Comprehensive Evaluation of Global Precipitation Products and Their Accuracy in Drought Detection in Mainland China

HUIHUI ZHANG,[a,c] HUGO A. LOAICIGA,[b] QINGYUN DU,[c] AND TOBIAS SAUTER[a]

[a] *Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany*
[b] *Department of Geography, University of California, Santa Barbara, Santa Barbara, California*
[c] *School of Resource and Environmental Sciences, Wuhan University, Wuhan, China*

(Manuscript received 21 December 2022, in final form 29 July 2023, accepted 1 August 2023)

ABSTRACT: Thorough evaluations of satellite precipitation products are necessary for accurately detecting meteorological drought. A comprehensive assessment of 15 state-of-the-art precipitation products (i.e., IMERG_cal, IMERG_uncal, GSMaP-G, CPC-Global, TRMM3B42, CMORPH-CRT, PERSIANN-CDR, PERSIANN, PERSIANN-CCS, SM2RAIN, CHIRPS, ERA5, ERA-Interim, MERRA-2, and GLDAS) is herein conducted for the period 2010–19 giving special attention to their performance in detecting meteorological drought over mainland China at 0.25° spatial resolution. The cited precipitation products are compared against China's gridded gauge-based Daily Precipitation Analysis (CGDPA) product, derived from 2400 meteorological stations, and their quality is assessed at daily, seasonal, and annual precipitation time scales. Meteorological droughts in the datasets are determined by calculating the standardized precipitation evapotranspiration index (SPEI). The performance of the precipitation products for drought detection with respect to the SPEI is assessed at three time scales (1, 3, and 12 months). The results show that the GSMaP-G outperforms other satellite-based datasets in drought detection and precipitation estimation. The MERRA-2 and the ERA5 are on average closer to the CGDPA reference data than other reanalysis products for precipitation estimation and drought detection. These products capture well the spatial and temporal pattern of the SPEI in southern and eastern China having a probability of detection (POD) above 0.6 and a correlation coefficient (CC) above 0.65. CPC-Global, IMERG, and the ERA5 reanalysis product are ideal candidates for application in western China, especially in the Qinghai–Tibetan Plateau and the Xinjiang Province. Generally, the accuracy of precipitation products for drought detection is improved with longer time scales of the SPEI (i.e., SPEI-12). This study contributes to drought-hazard detection and hydrometeorological applications of satellite precipitation products.

SIGNIFICANCE STATEMENT: The purpose of this study is to comprehensively evaluate the quality of 15 global satellite-based, gauge-based, and reanalysis precipitation products for meteorological drought detection at 1-, 3-, and 12-month time scales. This work systematically evaluates these products' capacity to capture precipitation occurrence and intensity in different seasons. This is followed by a comparison of the precipitation products' performance in drought detection. This work's findings and evaluation results will improve the ability of those who develop precipitation products in identifying error sources and further improving retrieval algorithms. This paper's results will serve as a valuable reference for end users seeking to better understand the application of precipitation products to drought detection.

KEYWORDS: Precipitation; Drought; Satellite observations; Reanalysis data

1. Introduction

Droughts are spatially complex hazards that have severe socioeconomic and environmental impacts (Bevacqua et al. 2021; Dikshit et al. 2022; Zhang et al. 2022). These impacts exacerbate water scarcity, affecting surface water and groundwater resources. Reduced water supply leads to crop failure and degraded aquatic habitats, severely impacting socioeconomic sectors (Mishra and Singh 2011; Bevacqua et al. 2021; Dikshit et al. 2022). Thus, from economic and environmental perspectives there is an urgent need to monitor drought under climatic change (Guo et al. 2022). China, for instance, has experienced extreme drought in the last decades, with the highest seasonal-mean temperature recorded since 1961 (Liu et al. 2022). These droughts inflicted large economic losses with average annual losses of $7 billion between 1984 and 2017, which could rise to $47 billion under global warming of 1.5°C (Su et al. 2018).

Drought prediction and quantification are imperative for reducing adverse environmental impacts and economic losses. Drought assessment necessitates the definition of an appropriate drought index to measure drought severity. Drought can be broadly categorized as meteorological, agricultural, hydrological, and socioeconomic drought (Wilhite and Glantz 1985; Mishra and Singh 2010). Meteorological drought is the precursor of hydrological and agricultural droughts. The severity

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-22-0233.s1.

Corresponding author: Tobias Sauter, tobias.sauter@geo.hu-berlin.de

DOI: 10.1175/JHM-D-22-0233.1

of meteorological drought is often given by the well-known standardized precipitation evapotranspiration index (SPEI), which quantifies the probability of drought occurring for a given climatic event (Vicente-Serrano et al. 2010; Abbasi et al. 2019).

The quality of the meteorological drought detection and monitoring depends mainly on the accuracy of precipitation data (Bai et al. 2019; Zhong et al. 2019). Gauge precipitation data are used as a reference in the assessment of droughts, climate trends, and variability, and for agricultural and hydrological applications (Sun et al. 2018; Zhang et al. 2020). However, gauge observations in mountainous regions are sparse and unevenly distributed, which makes it difficult to accurately assess the spatial distribution of precipitation (Xu et al. 2015). Given this problem, satellite-based and reanalysis precipitation products with wide temporal and spatial coverage are often more useful in drought monitoring than in situ observations. It is therefore important to assess the quality of modern precipitation products in precipitation estimation before considering their potential applications for drought monitoring.

The comprehensive quality assessment of multisource satellite-based and reanalysis precipitation products in precipitation estimation has attracted much attention in recent years. Trinh-Tuan et al. (2019) conducted a comprehensive validation of three satellite precipitation datasets for the period 2001–10, which were the Climate Prediction Center morphing technique (CMORPH), the Global Satellite Mapping of Precipitation (GSMaP) Reanalysis, and the Tropical Rainfall Measuring Mission satellite (TRMM3B42) products over central Vietnam. Derin et al. (2019) compared five precipitation products, namely, the Global Precipitation Measurement (GPM)-based Integrated Multi-satellitE Retrievals (IMERGV05B, IMERGV06B), the CMORPH, the GSMaPV07, and the Multi-Source Weighted-Ensemble Precipitation (MSWEP) V2.2 over mountainous regions worldwide. Liu et al. (2020) found that IMERG products exhibited better performance than the GSMaP and the Climate Hazards Center Infrared Precipitation with Station data (CHIRPS) on Bali Island from 2015 through 2017. These studies evaluated the cited precipitation products based on a few ground stations that are sparsely distributed in mountainous and high-altitude regions. The evaluation of precipitation products based on a few ground stations is prone to uncertainty because observations are sparsely distributed in mountainous and high-altitude regions. The triple collection (TC) method is an ideal candidate for error quantification in those regions with scarce precipitation observations. This method was successfully applied for evaluating soil moisture products (Gruber et al. 2016), leaf area index (Fang et al. 2012), and land water storage (van Dijk et al. 2014) when meteorological observations are lacking.

Previous research has largely focused on the applications or comparisons of a few selected datasets used for precipitation estimation. Few studies provide a comprehensive overview of the existing precipitation products in various topographic settings, time scales, and global climate types. Sun et al. (2018) reviewed several global precipitation datasets with respect to multiple time scales. Their evaluation did not incorporate the latest precipitation products such as IMERGV06B, the fifth generation of the European Centre for the Medium-Range Weather Forecasts (ECMWF) atmospheric reanalyses of the global climate (ERA5), a rainfall dataset derived from soil moisture through the SM2RAIN algorithm (SM2RAIN), and the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). The accurate evaluation of precipitation datasets remains a major challenge. For example, the density and distribution of the station network affect the accuracy of precipitation estimates, and the differences in the coordinated universal time (UTC) of the 24-h accumulated values of daily gauge reports potentially bias the results (Beck et al. 2017). Overall, the knowledge about the quality of precipitation products in complex terrain and multiple time scales, and about the potential error source of product algorithm is very limited, which complicates the practical application and algorithmic improvement of global precipitation products. Uncertainty in the precipitation datasets further propagates to drought detection algorithms and may lead to unreliable conclusions.

Only a few studies have investigated the adequacy of precipitation products for global drought analysis (Golian et al. 2019; Hinge et al. 2021). Tang et al. (2020) evaluated ten popular precipitation products for the period 2000–18 in three typical subregions of China, i.e., the Qinghai–Tibetan Plateau (TB), the Xinjiang Province (XJ), and northeastern China (NE). However, the latter study primarily focused on evaluating the performance of products in precipitation estimation. There has been little comprehensive evaluation of the performance of different global precipitation products, particularly reanalysis precipitation products, to guide future drought monitoring applications. Therefore, a systematic assessment of the accuracy and performance of satellite-based and reanalysis precipitation products in detecting meteorological drought is timely.

All the evaluations performed in this work use China's gridded gauge-based Daily Precipitation Analysis (CGDPA) product derived from 2400 meteorological stations as the reference dataset. This paper presents a comprehensive assessment of the quality of 15 major satellite-based, gauge-based, and reanalysis precipitation products during 2010–19 in different subregions of China at daily, seasonal, and annual scales using classical statistical metrics. The high-altitude and data-sparse subregions (TB and XJ) are further analyzed by applying the multiplicative triple collocation (MTC) method which is independent of the quality of the reference dataset. In addition, this work focuses on investigating precipitation products' reliability in identifying meteorological drought characteristics. The precipitation products' accuracy in drought assessment is evaluated using the SPEI index. The SPEI is calculated at the monthly, seasonal, and annual time scales, denoted by SPEI-1, SPEI-3, and SPEI-12 respectively.

2. Study area and datasets

a. Study area

The study covers mainland China for the period from 2010 to 2019. Seasonal analysis in this work is presented as spring (March–May), summer (June–August), autumn (September–November), and winter (December–February). Most precipitation

falls in the summer, and annual precipitation decrease from the wet southeast to the dry northwest (Wang et al. 2016). About 70% of gauging stations are relatively dense and uniformly distributed in the eastern and southern parts of China, with distances between stations ranging from 20 to 40 km. Fewer gauges are established in the western part of China [Fig. 1a(1)] where the interstation distance ranges between 50 to 100 km. China is divided into eight subregions based on topoclimatic conditions (Tang et al. 2016; Wang et al. 2018). The subregions are shown in Fig. 1b, and they are as follows. 1) The plain region of the Yangzi River (CJ) has a few isolated hills, but most of the plains have low relief and are lower than 45 m above sea level. 2) Southeast China (SE) features a humid subtropical climate with a hot summer and mild winter. The rainiest period of CJ and SE is from April through September, while precipitation decreases sharply in October, even though the weather remains warm. 3) Northern China (NC) lies north of the Qingling–Huaihe line. The latter is a reference line used by geographers to distinguish between northern and southern China, corresponding roughly to the 33rd parallel. Qinling refers to the Qin Mountains, and Huaihe refers to the Huai River. Most of the northern parts of China, including its northwest and northeast regions, feature a temperate continental climate, except for some areas that have a plateau climate. It is cold in winter and warm in summer in the NC subregion, with a large temperature difference between winter and summer, and between day and night. The NC subregion features relatively low precipitation with a maximum in summer. 4) Northwestern China (NW) is bounded approximately by the 400-mm annual isohyet (Shi et al. 2007). 5) NE comprises Heilong Jiang, Jilin, and Liaoning Provinces, and is located at a relatively high altitude. 6) The Qinghai–Tibetan Plateau is known as "the Roof of the World" with an extremely complex environment and high precipitation variability. The plateau is dominated by the plateau mountain climate with a high-altitude arid steppe interspersed with mountain ranges and large brackish lakes. 7) The Yungui Plateau (SW) subregion rises roughly 1000–2500 m above sea level, and the climate patterns in its western and eastern areas differ from each other due to their distinct terrain features. 8) The XJ subregion is distant from the ocean and enclosed by high mountains. It features a continental, dry climate with high interannual variability.

b. *Datasets*

The 15 global precipitation products and CGDPA reference datasets are available to the public (see Table S1 in the online supplemental material). The metadata of the 15 global precipitation products are summarized in Table 1. To match the gauge observations, all satellite-based, gauge-based, and reanalysis products are resampled to 0.25°. The preprocessing methods are divided into two categories based on the resolution of the product. The first linear interpolation method is used for products with resolutions coarser than 0.25°. The second method consists of the area-weighted mean resampling method, which is applied for the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN)-Cloud Classification System product (PCSS) at 0.04° resolution, SM2RAIN at 10 km (nearly 0.1°) resolution, and IMERG and GSMaP at 0.1° resolution. The precipitation estimates are calculated as the area-weighted mean of the data from all 0.04°/0.1° grid cells that are fully or partly contained within the 0.25° cell (Tang et al. 2020).

1) IN SITU OBSERVATION PRODUCT

The daily gridded ground-based precipitation dataset, CGDPA, developed by the National Meteorological Science Data Center of China is used as the reference dataset for the assessment of the satellite-based and reanalysis products. This dataset includes observations from 2400 gauge stations and has high accuracy (Shen and Xiong 2016). This precipitation dataset has undergone rigorous quality control and has been used in several studies to evaluate satellite-based precipitation datasets (Tang et al. 2016; Li et al. 2018; Tang et al. 2020; Zhang et al. 2020). This study assesses various products from 2010 through 2019 at 0.25° × 0.25° grid cells with at least one gauge. All datasets are transformed to the period 0000–2359 UTC (Beck et al. 2017; Baez-Villanueva et al. 2020). Missing days in the CGDPA product are removed from the satellite-based and reanalysis precipitation products. Some satellite and reanalysis products have gauge data (i.e., referring to the gauge data assimilated into products or used to adjust products) integrated into their precipitation estimates. However, our reference dataset (CGDPA) incorporates more rain gauges than these products and permits representative intercomparison. Table 1 presents the overlap between the gauge stations of the precipitation products and the CGDPA. The overlap is insignificant in the 15 global precipitation products so that our evaluation results can be considered independent. Figure 1a displays the main gauge sources with an overlap greater than 20% (GHCN monthly and CPC-Global) and their gauge locations.

2) SATELLITE-BASED AND GAUGE-BASED PRECIPITATION PRODUCTS

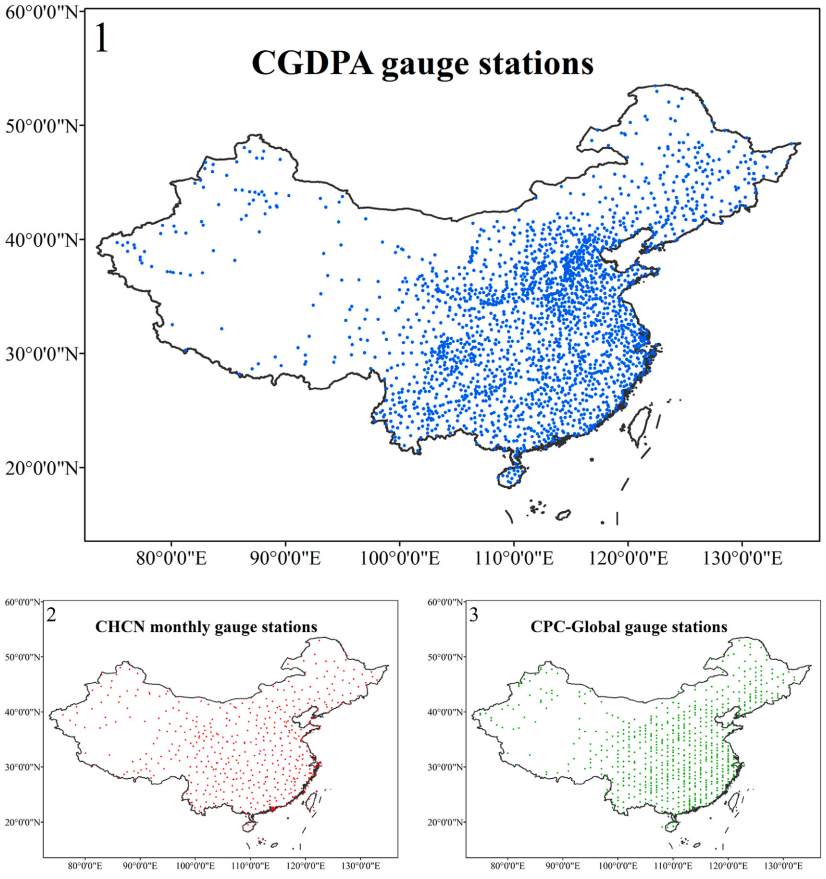
Ten satellite-based and one gauge-based precipitation products are analyzed in this study (see Table 2). This study focuses on gauge-incorporated satellite products since they are in general more accurate than satellite-only products (Tang et al. 2016; Beck et al. 2017). The SM2RAIN product uses soil moisture for obtaining accumulated rainfall estimates based on a "bottom-up" approach. It is included to complement the traditional ("top-down") retrieval approaches (Brocca et al. 2019).

3) REANALYSIS PRECIPITATION PRODUCTS

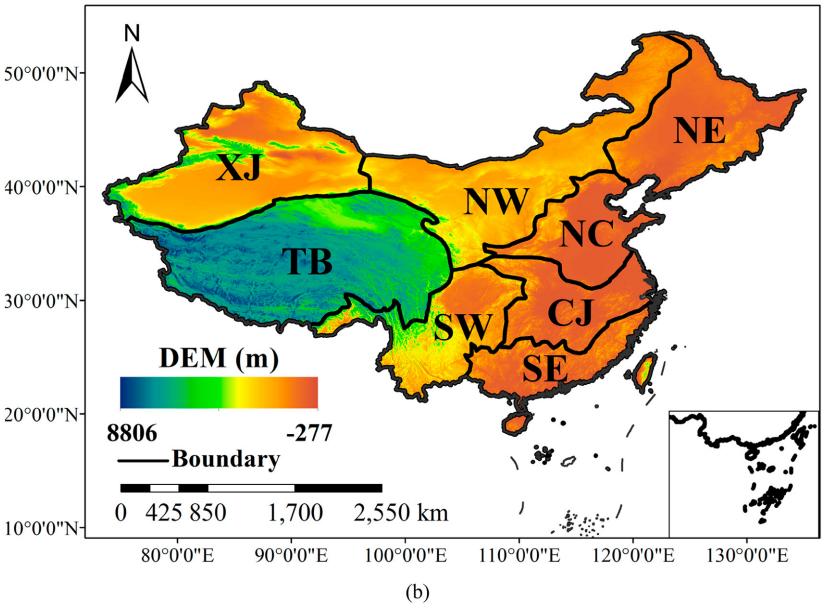
Four reanalysis products (Table 2) are analyzed as they are important sources for estimating precipitation in high-elevation areas and remote regions (Beck et al. 2019). Reanalysis products merge model outputs, remote sensing observations, and in situ measurements through a data assimilation procedure to produce a retrospective estimation of meteorological variables.

