Article ID: 1006-8775(2025)05-0497-14

A Precipitation Classification Scheme for China with Special Consideration of Extreme Events

MA Ya-yu (马亚宇)[1,2,3], WANG Jing-song (王劲松)[2,3], ZHAO Liang (赵 亮)[4], LI Wen-juan (李文娟)[5], YANG Hao (杨 昊)[1], WEN Wu (文 武)[1], ZHOU Hang (周 航)[1], CHEN Min (陈 敏)[6]

(1. School of Computer Science, Chengdu University of Information Technology, Chengdu 610225 China; 2. Key Laboratory of Space Weather, National Satellite Meteorological Center (National Center for Space Weather), China Meteorological Administration, Beijing 100081 China; 3. Innovation Center for FengYun Meteorological Satellite (FYSIC), Beijing 100081 China; 4. LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029 China; 5. Shanxi Meteorological Service Center, Taiyuan 030002 China; 6. Department of Student Affairs, Chengdu University of Information Technology, Chengdu 610225 China)

Abstract: Due to global warming, extreme weather and climate events are becoming more frequent, highlighting the need to explore the changing characteristics of precipitation in China, including extreme precipitation. A clustering algorithm was developed to classify summer (June, July, and August) daily precipitation in China from 1961 to 2020, considering spatial distribution, standard deviations, and frequency of extreme precipitation events. The results reveal six distinct precipitation climate zones, a classification that differs from previous divisions. While overall precipitation has decreased in most regions, the frequency of extreme precipitation events has increased across all clusters, indicating a shift in precipitation distribution patterns. Analysis shows that the weakened Lake Baikal blocking high and strengthened Mongolian cyclone influence the arid region in northwest China (Cluster 1), which is characterized by the lowest precipitation. The transition zone between the monsoon and arid region (Cluster 2) is affected by the Mongolian cyclone, water vapor transport from the Indian Ocean, and shifts in the monsoon boundary. Clusters 3 and 4 represent areas associated with advancement and retreat of the summer monsoon. In the Meiyu region, two distinct subregions have been identified exist. Cluster 4 is primarily influenced by the East Asia-Pacific wave train. Despite sharing similar climate drivers and proximity, Clusters 4 and 5 differ significantly due to topographic variations and disparate levels of urbanization. Cluster 5 exhibits a higher average precipitation, greater variability, and more frequent extreme events. Cluster 6 exhibits the highest overall precipitation in the coastal areas of Guangdong and Guangxi, where abundant water vapor contributes to a higher frequency of extreme precipitation. In addition, anthropogenic activities and urbanization significantly influence precipitation in Beijing-Tianjin-Hebei and Yangtze River Delta regions. This research proposes a precipitation classification scheme integrating multiple precipitation parameters, providing support for risk management and mitigation strategies in the face of increasing extreme precipitation events.

Key words: precipitation characteristics; extreme precipitation; *K*-means clustering algorithm; precipitation classification scheme; risk management

CLC number: P426.6 **Document code:** A

Citation: Ma, Y. Y., J. S. Wang, L. Zhao, W. J. Li, H. Yang, W. Wen, H. Zhou, and M. Chen, 2025: A Precipitation Classification Scheme for China with Special Consideration of Extreme Events. *J. Trop. Meteor.*, **31**, 497–510, https://doi.org/10.3724/j.1006-8775.2025.041.

1 INTRODUCTION

Human activities have contributed to the warming of the atmosphere, oceans, and land. The warming of the climate system is projected to continue until the middle of the 21st century. Global warming does not only lead to higher temperatures but also intensifies the frequency and severity of extreme weather events (Kirchmeier-Young et al. 2020; Tabari 2020). According to the Clausius-Clapeyron relationship, atmospheric moisture increases with climate warming, leading to increased precipitation extremes (Held et al. 2006; O'Gorman et al. 2009). The uneven spatial and temporal distribution of precipitation may trigger regional drought or flooding and may affect the surface ecosystem (Ding et al. 2008). Therefore, analyzing the spatial and temporal distribution of precipitation not only enhances the understanding of the regional climate system but is also crucial for water resources management and risk assessment related to droughts and floods.

Under the background of global warming, precipitation distribution worldwide has exhibited new patterns of variability. Wang et al. (2010) studied the variability of summer precipitation in the southeast United States and found that droughts and anomalous rainfall

Received 2024-08-28; **Revised** 2025-06-04; **Accepted** 2025-08-05

Funding: National Natural Science Foundation of China (U2442202, 42274217, 62441501); Key Innovation Team of China Meteorological Administration (CMA2024ZD01); Scientific Research Foundation of CUIT (376278, KYTZ202158)

Biography: MA Ya-yu, Ph.D., primarily undertaking research on climate change and remote sensing.

Corresponding author: WANG Jing-song, e-mail: wangjs@cma.gov.cn

events during 1978–2007 were more intense than those during 1948–1977. The annual variability in precipitation and precipitation days increased from east to west. In Africa, a sharp decline in precipitation began around 1999 (Lyon et al. 2014). There was a significant negative trend (95% confidence level) in the total annual precipitation in most Mediterranean regions from 1901 to 2009. The number of precipitation days decreased by 20% in the eastern Mediterranean, while the decrease was not significant in the western and central Mediterranean (Philandras et al. 2011). Furthermore, early evidence of long-term rainfall reduction has also been observed in Africa (AghaKouchak 2015). In China, climate warming also has a significant impact on precipitation patterns (Ma et al. 2021). Each year, extreme precipitation events linked to climate change result in substantial property damage and loss of life (Burke et al. 2016; Zhang et al. 2020). Over the past 60 years, most regions have experienced an increase in extreme precipitation events, while some areas have seen a decline in total precipitation. Li et al. (2021) analyzed the statistical characteristics of light and extreme precipitation events in Xinjiang. Their results showed that light precipitation contributed greatly to the annual total number of rainy days, while extreme events contributed for a large share of total annual precipitation. Wang et al. (2015) found that the annual precipitation in North China and the eastern Northwest China showed a downward trend. Liang et al. (2011) studied the spatiotemporal variation characteristics of precipitation in semi-humid and semi-arid Northeast China and reported that from 1961 to 2008, summer precipitation in Northeast China showed a decreasing trend. Lu et al. (2021) analyzed the spatio-temporal characteristics of precipitation in South China over the past 52 years and found that only 8% of rainfall stations exhibited a significant increase in annual rainfall. Ren et al. (2000) observed that summer precipitation increased in the middle and lower reaches of the Yangtze River, while the Yellow River basin experienced a decrease. Overall, China's precipitation trends remain uncertain, but there are clear spatial differences. Precipitation has declined in central and southern Northeast China, North China, Central China, and Southwest China, while it has increased in southeast coastal areas, the lower reaches of the Yangtze River, the Qinghai-Tibet Plateau, and northwest China increased (Li et al. 2012; You et al. 2011).

Global warming has rendered the global climate more extreme, with extreme weather and climate events rising in both intensity and frequency. Studies have shown that over the past half-century, annual precipitation in China has shown obvious interdecadal and multi-decadal fluctuations (Feng et al. 2011; Zhai et al. 2005). Around the turn of the 21st century, many studies used observational data to analyze the spatial characteristics of extreme precipitation (Zhai et al. 1999; Ren et al. 2000). In the past two decades, however, reanalysis data have been increasingly used to investigate the frequency and trends of extreme precipitation, placing less emphasis on observational studies focused on its distribution and variability (Pendergrass et al. 2017). Although the global climate models projected a growing influence of anthropogenic forcing on extreme precipitation, detecting such changes in the observational record remains challenging. Therefore, it is necessary to further investigate the spatial distribution, frequency, and intensity changes of both mean and extreme precipitation in China.

Currently, many scholars have paid attention to the distribution and variation of mean precipitation (Huang et al. 2011), while the precipitation variability has received less attention, despite its important theoretical and practical significance. Besides, there are many researches focused on the climate classification (Zheng et al. 2010; Beck et al. 2018). However, these climate classifications primarily consider the mean values and trends of climate variables, while classifications based on precipitation characteristics—particularly summer extreme precipitation—have been largely overlooked. Generally, summer accounts for the largest share of annual precipitation in China (56.5%) (Yao et al. 2017). Extreme precipitation and the disasters such as floods and landslides caused by extreme precipitation occur most frequently in summer. Therefore, classifying precipitation based on its characteristics—especially extreme events—is essential for effective flood control and risk management.

