south of 34°N. Findings from this study can provide guidance to communities regarding the performance of different forcing data over mainland China.

1 Introduction

Precipitation and shortwave radiation are the fundamental water and energy sources of land surface biological, physical and chemical processes (Zhao and Zhu 2015; Zhang et al. 2010). They can affect the moisture and heat exchange between the atmosphere and the land surface (Pan et al. 2014; Tian et al. 2007; Fekete et al. 2004; Gottschalck et al. 2005). Also, these two variables are the basic meteorological forcing inputs for land process simulations such as crop simulation, hydrologic modeling, dryland expansion estimation and dust events analysis (Bart and Lettenmaier 2004; Tang et al. 2007, 2008; Huang et al. 2015; Kang et al. 2016). Therefore, accurate precipitation and shortwave radiation data are essential for studies of climate change and land surface processes.

Although conventional measurement at stations can obtain the "true value" of the measured variable, it can only represent local scale information (Maurer 2002, Bogh et al. 2003), and is unable to depict the characteristics of spatial variation completely and continuously due to the limited number and location of stations (Duan et al 2012). In the late 1980s, data assimilation technology was proposed to reconstruct high resolution forcing data of historical climate (Xie et al. 2011, Zhao et al. 2010), and this brought unprecedented opportunities for researchers. These forcing data sets, which usually include precipitation, shortwave radiation, temperature, specific humidity, wind speed, surface pressure, and other meteorological data, are derived by assimilating numerical weather forecast information, ground observation data and remote sensing data together (Xie et al. 2011, Zhao et al. 2010, Pan et al. 2010). There are many forcing data sets currently available, such as that from the National Centers for Environmental Prediction and the National Center for Atmospheric Research reanalysis (NCEP/NCAR, Kalnay et al. 1996; Kistler et al. 2001), Global Land Data Assimilation System (GLDAS, Rodell et al. 2004), European Center for Medium-range Weather Forecasts (ECMWF) reanalysis ERA-Interim (Dee et al. 2011), Japanese 55-year Reanalysis (JRA-55, Kobayashi et al. 2015). In recent years, Chinese researchers have made great progress in developing forcing data. Two forcing data sets, namely China Meteorological Forcing Dataset (CMFD, released by the Institute of Tibetan Plateau Research, Chinese Academy of Sciences, He and Yang, 2011)and the China Meteorological Administration (CMA) Land Data Assimilation System (CLDAS, Shi et al. 2014), which cover China have been produced. Also, a data set named CN05.1 (released by National climate center of CMA) which merely interpolating CMA gauge data was made (Wu and Gao 2013). These forcing data sets are widely used because of their high spatial resolution, long time span over large areas and convenience of obtaining and processing. For example, CMFD forcing data set has been used to simulate the permafrost and seasonally frozen ground conditions on the Tibetan Plateau (Gao and Wang, 2013), to analyze precipitation impacts on vegetation spring phenology (Shen et al. 2015), to model land surface water and energy cycles of a mesoscale watershed (Xue et al. 2013), and to assess the climate and human impacts on surface water resources in the middle reaches of the Yellow River (Hu et al. 2015). Additionally, CMFD and GLDAS forcing data sets have been used to improve land surface temperature modeling for dry land in China (Chen et al. 2011). GLDAS has also been applied to analyze the long-term terrestrial water storage variations in the Yangtze River basin (Huang et al. 2013), and the newly released CLDAS forcing data set has been adopted in a recent study of drought monitoring (Han, 2015). CN05.1 also has

been used in many fields, such as simulating climate change over China (Gao et al. 2013) and studying the shift of western Pacific subtropical high (Huang et al. 2015).

However, the forcing data, either generated by interpolation ground observations or derived from reanalysis data, generally have considerable uncertainties (Qian et al. 2006). The bias associated with a forcing data set can be propagated into model results (Wang et al, 2016), which in turn may show unrealistic results if the forcing data sets are not reliable (Cosgrove et al., 2003). For example, errors in precipitation and shortwave radiation have a great impact on simulations of soil moisture, runoff and heat fluxes (Luo et al., 2003). As a result, it is necessary to evaluate the accuracy of the data sets so that the bias of these forcing data sets are fully recognized before they can be applied to land surface studies (Pan et al. 2014).Some studies have been conducted to evaluate the forcing data sets. Wang et al. (2014, 2016) assessed the applicability of GLDAS monthly precipitation data in China from 1979 to 2012, and found that both GLDAS-1 and GLDAS-2 precipitation matched well with observation precipitation data by visual inspection. Wang et al. (2016) evaluated CMFD daily precipitation data over the Qinghai-Tibetan Plateau from 2009 to 2012 and found that it mainly showed an overestimation for more than 255 days of the year. Wang et al. (2011) validated the GLDAS-1 daily and monthly precipitation data in a mesoscale basin in northeast China and concluded that GLDAS was of high quality for daily and monthly precipitation during March 2003 to March 2006. Wang and Zeng (2012) evaluated six reanalysis products (i.e., MERRA, NCEP–NCAR, CFSR, ERA-40,ERA-Interim, and GLDAS-1) using in situ measurements at 63 weather stations over the Tibetan Plateau and the result showed that GLDAS had the best overall performance for both daily and monthly precipitation.

Though the quality of GLDAS has been relatively well evaluated in previous studies, it is not bias free and the credibility of GLDAS in recent years over continental China is still unclear. On the other hand, CN05.1, CMFD and CLDAS are developed and maintained by Chinese scientists and they are supposed to have high accuracy and reliability because more ground observation data have been put into them. However, so far, there is no comprehensive evaluation over mainland China are conducted around these forcing data sets. In this study, the performance of CN05.1, CMFD, CLDAS and GLDAS in terms of precipitation and shortwave radiation were inter-compared and evaluated against available in situ observation. Such inter-comparison will benefit researchers to select meteorological forcing data, and in turn help the producer to further improve the quality of forcing data.

2 Used data

2.1 Observation data

2.1.1 Station precipitation data

Precipitation observed by a rain gauge network maintained by the Hydrology Bureau in the Ministry of Water Resources (MWR) of China were used as reference data (Xu et al. 2016). In this study, we used the precipitation data observed by rain gauges located in Hubei, Hunan and Jiangxi province in 2014. It should be note that the MWR precipitation data is

independent of the forcing date sets, in which the ground stations are operating by CMA. After quality control procedures, there are 5490 station can be used in this study. Their location are shown in Fig. 1.

2.1.2 Shortwave radiation station data

We used two kinds of shortwave radiation station data to verify the forcing data sets. The first kind is from CMA, but attention must be paid to CMFD, as their radiation is estimated from a hybrid model and thus is not fully independent of the radiation data used for evaluation. The other kind is from independent station data which are not included in the forcing data sets. Fig. 2 shows the station distribution.

1) Shortwave radiation station data from CMA

A daily surface solar radiation data set updated to 2010 is offered by the Data Assimilation and Modeling Center for Tibetan Multi-spheres (http://dam.CMFD.ac.cn/). This data set is produced by two kind of data. The first is estimated by hybrid model (Yang et al. 2001, 2006) with the air temperature, air pressure, relative humidity, and sunshine duration at 716 CMA stations. The other is estimated by ANN-based (Artificial Neural Network) model at 96 radiation stations which, because of its high accuracy, was used to correct the hybrid model estimate dynamically at a monthly scale. The ANN-based model is trained with recent observation data to estimate the earlier periods at 96 radiation stations (Tang et al. 2013). In this paper, we selected 625 stations with full data for 2008-2010.

2) Independent station data

① Shortwave radiation station from CERN

The Chinese Ecosystem Research Network (CERN) was established in 1988 by the Chinese Academy of Sciences (Su et al. 2005). The 2008-2014 shortwave radiation observation data used in this paper are provided by 35 field experimental stations in CERN covering various ecosystems, including farmland, forest, grassland, lakes and the sea. As shown in Fig. 2, these stations are located evenly over mainland China and cover various climate types and land cover types. Meanwhile, CERN stations are independent of CMA stations which are partly used in CMFD and CLDAS. Therefore, CERN is a perfect reference to assess the performance of these three forcing data sets, although the gauge density is not so high.

② Shortwave radiation station from HiWATER

Shortwave radiation observation data from eight stations in the Heihe River have been collected from the Heihe Watershed Allied Telemetry Experimental Research (HiWATER, Li et al. 2013) and it is widely used for land surface process studies (Liu et al. 2016; Cheng et al. 2014). There are 2-3 sites distributed in the upper, middle and downstream of the Heihe River basin.

③Shortwave radiation station from the TPE Database

A daily shortwave radiation record from the Meteorological dataset of the Ngari Desert Observation and Research Station and from the Meteorological dataset of the Muztagh Ata Station for Westerly Environment Observation and Research, was obtained from the Third Pole Environment (TPE) Database (http://www.tpedatabase.cn), as shown in Fig. 2.These two

stations were used as a supplement to evaluate the performance of the three reanalysis data sets on the west Tibetan Plateau where there are very few CMA ground stations.

2.2 Forcing data sets

2.2.1 CN05.1

CN05.1 provides precipitation and daily mean, minimum and maximum temperature data (Wu and Gao 2013). In this paper, the 0.25°×0.25° gridded monthly precipitation data over mainland China was used. It was interpolated from more than 2000 gauge stations over mainland China, and an "anomaly approach" (New et al. 2000) was applied in the interpolation. Considering that the meteorological stations mainly distribute in the eastern China where the economy is more developed and the terrain is more flat, CN05.1 may have big uncertainties in western China.

2.2.2 CMFD

The CMFD forcing data set was developed by the Institute of Tibetan Plateau Research, Chinese Academy of Sciences (He and Yang, 2011). This product covers the region of 70.0°E -140.0°E and 15.0°N -55.0°N, and includes precipitation, downward shortwave radiation, downward longwave radiation, 2-meter air temperature, specific humidity, wind speed and surface pressure. Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation data is used as the background field of precipitation data. However, TRMM has no valid data in the north of 40°N in most of the time. Therefore, GLDAS is used in these regions to solve this problem. Gauge observation data from 740 stations of CMA are used to correct systematic deviations in background data. The Global Energy and Water cycle Experiment - Surface Radiation Budget (GEWEX-SRB) radiation data provide the background field for the shortwave radiation data of CMFD. Notably, GLDAS also used to replace GEWEX-SRB in its unavailable time and region. Shortwave radiation data estimated with CMA station data which has been mentioned in 2.1.2 is also used. Other basic information of the data is listed in Table 1.

2.2.3 CLDAS

The version we evaluated is CLDAS-V2.0 (abbreviated as CLDAS in this paper), it was developed by CMA (Shi et al. 2014) and its spatial coverage is 60°E -160°E, 0°N -65°N. This is hourly gridded data with a spatial resolution of 0.0625°×0.0625°. CLDAS includes land surface forcing data, such as precipitation, shortwave radiation, temperature, specific humidity, wind speed and surface pressure, as well as soil status variables. It is a relatively new product, with temporal coverage from 2008 to 2017. Precipitation is combined and interpolated from two products, one is the Climate Prediction Center Morphing Technique (CMORPH) product and the other is an hourly merged precipitation product (V1.0) made by CMA which based on the observation data from automatic weather stations in China and CMORPH products through probability density function (PDF) and optimal interpolation (OI) merging algorithm (Shen et al. 2014). Shortwave radiation is retrieved from the FY-2C/E series of geostationary meteorological satellites. The Discrete Ordinates Radiative Transfer Program for a

Multi-Layered Plane-Parallel Medium (DISORT) method is used in the retrievals for radiation transfer calculations (Shi et al. 2011).