4) GRIDDED TEMPERATURE PRODUCTS FOR DROUGHT INDEXING

The drought index (SPEI) is computed from precipitation and air temperature. The SPEI is computed from each precipitation



(a)



(b)

FIG. 1. (a) The spatial distribution of rain gauges in China, 1) CGDPA gauge locations, 2) Global Historical Climatology Network (GHCN) monthly gauge locations, 3) CPC Global Unified Gauge-Based Analysis of Daily Precipitation (CPC-Global) gauge stations; (b) the geographical setting of China's eight subregions.

TABLE 1. Summary of precipitation products that incorporate gauges and their main gauge sources.

Main gauge sources	Precipitation products	Number of gauges used in mainland China	Overlap (%)
China Meteorological Administration (CMA)	CPC-Global	~700	~29.2
Global Telecommunication System (GTS)	CHIRPS	~200	~8.3
GHCN daily	CHIRPS	~228	~9.5
GHCN monthly	CHIRPS	~544	~22.7
Global Summary of the Day (GSOD)	CHIRPS	~251	~10.5
CPC-Global	MERRA-2	~700	~29.2
Global Precipitation Climatology Project (GPCP)	CMORPH-CRT GSMaP GLDAS-2.1 PERSIANN-CDR	—	—
Global Precipitation Climatology Centre (GPCC)	TRMM3B42 IMERG_cal	—	—

product using its respective precipitation values and air temperature data from the ERA5 reanalysis. We consider the ERA5 air temperature as the "ground truth" since previous studies have demonstrated its high quality (Tarek et al. 2020).

3. Methodology

a. Statistical metrics

The classification of precipitation intensities into categories is based on a standard of the World Meteorological Organization (WMO), which was adapted for the study area: 1) 1–3 mm day^{-1} (light precipitation); 2) 3–5 mm day^{-1} (low moderate precipitation); 3) 5–10 mm day^{-1} (high moderate precipitation); 4) 10–20 mm day^{-1} (low heavy precipitation); 5) 20–50 mm day^{-1} (high heavy precipitation); 6) >50 mm day^{-1} (extreme precipitation) (Zhou et al. 2020).

Ten indexes are used to evaluate the performance of target precipitation (i.e., satellite-based, gauge-based, and reanalysis) products against the reference precipitation datasets (i.e., CGDPA) (see Table 3).

The notation for symbols appearing in Table 3 is as follows: n is number of samples; S_i is precipitation of the target dataset at the ith location; G_i is precipitation of the reference dataset at the ith location; $\overline{S}$ is the mean value of the target precipitation at all the locations with rain gauges; $\overline{G}$ is the mean value of the reference precipitation at all the locations; H is the number of positive precipitation events detected in the reference and target datasets, i.e., $G >$ threshold and $S >$ threshold; F is the number of false alarms in which the target dataset detects precipitation but the reference dataset does not, i.e., $G \leq$ threshold and $S >$ threshold; and M is the number of missed events, which is inverse to F, i.e., $G >$ threshold and $S \leq$ threshold. The terms $f(x)$ and $g(x)$ represent the probability distributions of S and G, with probability density functions f and g, respectively; log(.) denotes the natural logarithm; β is the bias ratio between the reference (abbreviated as G) and target (abbreviated as S) datasets; α is the variability ratio; CV is the coefficient of variation; and σ is the standard deviation.

The Pearson correlation coefficient is used to discern the linear statistical association between the reference (i.e., rain gauge) and the target (i.e., satellite and reanalysis) datasets. Though precipitation distribution is non-Gaussian, the Pearson and Spearman correlation coefficient have similar values for the evaluated products. The root-mean-square error (RMSE) is used to assess the overall error characteristics of the datasets. The probability of detection (POD), the false alarm ratio (FAR), and the critical success index (CSI) measure the precipitation occurrence detection capability of datasets (Diem et al. 2014; Zhou et al. 2020). The threshold for calculating the CSI and defining rainfall/nonrainfall events is set at 1 mm day^{-1} as in many other studies, i.e., >1 mm day^{-1} represents the occurrence of a rainfall event (Mantas et al. 2015; Zhou et al. 2020). The Kullback–Leiber divergence (KLD) is applied to measure the similarity between two probability distributions. This algorithm has been applied in image matching and to estimate the accuracy of satellite precipitation products (Prakash et al. 2018; Zhang et al. 2020). A detailed demonstration of the KLD method can be found in Zhang et al. (2020). The Kling–Gupta efficiency (KGE) statistic is also applied, which combines the contributions of correlation, bias, and variability terms (Gupta et al. 2009; Kling et al. 2012). The KGE has been widely used to evaluate the performance of precipitation products and in hydrological applications (Beck et al. 2017; Zambrano-Bigiarini et al. 2017; Wang et al. 2018; Baez-Villanueva et al. 2018, 2020). There are two advantages of the KGE evaluation index. First, it decomposes the total performance into three components (correlation, bias, and variability); thus, the mismatches between the reference and evaluated product can be better understood. Second, compared to the RMSE, it does not assign disproportional weights to mismatches in high precipitation values (Zambrano-Bigiarini et al. 2017; Baez-Villanueva et al. 2018, 2020). The Taylor diagram (Taylor 2001), the probability density function (PDF), and the performance diagram (Roebber 2009) are also applied in this work to demonstrate the regional and seasonal error characteristics of the precipitation datasets (Nashwan et al. 2020). The Taylor diagram integrates the CC, the standard deviation (STD), and the centered root-mean-square deviation (RMSD). The performance diagram integrates the CSI, the POD, the frequency bias (POD divided by SR), and the success ratio (1 − FAR). The Taylor diagram and the PDF focus on precipitation intensity evaluation and have been implemented in several studies to evaluate the quality of satellite precipitation products (Tang et al. 2020; Zhou et al. 2020; Noor et al. 2021).

TABLE 2. Summary of global satellite-based, gauge-based, and reanalysis precipitation products. Note: G = gauge; S = satellite; R = reanalysis. The letter G refers to gauge observations that are assimilated into precipitation products or used to correct them; G does not represent soil moisture gauge observations and precipitation observations that are only used for the calibration and validation of the precipitation products.

Dataset (abbreviation)	Full name of the dataset	Resolution	Period	Coverage	Reference	Data sources
			Satellite-based precipitation dataset			
IMERG_cal	Integrated Multi-satellitE Retrievals for GPM Final run V06B—The calibrated product based on the GPCC monthly gauge analysis	0.1°/0.5 h	June 2006–present	60°S–60°N	Huffman et al. (2019)	G, S
IMERG_uncal	Integrated Multi-satellitE Retrievals for GPM Final run V06B—the microwave–infrared estimates without gauge adjustment	0.1°/0.5 h	June 2006–present	60°S–60°N	Huffman et al. (2019)	G, S
TRMM3B42	TRMM Multisatellite Precipitation Analysis (TMPA) 3B42 V7	0.25°/3 h	December 1997–December 2019	50°S–50°N	Huffman et al. (2007)	G, S
GSMaP	Gauge-adjusted Global Satellite Mapping of Precipitation V6	0.1°/1 h	January 2000–present	60°S–60°N	Mega et al. (2014)	G, S
CMORPH	Climate Prediction Center (CPC) morphing technique bias corrected (CRT)	0.25°/3 h	January 1998–present	60°S–60°N	Joyce et al. (2004)	G, S
CHIRPS	Climate Hazards Center Infrared Precipitation with Stations	0.05°/1 day	January 1981–present	50°S–50°N	Funk et al. (2015)	G, S, R
SM2RAIN	SM2RAIN based on ASCAT Soil Moisture Observations	10 km/1 day	January 2007–August 2019	Global	Brocca et al. (2019)	S
PERSIANN	Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks	0.25°/1 day	March 2000–present	60°S–60°N	Nguyen et al. (2019)	S
PCSS	PERSIANN-Cloud Classification System	0.04°/1 day	January 2003–present	60°S–60°N	Nguyen et al. (2019)	S
PCDR	PERSIANN-Climate Data Record	0.25°/1 day	January 1983–present	60°S–60°N	Ashouri et al. (2015)	G, S
			Gauge-based precipitation dataset			
CPC-Global	CPC Global Unified Gauge-Based Analysis of Daily Precipitation	0.5°/1 day	January 1979–present	89.75°N–89.75°S, 0.25°–359.75°E	Xie et al. (2007)	G
			Reanalysis precipitation product			
MERRA-2	Modern-Era Retrospective Analysis for Research and Applications, version 2 (FLX)	0.5° × 0.625°/1 h	January 1980–present	Global	Gelaro et al. (2017)	S, R
GLDAS-2.1	Global Land Data Assimilation System (Noah model)	0.25°/3 h	January 2000–present	60°S–90°N	Rodell et al. (2004)	G, S, R
ERA5	Fifth generation of ECMWF atmospheric reanalyses of the global climate	0.25°/1 h	January 1979–present	Global	C3S (2017)	S, R
ERA-Interim	ECMWF interim reanalysis	0.75°/3 h	January 1979–August 2019	Global	Dee et al. (2011)	S, R

TABLE 3. The statistical indexes used in the evaluation of precipitation products.

Statistical index	Equation	Optimal value	Unit	Index
Pearson correlation coefficient (CC)	$\mathrm{CC}=\frac{\sum_{i=1}^{n}(S_i-\overline{S})(G_i-\overline{G})}{\sqrt{\sum_{i=1}^{n}(S_i-\overline{S})^2}\sqrt{\sum_{i=1}^{n}(G_i-\overline{G})^2}}$	1	—	(1)
Root-mean-square error (RMSE)	$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(S_i-G_i)^2}{n}}$	0	mm	(2)
Probability of detection (POD)	$\mathrm{POD}=\frac{H}{H+M}$	1	—	(3)
False alarm ratio (FAR)	$\mathrm{FAR}=\frac{F}{H+F}$	0	—	(4)
Critical success index (CSI)	$\mathrm{CSI}=\frac{H}{H+M+F}$	1	—	(5)
Kling–Gupta efficiency (KGE)	$\mathrm{KGE}=1-\sqrt{(\mathrm{CC}-1)^2+(\beta-1)^2+(\alpha-1)^2}$	1	—	(6)
Kullback–Leiber divergency (KLD)	$\mathrm{KL}(f,g)=\int_{-\infty}^{+\infty}f(x)\log\frac{f(x)}{g(x)}dx$	0	—	(7)
Centered root-mean-square deviation (RMSD)	$\mathrm{RMSD}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}[(G_i-\overline{G})-(S_i-\overline{S})]^2}$	0	mm	(8)
Bias ratio (β)	$\beta=\frac{\overline{S}}{\overline{G}}$	1	—	(9)
Variability ratio (α)	$\alpha=\frac{\mathrm{CV}_S}{\mathrm{CV}_G}=\frac{\sigma_S/\overline{S}}{\sigma_G/\overline{G}}$	1	—	(10)

b. Error decomposition

This work decomposes the total error (i.e., mean square error) into systematic and random components, a procedure that has proven effective in characterizing the errors of satellite precipitation products (Willmott 1981; AghaKouchak et al. 2012). The systematic (P_{Syst}; %) and random (P_{Rand}; %) errors are derived from the total (P_{tot}; mm) error as follows:

$$P_{\mathrm{tot}}=\frac{1}{n}\sum_n(S-G)^2=\frac{1}{n}\sum_n(S^*-G)^2+\frac{1}{n}\sum_n(S-S^*)^2, \quad (11)$$

$$P_{\mathrm{Syst}}=\sum_n(S^*-G)^2\Big/\sum_n(S-G)^2, \quad (12)$$

$$P_{\mathrm{Rand}}=\sum_n(S-S^*)^2\Big/\sum_n(S-G)^2, \quad (13)$$

where S and G denote target precipitation datasets, and reference precipitation datasets, respectively. The term S^* is defined by the least squares linear regression function $S^*=a\times G+b$, with a and b denoting the slope and intercept, respectively.

c. Cross evaluation in gauge station sparse areas

The TC method is proposed to estimate the RMSE of three independent inputs in the absence of "ground truth" data. The MTC method is an extended TC approach to derive the RMSE and CC of a triplet at the same time with better performance than the TC method in characterizing precipitation products error (Alemohammad et al. 2015; Li et al. 2018; Tang et al. 2020; Lu et al. 2021). Testing a diverse combination of products serves as a basis for evaluating the sensitivity of the MTC method to the composition of the triplet. The MTC is calculated according to

$$\begin{cases}\sigma_{r_1}^2=C_{11}-\dfrac{C_{12}C_{13}}{C_{23}}\\ \sigma_{r2}^2=C_{22}-\dfrac{C_{12}C_{23}}{C_{13}}\\ \sigma_{r3}^2=C_{33}-\dfrac{C_{13}C_{23}}{C_{12}}\end{cases}, \quad (14)$$

$$\begin{cases}\rho_{t,1}^2=\dfrac{C_{12}C_{13}}{C_{11}C_{23}}\\ \rho_{t,2}^2=\dfrac{C_{12}C_{23}}{C_{22}C_{13}}\\ \rho_{t,3}^2=\dfrac{C_{13}C_{23}}{C_{33}C_{12}}\end{cases}, \quad (15)$$

where C_{ij} ($i=1,2,3$; $j=1,2,3$), $\sigma_{r_i}^2$ and $\rho_{t,i}^2$ denote, respectively, the covariance between products i and j, the RMSE, and the CC. More details about the MTC method can be found in Duan et al. (2021). This work applies the MTC method to the precipitation products evaluated in the XJ and TB subregions where in situ gauge stations are scarce (see Fig. 1).

d. The SPEI index

The SPEI index considers precipitation and potential evapotranspiration (PET) in determining the onset, duration,

TABLE 4. Drought classification based on SPEI values.

Drought category	SPEI value
Moderate drought	$-1.0 < \text{SPEI} \leq 0$
Severe drought	$-1.5 < \text{SPEI} \leq -1.0$
Extreme drought	$\text{SPEI} \leq -1.5$

and magnitude of drought conditions (Vicente-Serrano et al. 2010). The SPEI is a useful index for meteorological drought detection (Dikshit et al. 2021). This work extends the SPEI procedure from the gauge-station scale to the regional scale. The latitude, total monthly precipitation, and mean temperature information are used to calculate the SPEI index. PET is estimated based on the Thornthwaite equation (Thornthwaite 1948). The SPEI is computed using its respective precipitation values and air temperature data from the ERA5 reanalysis for each precipitation product.

The SPEI is calculated at the 0.25° spatial scale for three time scales (i.e., $n = 1$, 3, and 12 months). The SPEI value serves to categorize the severity degree of drought. The classification of the SPEI values is displayed in Table 4 (Shiru et al. 2018).

4. Results

a. *Overall performance at a daily scale*

The boxplots of five metrics (CC, RMSE, KLD, KGE, and CSI) over mainland China at the daily scale are shown in Fig. 2. The lower and upper edge of the box is the 25th and 75th percentiles, and the horizontal line in the box is the median. The spatial distribution of KGE indexes of 15 precipitation products in various regions of China is depicted in Fig. 3.

In summary, the metrics presented in Fig. 2 demonstrate that the CPC-Global gauge-based product has the highest accuracy among the 15 global precipitation products. In the following, this paper presents the performance analysis of satellite-based and reanalysis precipitation products separately.

1) SATELLITE-BASED AND GAUGE-BASED PRECIPITATION PRODUCTS

The red color in Fig. 2 represents satellite-based and CPC-Global gauge-based precipitation products. It is seen in Fig. 2 that most satellite-based products moderately correlate ($0.40 < \text{CC} < 0.80$) with the CGDPA reference datasets, while PCSS (CC = 0.14) and PERSIANN (CC = 0.23) have a poor correlation. The performance of the IMERG_cal product is better than that of the IMERG_uncal product with a smaller KLD. However, there is only a slight improvement with respect to the CSI (IMERG_cal: 0.41; IMERG_uncal: 0.39). This implies that the gauge-corrected IMERG product is more effective in improving intensity estimation than occurrence detection. As for the PERSIANN family of precipitation products, namely, PERSAINN, PCSS, and PCDR, the performance metrics of PCDR are better than those of PCSS and PERSIANN. This is not surprising because PCDR is a bias-adjusted product utilizing the GPCP data. This highlights the ability of bias correction to improve the accuracy of satellite-

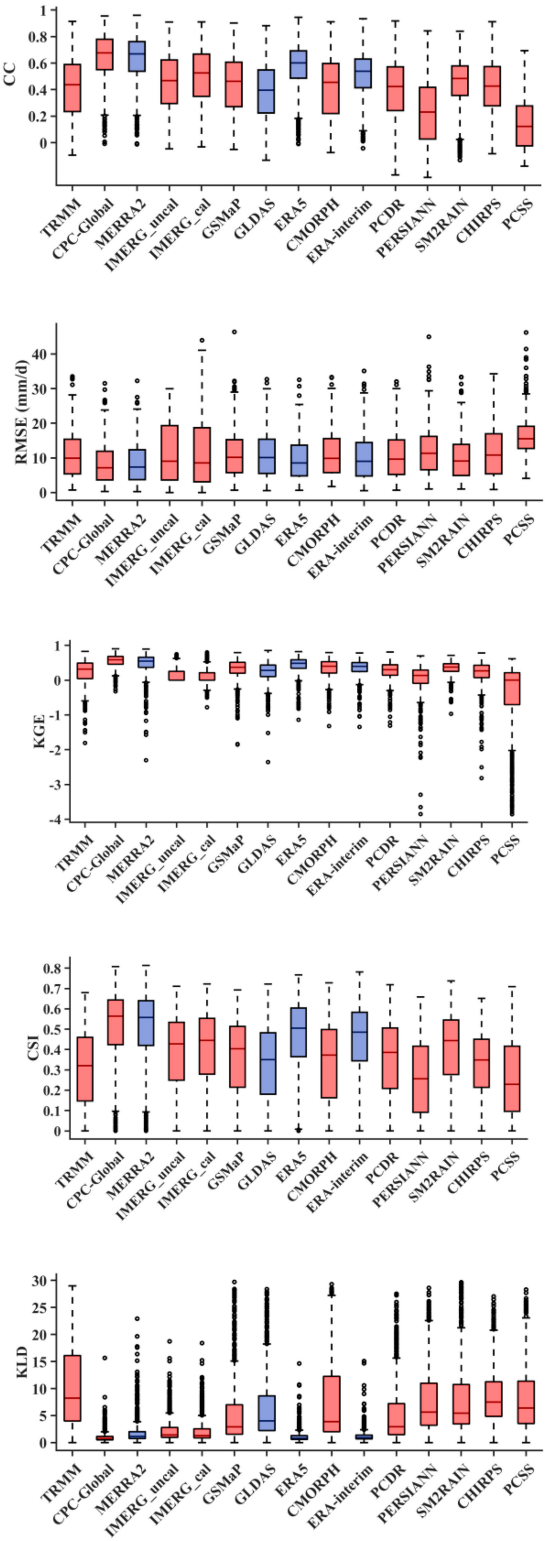


FIG. 2. Red (satellite- and gauge-based) and blue (reanalysis) boxplots of five metrics comparing 15 products with CGDPA datasets at the daily scale from 2010 through 2019. The bottom and top edges of the boxes in the figure represent the 25th and 75th percentiles, respectively, and the horizontal line in the middle of the boxes represents the median. The dots are the outliers for a single 0.25° grid cell.