This paper proposes a nove approach to analyze the distribution characteristics of precipitation and extreme precipitation events during the summer season in China. By considering the mean, standard deviation, and frequency of extreme precipitation, the study employs the *K*-means clustering algorithm to classify daily precipitation patterns, thereby developing a new precipitation classification scheme. The main objectives are to objectively describe precipitation distribution, comprehensively assess the severity of abnormal precipitation, and accurately depict the spatial and temporal patterns of extreme precipitation frequency and intensity. Furthermore, the paper explores the physical mechanisms underlying the identified classification results.

2 METHODS

2.1 *K-means clustering of daily precipitation using mean, standard deviation, and extreme event frequency*

As one of the traditional algorithms in machine learning, clustering has been widely used in many fields because of its practicality, efficiency, and intuitive understanding of data structure (Sardar et al. 2018; Tleis et al. 2017). The core theory of clustering is to reveal the valuable information hidden in the dataset (Ashabi et al. 2020). The *K*-means algorithm is a classical clustering method, which is widely used because of its simplicity, efficiency, and interpretability (Yokoi et al. 2011; Carvalho et al. 2016; Falkena et al. 2020). As an unsupervised learning algorithm, the *K*-means was firstly introduced in

signal processing by MacQueen (1967).

The purpose of *K*-means clustering is to divide *N* observations into *K* clusters, and each observation belongs to the cluster with the closest mean. Euclidean distance is typically used to measure the similarity between data points, where a smaller distance indicates greater similarity. The algorithm begins by randomly selecting *K* initial centroids. Based on the similarity between each data point and the cluster centroids, the centroids are iteratively updated to minimize the Sum of Squared Errors (SSE) within each cluster. When the SSE stabilizes or the objective function converges, the centroid becomes fixed, and the clustering process is complete.

With the intensification of climate change, the frequency and intensity of extreme precipitation events have been increasing, significantly impacting socio-economic activities and ecological environments during summer (June, July, and August). Traditional climate zoning methods, which primarily focus on the mean values and trends of meteorological parameters, often neglect the characteristics of extreme precipitation, including its frequency, intensity, and spatio-temporal distribution. These approaches no longer adequately meet the needs of society. In general, daily precipitation exceeding 50 mm is defined as a rainstorm (RS50), while precipitation above 100 mm is considered a heavy rainstorm (RS100). To address the emerging climate challenges, this study incorporates not only mean precipitation, but also its variations and extreme events (RS50/RS100). It develops an approach for precipitation climate zoning using the *K*-means cluster algorithm.

Besides, datasets are often preprocessed to make them clean, consistent, and noise-free. Because Euclidean distance is highly sensitive to differences in the magnitude of variables, data normalization is an essential step before clustering (Visalakshi and Thangavel 2009). To eliminate data redundancy and improve the efficiency of the clustering algorithm, the *Z*-score method (also known as zero-mean normalization) is used to make the mean value of data equal to 0 and standard deviation equal to 1 before *K*-means clustering.

2.2 *Elbow method*

When performing *K*-means clustering, the cost function is defined as the Within-Cluster Sum of Squares (WCSS). WCSS is the sum of squared Euclidean distances between each data point and the respective centroid of its assigned cluster. The expected value of WCSS decreases with successive iterations, and the algorithm's convergence depends on the dimension of the dataset and the number of centroids. The elbow method, suitable for multivariate datasets, is commonly used to determine the optimal number of cluster centers. It involves fitting the cluster center value to the data. In this paper, 2–20 cluster central values are tested. For each cluster count, the algorithm is run for 300 iterations, and the final WCSS value is reported. The "elbow point" is identified where WCSS shows a noticeable drop before leveling off.

3 DATA

The daily precipitation dataset for 699 stations in China from 1961 to 2020 was obtained from China Meteorological Information Center. Sea surface temperature data for the period 1960—2020, with a spatial resolution of 1°×1°, were provided by Earth System Research Laboratory Physical Sciences Division (ESRL/PSD) (Folland and Parker 1995; Ishii et al. 2005; Japan Meteorological Agency 2006). The water vapor flux data were sourced from ECMWF Reanalysis v5 (Hersbach et al. 2019; https://www.ecmwf.int/en/forecasts/datasets/search/ERA5), covering 1979–2020 at a resolution of 1°×1°.

Besides, data on geopotential height at 500 hPa, zonal wind (u), and meridional wind (v) at 700 hPa and 850 hPa were obtained from NCEP-NCAR reanalysis1 dataset (Kalnay et al. 1996), spanning 1948–2020 with a spatial resolution of 2.5°×2.5°.

4 RESULTS

4.1 *Spatio-temporal distribution of summer precipitation*

Figure 1a illustrates the temporal trend of the rainfall station number in the dataset. In this study, 648 stations with more than 80% data completeness during the summer months were selected for analysis. Fig. 1a depicts the temporal variation in the number of these selected stations that provided complete observations for each summer spanning 1961–2020. From 1961 to 1980, the number of rain stations in the dataset increased from 660 to 690. Between 1980 and 2000, the number remained relatively stable. After 2000, the dataset became essentially stable in terms of station count. As shown in the figure, the number of rainfall stations fluctuated slightly (<20 stations annually) from 1961 to 1970 before stabilizing in the subsequent period.

Figure 1b illustrates the temporal variation in the frequency of RS50 (red) and RS100 (green) events in

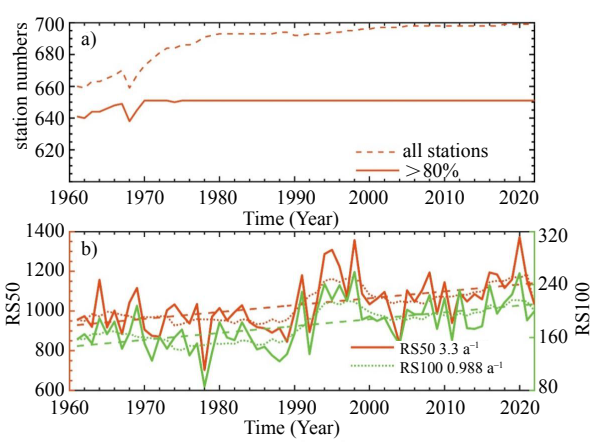


Figure 1. (a) The number of stations with complete observation data in summer (June, July, and August) in China from 1961 to 2020 (red dashed curve) and with observation time series integrity above 80% (red solid curve). (b) The frequency of RS100 (green)/RS50 (red) (solid curve), along with their trends (dashed line) and 10-year moving averages (dotted curve).

China from 1961 to 2020. The solid red curve shows an increasing trend in RS50 frequency. From 1961 to 1990, except for a minimum in 1978, the frequency remained relatively stable. During 1991–2000, there was a notable peak, with over 1200 RS50 events recorded in 1994, 1995, 1996, and 1998. During the first decade of the 21st century, the frequency remained generally stable, but from 2011 to 2020, a clear upward trend emerged. In 2020, the frequency of extreme precipitation reached 1371, the highest on record since 1961. The green curve represented RS100 frequency, which was significantly lower than RS50 but exhibited a similar temporal pattern. The 1991–2000 period was marked by a notably high frequency of both RS50 and RS100 events, indicating a rise in extreme precipitation. Notably, 2020 recorded 256 RS100 events, which was the highest annual total during the study period. These observational results provide strong evidence that the frequency of extreme precipitation events in China has been increasing over the past six decades.