2.2.4 GLDAS

The 0.25°×0.25° monthly GLDAS-1 forcing data from the NOAH model (abbreviated as GLDAS in this paper) is provided by the US National Aeronautics and Space Administration (NASA). From 2001 to the present, this version makes use of National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, which merged satellite data (IR and microwave) and gauge data. CMAP fields are spatially and temporally disaggregated by Global Data Assimilation System (GDAS) modeled precipitation fields. Air Force Weather Agency (AFWA) Agricultural Meteorology modeling system (AGRMET) provided cloud and snow products, and then the AFWA supplied procedure was adopted to calculate downward shortwave and longwave radiation fluxes (Rui and Beaudoing, 2017; Rodell et al. 2004).

3 Methodology

Precipitation and shortwave radiation were evaluated from various spatial and temporal scales. In terms of spatial scale, the patterns of each data were compared and three metrics were computed, i.e. average precipitation over mainland China (Mean), standard deviation (SD) and coefficient of variation (CV), the latter two can reflect the degree of dispersion of the data set itself. For temporal scale, monthly anomalies of precipitation were derived by subtracting the seven-year monthly climatology from each data element. Temporal coefficient variation (TCV) was computed to demonstrate the fluctuation characteristics of the forcing data sets in time series. The larger this value, the greater the temporal difference. In order to make a further comparison, CMFD and CLDAS were resampled to 0.25°×0.25° (same as the CN05.1 and GLDAS) by the bilinear resampling method. To verify the accuracy of each forcing data sets using gauge observations data, a pixel-point method (Chen et al. 2013) was applied by pairing gauge observation data from the corresponding grids of these three forcing products. Root Mean Squared Error (RMSE) and bias were selected as evaluation metrics of the pixel-point comparison. Bias reflects the degree to which the measured value is over- or under-estimated. To further describe the degree of correspondence between forcing data and observation data, a Taylor diagram (Taylor 2001) is used. It shows the ratio of standardized deviations, correlation coefficient and unbiased RMSE between forcing data and observation data, and these statistics can quantify how closely the forcing data resembles the observation. The formulations of these metrics are as follows:

$$\mathrm{SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (1)$$

$$\mathrm{CV} = \frac{SD}{\bar{x}} \quad (2)$$

$$\text{Correlation coefficient} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2 \sum_{i=1}^{N}(y_i-\bar{y})^2}} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-y_i)^2}{N}} \quad (4)$$

$$\text{Bias} = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} y_i \quad (5)$$

$$\text{Relative bias} = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} y_i} - 1 \quad (6)$$

$$\text{TCV} = \frac{\sqrt{\frac{1}{m}\sum_{j=1}^{m}(t_j-\bar{t})^2}}{\bar{t}} \quad (7)$$

where x_i is the element of data sets and $\bar{x}$ is the average value of this data set; y_i is the element of the reference data set and $\bar{y}$ is the average for data y_i; N is the number of points in the data; m is the number of months during 2008 to 2014; t_j is precipitation or shortwave radiation value of per month of each grid point in the data sets and $\bar{t}$ is the average of t_j.

4 Evaluation of precipitation data

4.1 Spatial distribution of precipitation

It is obvious that all the data sets reveal a gradual increasing pattern of annual mean precipitation from northwest to southeast in mainland China (Fig. 3). The distribution of precipitation from CMFD is similar to that of CN05.1. However, they still have difference in western China, such as in the Tibetan Plateau. Compared with CN05.1 and CMFD, the area where the annual mean precipitation is higher than 1500mm is smaller in CLDAS and GLDAS. Besides, the precipitation of CLDAS in north China is smaller than others. As shown in Table 2, the Mean value of CLDAS is significant lower than CN05.1 while CMFD has the highest Mean value. As for SD and CV, CN05.1, CMFD and GLDAS are closer.

4.2 Temporal variation of precipitation

Fig. 4 shows the time series of monthly mean precipitation anomalies averaged over mainland China. The forcing data sets match well with each other, which indicate that all forcing data sets can reflect the inter-annual and decadal variability of precipitation over mainland China. Noticeably, a relatively low negative anomaly after August in 2014 are shown in CMFD. The temporal standard deviation of each grid is divided by the average precipitation during the seven-year period to obtain the TCV as shown in Fig. 5. It can be seen that the spatial distribution of TCV of the four data sets are similar to each other in southeast China where generally have smaller TCV values. Also, all of they can reflect that in the north of China, precipitation changes greater in temporal. The TCV of CMFD is similar to that of CN05.1 while the TCV of CLDAS and GLDAS are higher in most areas of north China especially in dry regions (Xinjiang, Gansu and Inner Mongolia) where the

TCV are higher than 1.5. For CLDAS, the area where the TCV value are higher than 1.0 are larger than other forcing data sets and its Mean value is the biggest.

4.3 Compared with MWR station data

When compared the spatial distribution of four forcing data sets, it is clear that there are obvious differences among them in the middle and lower reaches of the Yangtze River (red polygon shown in Fig. 3). Therefore, we use the independent MWR observation data to further evaluate forcing data in this region. As shown in Fig. 6, the annual mean precipitation data of CMFD and CLDAS are more consistent with the MWR observation. However, the performance of GLDAS and CN05.1 are not as good as others. Their RMSE are about twice bigger than that of CLDAS. Also, it is obvious that the dispersion degree of GLDAS is the biggest compared with other data sets which indicate that GLDAS changes greatly in spatial. From the pattern of Fig. 6 (a) and the high bias of CN05.1, we can conclude that the annual mean precipitation in Hubei, Hunan and Jiangxi province are heavily overestimated by CN05.1. The evaluation results of monthly precipitation are listed in Table 3. According to the metrics, CLDAS performs best in most of the months of 2014. Fig. 7 also confirms that CLDAS performs well because the orange points representing CLDAS are concentrated together, located in a region where the correlation coefficient is between 0.6 and 0.9, the standardized deviation is close to 1 and the unbiased RMSE is low. This reflects that the quality of monthly precipitation of CLDAS is stable and reliable. However, the performance of other data sets in the monthly scale varies greatly, especially for CMFD and CN05.1.

5 Evaluation of shortwave radiation data

5.1 Comparison against ground measurements

A comparison between gauge observations and forcing data was carried out to examine which one was closest to the ground observations. What stands out in Fig. 8 (a), (d) and (g) is that compared with the stations from CMA, the points of CMFD are distributed around the diagonal evenly, while CLDAS and GLDAS have a much higher shortwave radiation value in about 96% points. Fig. 8 (a) and Fig. 9 show that, the unbiased RMSE, relative bias and RMSE between CMFD and CMA stations are the smallest and the correlation coefficient is the highest, the standardized deviation ratio is the closest to 1. These metrics of CLDAS perform better than GLDAS which indicate that CLDAS is more resemble to the observation than GLDAS. The high correlation may be due to the use of these site observations when producing CMFD, leading the comparison results do not convince enough.

To make up for this problem, we use the observation data from CERN, which is not merged into these forcing data sets. As shown in Fig. 8 (b), (e), (h), the results are similar to the evaluation results by CMA. The shortwave radiation of CMFD coincides well with the observation data, on the contrary, most stations are also significantly overestimated by CLDAS and GLDAS. Additionally, the metrics of CMFD shown in Fig. 9 perform the best which indicates that the estimation of CMFD

for shortwave radiation is more precise than for CLDAS and GLDAS in these areas, and GLDAS is worse compared with CLDAS.

Due to the fact that the distribution of CERN observation stations relatively sparse in western China, this paper supplements the data of eight observation stations in the Heihe River basin and two observation stations in the Tibetan Plateau to validate the three forcing data sets. As shown in Fig. 8 (c), (f) and (i), in these ten stations, CMFD is closer to the in situ observation, while both CLDAS and GLDAS show an obvious overestimation of shortwave radiation against gauge observation. As for other statistical indicators, CLDAS has the smallest unbiased RMSE and the closest standardized deviation compared to the observation. The correlation coefficient between GLDAS and observation is the highest followed by CMFD and CLDAS, while the RMSE and relative bias of CLDAS and GLDAS are about 2 and 10 times that of CMFD, respectively. In the study of Qi et al. (2015), GLDAS was also found to overestimates shortwave radiation from March 2000 to December 2007 in the Biliu Basin which is located in a coastal region of China. Besides, Wang et al. (2011) proved that shortwave radiation at the Changchun, Shenyang and Yanji stations in China was also overestimates by GLDAS from 2000 to 2006.

5.2 Spatial distribution of shortwave radiation

As mentioned above, the shortwave radiation data of CMFD matches well with the station data so we used this as reference data to evaluate the performance of CLDAS and GLDAS over mainland China.

As can be seen from Fig. 10, the distribution of shortwave radiation of the three forcing data sets have some common characteristics. They all have a similar spatial pattern showing that the shortwave radiation in western China is higher than in the east, while the largest shortwave radiation value appears in the Tibetan Plateau and the value in northeast China is relatively low. It is clear from Fig. 10 (a) that for CMFD's estimation, the regions where the shortwave radiation are more than 200W m^{-2} only locate in seven provinces (i.e. Xinjiang, Inner Mongolia, Qinghai, Tibet, Sichuan and Yunnan), and these areas are relatively small except Tibet and Qinghai. In contrast, CLDAS is similar to GLDAS, the areas where shortwave radiation are more than 200W m^{-2} are very large and they extend to north China such as Hebei and Shandong province. Especially, for GLDAS, this phenomenon even appears in southeast China. In addition, the Mean values of CLDAS and GLDAS are higher than CMFD, while its degree of dispersion was relatively larger as indicated by a smaller SD and CV (Table 4).

The difference among the three forcing data sets is shown in Fig. 11. It can be found that CLDAS and GLDAS have a much higher shortwave radiation than CMFD in most regions. More than 95% of the area over mainland China shows a positive difference for CLDAS when compared with CMFD, especially in some areas of Xinjiang province and the area 24°N-44°N, 105°E-120°E. When compared with CMFD, GLDAS is also significantly higher except over the Tibetan Plateau. In terms of the statistical metrics shown in Table 4, the absolute value of average difference, RMSE, and the relative bias between CMFD and CLDAS are smaller and the correlation coefficient of the three forcing data sets are all around 0.9. Overall, CLDAS and GLDAS are similar, both being higher than CMFD in most regions of mainland China.

5.3 Temporal variation of shortwave radiation

As shown in Fig. 12, the change trend of anomaly of CMFD and CLDAS match well with each other, but the anomaly amplitudes of these two data sets are different. The magnitude of the fluctuation of GLDAS is larger than others. There are 52 months when the difference between GLDAS and CMFD is greater than that of CLDAS. The anomaly of CLDAS is always positive after year 2013, which indicates that the shortwave radiation estimated by this product is clearly higher than the climatology mean value. This phenomenon also appears in GLDAS in most of the months of 2012-2014. As for the statistical metrics shown in Table 5, the RMSE and relative bias between CMFD CLDAS are smaller than the values between CMFD and GLDAS and their correlation coefficient are not high.