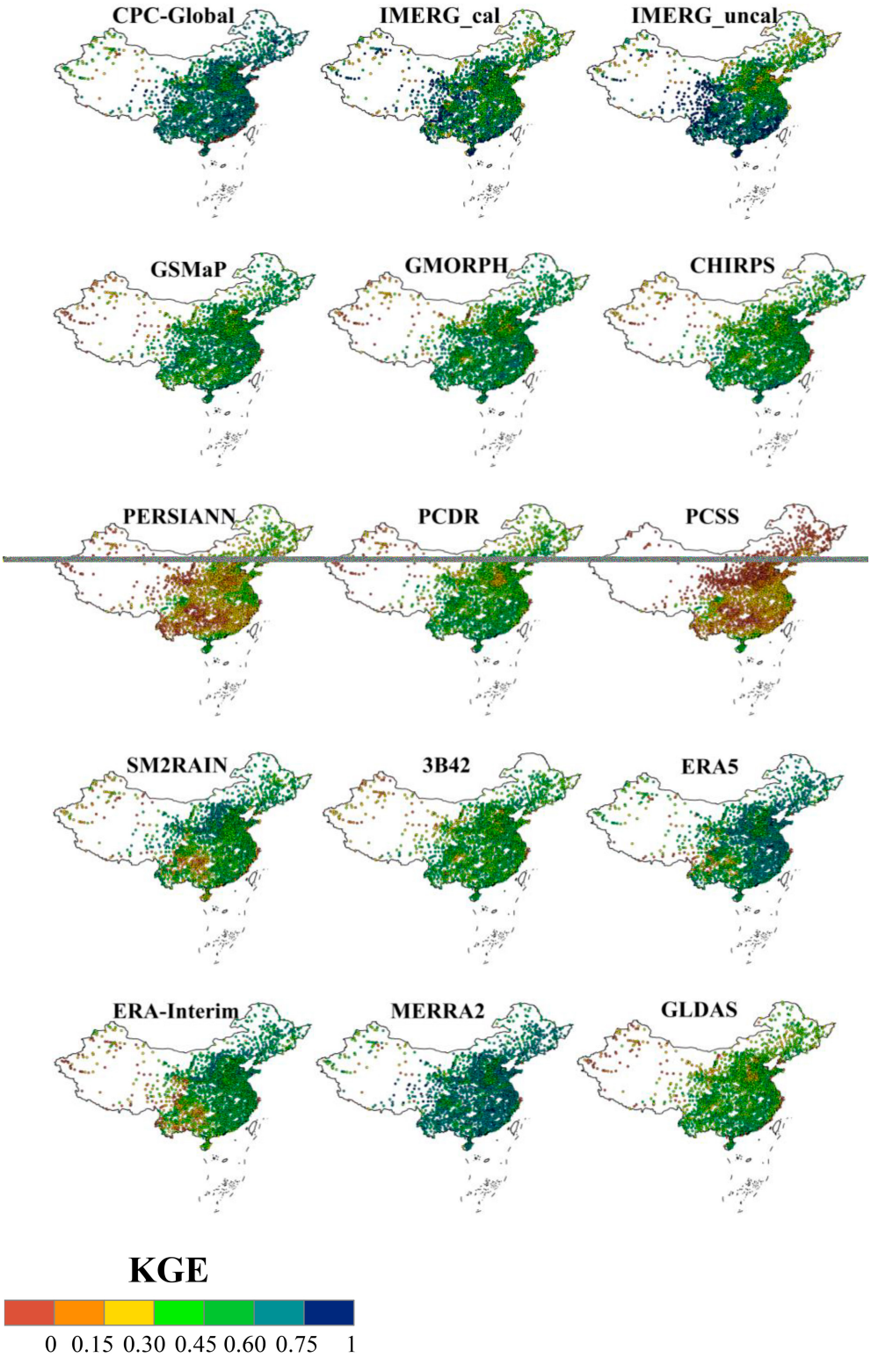


FIG. 3. Spatial distributions of KGE at the daily scale from 2010 to 2019 for various precipitation products in mainland China only displaying grids with gauge stations.

based precipitation. The comprehensive performance of the PERSIANN product (RMSE: 11.92 mm day^{-1}; KLD: 7.85) is superior to that of the PCSS (RMSE: 16.23 mm day^{-1}; KLD: 8.32) product, but the time lag of the PERSIANN product is 2 days compared to the nearly real-time estimation of the PCSS product. The lag time hinders the capacity of the PERSIANN product in near-real-time applications such as flood simulation and drought monitoring at a daily scale. The CHIRPS and PCDR products are both long-term climatological data records that could be used in the studies of atmosphere and weather patterns through time. The metrics of the CHIRPS (KGE: 0.20; CSI: 0.33; RMSE: 11.49 mm day^{-1}; KLD: 8.63) and PCDR (KGE: 0.26; CSI: 0.36; RMSE: 10.58 mm day^{-1}; KLD: 5.27) products exhibit minor differences, yet the PCDR is slightly better than CHIRPS's, which could be a better choice for climatological studies in China.

Figure 3 shows that the accuracy of precipitation products is questionable, particularly in arid regions such as the TB and XJ subregions. Among the 10 satellite-based and gauge-based precipitation products, the performance of the PERSIANN family is relatively poor. The overall quality of the PERSIANN family products is the worst in terms of the KGE values, with approximately 90% of the domain exhibiting KGE values lower than 0.50, particularly for the PCSS product that has poor accuracy. However, the significant advantages of PCSS products are their short time lag (1 h) and high spatial resolution (0.04°). The PCSS product is effective in monitoring the variability of near-real-time precipitation events. The PCDR is the best in the PERSIANN family products. This is probably because PCDR uses GPCP data for postadjustment that contains long-term historical precipitation datasets (Table 1). The PERSIANN products use passive microwave precipitation (PMW) data for training, but the PMW data lack the predictive skill to detect precipitation in winter.

The SM2RAIN product exhibited the second worst performance. It appears that the SM2RAIN product is generated based on the unstable link between soil moisture and precipitation. The KGE value of the SM2RAIN product is the largest in Inner Mongolia (KGE $>$ 0.60) where precipitation is low, and the KGE is the smallest in the SE (KGE $<$ 0.45) subregion where precipitation is abundant. This is due to the fact that the soil in humid areas becomes easily saturated and cannot reveal the characteristics of precipitation changes. Besides, the snow and the permafrost in the TB region significantly influence the quality of the SM2RAIN product.

The CPC-Global and IMERG_cal precipitation products are found to be the best precipitation data among the 15 precipitation products for daily precipitation estimation (see Figs. 2 and 3), even in the XJ (KGE $\sim$ 0.30) and TP (KGE $\sim$ 0.45) regions where the accuracy of most products is poor (KGE $<$ 0.25). The corrected IMERG_cal product has a higher KGE value than the IMERG_uncal product. The quality of the IMERG_cal gauge adjusted product is improved in the XJ, TB, and NE subregions compared to the IMERG_uncal, demonstrating the positive contribution of the gauge-station correction.

The performances of the GSMaP and CMORPH products are poorer than that of the CPC-Global and IMERG_cal products based on the spatial KGE and five statistical metrics, but they are more accurate than the rest of the satellite-based precipitation products. The high quality of the CPC-Global, IMERG_cal, GSMaP, and CMORPH are all related to CPC and GPCC datasets, which incorporate gauge observations in China (Table 1). This reveals the importance of the in situ correction for improving the accuracy of precipitation products.

2) Reanalysis precipitation products

The statistical metrics diagram (Fig. 2, blue color) shows that the MERRA-2 products have superior performance in detecting precipitation occurrence compared to the other three reanalysis precipitation products. Most reanalysis products have moderate correlations (0.40 $<$ CC $<$ 0.80) with the CGDPA reference datasets, while GLDAS-2.1 shows a poor correlation with CC = 0.38. Overall, MERRA-2 has high CC (0.65), and CSI (0.51) values which exceed those of ERA5, ERA-Interim, and GLDAS. However, they are lower than CPC-Global's.

The MERRA-2 data show high KGE values in the eastern part of China (with mean regional KGE: 0.63), followed by the ERA5, ERA-Interim, and GLDAS products. In terms of KGE, the quality of the ERA5 product (with mean KGE: 0.59) was better than the ERA-Interim product's (with mean KGE: 0.53). The reason is various newly reprocessed datasets and recent instruments that could not be imbedded in ERA-Interim but are imbedded in ERA5, including ground-based radar datasets. Four reanalysis products exhibited a significantly lower level of performance in the western regions of China. This is probably because of the rugged topography characteristic and complex climate (such as Gansu Province, which includes subtropical monsoon, temperate monsoon, temperate continental, and alpine climate types) in the western regions. Notably, the GLDAS product exhibited the lowest accuracy in high altitudes and steep-slope regions of TB. The quality of the GLDAS product depends on model-driven data, and assimilation techniques. GLDAS simulations are forced with a combination of NCEP's Global Data Assimilation System (GDAS), disaggregated daily GPCP precipitation, and Air Force Weather Agency (AFWA) radiation datasets. The GLDAS product solves the temporal continuity problem, but its quality must be improved as a reliable data source of precipitation for hydrological simulation. The performance of the MERRA-2 precipitation product is poorer in the TB subregion, while it seems to perform better in the low-lying Sichuan basin in the eastern TB subregion.

b. Seasonal performance at the daily scale

The division of seasons is described in section 2a. Figure 4 shows Taylor diagrams with the correlations and standard deviations of the products. The closer the product is to the Buoy point (reference dataset), the higher the accuracy of a precipitation product.

Figure 5 presents the performance diagrams for mainland China during the study period. The higher the 1 − FAR and

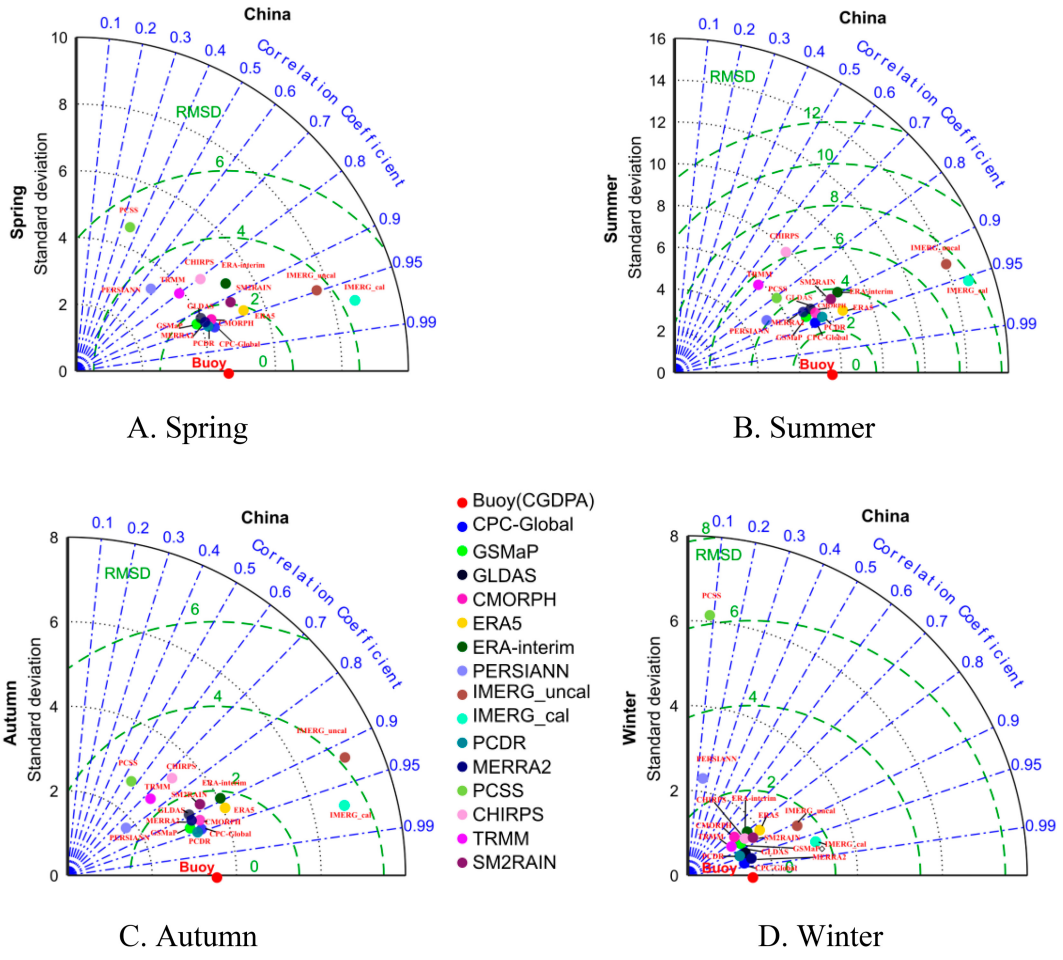


FIG. 4. Taylor diagrams at a daily scale from 2010 through 2019 for different seasons.

POD (upper-right corner), the better the quality of a precipitation product. Figure 6 shows the PDF, which represents the probability of the occurrence of precipitation events of different intensities. The closer to the top of the histogram, the higher the accuracy of a precipitation product (Wang et al. 2018; Zhou et al. 2020; Jiang et al. 2023; Lei et al. 2022).

There are large differences in the performance of the 15 precipitation products in different seasons. According to the Taylor diagrams (Fig. 4), in spring, summer, and autumn, the satellite-based precipitation products have higher correlations and lower errors, while reanalysis-based precipitation products performed better in winter. Figure 5 shows the performance diagram of the CPC-Global and MERRA-2 precipitation products. Both products reliably detect precipitation events in any season. The TRMM3B42 and CHIRPS products exhibited poor performance in all seasons. The PDF analysis (Fig. 6) indicates that the CPC-Global product is most consistent with the reference datasets. The PCSS product, however, has the worst performance. This is because the overestimation of low values and underestimation of high values is notable with the PCSS product. The following sections evaluate the performance of satellite-based and gauge-based precipitation products and reanalysis precipitation products separately.

1) SATELLITE-BASED AND GAUGE-BASED PRECIPITATION PRODUCTS

The Taylor diagrams imply that the CPC-Global, PCDR, and GSMaP products perform well in all seasons (low STD and RMSD, high CC). IMERG_cal and IMERG_uncal have a high RMSD in spring, summer, and autumn. The CDR exhibits the highest accuracy in all seasons in the PERSIANN family of products. The PCSS product had a significantly lower performance followed by the PERSIANN product in winter.

The performance diagram reveals that the IMERG_uncal and IMERG_cal products have strong detection power (high success ratio, CSI, and POD, low frequency bias) in spring, summer, and autumn, but weak detection power in winter. The IMERG_cal product is slightly better than IMERG_uncal product throughout the year. In spring, summer, and autumn, the overlap points corresponding to several precipitation products (Fig. 5), indicating that their skill in detecting precipitation is similar, while the distribution of points on the performance map in winter is scattered, indicating that the skills to detect precipitation are markedly different among products. In winter, the CMORPH, PCSS, PERSIANN, and SM2RAIN products are significantly worse

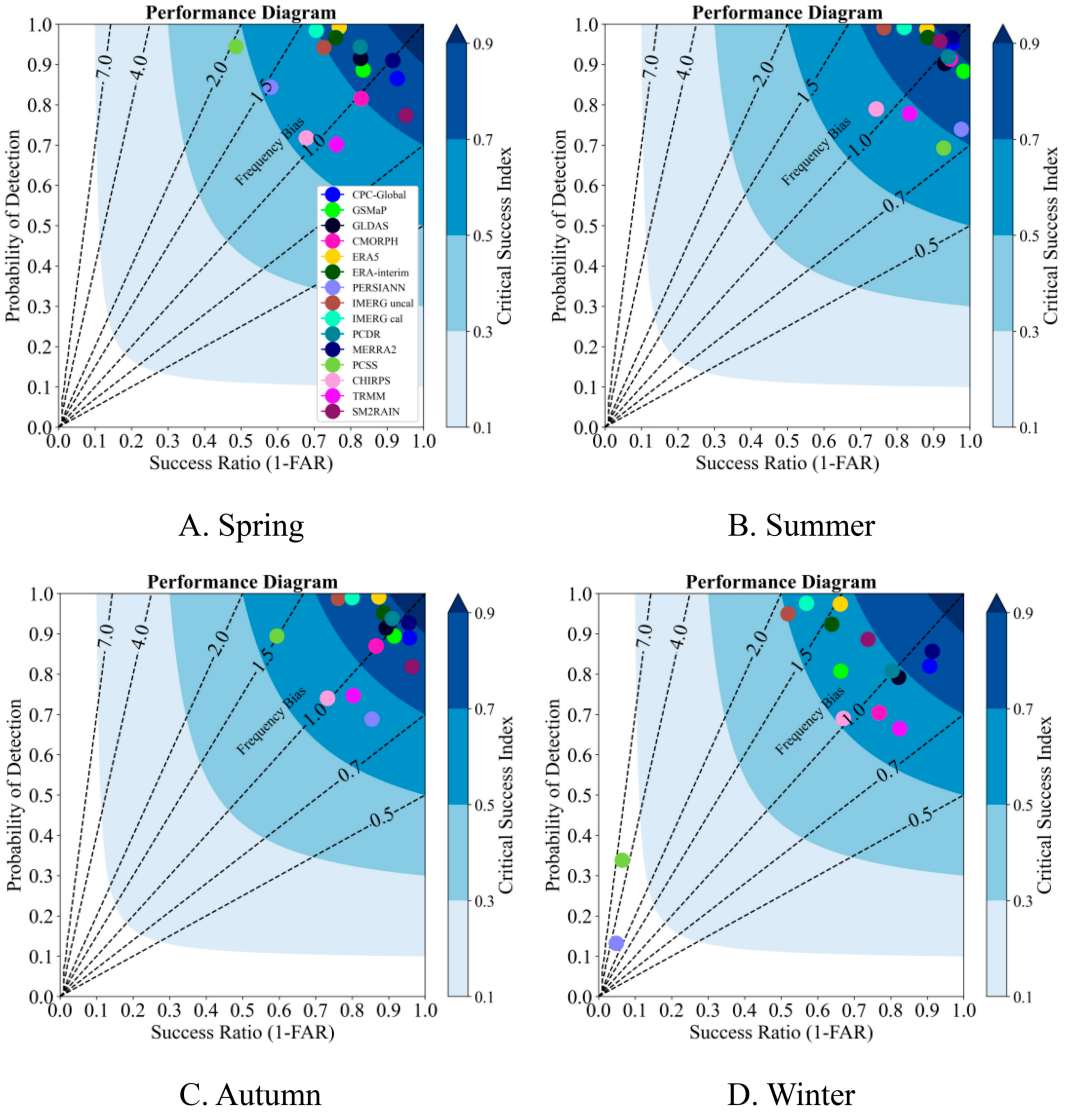


FIG. 5. Performance diagrams at a daily scale from 2010 to 2019 for different seasons.

at detecting precipitation than in other seasons. Notably, the PCSS and PERSIANN products showed large FAR, large bias, and small CSI values. The exception is the PCDR product, which also performs well in winter.