Understanding regional precipitation characteristics requires considering not only the mean precipitation but also the variability and frequency of extreme events, providing a more comprehensive depiction of precipitation behavior. Fig. 2a displays the spatial distribution of the frequency of daily precipitation exceeding 50 mm (RS50) during the summer months (June, July, and August) at each station from 1961 to 2020. In western China, RS50 events have occurred fewer than 40 times over the past 60 years. In Northeast China, the frequency ranges from 40 to 60 times. Notably, the Liaodong Peninsula, Bohai Bay, Beijing, and Tianjin exhibit significantly higher RS50 frequencies than other areas at similar latitudes. South of the Qinling Mountains-Huaihe River line, the frequency of extreme precipitation is generally high. In northern Jiangsu, the border of Hubei, Jiangxi, Anhui, and Guizhou, the frequency of the rainstorms exceeds 150 times. In the southern coastal areas of Guangdong and Guangxi, the frequency of RS50 is more than 180 times. Fig. 2b describes the spatial distribution of RS100 frequency from 1961 to 2020. As shown, daily RS100 events are rare in non-monsoon regions. However, along the Hubei-Jiangxi-Anhui border and in the coastal areas of Guangdong and Guangxi, the frequencies of RS100 exceed 40 events.

The mean precipitation reflects the climatological precipitation, while the standard deviation indicates the degree of variability. Both of them are essential for assessing the precipitation state. Fig. 2c illustrates the spatial distribution of mean daily summer precipitation across China. The distribution exhibits significant spatial variability, with a gradual increase from north to south. In Northwest China, daily precipitation is less than 2 mm. In Northeast China, North China, and Shanxi province, the average daily precipitation ranges from 2 mm to 5 mm. In contrast, areas surrounding Bohai bay and in eastern Liaoning province, the average daily precipitation in summer ranges from 6 mm to 7 mm. Northern Jiangsu, Jiangxi, Fujian, Chongqing, and Guizhou experience significantly higher precipitation compared to other provinces at the same latitude, with mean precipitation exceeding 7 mm. The highest mean precipitation is found in Guangxi, Guangdong, and southern Yunnan province.

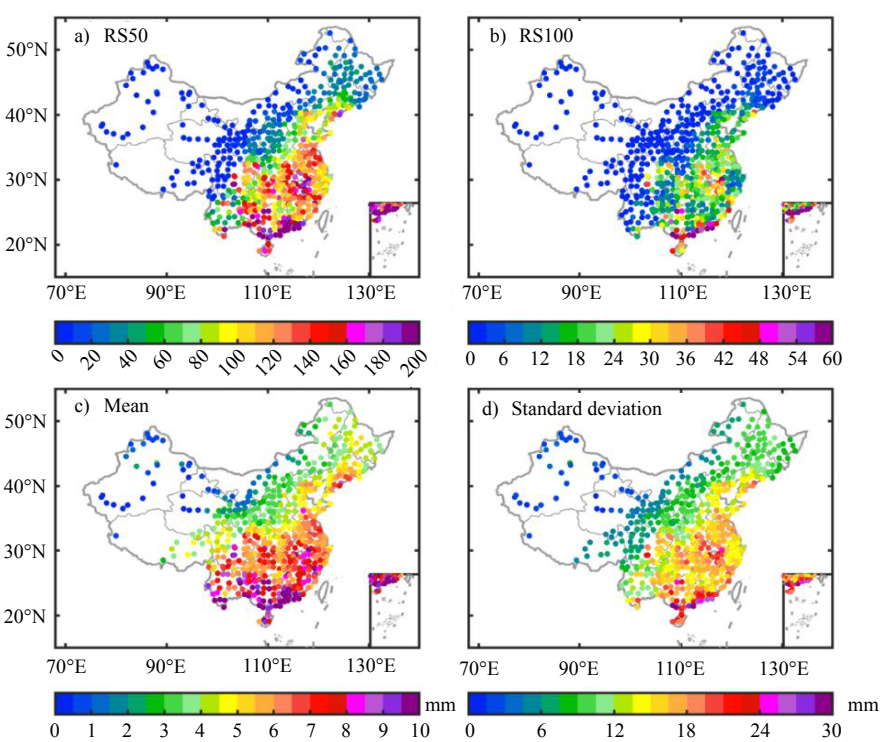


Figure 2. The spatial distributions of the frequency of (a) RS50 and (b) RS100 in summer (June, July, and August) in China from 1961 to 2020. (c) The spatial distribution of daily precipitation means and (d) standard deviation of daily precipitation.

Fig. 2d illustrates the spatial distribution of the standard deviation of summer daily precipitation. The spatial distribution of precipitation standard deviation is higher in the south than in the north, and higher in the east than in the west. In northwest China and North China, the standard deviation of precipitation is generally lower than 12 mm. In Bohai Bay and southern Liaoning province, it reaches around 15 mm. In southern China, the standard deviation values generally exceed 15 mm. However, the neighboring regions of Hubei, Anhui, and Jiangxi, as well as the south of Guangxi and Guangdong exhibit significantly large value centers, with the standard deviation surpassing 20 mm. It is worth mentioning that southwest China is characterized by high but stable summer precipitation, indicating a high mean with low variability.

4.2 *Distribution characteristics of daily precipitation using K-means clustering*

Based on the mean and standard deviation of daily precipitation, as well as the spatial distribution of the frequency of RS50/RS100 in the summer from 1961 to 2020 in China (Fig. 2), the *K*-means clustering algorithm was used to cluster the daily precipitation in China. As shown in Fig. 3a, an elbow-shaped inflection point is observed at the optimal number of clusters when the cluster center is set to 6. Beyond this point, the WCSS changes less significantly.

The unsupervised *K*-means clustering method is used to cluster the daily precipitation in China (Fig. 3b). The clustering result based on the precipitation characteristics is obviously different with existing climatic division in China (Zheng et al. 2010; Liu et al. 2022). In Fig. 3b, Cluster 1 and Cluster 2 are mainly distributed in northwest China, northeast China, north China, and the eastern Qinghai-Tibet Plateau. Although western China generally experiences less extreme precipitation, distinct classifications are observed, highlighting the region's distribution characteristics. The Liaodong Peninsula, the Beijing-Tianjin-Hebei region, the estuary of the Yangtze River, and parts of southwest China are mainly classified as Cluster 3 (purple). Cluster 4 and Cluster 5 are interspersed mainly across East China, Central China, and South China. A few rain stations classified as Cluster 6 are mainly distributed along the coastal area of South China. However, there are differences compared to the traditional climatic division. Cluster 1, Cluster 2, and Cluster 3 basically show a northeast-southwest pattern, representing arid and semi-arid areas. This pattern may be related to the topography and the boundary of the monsoon. On the other hand, Cluster 4, Cluster 5, and Cluster 6 belong to subtropical and tropical monsoon regions, and no longer follow the northeast-southwest trend. This difference may be due to the fact that these clusters lie within the monsoon region, and their distribution trend is less affected by the monsoon boundary.

This research focuses on summer precipitation in China, as it accounts for over 56.5% of the annual total rainfall (Yao et al. 2017), with most heavy-rainfall and flood events occurring during this season. While some studies use minute-scale precipitation data, such a short scale may not capture extreme events (Zhan et al. 2021). Conversely, traditional analyses using monthly or annual data may obscure key features of summer precipitation (Cao et al. 2013; Ren et al. 2015). By analyzing mean precipitation, frequency of extremes, and variability, this study aims to identify regional charateristics of precipitation-relatd risk to support more targeted disaster risk management.

Compared with traditional classification methods based on probability distributions or trends, this study offers greater objectivity and practicality. Previous methods often relied on a single metric, such as mean precipitation, which fail to fully capture the complexity of regional differences. For example, Yunnan has a mean precipitation similar to parts of central and southern China (Fig. 2c), but its lower variability and extremes distinguish it from regions such as Hebei and Henan, which traditional methods may incorrectly group together (Fig. 3b). While some studies consider multiple indicators, the final classifications often depend on subjective judgment, introducing potential bias (Zhan et al. 2021). In contrast, this study applies the *K*-means clustering algorithm to automatically detect similarities among precipitation features, reducing human intervention and improving objectivity. Despite its simplicity, the method performs well and is easy to implement, providing more accurate and practical classification results for disaster risk management.