The spatial pattern of TCV for shortwave radiation of the three forcing data sets have some common characteristics, the highest TCV appears in northwest and northeast China while the smallest TCV can be found in southwest China. For the south of 34°N, the TCV of CLDAS is lower than 0.25 while CMFD is higher than 0.25 in the southeast. In addition, the estimation of TCV in CMFD and GLDAS has an obviously higher value in the vicinity of Sichuan province and Chongqing province than the surrounding areas. The TCV of the three data sets are similar to each other north of 34°N and the difference mainly lies in the south of China. The characteristics mentioned above are shown in Fig. 13.

6 Discussion

Based on the preceding analysis, we can see that though these forcing data sets have some common characteristics and can reflect the features of precipitation and shortwave radiation over mainland China, they have many difference due to different resolution, the various data they merged and the diverse algorithm they used. For precipitation, the spatial distribution of forcing data sets were compared and their quality in the middle and lower reaches of the Yangtze River were evaluated. CLDAS performs better both in annual and monthly scale, this is not surprise because CLDAS merges data at more than 30000 stations which can improve the data quality greatly. The precipitation of CMFD performs well at annual mean but not so at monthly scale, and its heavily decrease of precipitation after August 2014 is abnormal. As far as we concern, the CMFD used less precipitation station data than CN05.1 and CLDAS, which influences its quality in 2014. GLDAS as a global data, the precision in mainland China is limited due to the observation data of China merged in it may be not enough. Though both CMFD and GLDAS merged remote sensing data, they are not similar because the station data they used are different. As for CN05.1 which was made by purely station data and mathematical method, it is reasonable that it performs worse than other forcing data sets in station-sparse regions.

Comparing the forcing data sets and observation data, it was found that the shortwave radiation of CMFD perform better than the other two. The reason is that there are only about 100 radiation stations that were sparsely deployed in China since 1961 and the radiation observation data may be unusable because it often include erroneous values and missing data (Shi et al. 2008), therefore, the radiation observation data merged in CLDAS and GLDAS are limited. However, for CMFD, it merged the 50-year data set of daily surface solar radiation at 716 CMA stations which was aforementioned in section 2.1.2.

Though this data set is estimated by model, it is widely validated and its performance is pretty well. As a result, shortwave radiation of CMFD is closer to observation data.

7 Conclusions

In recent years, increasingly forcing data sets have been developed aiming to provide better support for climate, agriculture and hydrological researches. In this study, precipitation and shortwave radiation data provided by CN05.1, CMFD, CLDAS and GLDAS were inter-compared and evaluated over mainland China which can help users to choose a better data sets. For precipitation, all the four forcing data sets reflect similar spatial distribution characteristics, i.e. a gradual increase from northwest to southeast in mainland China. The results also indicate that precipitation estimated by CLDAS is smaller than other data sets in most area of mainland China. When considering temporal variability, the monthly mean precipitation anomalies of the four forcing data sets match well with each other, but CMFD has abnormal performance after August 2014. The TCV are closely related to the amount of precipitation, the values of dry region usually higher than wet region. In addition, the mean value of CLDAS and the area where the TCV is higher than 1.0 is larger than the others. Compared with the MWF observation data of 2014 in the middle and lower reaches of the Yangtze River, CLDAS shows the best performance in both annual and monthly scale with the lowest RMSE and highest correlation coefficient. The annual mean precipitation of CMFD also match well with station data, while its quality at monthly scale is to be improved. We also found that CN05.1 heavily overestimates the precipitation in this region.

In terms of shortwave radiation, comparisons against ground shortwave radiation observation show that the shortwave radiation value is significantly overestimated by CLDAS and GLDAS, and usually they have a much higher RMSE and relative bias value, and relatively low correlation coefficient. Compared with GLDAS, CLDAS performs better. In contrast, CMFD is closer to the observation data and most of its statistical metrics perform much better. All forcing data sets show higher values in western China than in eastern China, and the largest shortwave radiation value exists in the Tibetan Plateau. Also, the spatial characteristics of CLDAS and GLDAS are similar, with both of their values being higher than CMFD in most areas of mainland China, while the metrics between CLDAS and CMFD are better than those of GLDAS. The time series of anomaly shows that GLDAS fluctuates heavily while CMFD is more stable. The temporal variability of the three forcing data sets is more similar north of 34°N, while the difference in the south is larger. All the results reflect the fact that temporal variation of shortwave radiation in the north of China is larger than that in the south.

To sum up, there is no data sets that performs the best in terms of both precipitation and shortwave radiation. CLDAS has the highest spatial and temporal resolution, and it performs best in terms of precipitation. However, the shortwave radiation is obviously overestimated by CLDAS. CMFD also has high resolution and its shortwave radiation data matches well with the station data; its annual mean precipitation is reliable but its monthly precipitation needs improvements. Both precipitation and shortwave radiation data from GLDAS over mainland China need to be improved, so does to the precipitation of CN05.1. As these products are widely used and being developed, our results could benefit researchers for forcing data selection and

uncertainty quantification and also could provide clues for data producers to further improve their data sets. Meanwhile, the results of this inter-comparison highlights that big uncertainties exist in the currently available forcing data, especially in the west region of mainland China where the density of ground stations is low, and where there is a need to improve the quality of forcing data in these regions.

Acknowledgements

This work was jointly supported by the National Basic Research Program of China (No.2015CB953703), the National Key Research and Development Program of China (2016YFA0601603), and the National Natural Science Foundation of China (41371328 & 91537210). We are grateful for CN05.1 provided by Dr. Xuejie Gao at IAP，CAS and CLDAS provided by Dr. Chunxiang Shi at CMA. The GLDAS data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The authors also wish to thank the Third Pole Environment Database, CERN Database and HiWATER Project for provide shortwave radiation station data. The computation for this work is supported by Tsinghua National Laboratory for Information Science and Technology. We acknowledge reviewers for their insightful and constructive comments which improve the manuscript substantially.

References

Bart, N., and Lettenmaier, D. P.: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, Journal of Geophysical Research Atmospheres, 109, 265-274, doi:10.1029/2003JD003497, 2004.

Boegh, E., Thorsen, M., Butts, M. B., Hansen, S., Christiansen, J. S., Abrahamsen, P., Hasager, C. B., Jensen, N. O., van der Keur, P., Refsgaard, J. C., Schelde, K., Soegaard, H., and Thomsen, A.: Incorporating remote sensing data in physically based distributed agro-hydrological modelling, Journal of Hydrology, 287, 279-299, doi:10.1016/j.jhydrol.2003.10.018, 2004.

Cheng, G., Zhao, W., Feng, Z., Xiao, S., and Xiao H.: Integrated study of the water–ecosystem–economy in the Heihe River Basin, National Science Review, 1, 413-428, doi:10.1093/nsr/nwu017, 2014.

Chen, Y. Y., Yang, K., He, J., Qin, J., Shi, J. C., Du, J. Y., and He, Q.: Improving land surface temperature modeling for dry land of China, Journal of Geophysical Research-Atmospheres, 116, D20104, doi:Artn D2010410.1029/2011jd015921, 2011.

Chen, Y. Y., Yang, K., Qin, J., Zhao, L., Tang, W. J., and Han, M. L.: Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau, Journal of Geophysical Research-Atmospheres, 118, 4466-4475, doi:10.1002/jgrd.50301, 2013.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q. Y., Luo, L. F., Higgins, R. W., Pinker, R. T., Tarpley, J. D., and Meng, J.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, Journal of Geophysical Research-Atmospheres, 108, 1887-1902, doi:Artn 884210.1029/2002jd003118, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553-597, doi:10.1002/qj.828, 2011.

Fekete, B. M., Vorosmarty, C. J., Roads, J. O., and Willmott, C. J.: Uncertainties in precipitation and their impacts on runoff estimates, Journal of Climate, 17, 294-304, doi:10.1175/1520-0442(2004)017<0294:Uipati>2.0.Co;2, 2004.

Gao, X. J., Wang, M. L., and Giorgi, F.: Climate Change over China in the 21st Century as Simulated by BCC_CSM1.1-RegCM4.0, Atmospheric and Oceanic Science Letters, 6, 381-386,doi:10.3878/j.issn.1674-2834.13.0029,2013.

Gottschalck, J., Meng, J., Rodell, M., and Houser, P.: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states, Journal of Hydrometeorology, 6, 573-598, doi:10.1175/Jhm437.1, 2005.

Guo, D., and Wang, H.: Simulation of permafrost and seasonally frozen ground conditions on the Tibetan Plateau, 1981—2010, Journal of Geophysical Research-atmospheres, 118, 5216-5230, doi:10.1002/jgrd.50457, 2013.

He, J., and Yang, K.: China Meteorological Forcing Dataset. Cold and Arid Regions Science Data Center, Lanzhou, doi:10.3972/westdc.002.2014.db, 2011.

Han S.: The simulation and evaluation using CLM3.5 and SSIB2 land surface model based CLDAS forcing data with drought monitoring. Nanjing University of Information Science and Technology, 2015.

Hu, Z. D., Wang, L., Wang, Z. J., Hong, Y., and Zheng, H.: Quantitative assessment of climate and human impacts on surface water resources in a typical semi-arid watershed in the middle reaches of the Yellow River from 1985 to 2006, International Journal of Climatology, 35, 97-113, doi:10.1002/joc.3965, 2015.

Huang, J. P., Yu, H. P., Guan, X. D., Wang, G. Y., and Guo, R. X.: Accelerated dryland expansion under climate change, Nature Climate Change, 6, doi:10.1038/Nclimate2837, 2016.

Huang, Y., Salama, M. S., Krol, M. S., van der Velde, R., Hoekstra, A. Y., Zhou, Y., and Su, Z.: Analysis of long-term terrestrial water storage variations in the Yangtze River basin, Hydrology and Earth System Sciences, 17, 1985-2000, doi:10.5194/hess-17-1985-2013, 2013.

Huang, Y., Wang, H., Fan, K., and Gao, Y.: The western Pacific subtropical high after the 1970s: westward or eastward shift? , Climate Dynamics, 44, 2035-2047,doi: 10.1007/s00382-014-2194-5,2015.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society, 77, 437-471, doi:10.1175/1520-0477(1996)077<0437:Tnyrp>2.0.Co;2, 1996.

Kang, L., Huang, J., Chen, S., and Wang, X.: Long-term trends of dust events over Tibetan Plateau during 1961–2010, Atmospheric Environment, 125, 188-198, doi:10.1016/j.atmosenv.2015.10.085 2016.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan, 93, 5-48, doi:10.2151/jmsj.2015-001, 2015.

Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y., Wen, J., Li, H., Zhu, G., Guo, J., Ran, Y., Wang, S., Zhu, Z., Zhou, J., Hu, X., and Xu, Z. : Heihe watershed allied telemetry experimental research (HiWATER): Scientific objectives and experimental design, Bulletin of the American Meteorological Society, 94, 1145-1160, doi:10.1175/BAMS-D-12-00154.1, 2013.

Liu, S. M., Xu, Z. W., Song, L. S., Zhao, Q. Y., Ge, Y., Xu, T. R., Ma, Y. F., Zhu, Z. L., Jia, Z. Z., and Zhang, F.: Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces, Agricultural and Forest Meteorology, 230, 97-113, doi:10.1016/j.agrformet.2016.04.008, 2016.

Luo, L. F., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Lohmann, D., Cosgrove, B., Wen, F. H., Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, Journal of Climate, 15, 3237-3251, doi:10.1175/1520-0442(2002)015<3237:Althbd>2.0.Co;2, 2002.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space–time climate variability. Part II: development of 1901–96 monthly grids of terrestrial surface climate, Journal of climate, 13, 2217-2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2 2000.