The probability of daily precipitation values is well captured throughout the year by the CPC-Global, CHIRPS, and CMORPH (Figs. 6a–d). The quality of SM2RAIN fluctuates with the seasons. PERSIANN and PCSS have poor detection power with respect to precipitation in all seasons. The IMERG_cal and IMERG_uncal significantly deviate from the reference dataset.

In spring (Fig. 6a), the CHIRPS product exhibits a good performance, and is slightly worse than CPC-Global precipitation product. Following CPC-Global and CHIRPS, the precipitation products CMORPH and GSMaP capture precipitation intensity reasonably well. The PCDR is superior to the other members of the PERSIANN family. The product SM2RAIN shows a strong underestimation of light and moderate precipitation events whose intensity is below 10 mm day^{-1}.

The heavy precipitation events (10–50 mm day^{-1}) increased significantly in summer (Fig. 6b). The CPC-Global and CMORPH products exhibited a perfect performance in capturing heavy precipitation events. The performance of the CMORPH product is slightly better than the CPC-Global product's based on the evaluation results. The GSMaP and MERRA-2 products are worse than the CPC-Global and CMORPH products, but better than the other products in summer. Among the nine satellite-based precipitation products IMERG_cal and IMERG_uncal show a poor accuracy especially in identifying high heavy precipitation (20–50 mm day^{-1}) and extreme precipitation (>50 mm day^{-1}). PCDR exhibits the best performance among the PERSIANN

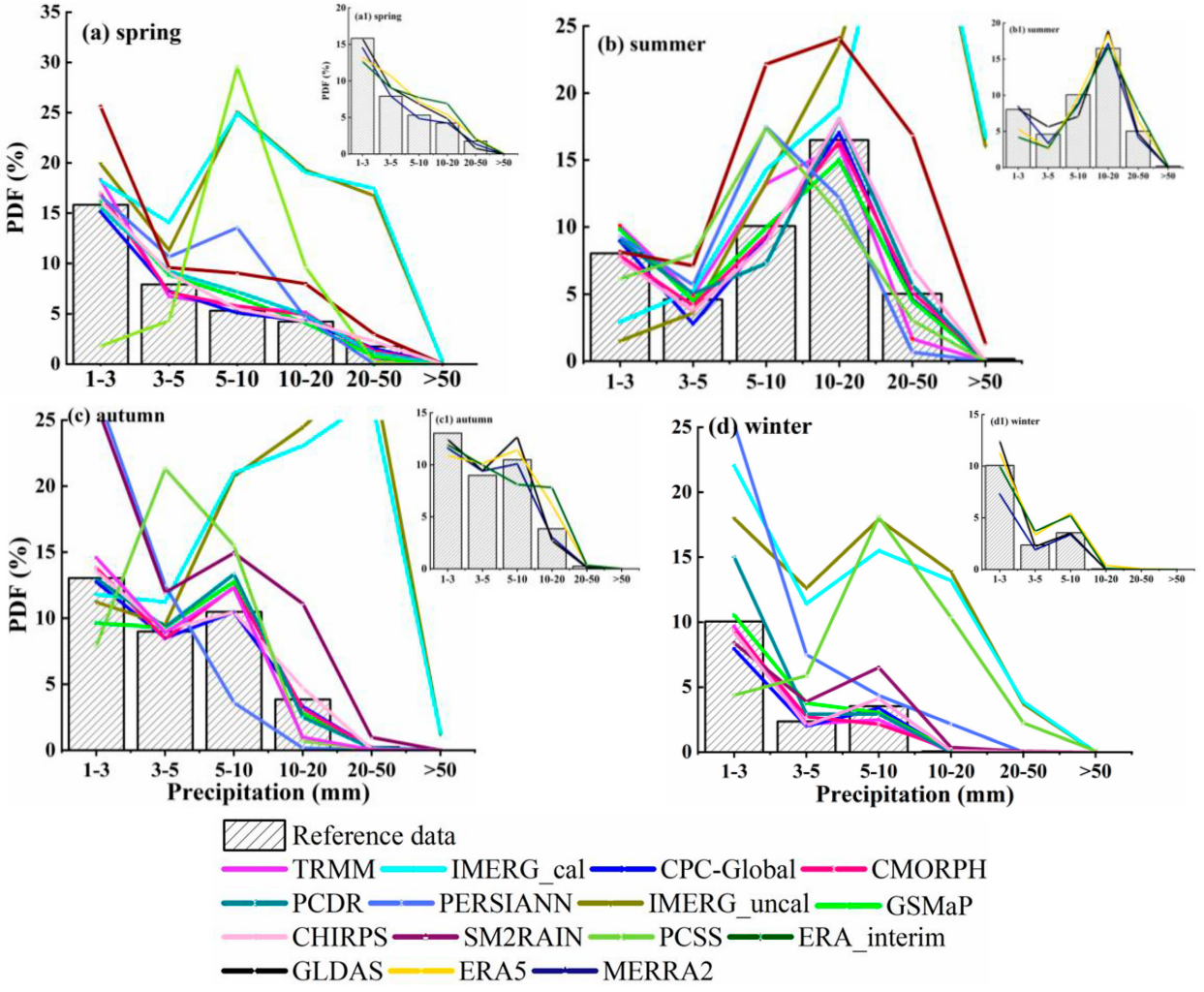


FIG. 6. The probability density function of daily precipitation from 2010 through 2019 for (a) spring, (b) summer, (c) autumn, and (d) winter for gauge- and satellite-based products. (a1)–(d1) The probability density functions for reanalysis precipitation products.

products, while PCSS is the worst. The performance of SM2RAIN precipitation products is notably improved in summer compared to its poor performance in the spring.

Figure 6c demonstrates that the CPC-Global product performed best, followed by CHIRPS and CMORPH in autumn. The calibrated IMERG_cal dataset also significantly overestimates precipitation events above 10 mm day^{-1}. PCDR detects events above 10 mm day^{-1} significantly better than the other members of the PERSIANN family.

CPC-Global and MERRA-2 perform similarly well in winter (Fig. 6d) and are more reliable than the CHIRPS product. The TRMM3B42 dataset tends to underestimate all precipitation events, especially light (1–3 mm day^{-1}) and moderate (3–10 mm day^{-1}) precipitation in winter.

2) REANALYSIS PRECIPITATION PRODUCTS

The Taylor diagram shows that the quality of the ERA5 product is higher than that of the ERA-Interim precipitation product except in winter. The performance diagram (Fig. 5) establishes that the four reanalysis products have good detection ability for precipitation events. The MERRA-2 product is the best among them, and the performances of the rest of the reanalysis products are not much different from each other.

The PDF diagram shows that MERRA-2 is the best reanalysis product in spring [Fig. 6a(1)] followed by GLDAS and ERA5. The ERA5 product underestimates light precipitation (1–3 mm day^{-1}) while overestimating other precipitation events. Nevertheless, it performed better than the ERA-Interim product in most cases. MERRA-2 is the best in detecting summer precipitation [Fig. 6b(1)], followed by ERA5 and GLDAS. The ERA5 and ERA-Interim products tend to underestimate light and moderate precipitation while overestimating heavy summer precipitation. In autumn [Fig. 6c(1)], the MERRA-2 product is close to the reference datasets. In contrast, the ERA-Interim product exhibits the worst performance. In winter the MERRA-2 shows a significant advantage [Fig. 6d(1)]. The GLDAS product performs better than the ERA5 and

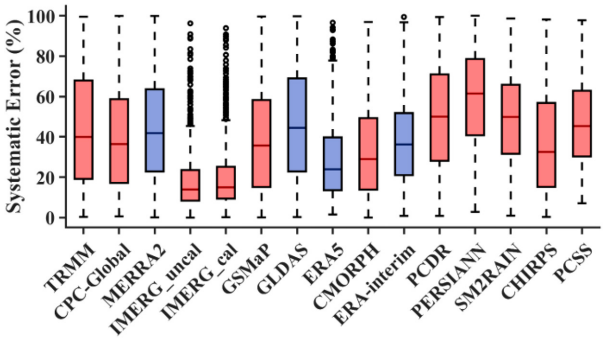


FIG. 7. The proportion of systematic errors for precipitation products in mainland China.

ERA-Interim products in detecting winter precipitation events, except for light precipitation. The ERA5 and ERA-Interim products show similar performance.

c. *Error component decomposition*

The error component decomposition is performed on daily precipitation. The mean error (ME) quantifies the different error sources. It may occur that the actual absolute bias of the product is large, but the ME of the evaluation results is small. The ME metric may not reliably reflect the performance of products. Error decomposition methods provide us with a more objective analysis of error sources. The errors of multiple products generally originate from the retrieval algorithm and reanalysis model.

The random error is not analyzed in this paper because it is inversely related to the systematic error described in section 3b. The error decomposition method proposed by Willmott (1981) (see Fig. 7) indicates that the systematic error of the IMERG_cal, IMERG_uncal, ERA5, CMORPH, and CPC-Global products is relatively small (lower than 35%) while that of the PCDR, PERSIANN, PCSS, and SM2RAIN products is large (exceed 50%). Besides, the systematic error tended to be more pronounced in ERA-Interim (38.2%) compared with the ERA5 product (28.7%). It was notable that the systematic errors of reanalysis-based precipitation products, except for the ERA5 product, are generally larger than the satellite-based products.

d. *Annual trends of indices*

The long-term trend of annual KGE and CSI values for precipitation products from 2010 to 2018 is shown in Fig. 8. The KGE and CSI indices are computed from daily precipitation and averaged over the year. This time-limited period is used in this study because the ERA-Interim, SM2RAIN, and PERSIANN datasets are partly unavailable in 2019. The KGE values show that the IMERG_cal product's KGE is better than IMERG_uncal product, demonstrating that the improvement caused by gauge adjustment is greater than the improvement caused by satellite sensors. It is important to have a continuous stable performance of a precipitation product over a long period. The KGE value of the PCSS product varies greatly, and the downward trend of its KGE indicates

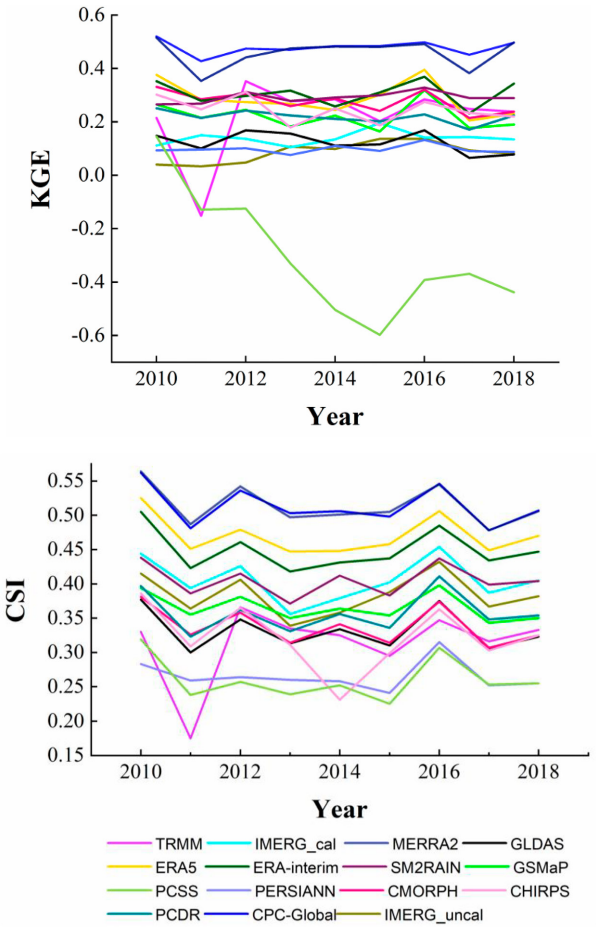


FIG. 8. The CSI and KGE for precipitation products from 2010 through 2018.

that the capacity to estimate precipitation intensity gradually decreases over time. The PCSS product has the highest resolution (0.04°) in the PERSIANN product family, but the accuracy of the cloud segmentation algorithm varies greatly from year to year. When analyzing long-term series special attention could be paid to the accuracy of the PCSS, TRMM3B42, and CHIRPS products because their quality exhibits large interannual variability.

e. *Cross evaluation in the gauge sparse subregions (TB and XJ)*

The cross evaluation is performed on daily precipitation. A careful selection of datasets is critical to ensure reliable MTC results. This is because the same type of products used in triplet collection method may contain cross-correlated errors due to their overlapping use of common input data and processing methods. For example, several satellite-based products that are gauge adjusted belong to the same types in this case. Therefore, they were not used to maximize the independence of products for MTC analysis. Instead, three different precipitation product types are used in each triplet: one satellite-based product ("top-down"), one reanalysis product, and the

TABLE 5. The daily precipitation in the Qinghai–Tibetan Plateau (triplets 1–12) and in the Xinjiang subregion (triplets 13–24) is estimated with different triplet combinations.

Triplet	Products	RMSE (mm day^{-1})			CC		
		Product A	Product B	Product C	Product A	Product B	Product C
		Qinghai–Tibetan Plateau					
1	SM2RAIN–MERRA-2–CPC-Global	3.204	2.311	0.551	0.326	0.789	0.911
2	SM2RAIN–MERRA-2–IMERG_cal	2.610	3.707	1.936	0.545	0.470	0.649
3	SM2RAIN–MERRA-2–IMERG_uncal	2.685	3.633	1.906	0.522	0.492	0.645
4	SM2RAIN–CGDPA–ERA5	3.059	2.933	1.702	0.385	0.657	0.612
5	SM2RAIN–MERRA-2–CGDPA	2.787	3.463	2.267	0.478	0.536	0.587
6	SM2RAIN–MERRA-2–GSMaP	2.404	3.876	2.167	0.598	0.423	0.554
7	SM2RAIN–MERRA-2–PCDR	3.124	2.854	2.627	0.372	0.694	0.296
8	SM2RAIN–CGDPA–GLDAS	3.099	2.620	3.168	0.366	0.703	0.278
9	SM2RAIN–MERRA-2–TRMM	2.713	3.536	3.191	0.503	0.517	0.274
10	SM2RAIN–MERRA-2–CMORPH	2.054	4.251	2.817	0.696	0.350	0.243
11	SM2RAIN–MERRA-2–CHIRPS	3.334	2.727	3.342	0.350	0.720	0.204
12	CGDPA–MERRA-2–CHIRPS	1.097	1.460	1.494	0.439	0.600	0.417
		Xinjiang					
13	SM2RAIN–MERRA-2–CPC-Global	2.717	3.163	0.957	0.112	0.640	0.830
14	SM2RAIN–MERRA-2–IMERG_cal	2.490	4.311	1.898	0.230	0.298	0.629
15	SM2RAIN–MERRA-2–IMERG_uncal	2.425	4.401	1.945	0.217	0.330	0.619
16	SM2RAIN–MERRA-2–GSMaP	2.378	4.455	2.117	0.258	0.272	0.586
17	SM2RAIN–MERRA-2–CGDPA	2.585	4.127	2.357	0.193	0.396	0.475
18	SM2RAIN–MERRA-2–CMORPH	2.282	4.550	2.469	0.289	0.253	0.408
19	SM2RAIN–CGDPA–GLDAS	2.680	3.746	2.600	0.136	0.504	0.397
20	SM2RAIN–CGDPA–ERA5	2.658	3.679	3.837	0.144	0.517	0.393
21	SM2RAIN–MERRA-2–TRMM	2.472	4.334	2.475	0.236	0.317	0.363
22	SM2RAIN–MERRA-2–PCDR	2.668	3.725	2.427	0.140	0.474	0.337
23	SM2RAIN–MERRA-2–CHIRPS	2.860	4.091	3.050	0.102	0.546	0.168
24	CGDPA–MERRA-2–CHIRPS	1.238	1.781	1.544	0.245	0.518	0.258

SM2RAIN product ("bottom-up"), or gauge-based interpolated products are used in this study for cross evaluation purposes. The SM2RAIN is considered as a truly independent source of precipitation that is widely used in MTC analysis (Massari et al. 2017). The PERSIANN, PCSS, and ERA-Interim products with poor quality in the evaluations of classic statistical metrics (section 4a) were discarded from the triplets.