Most precipitation-classification studies emphasize spatial contiguity, often grouping regions based on composite indicators such as mean, variability, and extremes, as seen in the well-known 400-mm Hu Line demarcating monsoon and non-monsoon zones (Hu 1935). However, in reality, even neighboring regions may exhibit distinct precipitation characteristics due to differences in land surface conditions and urbanization levels. The clustering results in this study reveal a more dispersed spatial pattern, highlighting the need for localized analysis. For instance, the Yangtze River Delta is grouped with northern Hebei and Shandong, suggesting that despite its higher rainfall (Fig. 2c), its lower variability implies reduced disaster risk. In contrast, the Poyang Lake region (Cluster 5), located at the junction of Jiangxi, Zhejiang, and Anhui, is surrounded by Cluster 4 (Fig. 3b), indicating significantly different precipitation characteristics and the need for special attention and tailored management strategies.

It can be challenging to discern the specific differences and commonalities among mean precipitation, its standard deviation, and extreme precipitation based solely on the spatial distribution of different clusters (Fig. 3b). To gain further insights, scatterplots of mean and standard deviation of summer precipitation, as well as the

cumulative frequency of extreme precipitation (>50 mm & >100 mm), are plotted in Figs. 3c and 3d. As shown in Fig. 3c, after spatial normalization, the mean and standard deviation of all clusters are approximately linearly related. The cumulative frequency of extreme precipitation increases gradually from northwest to southeast, while the cluster centroids show a nonlinear increase. Although Cluster 1 contains many rain stations, the frequencies of extreme precipitation at these stations are low and overlapping (Fig. 3d). The central values of the mean precipitation of each cluster (Table 1) show that Northwest China (Cluster 1) has the lowest mean and standard deviation of summer daily precipitation, at 1.3 mm and 3.8 mm, respectively. In contrast, the southern regions of Guangdong and Guangxi (Cluster 6) have the highest values, with a mean of 12.3 mm and a standard deviation of 25.9 mm. Notably, the sparser stations do not affect the classification results, as the *K*-means algorithm classifies the data based on their similarities. Table 1 shows the number of stations in each cluster. For Cluster 6, there are only 14 stations, despite its location in southern China, which has the highest density of rain stations.

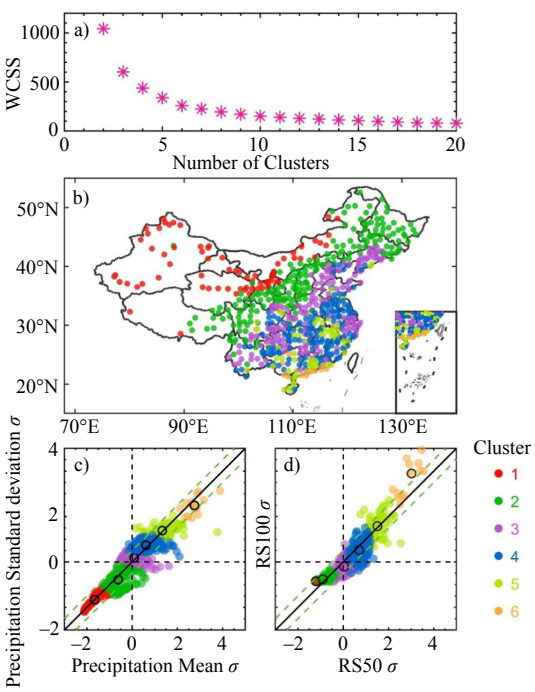


Figure 3. (a) Variation of WCSS and (b) clustering results from *K*-means analysis based on the spatial distribution of RS50, RS100, summer precipitation mean, and standard deviation in China from 1961 to 2020. (c) Scatter plot of precipitation mean vs. standard deviation, and (d) RS50 vs. RS100 frequency. Colored circles (red, green, purple, blue, light green, orange) represent Clusters 1–6 and black-bordered circles indicate cluster centers. All variables are spatially normalized. Green dashed lines mark the ±0.5σ threshold for standard deviation.

Figure 4a illustrates the cumulative probability distribution for each cluster, considering only observations with precipitation greater than 0.1 mm. The precipitation threshold of each cluster changes slightly as the cumulative probability ranges from 0 to 50%. For Cluster 1, the precipitation only increases by 5.5 mm as the cumulative probability increases from 50% to 95%. However, other clusters show much larger increases over the same range. For example, in Cluster 2 (Northeast China, North China, Yunnan-Guizhou Plateau, and surrounding areas), the 95th percentile precipitation threshold reaches 27.5 mm. The precipitation threshold for Cluster 6 exhibits a wide range of thresholds, varying from 7.1 mm to 76.4 mm as the cumulation probability increases from 50% to 95%. Fig. 4b shows the probability density distribution of precipitation for six clusters. In all clusters, the most probable rainfall amount is 0.1 mm, and the probability density decreases exponentially with increasing precipitation. Fig. 4c presents the log10-scaled precipitation density for values greater than 1 mm, enhancing the visualization of extreme precipitation events. For precipitation amounts below 8 mm, the probability densities decrease from Cluster 1 to Cluster 6. However, for precipitation exceeding 16 mm, the probability densities increase from Cluster 2 to Cluster 6. These patterns demonstrate that our precipitation classification effectively captures the distinct characteristics of extreme precipitation for each cluster.

We calculate the frequency and variation trend of RS50/RS100 in different clusters (Figs. 4d and 4e). To account for the varying number of samples in each cluster, we divide the frequency of extreme precipitation by the number of rain stations within each cluster. Our analysis reveals a decreasing trend in the frequency of extreme precipitation from south to north. Specifically, we observe a significant increasing trend of extreme precipitation in south Guangdong and Guangxi (Cluster 6), suggesting an intensification of extreme precipitation events in these areas. Although a larger slope generally corresponds to a smaller *p*-value, indicating a more significant trend, Cluster 6—despite exhibiting the largest slope—does not show a statistically significant trend. This is primarily due to high interannual variability, which reduces the stability of the trend and weakens statistical power. Clusters 3, 4, and 5—despite having lower extreme precipitation frequencies—exhibit more stable patterns and statistically significant upward trends. The northwest region (Cluster 1) and the semi-arid region (Cluster 2)

Table 1. Station numbers and final centroids of the mean (units: mm) and standard deviation (units: mm) of daily precipitation for each cluster as determined by *K*-means clustering analysis from 1961 to 2020.

Cluster	1	2	3	4	5	6
Station numbers	70	189	155	157	63	14
Mean	1.3	3.9	5.7	7.0	8.7	12.3
Standard deviation	3.8	8.5	13.6	16.6	20.0	25.9