Pan, X., Xin, L. I., and Chao, Z.: Review of Research of Forcing Data for Regional Scale Hydrological Model, Advances in Earth Science, 25, 1314-1324, doi:10.11867/j.issn.1001-8166.2010.12.1314, 2010.

Pan, X. D., Li, X., Yang, K., He, J., Zhang, Y. L., and Han, X. J.: Comparison of Downscaled Precipitation Data over a Mountainous Watershed: A Case Study in the Heihe River Basin, Journal of Hydrometeorology, 15, 1560-1574, doi:10.1175/Jhm-D-13-0202.1, 2014.

Qi, W., Zhang, C., Fu, G. T., and Zhou, H. C.: Global Land Data Assimilation System data assessment using a distributed biosphere hydrological model, Journal of Hydrology, 528, 652-667, doi:10.1016/j.jhydrol.2015.07.011, 2015.

Qian, T., Dai, A., Trenberth, K. E., and Oleson, K. W.: Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations, Journal of Hydrometeorology, 7, 953-975, doi: 10.1175/JHM540.1, 2006.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bulletin of the American Meteorological Society, 85, 381-394, doi:10.1175/Bams-85-3-381, 2004.

Rui, H. and Beaudoing, H., 2017: Readme document for GLDAS version 1 Data Products. Accessed 30 September 2017. [Available online at https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS_V1/README.GLDAS.pdf.].

Shen, M., Piao, S., Cong, N., Zhang, G., and Janssens, I.: Precipitation impacts on vegetation spring phenology on the Tibetan Plateau, Global Change Biology, 21, 3647-3656, doi:10.1111/gcb.12961, 2015.

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, doi:10.1002/2013JD020686, 2014.

Shi, C., Jiang, L., Zhang, T., Xu, B., and Han, S.: Status and Plans of CMA Land Data Assimilation System (CLDAS) Project, EGU General Assembly Conference Abstracts. 2014.

Sheffield, J., Duan, Q. Y., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains, Journal of Geophysical Research-Atmospheres, 108, 8843, doi:Artn 884310.1029/2002jd003246, 2003.

Shi, C., Xie, Z., Qian, H., Liang, M., Yang, X.: China land soil moisture EnKF data assimilation based on satellite remote sensing data, Science Chine Earth Sciences, 54, 1430-1440, doi: 10.1007/s11430-010-4160-3, 2011.

Shi, G., Hayasaka, T., Ohmura, A., Chen, Z., Wang, B., Zhao, J., Che, H, and Xu, L.: Data Quality Assessment and the Long-Term Trend of Ground Solar Radiation in China, Journal of Applied Meteorology & Climatology, 47, 1006-1016,doi: 10.1175/2007JAMC1493.1,2008.

Tang, Q. H., Oki, T., Kanae, S., and Hu, H. P.: Hydrological cycles change in the Yellow River basin during the last half of the twentieth century, Journal of Climate, 21, 1790-1806, doi:10.1175/2007jcli1854.1, 2008.

Tang, Q., and Hu, H.: The Influence of Precipitation Variability and Partial Irrigation within Grid Cells on a Hydrological Simulation, Cas-Twas-Wmo Forum,international Training Workshop on Regional Climate Change and ITS Impact Assessment, doi: 10.1175/JHM589.1, 2011.

Tang, W. J., Yang, K., Qin, J., and Min, M.: Development of a 50-year daily surface solar radiation dataset over China, Science China-Earth Sciences, 56, 1555-1565, doi:10.1007/s11430-012-4542-9, 2013.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research Atmospheres, 106, 7183-7192, doi: 10.1029/2000JD900719, 2001.

Tian, Y., Peters-Lidard, C. D., Choudhury, B. J., and Garcia, M.: Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications, Journal of Hydrometeorology, 8, 1165-1183, doi:10.1175/2007jhm859.1, 2007.

Wang, F. X., Wang, L., Koike, T., Zhou, H. C., Yang, K., Wang, A. H., and Li, W. L.: Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model, Journal of Geophysical Research-Atmospheres, 116, D21108, doi:Artn D2110810.1029/2011jd015990, 2011.

Wang, A. H., and Zeng, X. B.: Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau, Journal of Geophysical Research-Atmospheres, 117, 5102, doi:Artn D0510210.1029/2011jd016553, 2012.

Wang, Y., Nan, Z., Chen, H., and Wu, X.: Correction of daily precipitation data of ITPCAS dataset over the Qinghai-Tibetan Plateau with KNN model, Geoscience and Remote Sensing Symposium, doi: 10.1109/IGARSS.2016.7729148 2016, 593-596.

Wei, W., Hui, L., Yang, D., Khem, S., Yang, J., Gao, B., Peng, X., and Pang, Z.: Modelling Hydrologic Processes in the Mekong River Basin Using a Distributed Model Driven by Satellite Precipitation and Rain Gauge Observations, Plos One, 11, e0152229, doi:10.1371/journal.pone.0152229, 2016.

Wen, S., Guo, X., He, H.., and Yue, Y: Web-based Management and Sharing System of CERN Observational Data, Journal of Agricultural Mechanization Research, doi:10.13427/j.cnki.njyi.2005.04.083, 2005.

Wen, W., Wu. X., and Peng, W.: Assessing the applicability of GLDAS monthly precipitation data in China, Advances in Water Science (in Chinese), 25, 769-778, 2014.

Wu, J., and Gao, X.-J.: A gridded daily observation dataset over China region and comparison with the other datasets, Chin. J. Geophys., 56, 1102-1111, doi:10.6038/cjg20130406, 2013.

Xie, X., Jin-Hai, H. E., and Li, Q. I.: A review on applicability evaluation of four reanalysis datasets in China, Journal of Meteorology & Environment, 2011.

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., and Hou, A.: Ground Validation of GPM IMERG and TRMM 3B42V7 Rainfall Products over Southern Tibetan Plateau Based on a High-Density Rain-Gauge Network, Journal of Geophysical Research Atmospheres, 122,doi:10.1002/2016JD025418,2017.

Xue, B., Wang, L., Yang, K., Tian, L., Qin, J., Chen, Y., Zhao, L., Ma, Y., Koike, T., Hu, Z. Y., and Li, X. P.: Modeling the land surface water and energy cycles of a mesoscale watershed in the central Tibetan Plateau during summer with a distributed hydrological model, Journal of Geophysical Research-Atmospheres, 118, 8857-8868, doi:10.1002/jgrd.50696, 2013.

Yang, K., Huang, G. W., and Tamai, N.: A hybrid model for estimating global solar radiation, Solar Energy, 70, 13-22, doi:10.1016/S0038-092x(00)00121-3, 2001.

Yang, K., Koike, T., and Ye, B. S.: Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets, Agricultural and Forest Meteorology, 137, 43-55, doi:10.1016/j.agrformet.2006.02.001, 2006.

Yang, K., He, J., Tang, W. J., Qin, J., and Cheng, C. C. K.: On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau, Agricultural and Forest Meteorology, 150, 38-46, doi:10.1016/j.agrformet.2009.08.004, 2010.

Zhang, H., Liu, G., Ye, Y., and Huang, C.: Distributed modeling of shortwave solar radiation distribution over the Tibetan Plateau, J Natural Resources, 25, 811-821, doi:10.11849/zrzyxb.2010.05.012, 2010.

Zhao, T., Congbin, F. U., Zongjian, K. E., and Guo, W.: Global Atmosphere Reanalysis Datasets:Current Status and Recent Advances, Advances in Earth Science, 25, 242-254, doi:10.11867/j.issn.1001-8166.2010.03.0241, 2010.

Zhao, Y., and Zhu, J.: Assessing Quality of Grid Daily Precipitation Datasets in China in Recent 50 Years, Plateau Meteorology, doi:10.7522/j.issn.1000-0534.2013.00141, 2015.

Table 1. Basic information of the data sets used in this paper.

Name	Type	Analyzed period	Available period	Variables	Spatial resolution	Number of sites
CN05.1	forcing data set	2008-2014	1961-2014	precipitation	0.25°	
CMFD	forcing data set	2008-2014	1979-2016	precipitation; shortwave radiation	0.1°	
CLDAS	forcing data set	2008-2014	2008-2016	precipitation; shortwave radiation	0.0625°	
GLDAS	forcing data set	2008-2014	2000-2016	precipitation; shortwave radiation	0.25°	
Hydrology Bureau data	Observation data	2014	2014	precipitation		5490
CMA	observation data	2008-2010	Different at each site	shortwave radiation		625
CERN	observation data	2008-2014	Different at each site	shortwave radiation		35
HiWATER	observation data	Different at each site	Different at each site	shortwave radiation		8
TPE Database	observation data	Different at each site	Different at each site	shortwave radiation		2

Table 2. Spatial statistical metrics of annual mean precipitation from 2008 to 2014.

Metrics	CN05.1	CMFD	CLDAS	GLDAS
Mean(mm yr^{-1})	612.09	637.65	508.58	609.44
SD(mm yr^{-1})	497.61	511.09	429.50	506.55
CV	0.81	0.80	0.84	0.83

Table 3. Statistical metrics of monthly precipitation in 2014 between forcing data set and MWR rain gauge observations.

Time	Bias				RMSE			
	CN05.1	CMFD	CLDAS	GLDAS	CN05.1	CMFD	CLDAS	GLDAS
Jan.	23.62	2.37	-1.85	-0.48	29.53	19.70	8.90	8.26
Feb.	48.04	31.53	4.74	20.43	59.13	59.87	26.19	37.15
Mar.	61.37	14.81	-9.26	12.50	77.02	37.54	36.43	59.47
Apr.	62.83	17.93	-8.47	12.39	79.06	47.63	35.86	67.15
May	66.23	24.72	13.32	39.51	93.05	66.51	52.63	105.35
June	40.11	23.59	0.49	19.48	74.53	67.77	50.06	87.73
July	18.68	7.38	-0.24	1.04	67.86	78.26	59.45	92.06
Aug.	22.37	-53.29	3.57	11.01	62.51	78.93	47.46	74.30
Sept.	4.58	-21.18	-8.57	1.74	36.81	45.66	32.35	44.09
Oct.	26.07	-17.59	5.56	8.72	37.11	38.95	27.76	32.35
Nov.	26.73	-21.04	-2.91	14.36	37.71	34.76	25.95	40.79
Dec.	8.99	-5.43	-4.33	-1.65	14.90	16.97	9.40	10.84
Annual	421.38	273.68	234.39	427.71	403.16	3.80	-7.93	139.06

Table 4. Spatial statistical metrics of annual mean shortwave radiation and bias from 2008 to 2014 (CMFD was used as reference data when average difference, RMSE, relative bias and correlation coefficient were calculated).

Metrics	CMFD	CLDAS	GLDAS
Mean(W m^{-2})	178.60	202.26	203.13
SD(W m^{-2})	31.13	28.82	20.99
CV	0.17	0.14	0.10
Average difference (W m^{-2})	--	23.72	24.57
RMSE(W m^{-2})	--	27.55	28.61
Relative bias	--	0.14	0.15
Correlation coefficient	--	0.89	0.92

Table 5. Statistical metrics of monthly mean shortwave radiation deseasonalized anomalies among forcing data sets during 2008-2014.