Table 5 lists the triplets, and every row in the table represents a triplet (product A, B, and C). It is seen in Table 5 that MTC with 24 different triplets yielded a consistent overall ranking of products. MTC tended to yield higher CC estimates, compared to the results of the statistical metrics (section 4a). CPC-Global, IMERG_cal, IMERG_uncal, CGDPA, ERA5, GSMaP, MERRA-2, PCDR, SM2RAIN, GLDAS, TRMM, CMORPH, and CHIRPS ranked from best to worst in terms of the mean value of correlation with the unknown truth from all pixels. The relative rank displays a slight difference in TB and XJ regions (such as ERA5, SM2RAIN, and CMORPH). Similar findings were reported by Li et al. (2018), Massari et al. (2017), and Duan et al. (2021) using the MTC method. Overall, the gauge-based and GPM satellite products are more accurate even in the TB and XJ regions, and the reanalysis product ERA5 is of high quality and very close in performance to the gauge-based product. This suggests that MTC could be used to identify the "best" available precipitation products for poorly gauged areas where reference datasets are not available.

Figures 9(1–11) and 9(13–23) depict the spatial distribution of various triplets with different combinations of precipitation products in the TB and XJ subregions, respectively. The metric maps for SM2RAIN and MERRA-2 in Figs. 9(1–11) and 9(13–23) are provided by the first triplet and the thirteenth triplet, respectively. According to Table 5 and Fig. 9 the IMERG_cal product shows the second largest CC in most areas in the TB (0.649) and XJ (0.629) subregions. The IMERG_uncal product is only slightly worse than IMERG_cal. This demonstrates that the latest version of the GPM products performed well even in regions with scarce gauge stations. It is worth noting that the high quality of GPM products does not completely rely on gauge correction compared with the performance of other gauge adjusted products. The ERA5 reanalysis product exhibits better performance in the TB regions (CC: 0.612) than in the XJ regions. This means that the choice of precipitation products and the improvement of algorithms must be paid attention to in complex terrain subregions with few gauge stations.

f. Performance at seasonal scale

Precipitation products such as IMERG, TRMM3B42, CHIRPS, and PCDR have bias correction/adjustments based on monthly gauge precipitation, whereas products such as GSMaP use daily gauge precipitation. The bias may be different at daily and seasonal scales. Therefore, this work evaluates the performance of precipitation products at

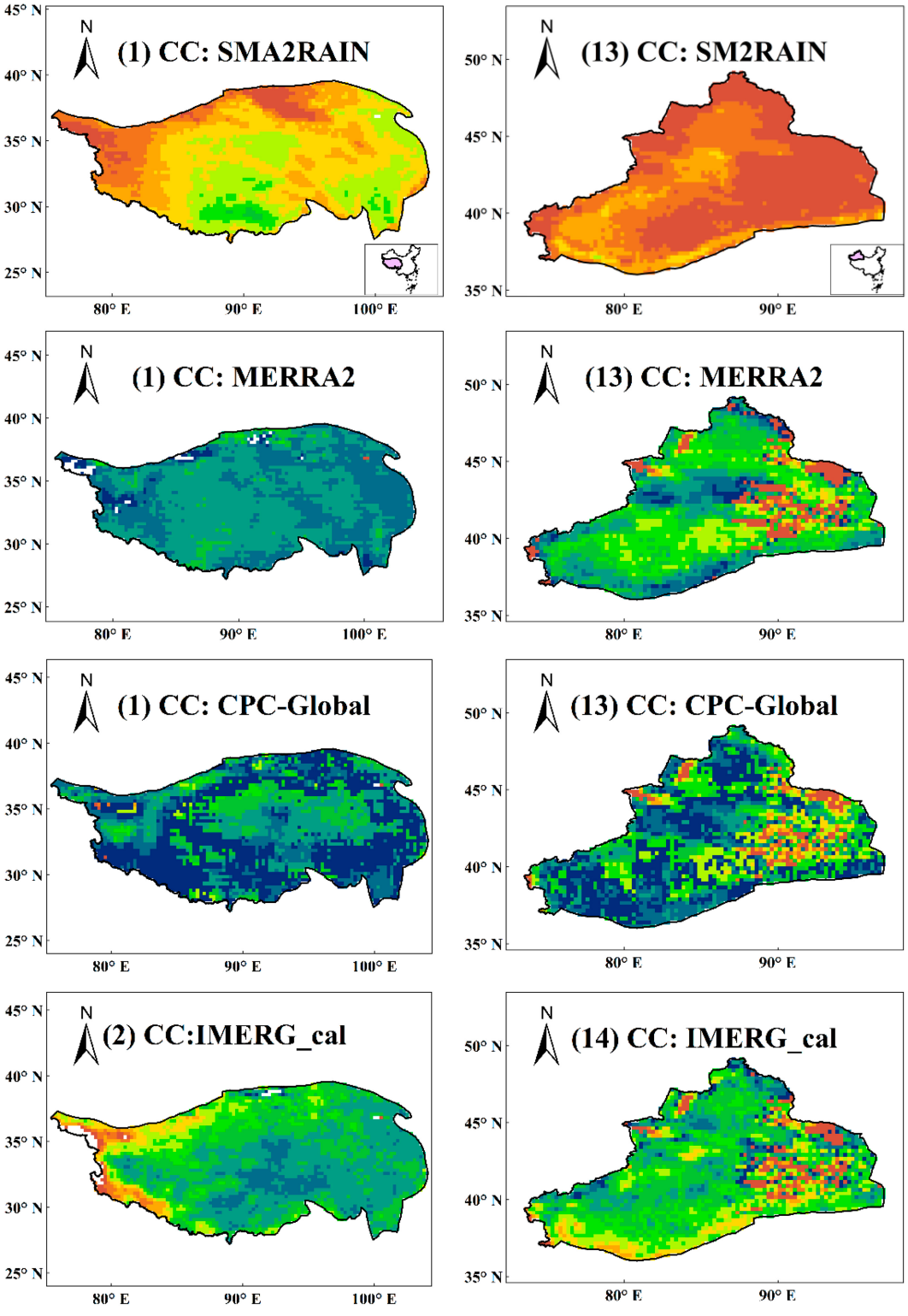


FIG. 9. The CC of the TB (1–11) and XJ (13–23) subregions based on MTC using data from 2010 through 2019. The number in parentheses corresponds to the triplet index in Table 5. For example, the metrics for the SM2RAIN and MERRA-2 products are from the first triplet for the TB region and the thirteenth triplet for the XJ region.

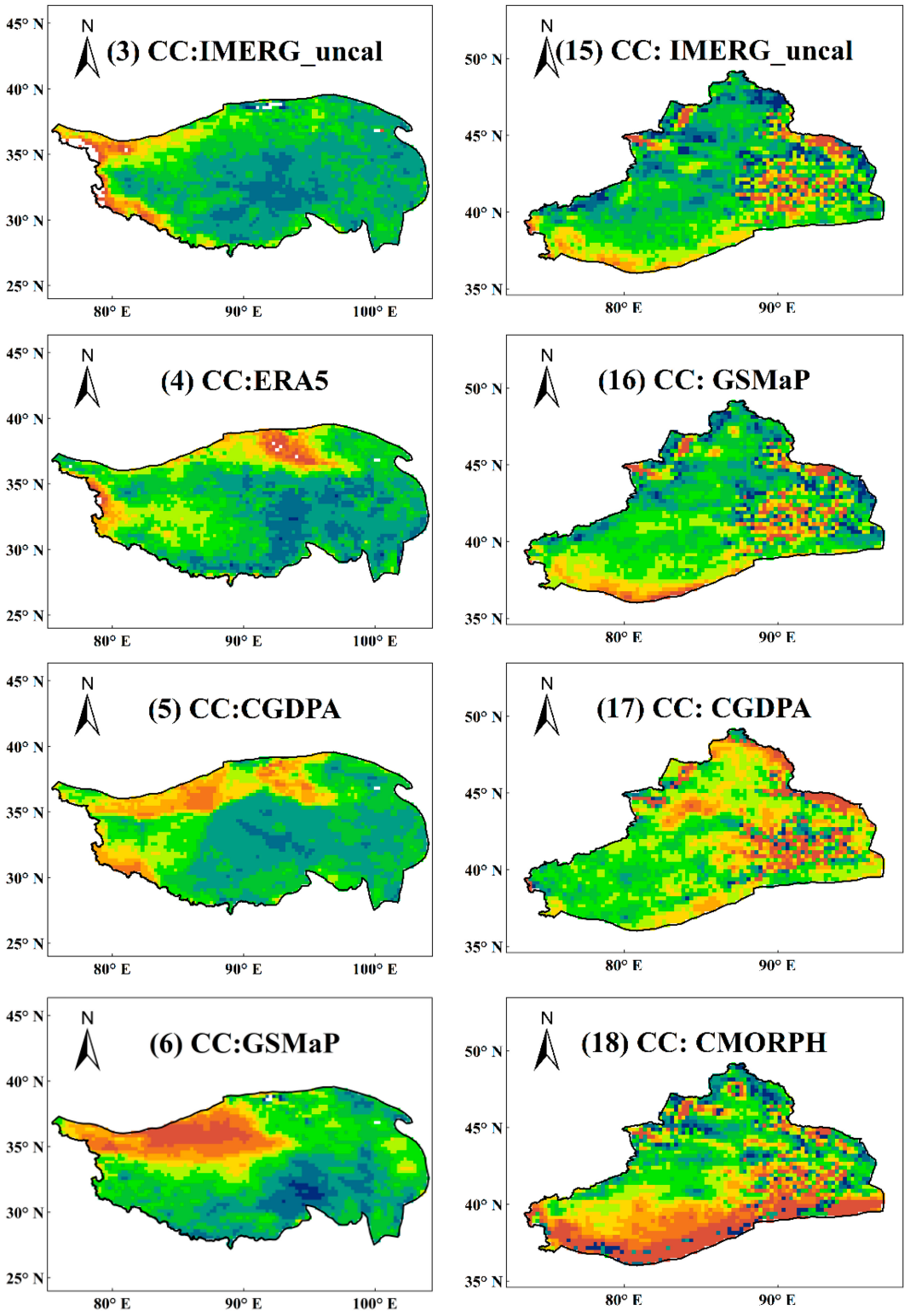


FIG. 9. (*Continued*)

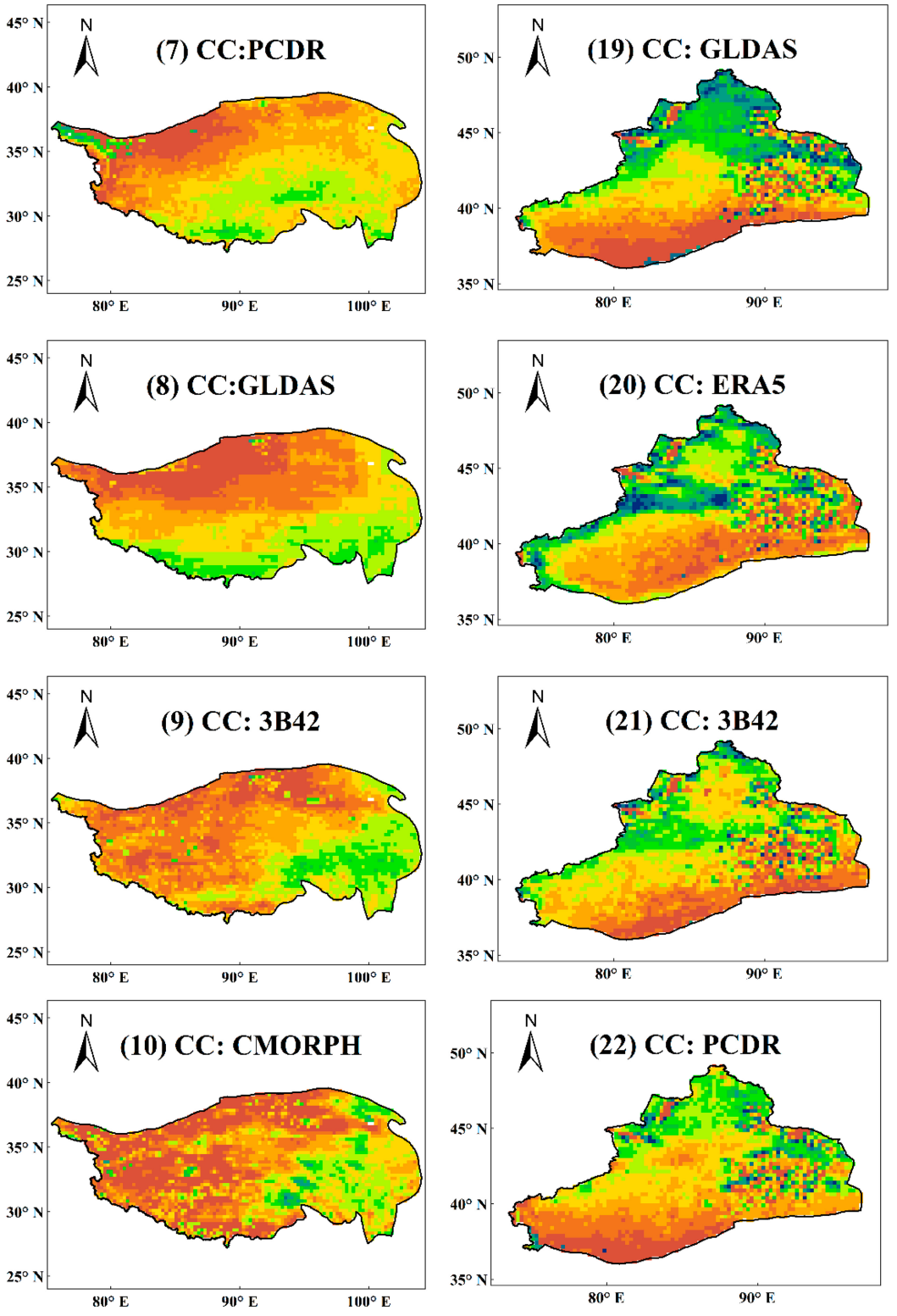


FIG. 9. (*Continued*)

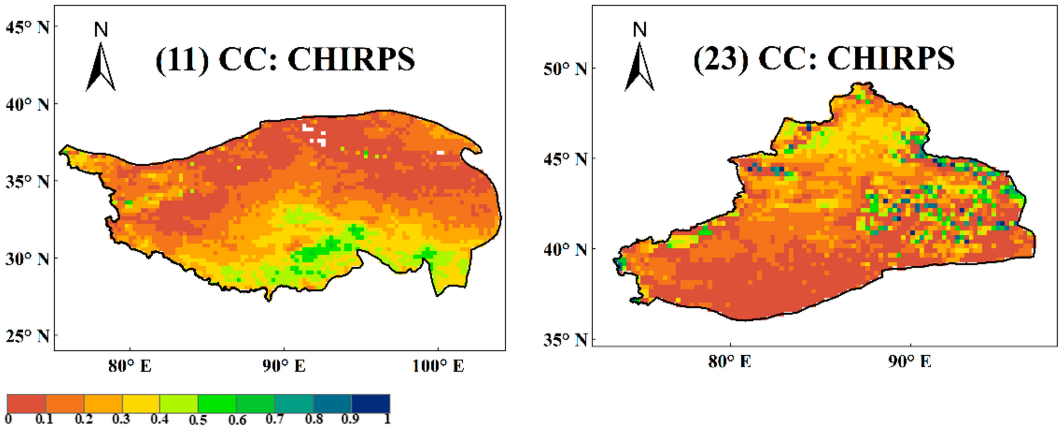


FIG. 9. (*Continued*)

the seasonal scale. The satellite-based and reanalysis precipitation products at the seasonal scale are compared in this paper by aggregating daily precipitation to seasonal precipitation.

Figure 10 illustrates the boxplots of CC and RMSE metrics for all seasons. Among the satellite-based products, GSMaP, IMERG, TRMM, PCDR, and CHIRPS exhibit stable performance with high CC and low RMSE in all seasons. GSMaP and IMERG outperform the other products. For reanalysis products, MERRA-2 performs the best, followed by ERA5. However, it is important to note that the quality of several products fluctuates with the seasons. Specifically, the accuracy of SM2RAIN and CMORPH significantly decreases in winter [Fig. 10a(4)], while the accuracy of GLDAS fluctuates in summer [Fig. 10a(2)].

This work compared the CC and RMSE at a daily scale displayed in Fig. 2 with the same statistic at a seasonal scale displayed in Fig. 10. The comparison reveals that IMERG_cal performs better (with high CC, and low RMSE) at seasonal time scales than daily time scales, which demonstrates that the evaluation results are dependent on time scales.

g. Drought monitoring performance

The evaluation of precipitation products shows that gauge-adjusted products perform relatively better than the rest. Only gauge-adjusted products of the PERSIANN family and IMERG series were used to assess the drought monitoring performance. ERA-Interim is not used among reanalysis products due to its poor performance compared to ERA5.

1) SPATIAL-SCALE ANALYSIS OF SPEI

The SPEI is calculated at multiple time scales (1, 3, and 12 months) to assess the effectiveness of precipitation products in drought monitoring. The spatial patterns of CC, POD, and FAR associated with the SPEIs computed from precipitation products are displayed in Figs. 11–13, respectively. CC, POD, and FAR are computed against the SPEI from the reference dataset (CGDPA). This study specifies the SPEI < -1.0 (severe drought) as the drought condition threshold for the computation of the POD and FAR.

It is seen in Fig. 11 that all products exhibit similar spatial patterns with high CC in eastern and southern China. However, they have inferior performances in monitoring drought over western China where there is small CC. The worst performance is observed in the TB subregion with scarce gauge stations for all precipitation products. The accuracy of precipitation products is affected by large-scale climatic variations (Yu et al. 2022). Nearly all precipitation products display better accuracy in eastern China (humid climate) compared to western China (arid climate). This may be due to three reasons. First, there are numerous mountainous and high-altitude regions in western China. The number of in situ gauge stations is very limited in the rugged terrain compared with that in the plains and valleys of eastern China. The spatial sparsity of the in situ gauge observation network improves the difficulty of error correction in western China. Second, a higher frequency of orographic precipitation easily prevails in mountainous regions which largely contributes to a complex convective system. The duration of extreme precipitation events in arid regions (low moisture) tends to be shorter than in humid regions (high moisture especially in the coastal area). The short-term extreme precipitation events are easy to overlook and difficult to be detected by satellite sensors. This constitutes a challenge for most precipitation products (Navarro et al. 2020). Third, the atmosphere is significantly colder (snowfall and frozen) in western China compared with that in eastern China. The performances of precipitation products fluctuated and were variable in winter (snowfall and frozen) compared with other seasons (see Figs. 4 and 5).