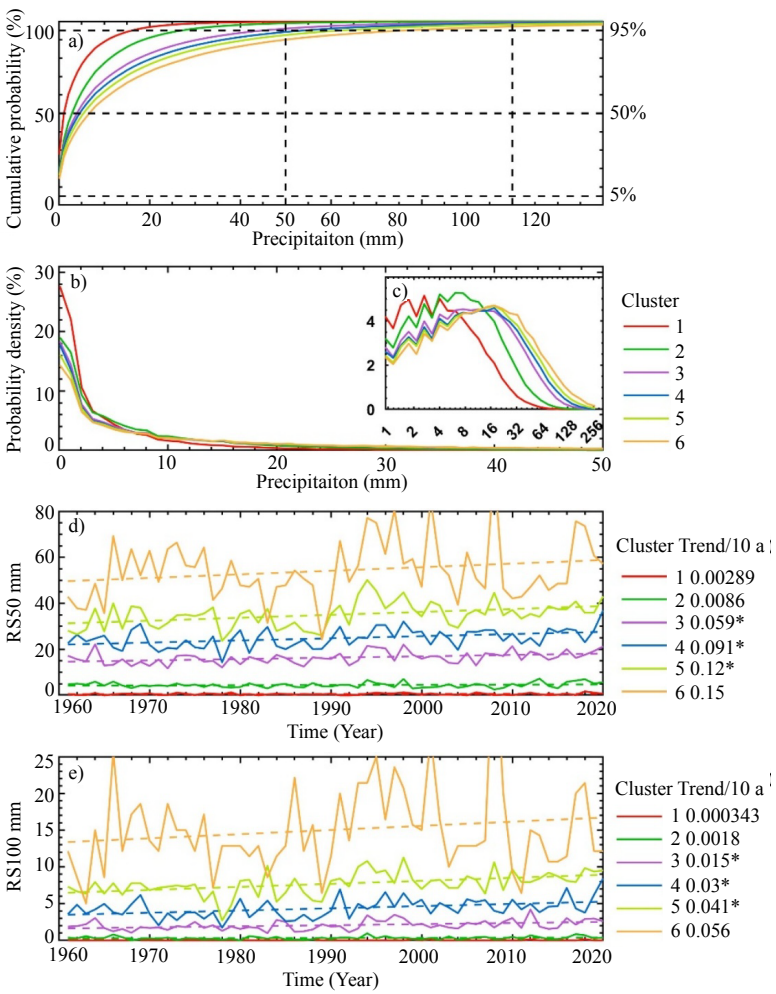


Figure 4. (a) Cumulative probability and (b) probability density distributions of daily precipitation for each cluster from 1961 to 2020 based on *K*-means clustering. (c) Same as (b), but with a logarithmic *x*-axis for precipitation >1 mm. Time series of (d) RS50 and (e) RS100 frequency (solid curves) with their trends (dashed lines) in each cluster. Frequencies are normalized by the number of stations for cluster. Asterisks indicate statistically significant trends at the 95% confidence level.

exhibit relatively weak increasing trends in extreme precipitation, which are not statistically significant. These observations indicate that the frequency of extreme precipitation is more pronounced in the southern regions, while the northwest and semi-arid areas experience less noticeable changes in extreme precipitation patterns.

These findings contribute to our understanding of regional precipitation characteristics and highlight the importance of considering spatial variations when analyzing extreme precipitation events. Further research is needed to investigate the underlying mechanisms driving these trends and their potential implications for local climate and water resource management.

5 DISCUSSION

5.1 *The exploration of the physical mechanism based on the clustering results*

Although the classification results of daily precipitation based on the *K*-means clustering algorithm are relatively objective, the physical meaning behind these statistical classification results is not clear. Therefore, the physical meaning of these six clusters is analyzed in this section.

Firstly, the time series of the summer daily precipitation spatial average in each cluster is obtained. The time series are standardized to analyze its change trend (Fig. 5). Over the past 60 years, only Jiangxi and South China (Cluster 5) showed slightly increased mean precipitation, while other regions experienced declines, particularly Northeast, North, and Southwest China (Cluster 2) with the most significant decreases. It is worth noting that although most clusters exhibit a decreasing trend in mean precipitation, only Clusters 2 and 3 show statistically significant declines. In other regions, large variability in mean precipitation weakens the significance of the trend. Fig. 4 illustrate an increasing trend in extreme precipitation across most areas, while Fig. 5 indicates a general decline in mean precipitation. These findings highlight the importance of using multiple indicators in precipitation assessments.

In order to investigate the physical reasons for the precipitation distribution characteristics in different areas,

correlation analysis is conducted between the time series of the precipitation spatial average in each cluster and summer sea surface temperature (SST) (Fig. 6). Additionally, regression analysis is carried out between the time-series of the precipitation spatial average and the geopotential height field (Fig. 7). The results reveal that for Cluster 1, summer precipitation is mainly affected by polar high and Mongolian cyclone (Fig. 7a). The persistent high-pressure system over northwest China is the main cause of drought in the region. In Fig. 7, the negative regression coefficient of Cluster 1 in the northwest region indicates that when the high-pressure system in the Baikal region is weak (or controlled by the low-pressure system), it is conducted to the formation of convergent upward movement and precipitation. Warm advection south of the low-pressure system triggers upward motion and promotes the development of low-pressure systems in the lower troposphere, ultimately leading to precipitation in northwest China. This confirms that Cluster 1 accurately represents the primary characteristics of the northwest arid zone in climatic regionalization.

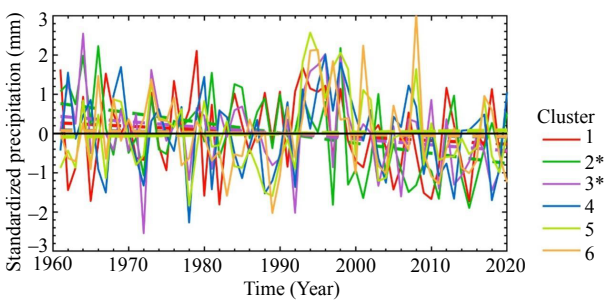


Figure 5. Standardized time series (curves) of the spatial average of summer daily precipitation and its variation trend (dotted lines) in each cluster based on the *K*-means clustering algorithm. Statistical significance at the 95% confidence level is denoted by an asterisk.

Cluster 2 (green) belongs to the transitional zone between monsoon and arid regions, which is affected not only by high latitude atmospheric circulation but also by the water vapor transport from the Indian Ocean (Figs. 6b, 7b). The increase in SSTs in the Indian Ocean and Pacific Ocean leads to a decrease in the land-sea thermal contrast, resulting in a weaker monsoon and a smaller affected area. This ultimately results in reduced interannual precipitation variability in Cluster 2. Therefore, Cluster 2 actually represents the boundary region affected by the monsoon. Due to the varying strength of the monsoon, this region shows a certain width.

China's northwest regions are located in the arid inland with high altitude, near the Qinghai-Tibet Plateau in the south. This region is controlled by a high-pressure system, which inhibits convection, resulting in low mean precipitation, small standard deviation, and low probability of extreme precipitation. The northernmost border of the

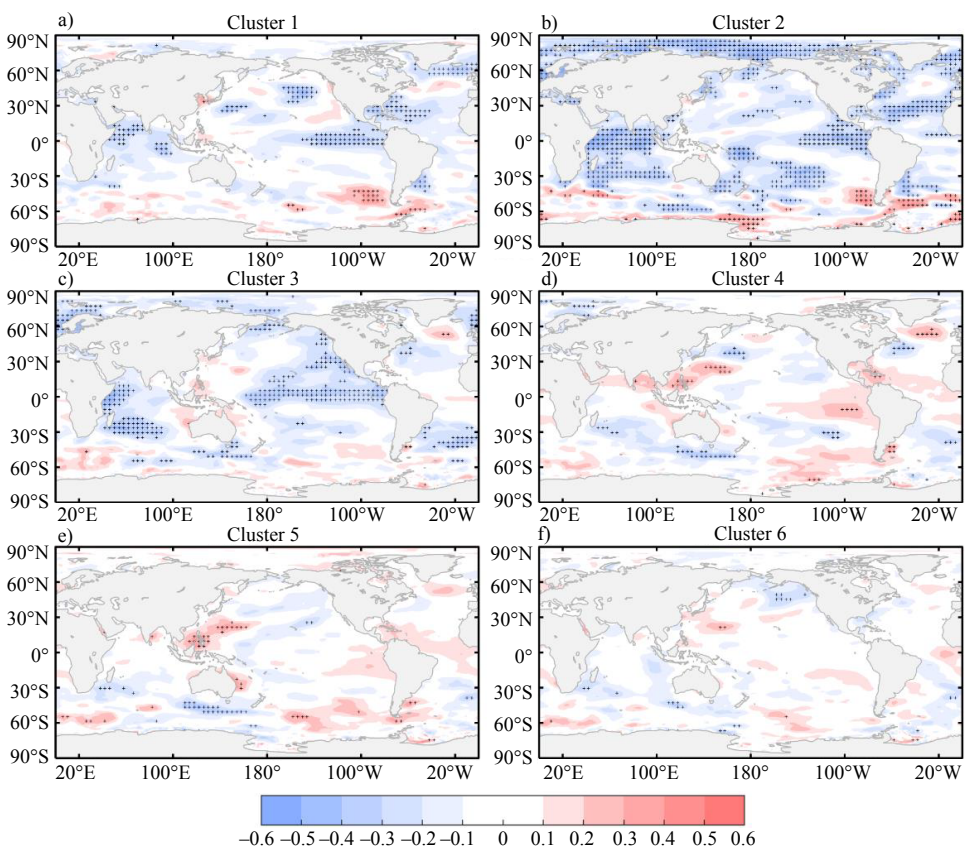


Figure 6. Correlation maps between the standardized daily precipitation mean and SST in each cluster. Grid points that are statistically significant at the 95% confidence level based on a two-tailed Student's *t*-test are denoted by small crosses.

summer monsoon, known as the northern edge of the monsoon, serves as the boundary between arid and subhumid regions (Li et al., 2013). Fig. 8b explores the variability of the wind field at 850 hPa, using the isolines of meridional wind (v) speed of 1.5 m s^{-1} (Fig. 8b, red contours) as the northern edge of monsoon. Except for northeast China, the northern edge of the monsoon is consistent with the boundary between Cluster 1 and Cluster 2. Therefore, the advance and retreat of monsoon lead to the classification of Cluster 1 and Cluster 2.