Metrics	CLDAS-CMFD	GLDAS-CMFD
RMSE(W m^{-2})	5.14	5.79
Relative bias	1.14	1.66
Correlation coefficient	0.50	0.62

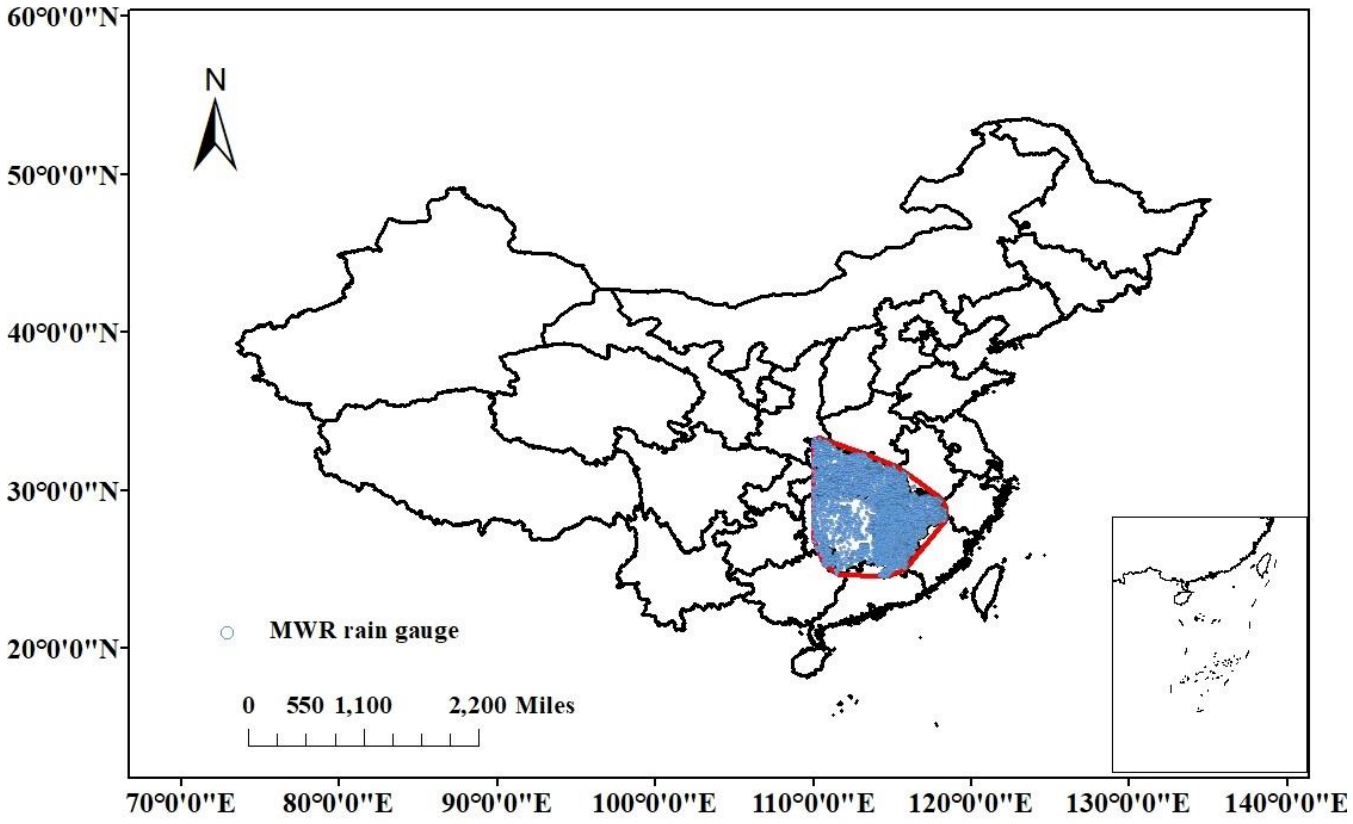

Figure 1: Locations of the rain gauges operated by the Ministry of Water Resources of China, used as an independent precipitation data source in this study.

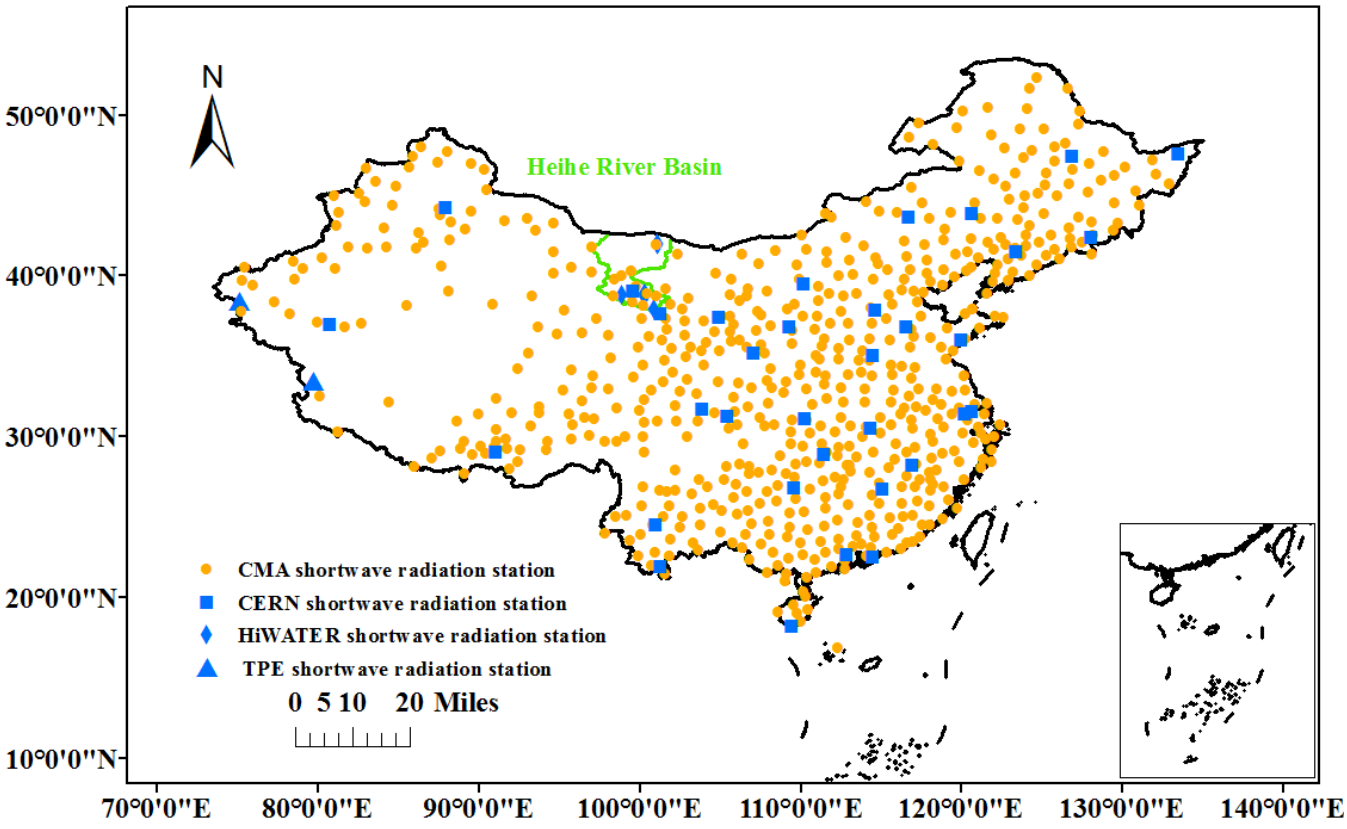

Figure 2: Locations of shortwave radiation stations over mainland China investigated in this study.

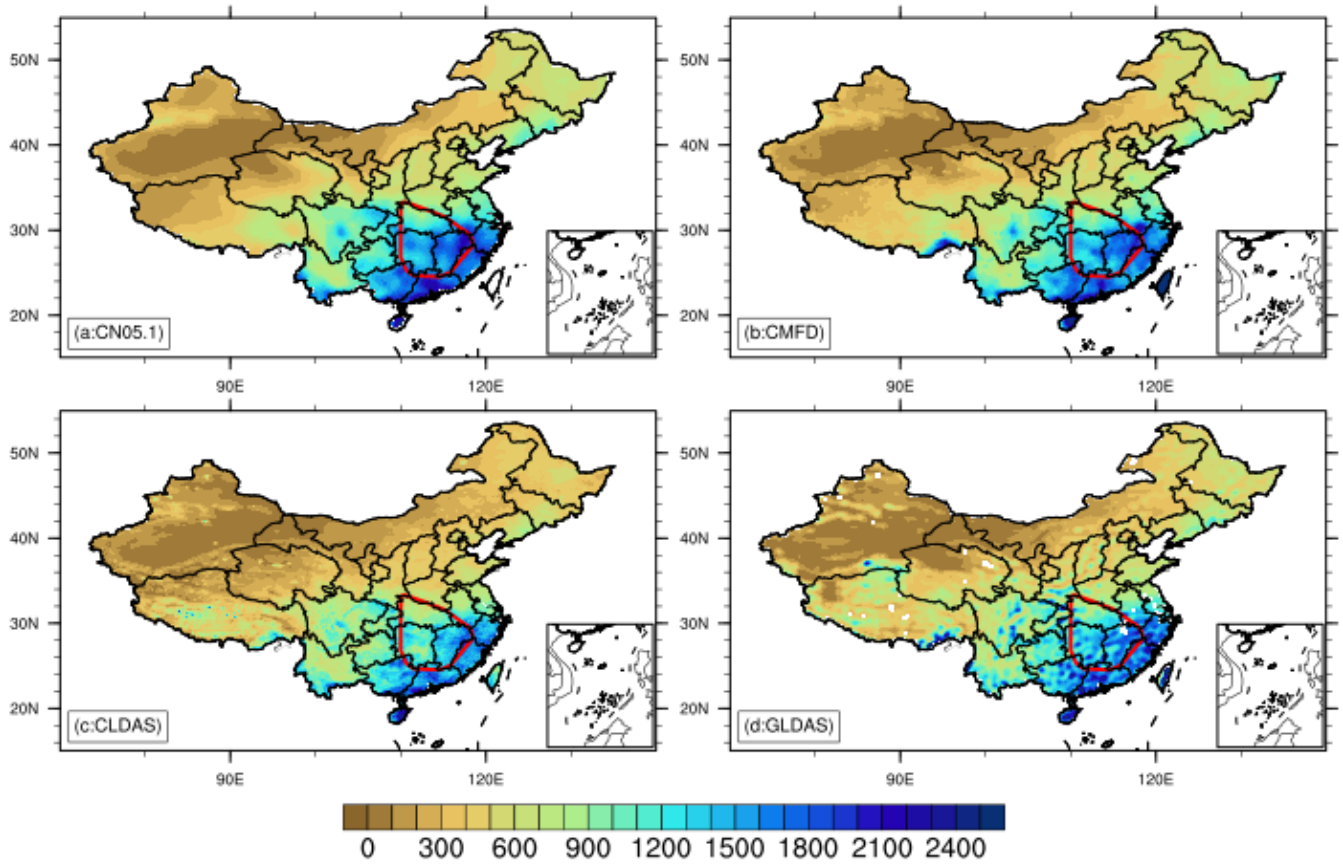

Figure 3: Spatial distribution of annual mean precipitation (over 2008-2014, unit: mm yr-1). (a)-(d) are the result for CN05.1, CMFD, CLDAS, GLDAS, respectively.