Generally, the GSMaP product exhibits a better performance in detecting drought events compared to other satellite precipitation products at multiple time scales, with the highest CC (above 0.65) in mainland China, followed by the IMERG_cal product. The CMORPH shows the worst performance in the XJ subregion and northeast China with respect to all evaluated products. Among four reanalysis precipitation products the performance of MERRA-2 is obviously better than ERA5 and GLDAS, especially in the TB subregion. The ERA5 is better than GLDAS in drought detection, while the ERA-Interim is no doubt the worst.

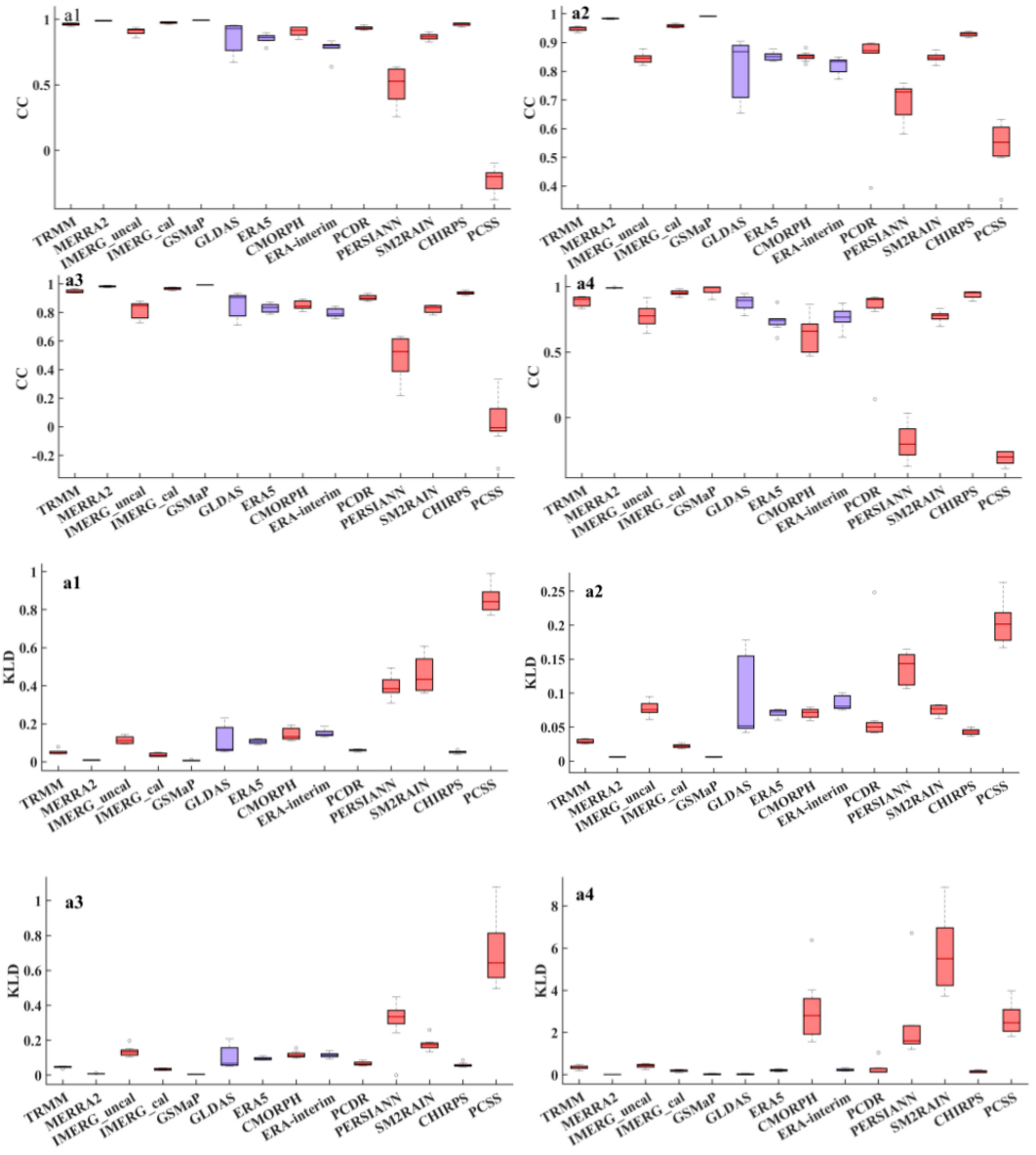


FIG. 10. Boxplots of CC, and RMSE metrics comparing satellite-based (red) and reanalysis products (blue) with CGDPA datasets at the seasonal scale: (a1) spring, (a2) summer, (a3) autumn, and (a4) winter.

The POD (Fig. 12) and FAR (Fig. 13) are used to estimate the true and false detection rates of drought, respectively. The spatial patterns of POD and FAR for all products are similar to that of CC (Fig. 11). More than 60% of drought months can be accurately detected, as illustrated by PODs usually above 0.60, by FARs usually (less than 0.30) in southern and eastern China. The rate of true detection is higher in southeastern China compared to that in northwestern, with the results shown in Fig. 12. The detection performances of the IMERG_cal, GSMaP, and MERRA-2 products compare well with the rest of the products. The lowest PODs are mostly found in the TB, XJ, and SW subregions, indicating that the quality of precipitation products is limited in those regions (Fig. 12). The drought detection accuracy of precipitation products is improved with longer time scales of SPEI (i.e., SPEI-12). This improvement may be due to the SPEI of longer time scales (e.g., 12-month time scales, SPEI-12) being more accurately predicted with smoother values (Anshuka et al. 2019; Dikshit and Pradhan 2021). A low FAR is generally observed in southern and eastern China (Fig. 13). The highest FAR is observed in the TB subregion, followed by parts of the XJ and NW subregions. This is consistent with the results observed with respect to the POD index. The GSMaP and MERRA-2 products exhibit a consistently low FAR in mainland China at three investigated time scales. The FAR of the reanalysis products is relatively high with exception of the

MERRA-2 products compared with satellite-based precipitation products. The MERRA-2 product performed well in drought monitoring in mainland China with respect to the investigated three temporal scales. The use of observations from newer microwave sounders and hyperspectral infrared radiance instruments improves the capacity of MERRA-2 in drought detection compared with other model-derived precipitation products. Moreover, MERRA-2 is the first satellite-era global reanalysis to assimilate space-based observations of aerosols and represent their interactions with other physical processes in the climate system. Yet, it is not possible to ascertain whether or not the original spatial resolution of the MERRA-2 product is suitable for drought monitoring in general because the MERRA-2 product was downscaled from 0.625° to 0.25° in this study.

2) TEMPORAL ANALYSIS OF SPEI

Temporal changes of the SPEI-1, SPEI-3, and SPEI-12 were calculated from precipitation products over mainland China [Figs. 14a–c,a(1)–c(1)]. Figure 14c shows that the SPEI curves become smoother with longer time scales. The PCDR product exhibits a significant overestimation between late 2015 and early 2017. Although there are differences in the magnitude of the SPEI values from the reference dataset, the precipitation products' estimates generally exhibit similar drought-intensity values.

Figures 14e–g display the difference values, in which the red line represents the reference dataset. For satellite-based products, the GSMaP, IMERG_cal, and CHIRPS products show a close match with the reference dataset at all time scales. The CMORPH product shows a relatively consistent similarity with the reference dataset with the exception of some months (e.g., between late 2017 and early 2019). The performance of the TRMM3B42 and PCDR products fluctuated concerning drought detection, with significant overestimation and underestimation.

The ERA5 and MERRA-2 products are highly consistent with the reference dataset at all time scales among the reanalysis products. Specifically, the MERRA-2 product shows the best match with the reference dataset (smallest SPEI difference values) at all time scales, which suggests that the MERRA-2 product is most suitable for detecting drought over mainland China, closely followed by the GSMaP and IMERG_cal products. The GLDAS tends to overestimate or underestimate drought events with unreliable performance.

5. Discussion and conclusions

This study assessed the quality of global precipitation products at the daily, seasonal, and annual time scales from 2010 to 2019, and their suitability for detecting drought within mainland China. The precipitation estimation of gauge-based CPC-Global and satellite-based GSMaP products performed well on a daily scale, indicating that the gauge incorporation could significantly improve the quality of precipitation products in detecting the occurrence and intensity of the precipitation. Among 11 satellite-based and gauge-based products, the CPC-Global, IMERG_cal, GSMaP, and CMORPH have a high quality, which is related to their incorporation of CPC and GPCC gauge datasets. The PCDR performance was found to be better than those of PCSS and PERSIANN in the PERSIANN family. Among the reanalysis products MERRA-2 proved superior to ERA5, ERA-Interim, and GLDAS. The precipitation evaluation indicates that precipitation products' accuracy varies from season to season when evaluated with respect to their seasonal performance. The accuracy of the precipitation products is usually worse in winter. The CMORPH, PCSS, PERSIANN, and SM2RAIN products perform poorly with respect to precipitation detection in winter. Three possible reasons for poor performance in winter are as follows: 1) PMW has trouble discerning ice and snow from clouds (CMORPH); 2) the frozen soil and soil saturation introduce large errors into the "bottom-up" algorithm (SM2RAIN); 3) the low occurrence of precipitation days introduces many zero values in the training datasets hindering the accurate prediction of the artificial neural network algorithm (PCSS and PERSIANN).

The cross evaluation performed with the MTC method revealed that the CPC-Global, IMERG_cal, IMERG_uncal, CGDPA, ERA5, GSMaP, MERRA-2, PCDR, SM2RAIN, GLDAS, TRMM, CMORPH, and CHIRPS products were ranked from best to worst, respectively, in terms of the mean value of correlation. The gauge-based and GPM satellite products are more accurate even in the gauge scarce regions, and the reanalysis product ERA5 is of high quality and very close in performance to the gauge-based products. The MTC method is useful to evaluate product performance in mountainous regions with few gauge observations.

The improvement of product quality is not completely explained by gauge correction. The product retrieval algorithm is also important. Specifically, although MERRA-2 and GLDAS incorporate gauge datasets, the distinct performances of the two products may be due to the gauge-data processing approach and their representation of model physics. In terms of the gauge-data processing approach MERRA-2 implements a sophisticated data assimilation method that combines observations from multiple sources with a numerical weather prediction model to produce a consistent, high-quality estimate of atmospheric variables, including precipitation. In contrast, GLDAS-2.1 represents merged, spatially, and temporally interpolated fields of GDAS, GPCP, and AFWA radiation fields. In addition, MERRA-2 relies on a more advanced atmospheric model and includes more sophisticated parameterizations of key processes such as cloud microphysics, which may result in better precipitation estimates. However, GLDAS-2.1 admits precipitation as an input instead of model physics simulation. This means the correct description of the complex physical process in the retrieval algorithm significantly improves the products' quality. The quality improvement of precipitation data is imperative because drought detection capacity depends heavily on the bias of the precipitation products.

Meteorological drought is detected by the SPEI index from global precipitation products. Our main findings are as follows: The SPEI calculated from the precipitation products exhibit similar spatial patterns, with large CC and POD in southern

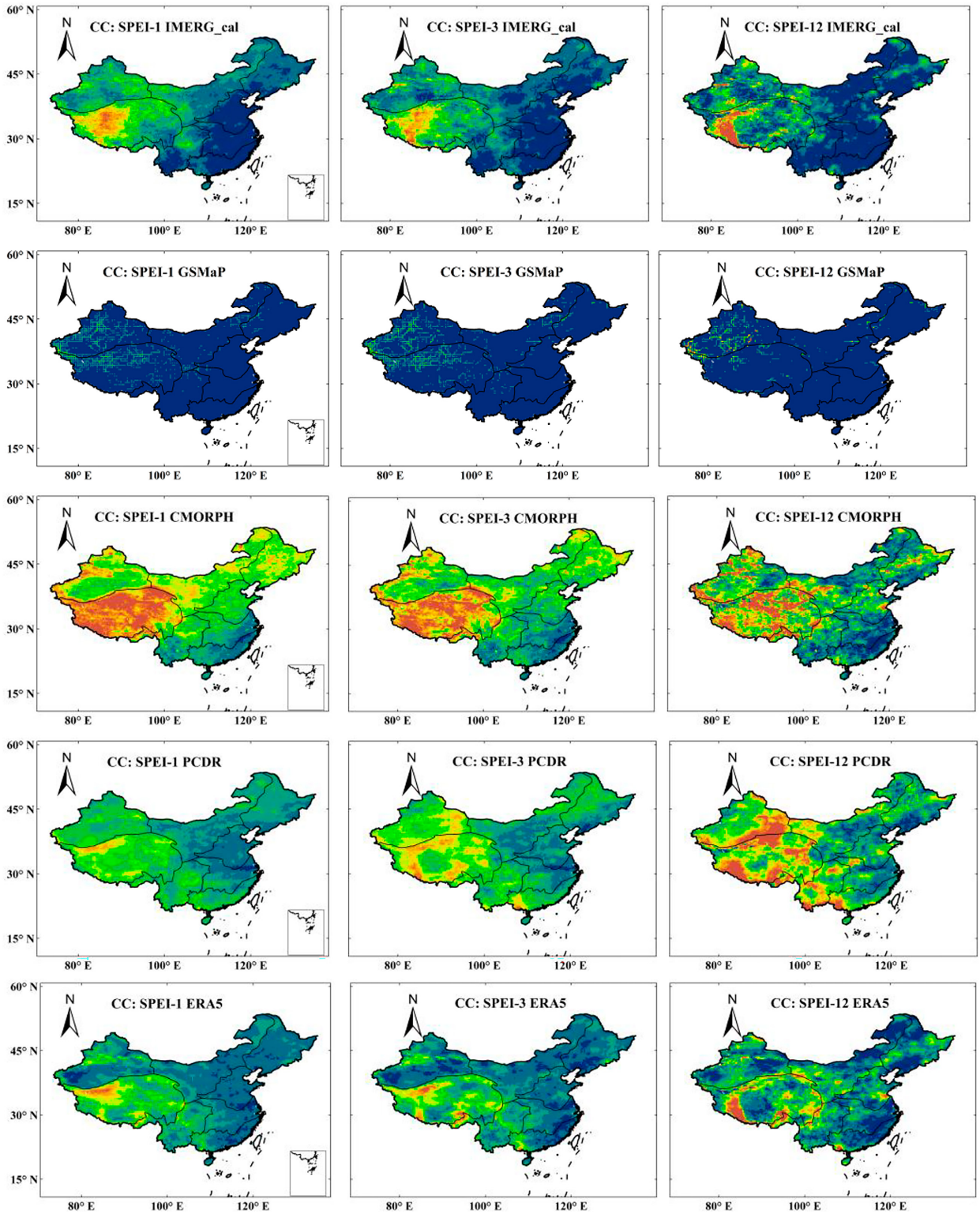


FIG. 11. The CC of the SPEI-1, SPEI-3, and SPEI-12 was calculated from the satellite-based and reanalysis precipitation products with the CGDPA reference dataset over mainland China.

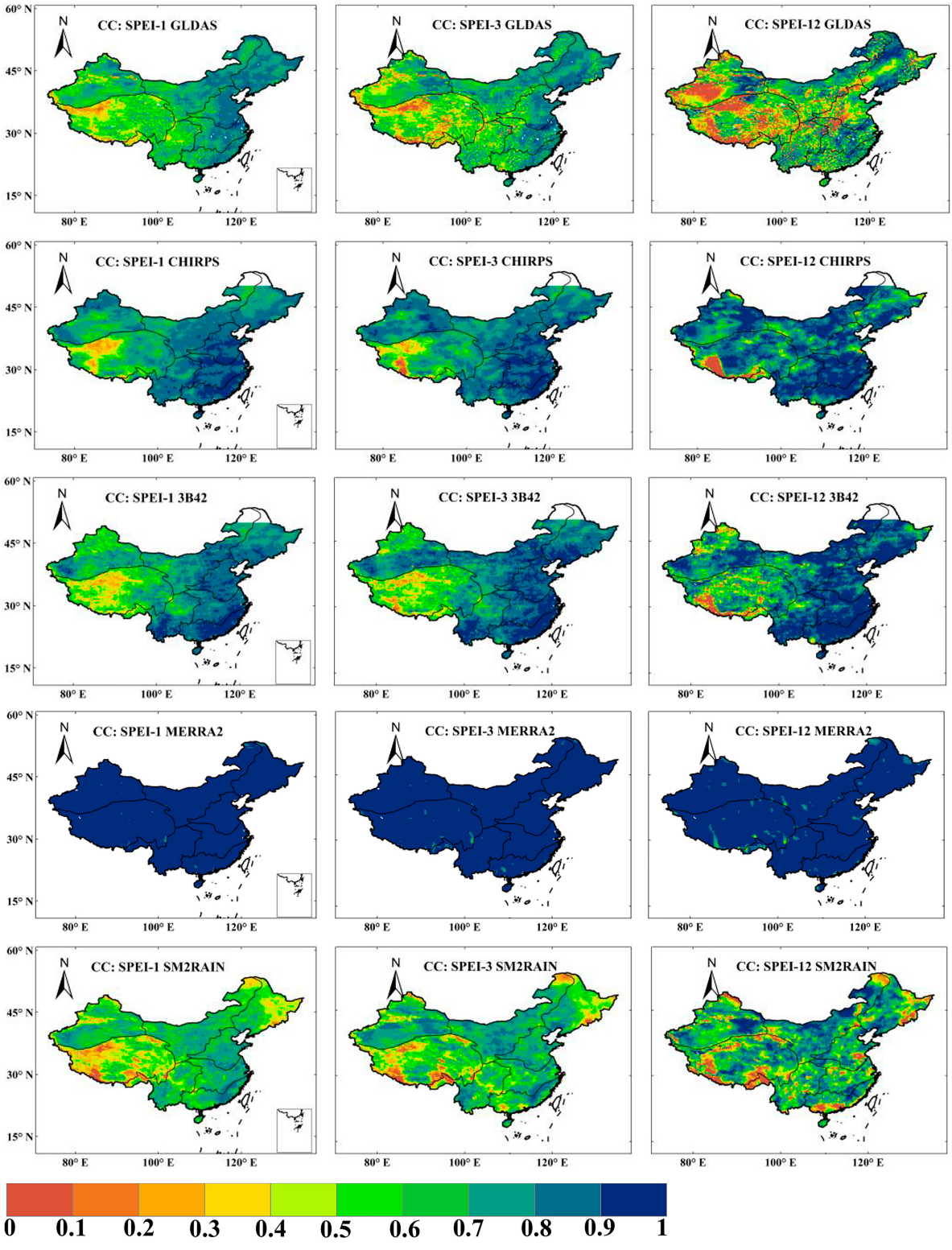


FIG. 11. (*Continued*)