The interannual SST variations in Clusters 3 and 4 are in anti-phase (Figs. 6c and 6d), but the correlation coefficient of the time series of precipitation mean is 0.62 (Table 2). This indicates that the difference between the two clusters is not mainly in the precipitation mean, but in the variability and extremes of precipitation. The changes in precipitation mean values in both clusters are affected by intensity and movement of the monsoon. The influence of SST on the precipitation mean in these two regions may be opposite, especially in the eastern Pacific. In addition, the geopotential height field of Cluster 4 indicates its association with the East Asian-Pacific wave train (EAP) originating from the tropics (Fig. 7d). This wave train represents the planetary wave anomaly and propagation induced by tropical convection. The EAP has a significant influence on the East Asian summer monsoon (Huang 1992) and forms the meridional geopotential height field anomaly ("+ — +"), reflecting the stable maintenance of the high-pressure system and low-pressure system (Fig. 7d). This configuration facilitates the interaction between warm and cold air in the areas of Cluster 4, forming a typical Meiyu front. On the other hand, the geopotential height field of Cluster 3 does not exhibit an EAP signal and is not easy to form persistent precipitation similar to Meiyu (Fig. 7c).

The variations of the SST field and geopotential height field of Cluster 4 and Cluster 5 are very similar and are related to the EAP wave train and the SST anomalies in the Western Pacific near the Chinese mainland. Geographically, Cluster 4 and Cluster 5 are in a staggered distribution. However, when analyzing the various precipitation characteristics, the precipitation mean, precipitation standard deviation, and extreme precipitation frequency of Cluster 5 are significantly higher than those of Cluster 4. This indicates that the difference between the two clusters is mainly due to the distinct local precipitation characteristics, which may have little connection with the upstream and downstream climate systems. The nature of local precipitation may be greatly affected by geographical factors. Cluster 5 is distributed in South China and at the border of Hubei, Jiangxi, and Anhui. The area of Cluster 5 in the lower

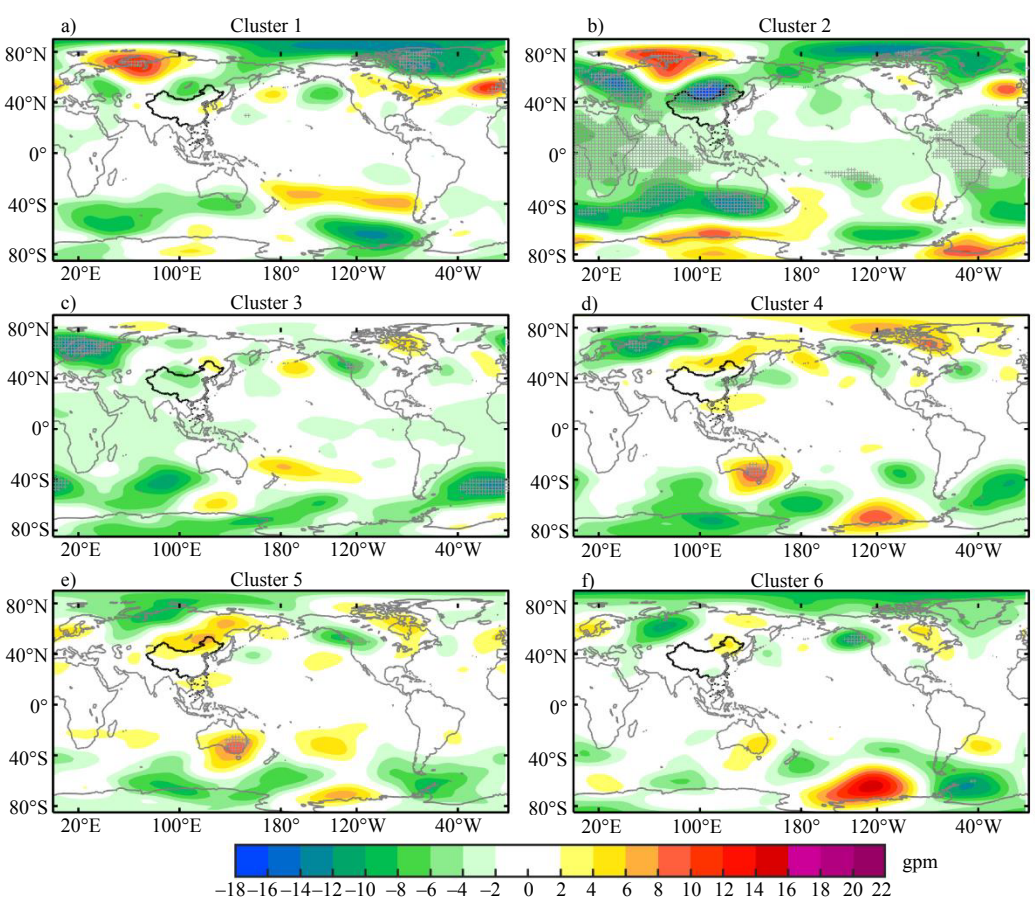


Figure 7. Regression maps of 500 hPa geopotential height against the spatial daily precipitation mean of each cluster (shading; units: gpm). Grid points that are statistically significant at the 95% confidence level based on a two-tailed Student's t-test are denoted by small crosses.

Table 2. Correlation coefficients of summer daily precipitation mean in each cluster based on the *K*-means clustering analysis. Values significant at the 95% confidence level are marked with asterisks.

Cluster	1	2	3	4	5	6
1	1	**0.51***	0.22	-0.14	0.11	0.18
2		1	**0.38***	-0.04	-0.04	-0.02
3			1	**0.62***	**0.55***	**0.3***
4				1	**0.67***	0.23
5					1	**0.68***
6						1

reaches of the Yangtze River has a developed water system, with numerous tributaries converging into the Yangtze River. Poyang Lake, the largest freshwater lake in China, supplies sufficient water vapor to the regions of Cluster 5. In addition, warm and humid water vapor is transported here by the southwest airflow from the Indian Ocean and the southerly airflow from the South China Sea (Fig. 8a). Moreover, the presence of Ta-pieh Mountains, Lu Mountain, Huangshan Mountain, and Sanqing Mountain further influences the local climate. When the water vapor reaches this region under the influence of southwest and southerly airflow (Fig. 8), it converges and rises under the force of topography, easily leading to the formation of the stable, heavy precipitation (Zheng et al. 2014). The special geographical factors and hydrological conditions contribute to higher precipitation totals and more frequent extreme precipitation in the region of Cluster 5 compared to the adjacent Yangtze River Delta region.

Cluster 6 is mainly distributed in the coastal areas of Guangdong and Guangxi. Statistical analysis indicates that the frequency of RS100 in Cluster 6 is significantly higher compared to other regions. There is no obvious signal in the corresponding SST field (Fig. 6f), while the atmospheric circulation field shows signals in the Ural Mountains and polar regions (Fig. 7f). Tropical monsoon and subtropical monsoon are important in the East Asian monsoon system (Chen et al. 2001). The coastal area of South China is controlled by the tropical monsoon system, where precipitation is mainly caused by the abundant water vapor brought by the tropical southwest monsoon (Fig. 8a). The convergence of warm and cold air, along with the typhoons landing in South China, make the region prone to stable heavy precipitation in summer.