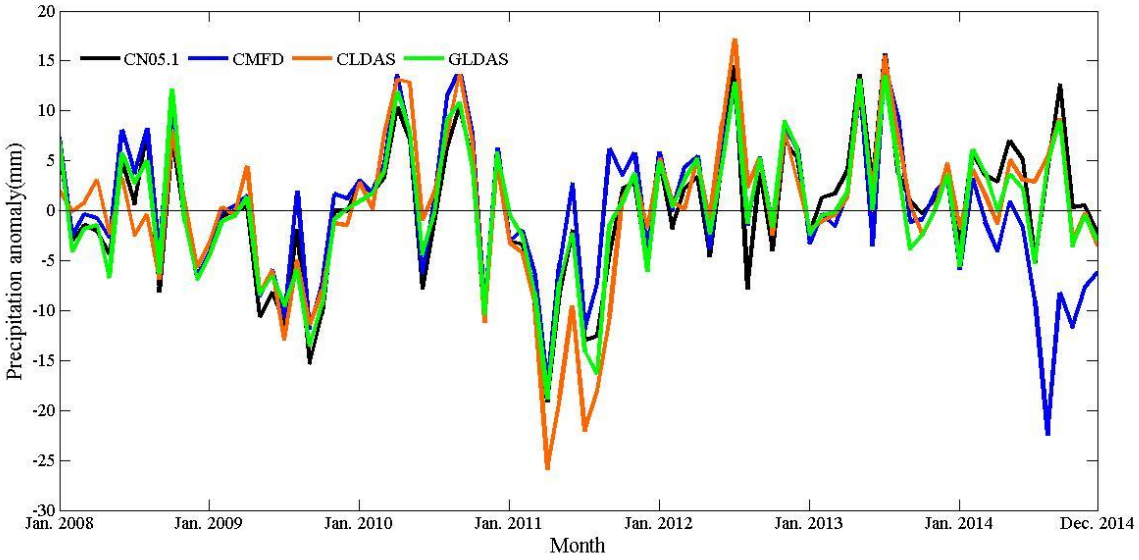

Figure 4: Time series of monthly mean precipitation anomalies from CN05.1 (black), CMFD (blue), CLDAS (orange) and GLDAS (green).

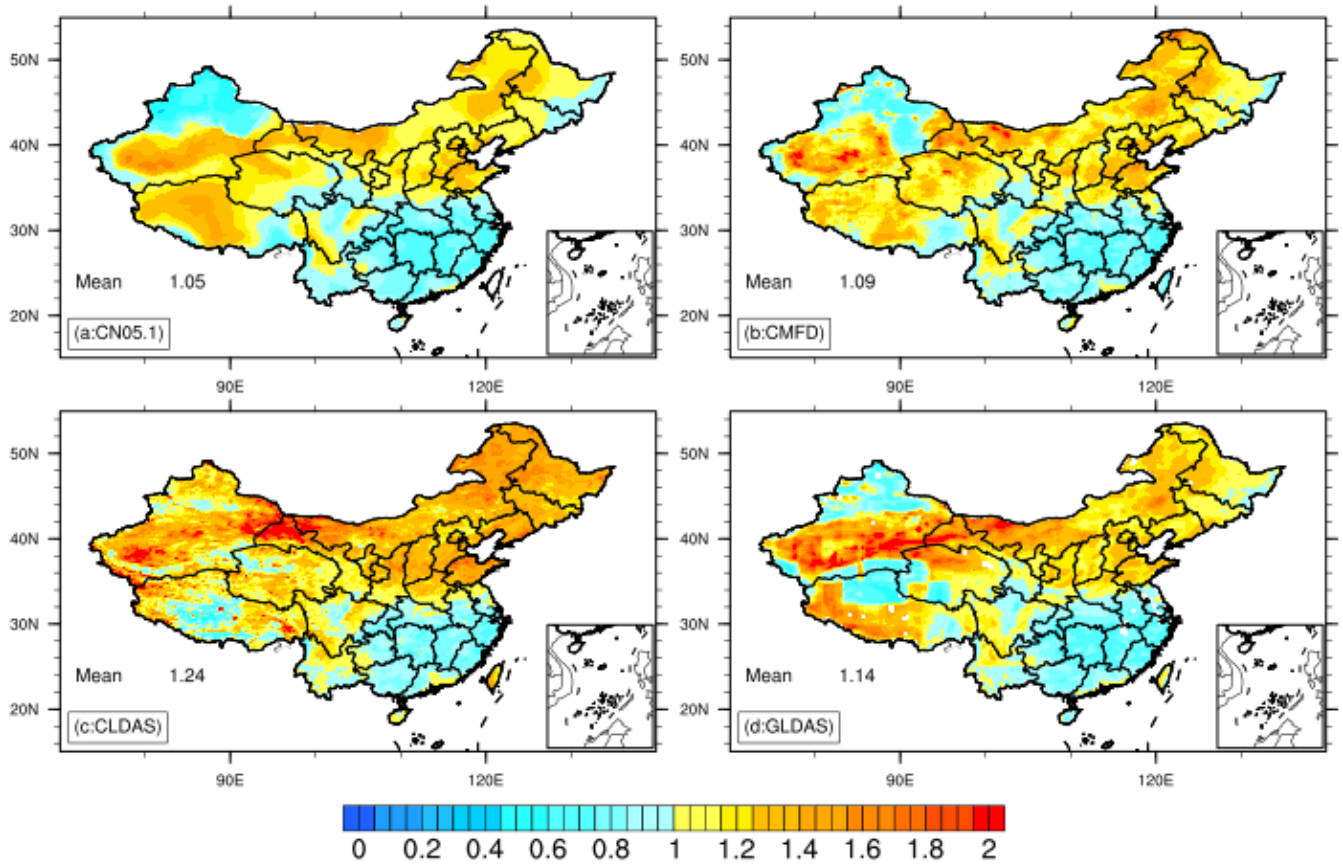

Figure 5: Distribution of TCV of (a) CN05.1, (b) CMFD, (c) CLDAS, and (d) GLDAS precipitation from 2008 to 2014.

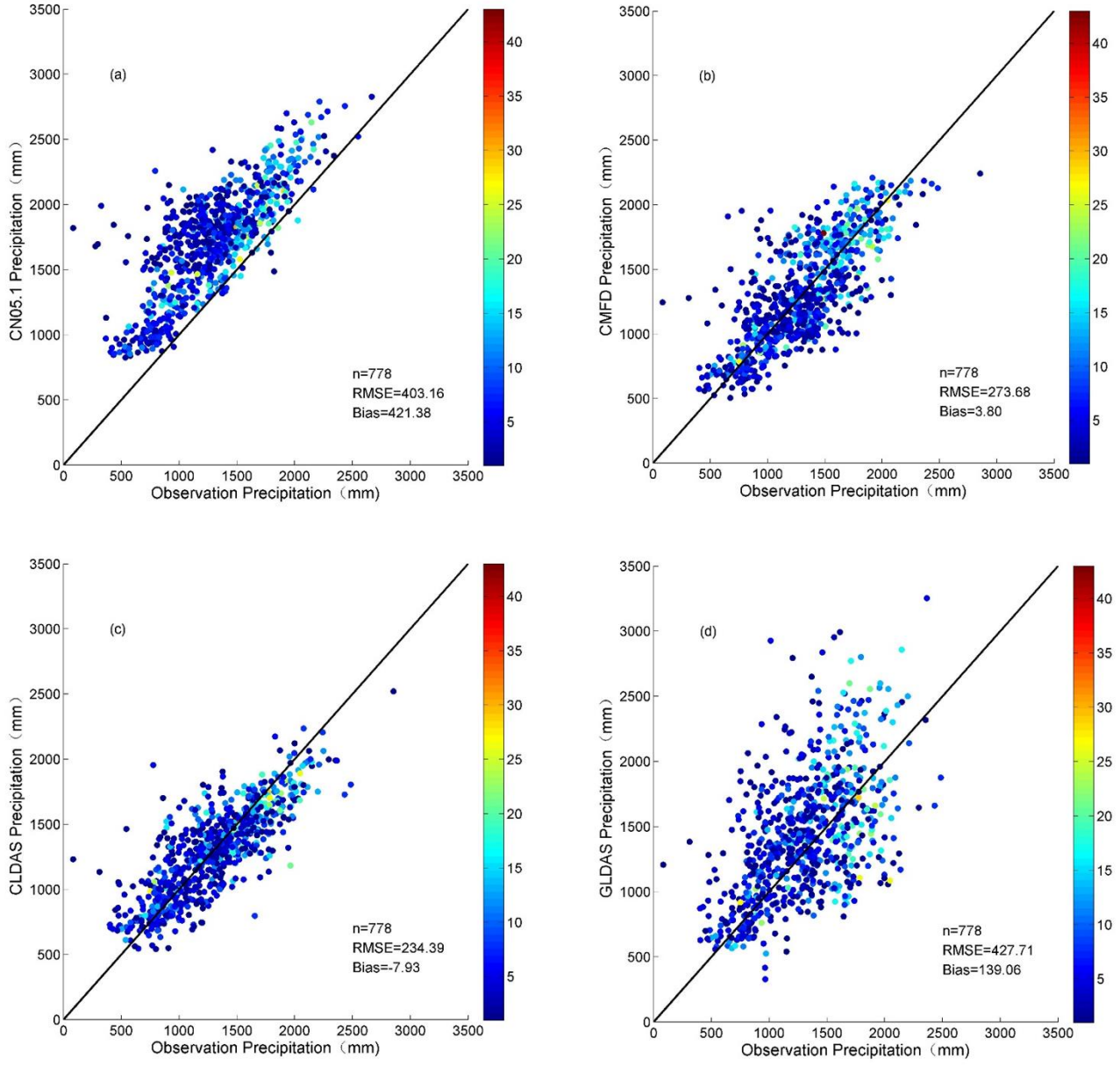

Figure 6: Comparison of the precipitation from (a) CN05.1, (b) CMFD, (c) CLDAS, and (d) GLDAS against MWR rain gauge observation. The color bar on the right indicates the number of MWR rain gauges in one 0.25°×0.25° grid.