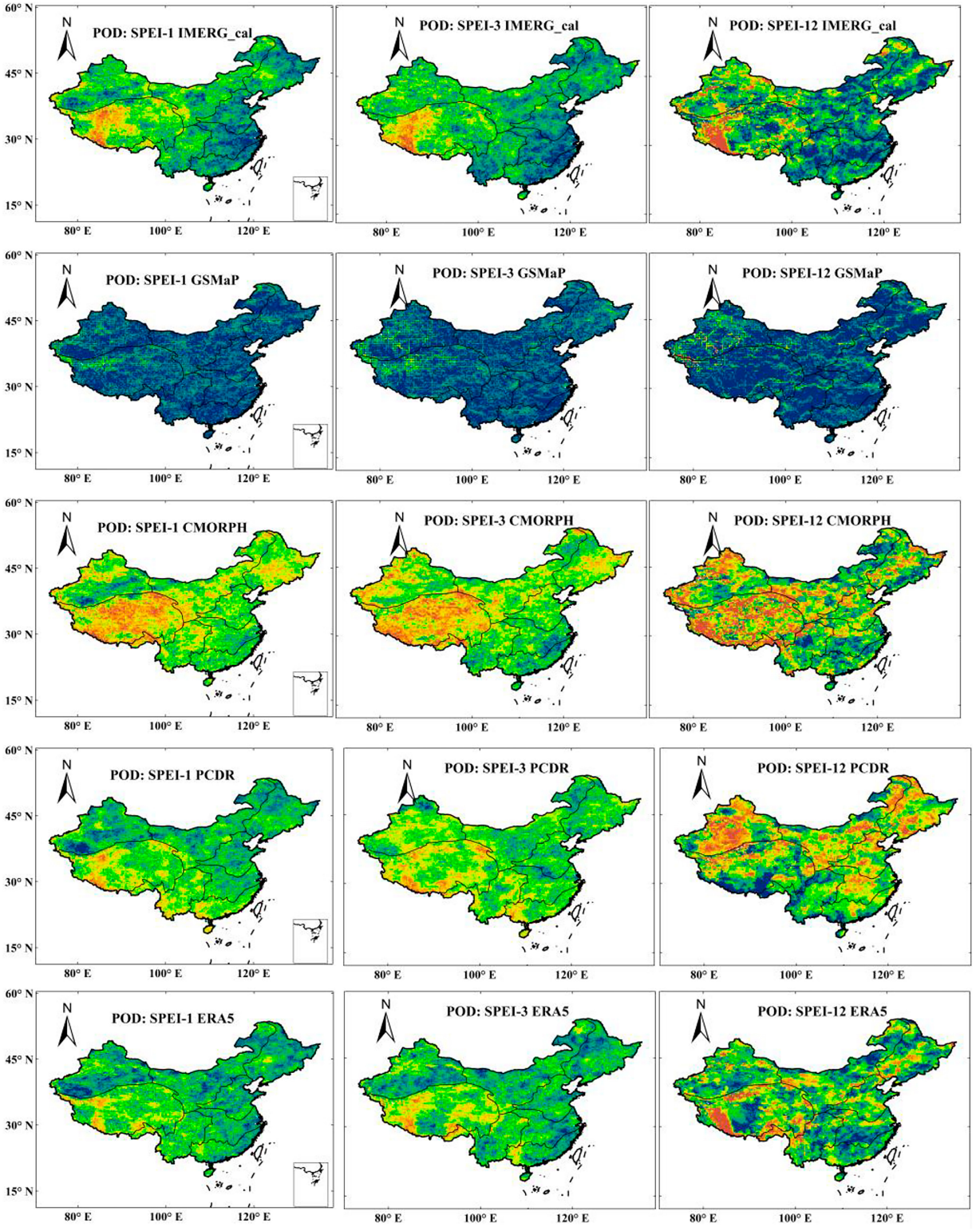


FIG. 12. The POD of the SPEI-1, SPEI-3, and SPEI-12 was calculated from the satellite-based and reanalysis precipitation products with CGDPA reference dataset over mainland China.

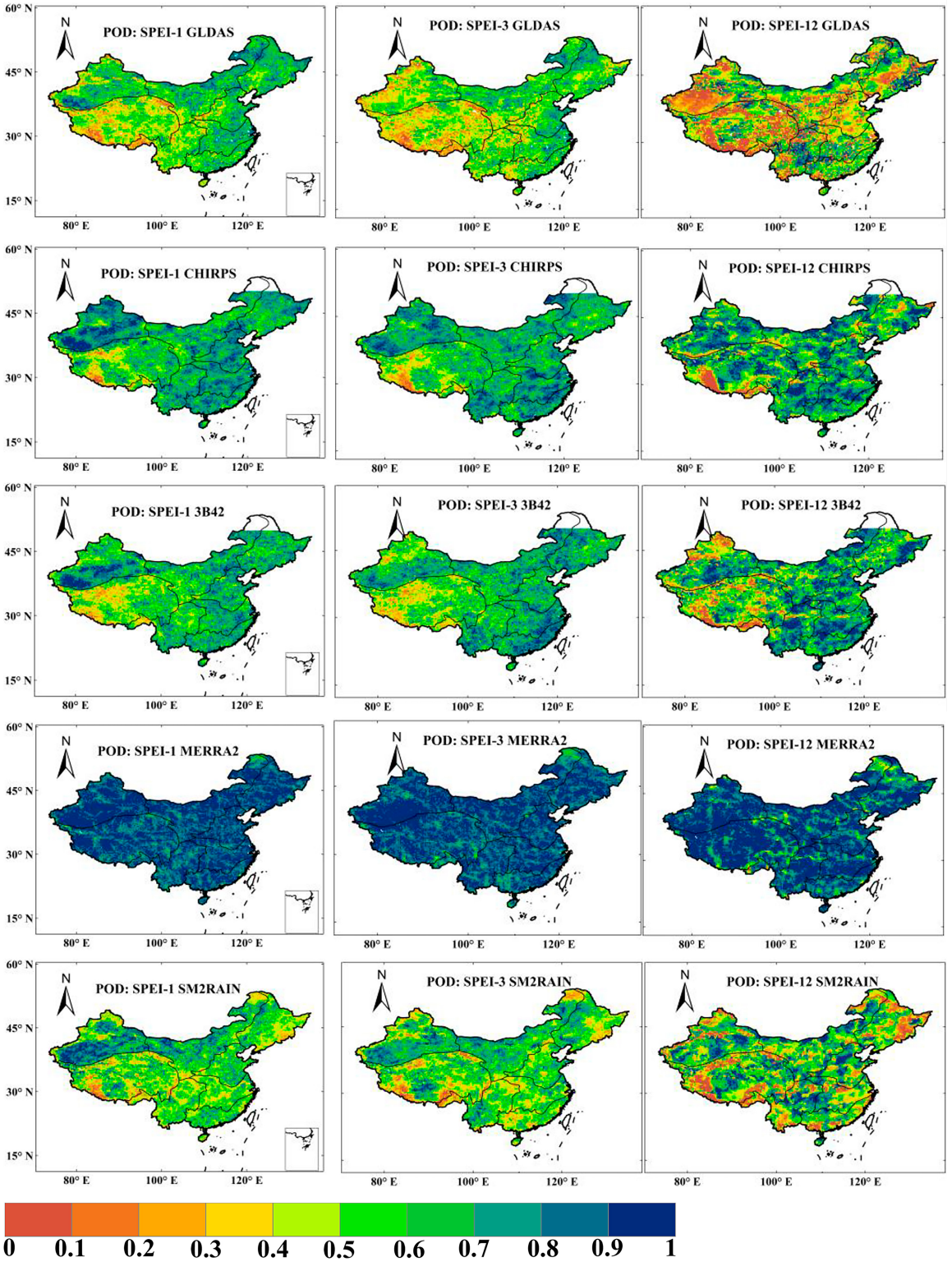


FIG. 12. (*Continued*)

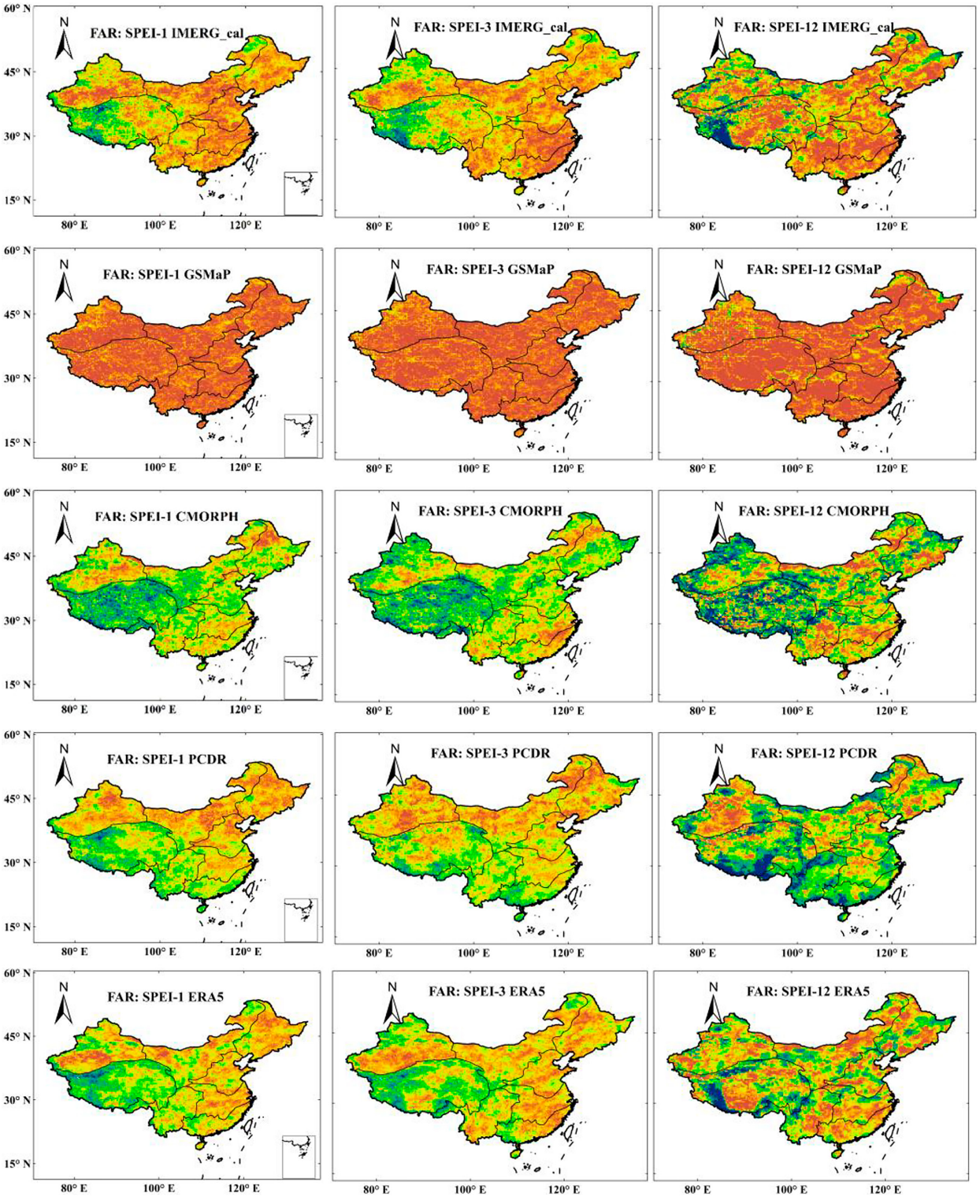


FIG. 13. The FAR of the SPEI-1, SPEI-3, and SPEI-12 was calculated from the satellite-based and reanalysis precipitation products with CGDPA reference dataset over mainland China.

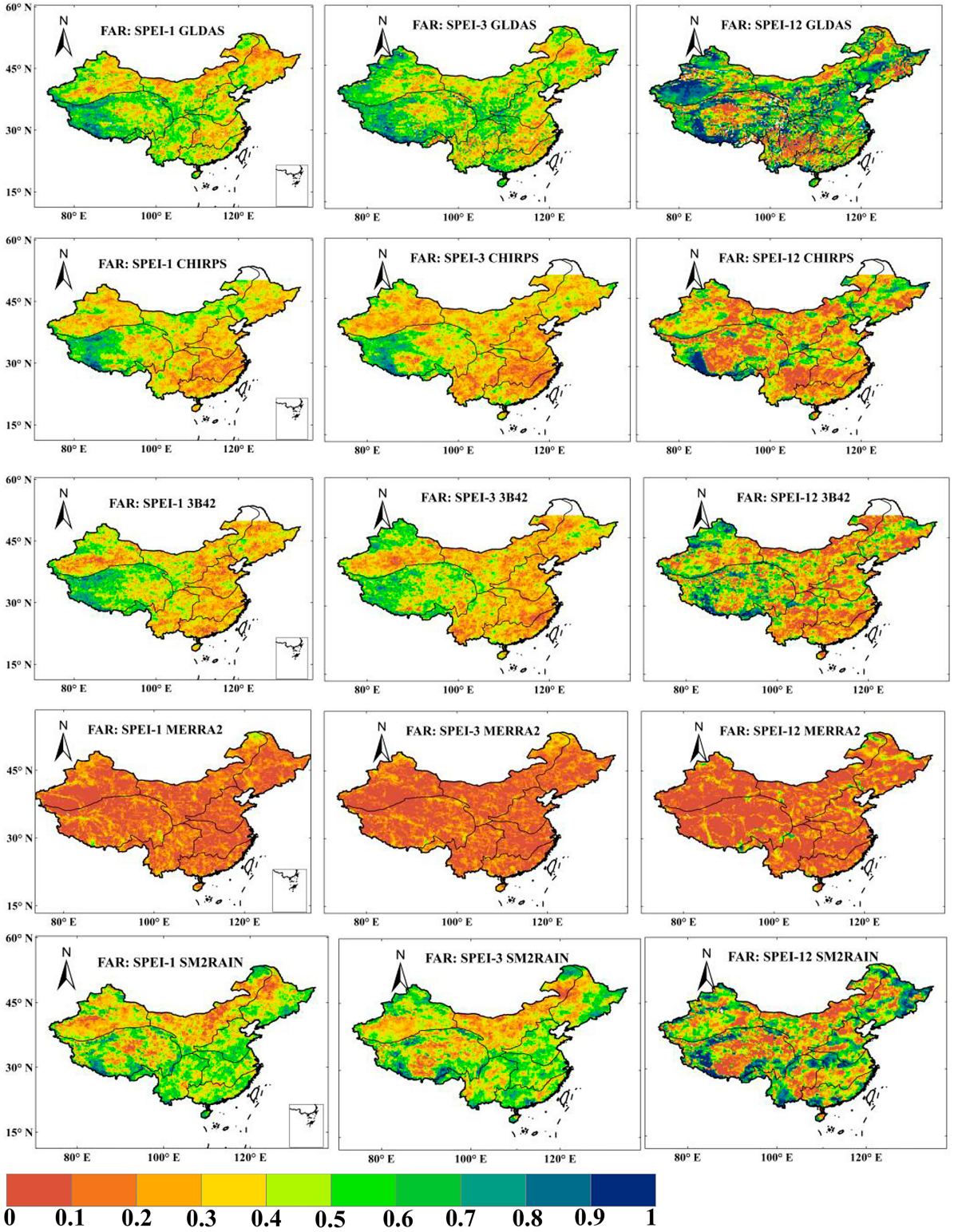


FIG. 13. (*Continued*)