5.2 *The physical mechanism exploration of local precipitation characteristics*

China has a vast territory that encompasses various landforms, including basins, mountains, hills, plateaus, plains, and deserts. The topography is high in the west and low in the east, with a three-step terrain from west to east. The Qinghai-Tibet Plateau in the southwest, known as the "Roof of the world", has an average altitude of over 4000 m. The eastern region consists of hilly, mountainous, and plain terrain, with an average altitude of less than 500 m. China's territory extends to 53°N, near the Mongolian Plateau and Siberia. It borders the western Pacific Ocean to the east, while the South China Sea lies to the south. In summer, water vapor from the western Pacific Ocean and the South China Sea brings substantial precipitation to central and eastern China. Additionally, the southwestern monsoons influence the distribution of water vapor from the Indian Ocean, impacting precipitation patterns in southwest China (Fig. 8). The complex geographical environment and the land-sea distribution create special dynamic and thermodynamic conditions, which jointly affect the distribution characteristics of summer precipitation in China. Precipitation distribution characteristics (Fig. 3b) present obvious east-west and north-south variations. Besides, in Eastern China, the frequency of precipitation and extreme precipitation gradually increases from northwest to southeast.

In addition, particular attention should be paid to the

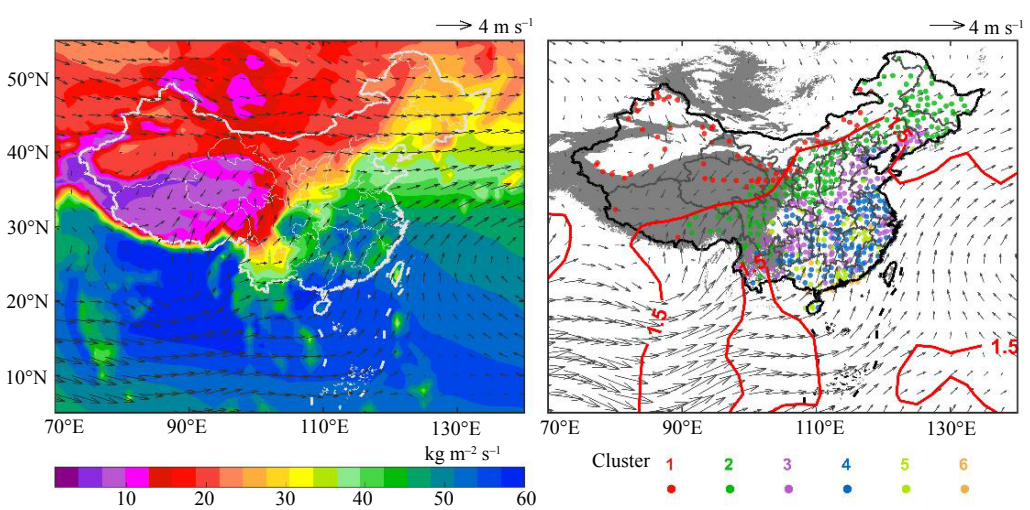


Figure 8. (a) The 700 hPa mean wind field (u, v; vectors; m s^{-1}), and water vapor flux (shading; kg m^{-2} s^{-1}). (b) 850 hPa mean wind field (u, v; vectors; m s^{-1}) and *K*-means clustering result based on the precipitation means, standard deviation, and extreme precipitation (scatter). The 1.5 m s^{-1} meridional mean wind at 850 hPa is denoted by red contours. Areas above 1500 m are shaded by grey.

middle and lower reaches of the Yangtze River. The region is classified into two clusters, Cluster 5 (light green) representing the Poyang Lake region and Cluster 3 (purple) for the Yangtze River Delta region. By analyzing the annual mean summer precipitation of the two regions (Fig. 9a), it is found that the daily-mean summer precipitation of Cluster 3 and Cluster 5 shows an upward trend. However, before 1988, the mean precipitation in Cluster 3 was lower than that of Cluster 5. After that, the situation reversed, and the precipitation mean in Cluster 3 was higher than that in Cluster 5. Although both regions show an increasing trend in mean precipitation, the upward trend in Cluster 3 is significant, whereas that in Cluster 5 is not. This indicates that the daily mean precipitation in Cluster 3 of the Yangtze River Delta is higher than that in Cluster 5 of Poyang Lake. Besides, the difference in extreme precipitation frequency between these two areas is also analyzed. Figs. 9b and 9c show the temporal evolution of extreme precipitation frequency in Cluster 3 and Cluster 5 along the Yangtze River. It is observed that the frequency of extreme precipitation in Cluster 3 is generally lower than that in Cluster 5. However, the two curves are getting closer, indicating that the increasing trend of extreme precipitation in the Yangtze River Delta area (Cluster 3) is higher than that in the Poyang Lake area (Cluster 5). Both regions exhibit a significant increasing trend in both RS50 and RS100, further confirming the rise in extreme precipitation.

These results indicate that the Yangtze River Delta region and the Poyang Lake area exhibit distinctly different precipitation characteristics. Since the reform and opening-up policy, the Yangtze River Delta has experienced a much faster rate of economic development compared to other regions, resulting in a high level of urbanization. The discharge of pollutants from human activities in this region serves as condensation nuclei, making it relatively difficult for large raindrops to form. Moreover, the abundant water vapor along the Yangtze River contributes to continuous rainfall during the summer in the Yangtze River Delta (Wen et al. 2020). On the other hand, Poyang lake region exhibits a lower level of urbanization compared to the Yangtze River Delta region, with less anthropogenic influence. In addition, the Poyang Lake region is surrounded by mountains on three sides, with a plain only in the southwest. The trumpet-shaped terrain facilitates the convergence of southwest monsoon airflow, making natural factors essential in shaping the characteristics of precipitation. The water vapor flux in summer (Fig. 8a) indicates that the Poyang Lake region experiences significantly higher water vapor flux compared to the Yangtze River Delta region.

Apart from the Yangtze River coastal region, urban patterns in other regions also affect the precipitation distribution characteristics. The daily distribution characteristics of precipitation reveals that the summer precipitation in the Beijing-Tianjin-Hebei region (Cluster 4) is significantly higher than that in the surrounding areas (Cluster 3).

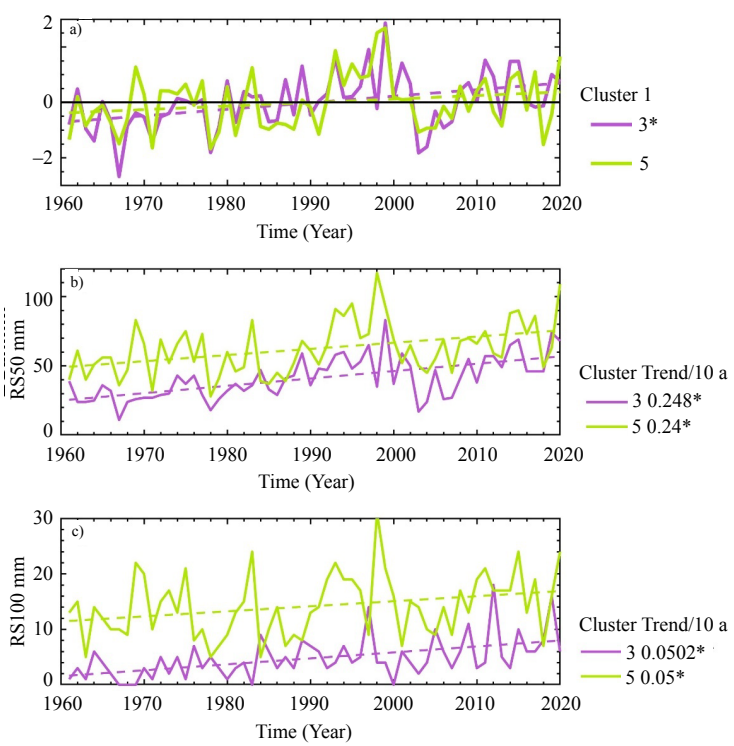


Figure 9. (a) Standardized time series (solid curves) and trends (dotted lines) of summer daily precipitation averages for Cluster 3 (purple) and Cluster 5 (green) in the middle and lower Yangtze River region based on *K*-means clustering. Time series of (b) RS50 and (c) RS100 frequency (solid curves) and their trends (dashed lines) for Cluster 3 (Yangtze River Delta) and Cluster 5 (Poyang Lake region). Asterisks indicate statistically significant trends at the 95% confidence level.