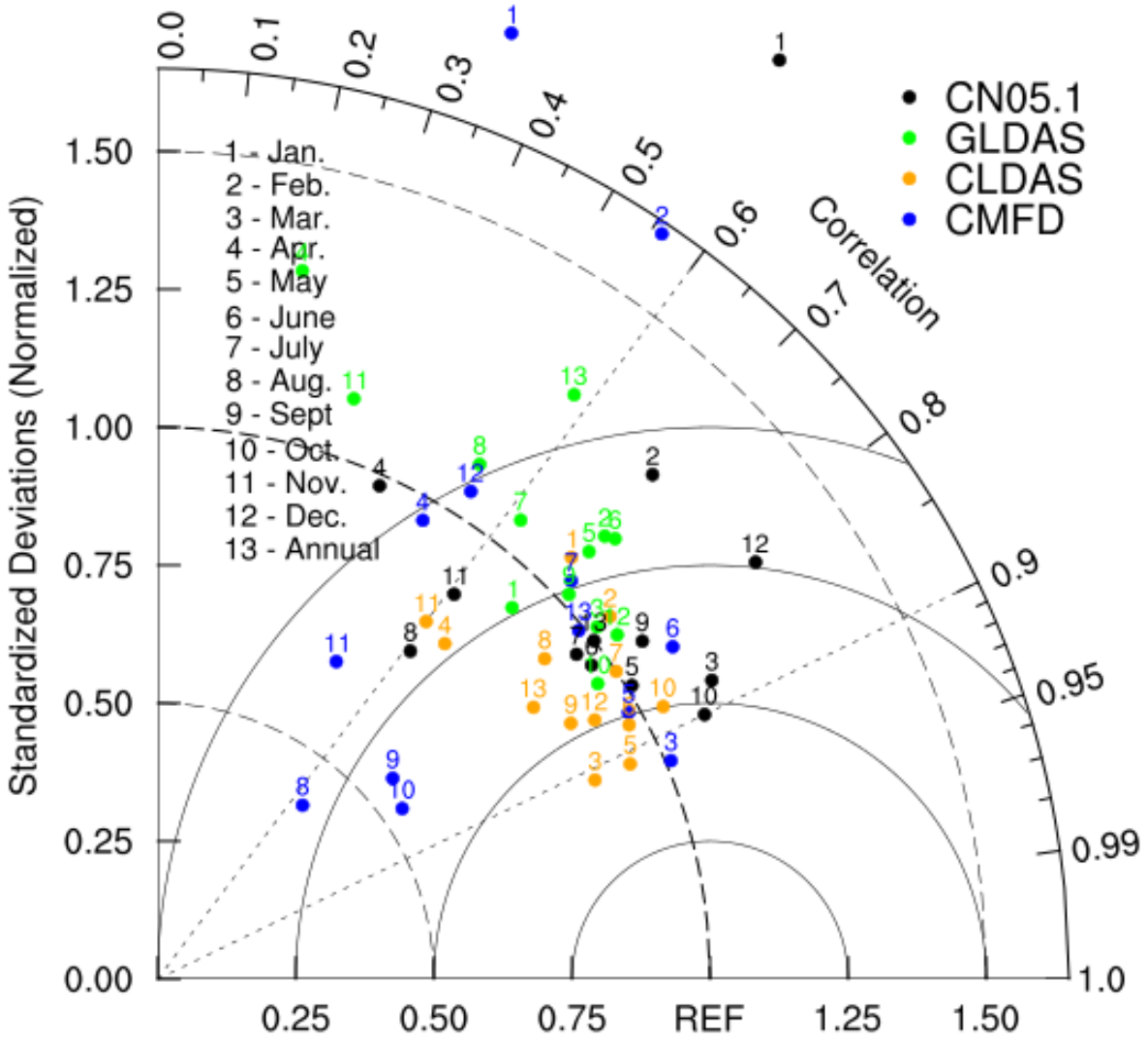

Figure 7: Taylor diagram for the monthly/annual precipitation of CN05.1 (black), CMFD (blue), CLDAS (orange), and GLDAS (green).

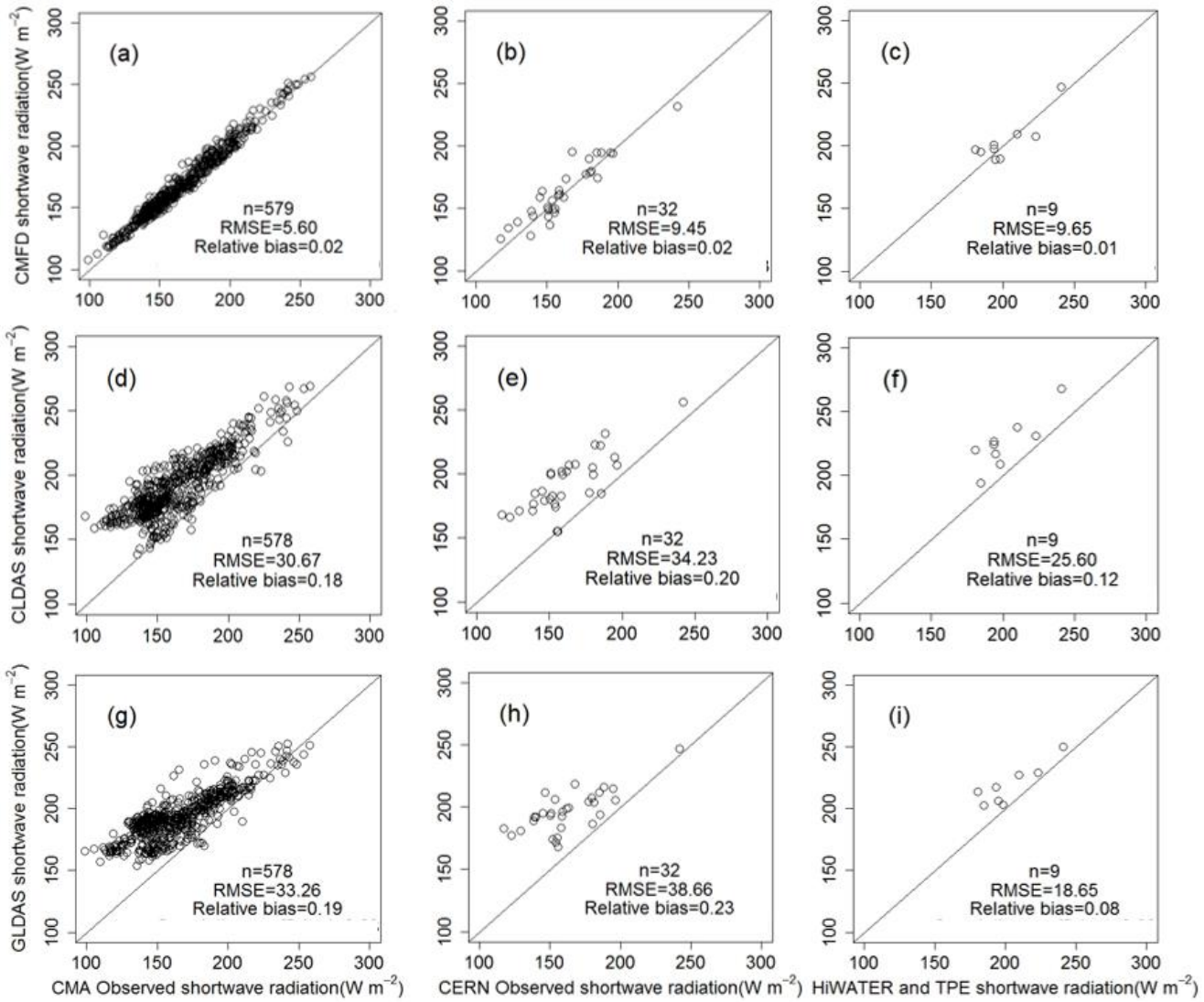

Figure 8: The relationship between shortwave radiation from forcing data and observation data.

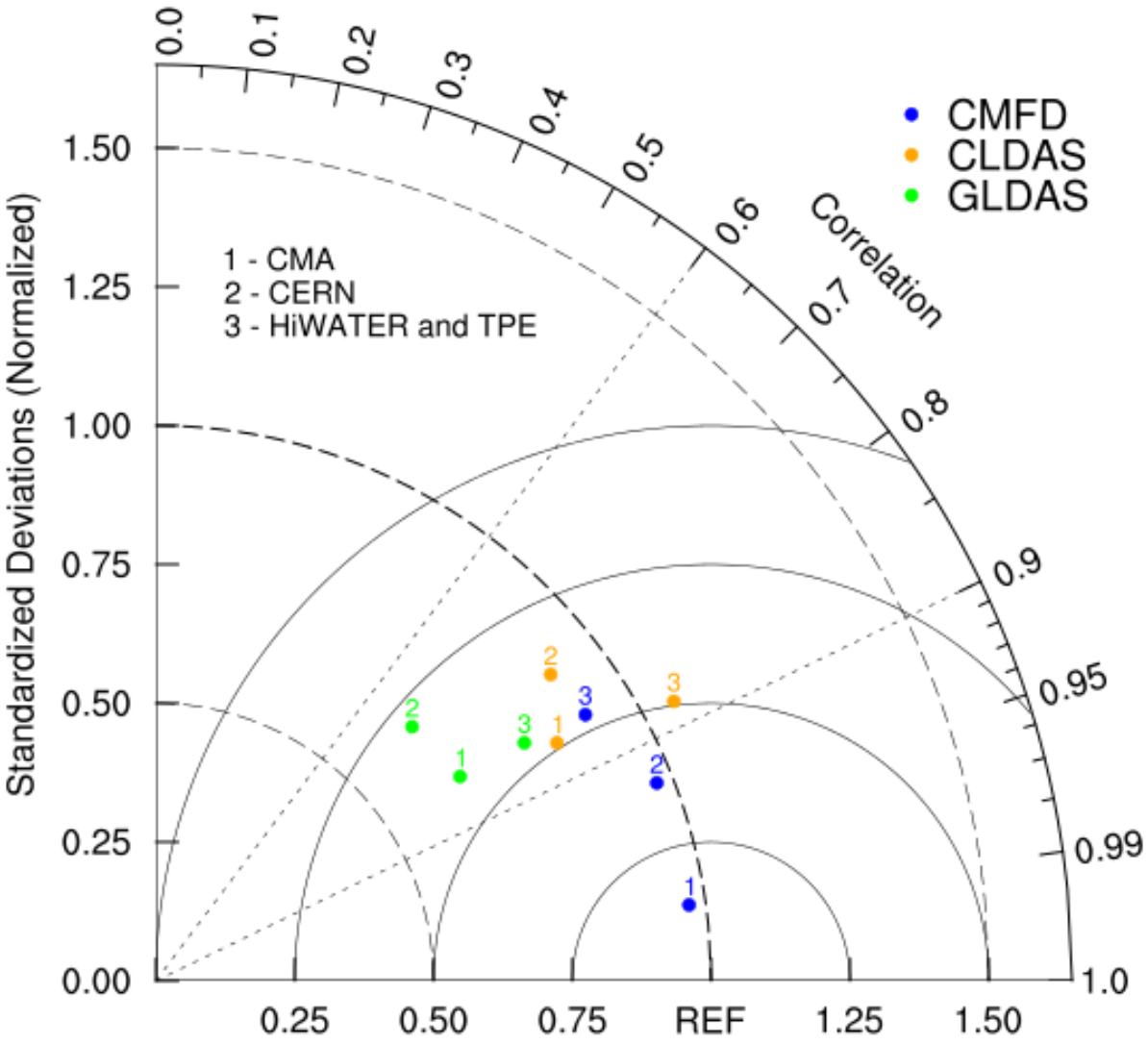

Figure 9: Taylor diagram for the shortwave radiation of CMFD (blue), CLDAS (orange), and GLDAS (green).