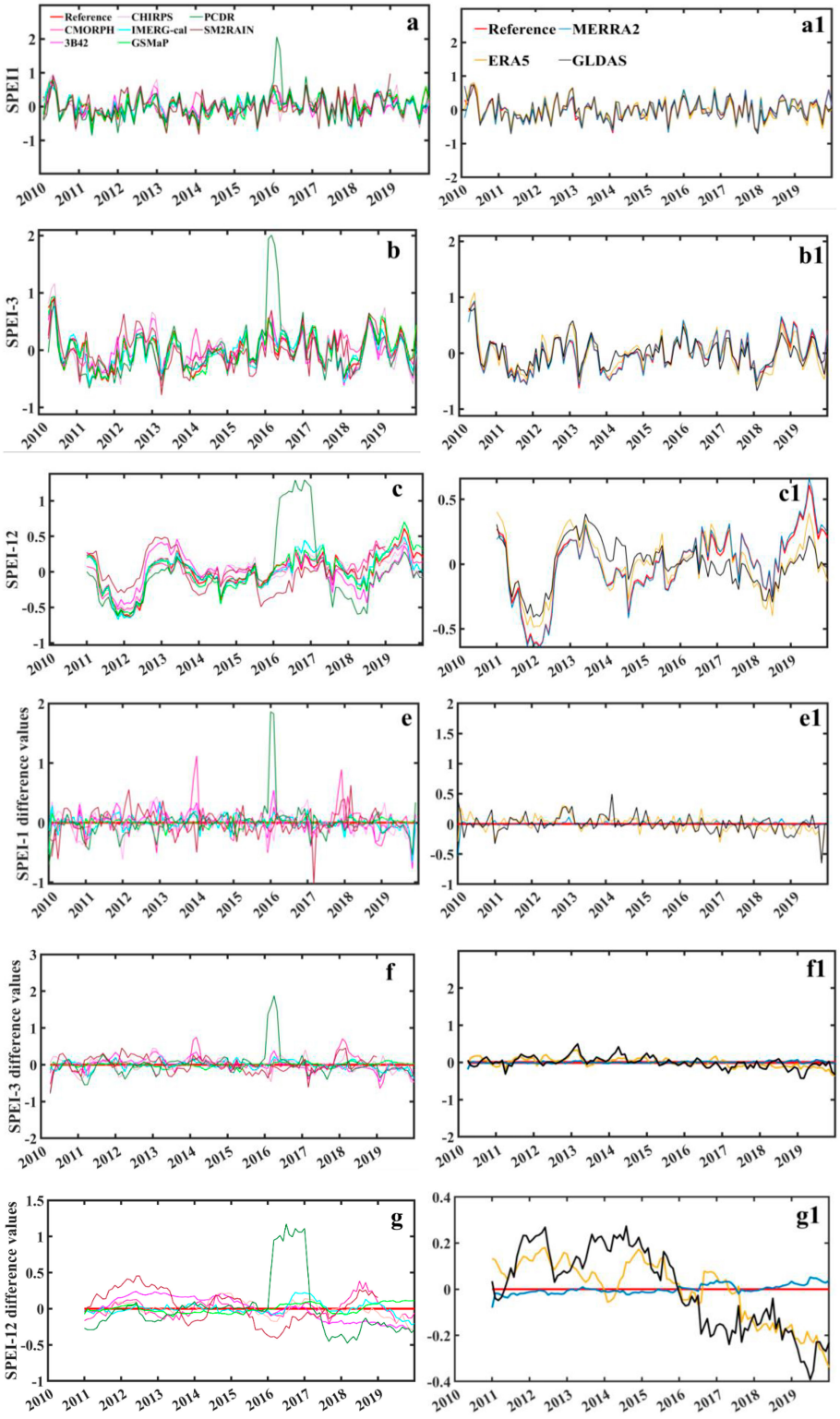


FIG. 14. Temporal changes of SPEI-1, SPEI-3, and SPEI-12 were calculated from the CGDPA and (a)–(c) satellite-based and (a1)–(c1) reanalysis products in mainland China. The difference values in SPEI-1, SPEI-3, and SPEI-12 between the CGDPA and (e)–(g) satellite-based and (e1)–(g1) reanalysis products over mainland China.

and eastern China, and small CC and high FAR observed in western China. The complex topography, high altitude, sparse distribution of in situ gauge stations, and arid atmosphere constitute a challenge for drought detection in western China, especially in the Qinghai–Tibetan Plateau. MERRA-2 and GSMaP perform much better than other products in terms of CC, POD, and FAR with respect to the products capacity to detect drought at various spatial and temporal scales. They are followed by IMERG_cal, CHIRPS, and ERA5 in drought detection performance. The reanalysis MERRA-2 product performs well in drought detection, which is largely due to it assimilating space-based observations of aerosols and representing their interactions with other physical processes in the climate system. The good performance of CHIRPS is not surprising, because the data were created for early warning systems and drought monitoring.

Precipitation products are necessary for characterizing meteorological drought. The precipitation evaluation results depend on time scales, such as is the case with IMERG, which performs better at seasonal scales than daily scales. Moreover, exploring the bias that may be introduced by the Thornthwaite PET algorithm in the SPEI drought evaluation index is necessary. High-quality, high-spatial resolution, and near-real-time long-term (over 30 years) datasets of precipitation are required for assessing and monitoring drought. However, currently, there is still a lack of such datasets. Therefore, downscaling and developing long-term real-time satellite products with high spatial–temporal resolution at a global scale deserve further study.

This study's methodology and results could provide a valuable reference for end users seeking to better understand the application of precipitation products to drought detection under different scenarios by considering terrain characteristics, seasonality, and precipitation intensity. It is hoped these evaluation results will benefit developers of precipitation products in identifying error sources and further improving retrieval algorithms.

Acknowledgments. This work was funded by the Einstein Research Unit "Climate and Water under Change" from the Einstein Foundation Berlin and Berlin University Alliance (ERU-2020-609). We are grateful to the editorial team and three anonymous reviewers for their efforts to improve this manuscript. We also thank China's Meteorological Administration for providing CGDPA products. The authors report no potential conflicts of interest.

Data availability statement. All precipitation products used during this study are available upon reasonable request. Their access approaches are displayed in Table S1. Supplemental data to this article can be found in the online supplemental material.

REFERENCES

Abbasi, A., K. Khalili, J. Behmanesh, and A. Shirzad, 2019: Drought monitoring and prediction using SPEI index and gene expression programming model in the west of Urmia Lake. *Theor. Appl. Climatol.*, **138**, 553–567, https://doi.org/10.1007/s00704-019-02825-9.

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

Alemohammad, S. H., and Coauthors, 2015: Characterization of precipitation product errors across the United States using multiplicative triple collocation. *Hydrol. Earth Syst. Sci.*, **19**, 3489–3503, https://doi.org/10.5194/hess-19-3489-2015.

Anshuka, A., F. F. van Ogtrop, and R. W. Vervoort, 2019: Drought forecasting through statistical models using standardised precipitation index: A systematic review and meta-regression analysis. *Nat. Hazards*, **97**, 955–977, https://doi.org/10.1007/s11069-019-03665-6.

Ashouri, H., and Coauthors, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1.

Baez-Villanueva, O. M., M. Zambrano-Bigiarini, L. Ribbe, A. Nauditt, J. D. Giraldo-Osorio, and N. X. Thinh, 2018: Temporal and spatial evaluation of satellite rainfall estimates over different regions in Latin-America. *Atmos. Res.*, **213**, 34–50, https://doi.org/10.1016/j.atmosres.2018.05.011.

——, and Coauthors, 2020: RF-MEP: A novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sens. Environ.*, **239**, 111606, https://doi.org/10.1016/j.rse.2019.111606.

Bai, X., X. Wu, and P. Wang, 2019: Blending long-term satellite-based precipitation data with gauge observations for drought monitoring: Considering effects of different gauge densities. *J. Hydrol.*, **577**, 124007, https://doi.org/10.1016/j.jhydrol.2019.124007.

Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

——, E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1.

Bevacqua, A. G., P. L. B. Chaffe, V. B. P. Chagas, and A. AghaKouchak, 2021: Spatial and temporal patterns of propagation from meteorological to hydrological droughts in Brazil. *J. Hydrol.*, **603**, 126902, https://doi.org/10.1016/j.jhydrol.2021.126902.

Brocca, L., and Coauthors, 2019: SM2RAIN–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations. *Earth Syst. Sci. Data*, **11**, 1583–1601, https://doi.org/10.5194/essd-11-1583-2019.

C3S, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store, accessed 21 May 2022, https://doi.org/10.24381/cds.adbb2d47.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Derin, Y., and Coauthors, 2019: Evaluation of GPM-era global satellite precipitation products over multiple complex terrain regions. *Remote Sens.*, **11**, 2936, https://doi.org/10.3390/rs11242936.

Diem, J. E., J. Hartter, S. J. Ryan, and M. W. Palace, 2014: Validation of satellite rainfall products for western Uganda. *J.*

Hydrometeor., **15**, 2030–2038, https://doi.org/10.1175/JHM-D-13-0193.1.

Dikshit, A., and B. Pradhan, 2021: Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.*, **801**, 149797, https://doi.org/10.1016/j.scitotenv.2021.149797.

——, ——, and A. Huete, 2021: An improved SPEI drought forecasting approach using the long short-term memory neural network. *J. Environ. Manage.*, **283**, 111979, https://doi.org/10.1016/j.jenvman.2021.111979.

——, ——, and M. Santosh, 2022: Artificial neural networks in drought prediction in the 21st century—A scientometric analysis. *Appl. Soft Comput.*, **114**, 108080, https://doi.org/10.1016/j.asoc.2021.108080.

Duan, Z., E. Duggan, C. Chen, H. Gao, J. Dong, and J. Liu, 2021: Comparison of traditional method and triple collocation analysis for evaluation of multiple gridded precipitation products across Germany. *J. Hydrometeor.*, **22**, 2983–2999, https://doi.org/10.1175/JHM-D-21-0049.1.

Fang, H., S. Wei, C. Jiang, and K. Scipal, 2012: Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sens. Environ.*, **124**, 610–621, https://doi.org/10.1016/j.rse.2012.06.013.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10.1038/sdata.2015.66.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

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

Gruber, A., C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45**, 200–211, https://doi.org/10.1016/j.jag.2015.09.002.

Guo, X., A. Zhu, Q. Li, Z. Xia, and R. Chen, 2022: Long-term solutions for China's heat and drought. *Science*, **378**, 1061, https://doi.org/10.1126/science.adf6012.

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

Hinge, G., M. M. Mohamed, D. Long, and M. A. Harmouda, 2021: Meta-analysis in using satellite precipitation products for drought monitoring: Lessons learnt and way forward. *Remote Sens.*, **13**, 4353, https://doi.org/10.3390/rs13214353.

Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

——, and Coauthors, 2019: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). NASA Algorithm Theoretical Basis Doc., version 6, 38 pp., https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf.

Jiang, S. H., L.-y. Wei, L.-l. Ren, L.-q. Zhang, M.-h. Wang, and H. Cui, 2023: Evaluation of IMERG, TMPA, ERA5, and CPC precipitation products over mainland China: Spatiotemporal patterns and extremes. *Water Sci. Eng.*, **16**, 45–56, https://doi.org/10.1016/j.wse.2022.05.001.

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

Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424–425**, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011.

Lei, H., H. Zhao, and T. Ao, 2022: Ground validation and error decomposition for six state-of-the-art satellite precipitation products over mainland China. *Atmos. Res.*, **269**, 106017, https://doi.org/10.1016/j.atmosres.2022.106017.

Li, C., G. Tang, and Y. Hong, 2018: Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the triple collocation method across mainland China. *J. Hydrol.*, **562**, 71–83, https://doi.org/10.1016/j.jhydrol.2018.04.039.

Liu, C. Y., P. Aryastana, G.-R. Liu, and W.-R. Huang, 2020: Assessment of satellite precipitation product estimates over Bali Island. *Atmos. Res.*, **244**, 105032, https://doi.org/10.1016/j.atmosres.2020.105032.

Liu, Y., Z. Z. Hu, R. Wu, and X. Yuan, 2022: Causes and predictability of the 2021 spring southwestern China severe drought. *Adv. Atmos. Sci.*, **39**, 1766–1776, https://doi.org/10.1007/s00376-022-1428-4.

Lu, X., G. Tang, X. Liu, X. Wang, Y. Liu, and M. Wei, 2021: The potential and uncertainty of triple collocation in assessing satellite precipitation products in central Asia. *Atmos. Res.*, **252**, 105452, https://doi.org/10.1016/j.atmosres.2021.105452.

Mantas, V. M., Z. Liu, C. Caro, and A. J. S. C. Pereira, 2015: Validation of TRMM Multi-Satellite Precipitation Analysis (TMPA) products in the Peruvian Andes. *Atmos. Res.*, **163**, 132–145, https://doi.org/10.1016/j.atmosres.2014.11.012.

Massari, C., W. Crow, and L. Brocca, 2017: An assessment of the performance of global rainfall estimates without ground-based observations. *Hydrol. Earth Syst. Sci.*, **21**, 4347–4361, https://doi.org/10.5194/hess-21-4347-2017.

Mega, T., T. Ushio, T. Kubota, M. Kachi, K. Aonashi, and S. Shige, 2014: Gauge adjusted global satellite mapping of precipitation (GSMaP_Gauge). *2014 31th URSI General Assembly and Scientific Symp.*, Beijing, China, IEEE, https://doi.org/10.1109/URSIGASS.2014.6929683.

Mishra, A. K., and V. P. Singh, 2010: A review of drought concepts. *J. Hydrol.*, **391**, 202–216, https://doi.org/10.1016/j.jhydrol.2010.07.012.

——, and ——, 2011: Drought modeling—A review. *J. Hydrol.*, **403**, 157–175, https://doi.org/10.1016/j.jhydrol.2011.03.049.

Nashwan, M. S., S. Shahid, A. Dewan, T. Ismail, and N. Alias, 2020: Performance of five high resolution satellite-based precipitation products in arid region of Egypt: An evaluation. *Atmos. Res.*, **236**, 104809, https://doi.org/10.1016/j.atmosres.2019.104809.

Navarro, A., E. García-Ortega, A. Merino, and J. L. Sánchez, 2020: Extreme events of precipitation over complex terrain derived from satellite data for climate applications: An evaluation of the southern slopes of the Pyrenees. *Remote Sens.*, **12**, 2171, https://doi.org/10.3390/rs12132171.

Nguyen, P., and Coauthors, 2019: The CHRS data portal, an easily accessible public repository for PERSIANN global satellite

precipitation data. *Sci. Data*, **6**, 180296, https://doi.org/10.1038/sdata.2018.296.

Noor, M., T. Ismail, S. Shahid, M. Asaduzzaman, and A. Dewan, 2021: Evaluating intensity-duration-frequency (IDF) curves of satellite-based precipitation datasets in peninsular Malaysia. *Atmos. Res.*, **248**, 105203, https://doi.org/10.1016/j.atmosres.2020.105203.

Prakash, S., A. K. Mitra, A. AghaKouchak, Z. Liu, H. Norouzi, and D. S. Pai, 2018: A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.*, **556**, 865–876, https://doi.org/10.1016/j.jhydrol.2016.01.029.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, https://doi.org/10.1175/BAMS-85-3-381.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

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

Shi, Y., Y. Shen, E. Kang, D. Li, Y. Ding, G. Zhang, and R. Hu, 2007: Recent and future climate change in northwest China. *Climatic Change*, **80**, 379–393, https://doi.org/10.1007/s10584-006-9121-7.

Shiru, M. S., S. Shahid, N. Alias, and E.-S. Chung, 2018: Trend analysis of droughts during crop growing seasons of Nigeria. *Sustainability*, **10**, 871, https://doi.org/10.3390/su10030871.

Su, B., and Coauthors, 2018: Drought losses in China might double between the 1.5°C and 2.0°C warming. *Proc. Natl. Acad. Sci. USA*, **115**, 10 600–10 605, https://doi.org/10.1073/pnas.1802129115.

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

Tang, G., Y. Ma, D. Long, L. Zhong, and Y. Hong, 2016: Evaluation of GPM day-1 IMERG and TMPA version-7 legacy products over mainland China at multiple spatiotemporal scales. *J. Hydrol.*, **533**, 152–167, https://doi.org/10.1016/j.jhydrol.2015.12.008.

——, M. P. Clark, S. M. Papalexiou, Z. Ma, and Y. Hong, 2020: Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sens. Environ.*, **240**, 111697, https://doi.org/10.1016/j.rse.2020.111697.

Tarek, M., F. P. Brissette, and R. Arsenault, 2020: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.*, **24**, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Thornthwaite, C. W., 1948: An approach toward a rational classification of climate. *Geogr. Rev.*, **38**, 55–94, https://doi.org/10.2307/210739.

Trinh-Tuan, L., J. Matsumoto, T. Ngo-Duc, M. I. Nodzu, and T. Inoue, 2019: Evaluation of satellite precipitation products over central Vietnam. *Prog. Earth Planet. Sci.*, **6**, 54, https://doi.org/10.1186/s40645-019-0297-7.

van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning, 2014: A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrol. Earth Syst. Sci.*, **18**, 2955–2973, https://doi.org/10.5194/hess-18-2955-2014.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1.

Wang, C., G. Tang, Z. Han, X. Guo, and Y. Hong, 2018: Global intercomparison and regional evaluation of GPM IMERG version-03, version-04 and its latest version-05 precipitation products: Similarity, difference and improvements. *J. Hydrol.*, **564**, 342–356, https://doi.org/10.1016/j.jhydrol.2018.06.064.

Wang, W., W. Cui, X. Wang, and X. Chen, 2016: Evaluation of GLDAS-1 and GLDAS-2 forcing data and Noah model simulations over China at the monthly scale. *J. Hydrometeor.*, **17**, 2815–2833, https://doi.org/10.1175/JHM-D-15-0191.1.

Wilhite, D. A., and M. H. Glantz, 1985: The drought phenomenon: The role of definitions. *Water Int.*, **10**, 111–120, https://doi.org/10.1080/02508068508686328.

Willmott, C. J., 1981: On the validation of models. *Phys. Geogr.*, **2**, 184–194, https://doi.org/10.1080/02723646.1981.10642213.

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

Xu, S., C. Wu, L. Wang, A. Gonsamo, Y. Shen, and Z. Niu, 2015: A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sens. Environ.*, **162**, 119–140, https://doi.org/10.1016/j.rse.2015.02.024.

Yu, L., G. Leng, and A. Python, 2022: A comprehensive validation for GPM IMERG precipitation products to detect extremes and drought over mainland China. *Wea. Climate Extremes*, **36**, 100458, https://doi.org/10.1016/j.wace.2022.100458.

Zambrano-Bigiarini, M., A. Nauditt, C. Birkel, K. Verbist, and L. Ribbe, 2017: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrol. Earth Syst. Sci.*, **21**, 1295–1320, https://doi.org/10.5194/hess-21-1295-2017.

Zhang, H., H. A. Loáiciga, D. Ha, and Q. Du, 2020: Spatial and temporal downscaling of TRMM precipitation with novel algorithms. *J. Hydrometeor.*, **21**, 1259–1278, https://doi.org/10.1175/JHM-D-19-0289.1.

Zhang, Q., R. Shi, V. P. Singh, C.-Y. Xu, H. Yu, K. Fan, and Z. Wu, 2022: Droughts across China: Drought factors, prediction and impacts. *Sci. Total Environ.*, **803**, 150018, https://doi.org/10.1016/j.scitotenv.2021.150018.

Zhong, R., X. Chen, C. Lai, Z. Wang, Y. Lian, H. Yu, and X. Wu, 2019: Drought monitoring utility of satellite-based precipitation products across mainland China. *J. Hydrol.*, **568**, 343–359, https://doi.org/10.1016/j.jhydrol.2018.10.072.

Zhou, Z., B. Guo, W. Xing, J. Zhou, F. Xu, and Y. Xu, 2020: Comprehensive evaluation of latest GPM era IMERG and GSMaP precipitation products over mainland China. *Atmos. Res.*, **246**, 105132, https://doi.org/10.1016/j.atmosres.2020.105132.