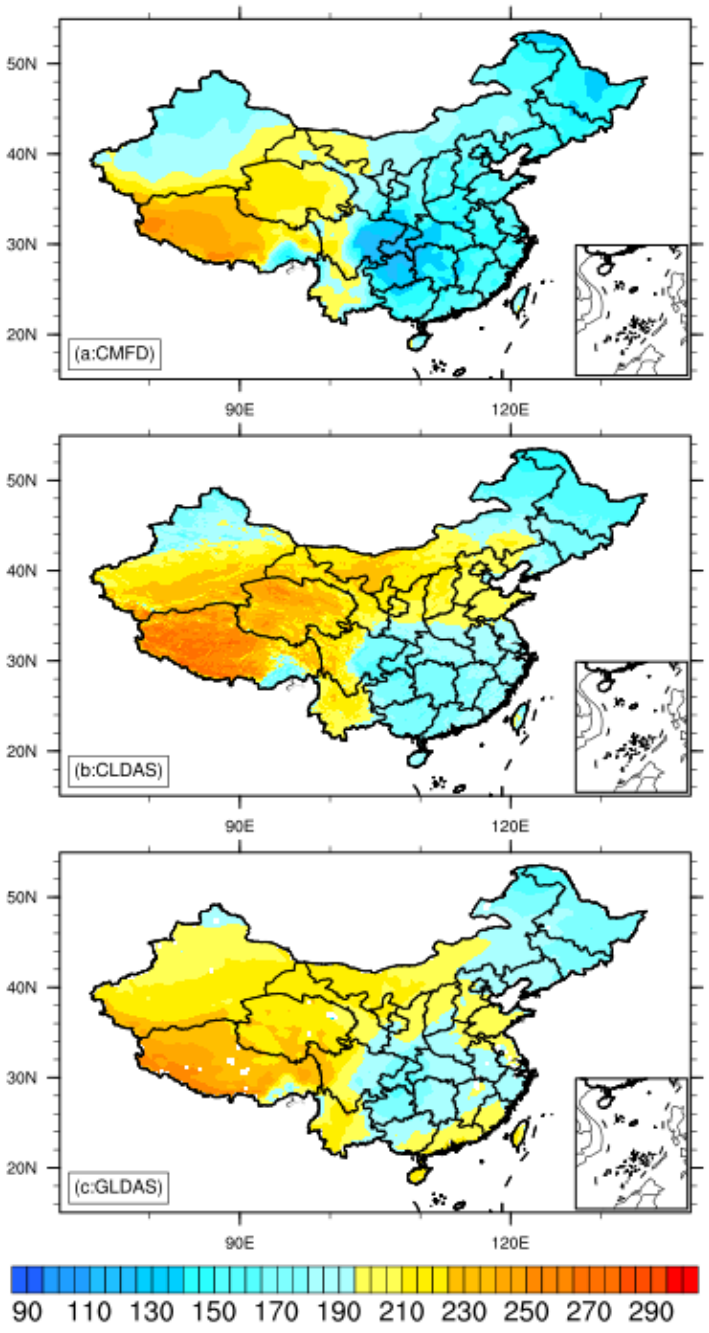

Figure 10: Spatial distribution of annual mean shortwave radiation of three forcing data set (over 2008-2014, unit: W m^{-2}).

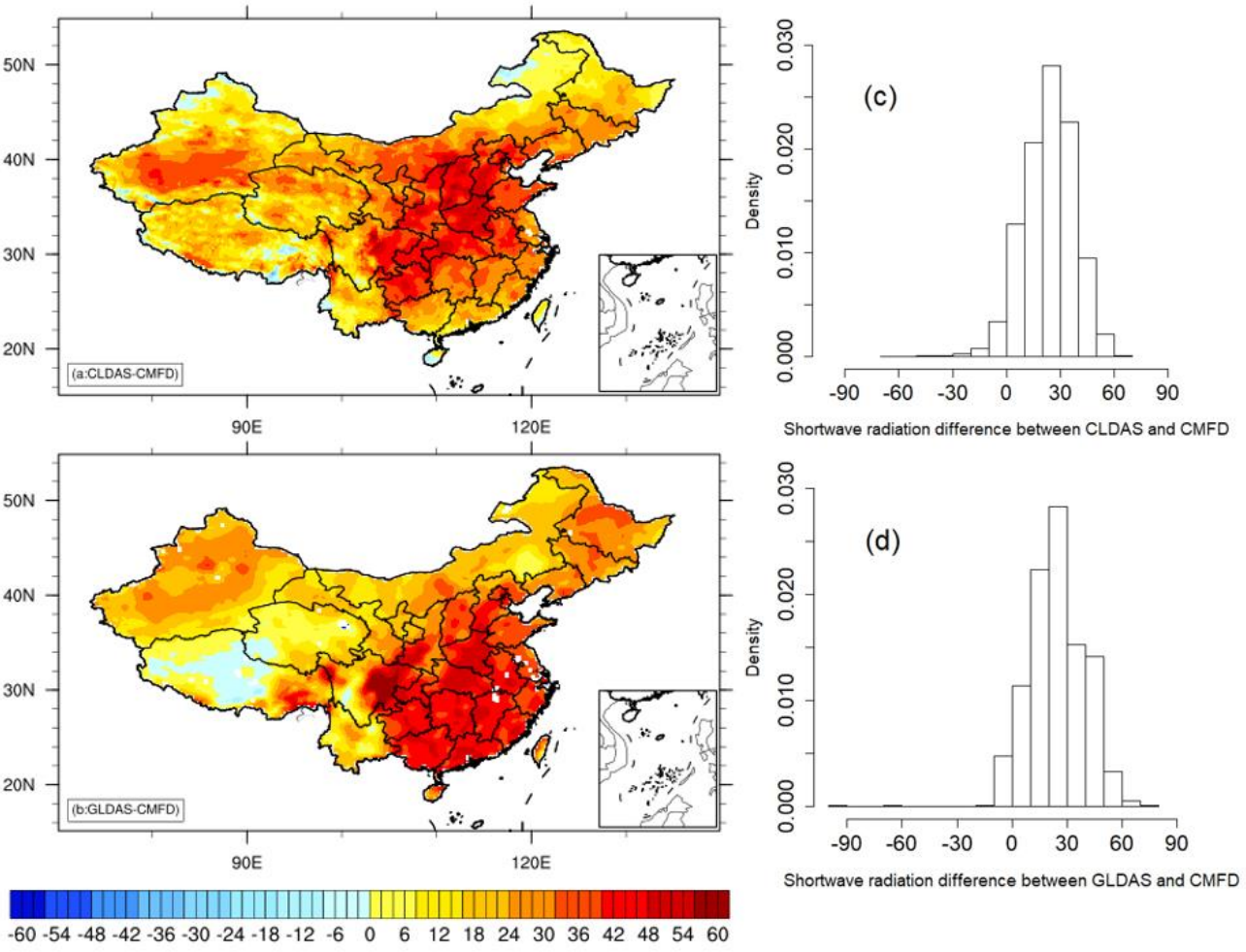

Figure 11: Shortwave radiation difference among three forcing data sets and corresponding histogram from CMFD and CLDAS (a and c) and GLDAS and CMFD (b and d).

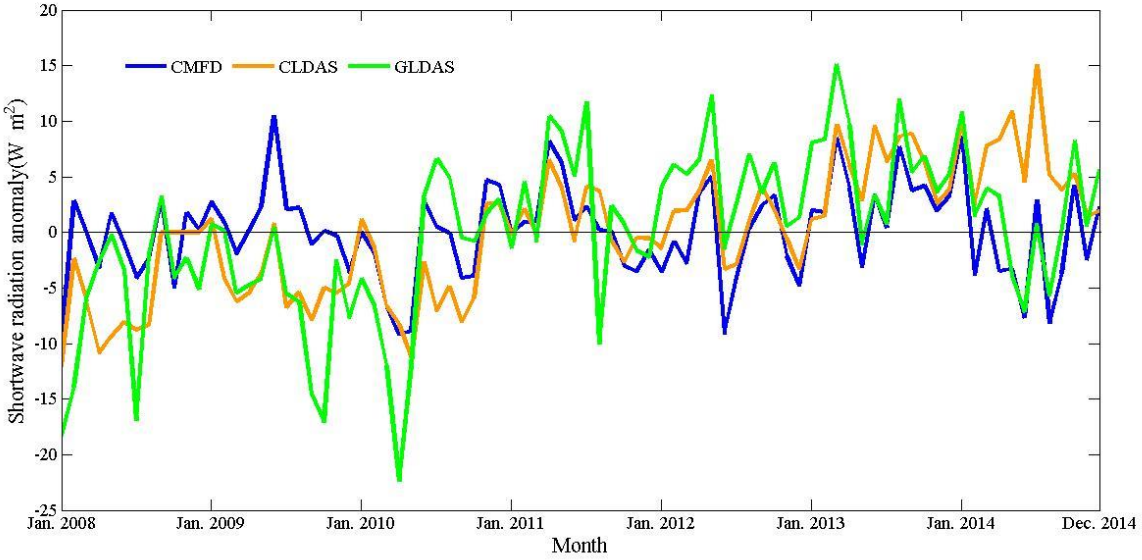

Figure 12: Time series of monthly mean shortwave radiation anomalies from CMFD (blue), CLDAS (orange) and GLDAS (green) (unit: W m^{-2}).

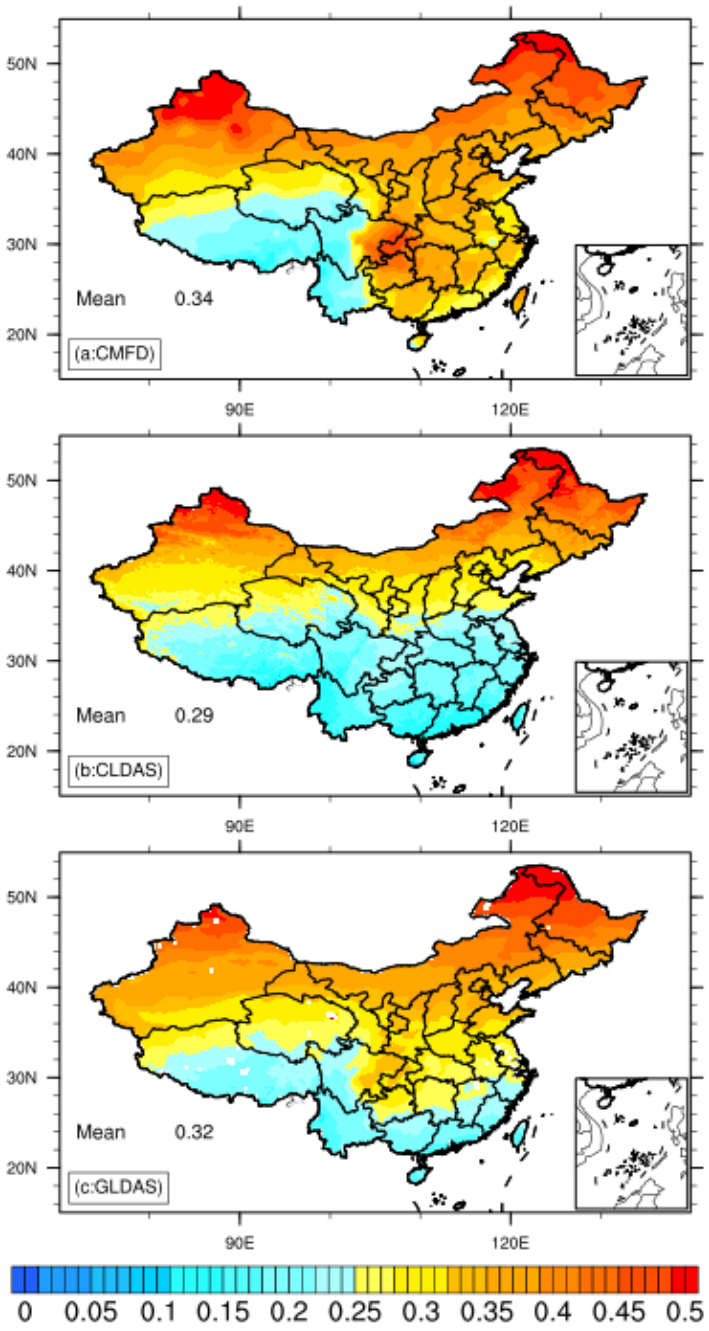

Figure 13: Distribution of TCV of shortwave radiation from forcing data sets.