6 CONCLUSIONS

This study analyzes the spatiotemporal distribution of summer precipitation across China from 1961 to 2020 using daily precipitation data. By incorporating precipitation mean, standard deviation, and extreme event frequency, we apply *K*-means clustering to develop an objective precipitation climate zoning system. This approach addresses limitations of traditional single-factor or subjective methods by automatically identifying similarities among precipitation characteristics, enhancing classification accuracy. We identify six distinct precipitation zones that differ from conventional climate divisions. For each zone, we analyze summer daily precipitation patterns and explore underlying physical mechanisms. The results offer a scientific basis for precise delineation of extreme precipitation disaster zones, supporting improved risk management strategies.

Based on the observation data, this study proves the emergence of new characteristics in the precipitation distribution pattern. The research confirms the fact that extreme precipitation events occur frequently. The frequency of extreme precipitation events has shown a notable increase since 2010. Although some regions (notably the middle and lower reaches of the Yangtze River) demonstrate increasing mean precipitation, the majority of China shows declining trends, indicating an overall decrease in total precipitation (Fig. 5). Among them, the most rapid decrease in precipitation is observed in the transition zone between monsoon and non-monsoon regions (Cluster 2), while the most pronounced increase in extreme precipitation frequency in the coastal area of South China (Cluster 6).

The physical mechanism analysis reveals the mechanism based on the clustering results. The arid region in northwest China, influenced by the high pressure in Lake Baikal region, is classified as Cluster 1. The mean and standard deviation of precipitation in this region are the lowest, and the frequency of extreme precipitation is very low. When the Baikal high weakens and the Mongolian cyclone intensifies, convergent uplift increases summer precipitation in this region. Influenced by the Mongolian cyclone, the transitional zone between the monsoon and the arid region is classified as Cluster 2. This zone represents the northern boundary of the monsoon due to the discrepancies of water vapor supply and land-sea thermal contrast in the Indian Ocean. The frequency of extreme precipitation in this region is low. The East Asian monsoon area, which is related to monsoon intensity (advance and retreat) is judged as Cluster 3 (purple) and Cluster 4 (blue). Cluster 4 represents a typical Meiyu region, mainly influenced by the EAP wave train from the tropics and the SST anomaly pattern in the Middle East Pacific. There are two special regions within the Meiyu regions: Cluster 4 and Cluster 5. These regions have similar climatic impact factors and are spatially interleaved with each other. Cluster 5 is classified as a different type from Cluster 4 because the precipitation properties (including mean value, precipitation variance, and extreme precipitation frequency) of Cluster 5 are significantly higher. This difference may be related to topography and urbanization factors. Cluster 6 mainly covers the coastal areas of Guangdong and Guangxi, with a much higher frequency of extreme precipitation compared to other areas. Although clear climatic drivers are not identified, this may result from the abundant water vapor in tropical coastal zones.

In general, large-scale geographical features such as geographical location, topography, and land-sea distribution contribute to the differences in the precipitation distribution characteristics between the eastern and western regions, as well as between the southern and northern regions. The topography and land-sea configurations, the atmospheric circulations, and SST field configurations, especially in eastern China, result in the variations in thermodynamic, dynamic conditions, and water vapor conditions across different parts of China (Ren et al. 2025). These differences finally lead to various weather situations and cause the boundary of each cluster to roughly follow the northeast-southwest trend. Specifically, from northwest to southeast, there is a gradually increase in the mean and standard deviation of precipitation, as well as the frequency of extreme precipitation. In addition, the distribution characteristics of the underlying surface (such as rivers, cities, mountains, etc.) further influence the local precipitation distribution. Specifically, the Poyang Lake region and the Yangtze River Delta region exhibit different precipitation distribution characteristics due to the trumpet-shaped mountain terrain around Poyang Lake and the urbanization in the Yangtze River Delta. Notably, with the economic development of the Beijing-Tianjin-Hebei region, human activities has an impact on the precipitation characteristics of the region, resulting in significantly higher mean and standard deviation of precipitation compared to the surrounding areas.

Acknowledgements: The authors wish to thank Dr. Nicolas Viltard at the Laboratoire Atmosphère, Milieux et Observations Spatiales (LATMOS) for the insightful discussion.

REFERENCES

AghaKouchak, A., 2015: A multivariate approach for persistence-based drought prediction: Application to the 2010–2011 East Africa drought. *J. Hydrol.*, **526**, 127–135, https://doi.org/10.1016/j.jhydrol.2014.09.063.

Ashabi, A., S. Bin Sahibuddin, and H. Mehdi Salkhordeh, 2020: The systematic review of K-means clustering algorithm. *2020 The 9th International Conference on Networks, Communication and Computing*, NY, USA, ACM, 13–18, https://doi.org/10.1145/3447654.3447657.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, 2018: Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data*, **5**, 180214, https://doi.org/10.1038/sdata.2018.214.

Burke, C., P. Stott, A. Ciavarella, and Y. Sun, 2016: Attribution of extreme rainfall in southeast China during May 2015. *Bull. Amer. Meteor. Soc.*, **97**, S92–S96, https://doi.org/10.1175/bams-d-16-0144.1.

Cao, J., Z. Jiang, F. Ren, and Z. Xn, 2013: An application of the generalized linear statistical downscaling method to simulating daily precipitation. *Acta Meteor. Sin.*, **71**, 167–175, https://doi.org/10.11676/qxxb2013.014. (in Chinese)

Carvalho, M. J., P. Melo-Goncalves, J. C. Teixeira, and A. Rocha, 2016: Regionalization of Europe based on a K-means cluster analysis of the climate change of temperatures and precipitation. *Phys. Chem. Earth*, *Parts A*/*B*/*C*, **94**, 22–28, https://doi.org/10.1016/j.pce.2016.05.001.

Chen, L., W. Li, P. Zhao, and S. Tao, 2000: On the process of summer monsoon onset over East Asia. *Clim. Environ. Res*., **5**, 345–355, https://doi.org/10.3878/j.issn.1006-9585.2000.04.02. (in Chinese)

Choi, W.-Y., S.-K. Oh, and H.-K. Kim, 2014: Design of RBFNN-based pattern classifier for the classification of precipitation/non-precipitation cases. *J. Korean Inst. Intell. Syst.*, **24**, 586–591, https://doi.org/10.5391/jkiis.2014.24.6.586.

Dhanachandra, N., K. Manglem, and Y. J. Chanu, 2015: Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Comput Sci*, **54**, 764–771, https://doi.org/10.1016/j.procs.2015.06.090.

Ding, Y., Z. Wang, and Y. Sun, 2008: Inter-decadal variation of the summer precipitation in East China and its association with decreasing Asian summer monsoon. Part I: Observed evidences. *Int. J. Climatol.*, **28**, 1139–1161, https://doi.org/10.1002/joc.1615.

Falkena, S. K. J., J. Wiljes, A. Weisheimer, and T. G. Shepherd, 2020: Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. *Quart. J. Roy. Meteor. Soc.*, **146**, 2801–2814, https://doi.org/10.1002/qj.3818.

Feng, L., T. Zhou, B. Wu, T. Li, and J.-J. Luo, 2011: Projection of future precipitation change over China with a high-resolution global atmospheric model. *Adv. Atmos. Sci.*, **28**, 464–476, https://doi.org/10.1007/s00376-010-0016-1.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart*. *J. Roy*. *Meteor*. *Soc*., **121**, 319–367, https://doi.org/10.1002/qj.49712152206.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, https://doi.org/10.1175/JCLI3990.1.

Hersbach, H., B. Bell, and P. Berrisford, 2019: ERA5 monthly averaged data on single levels from 1979 to present., accessed 17 May 2020, https://doi.org/10.1002/qj.3803.

Hu, H., 1935: The distribution of China's population and its causes. *Acta Geogr. Sin*. **2**, 33–74. (in Chinese)

Huang, R., 1992: The east Asia/Pacific pattern teleconnection of summer circulation and climate anomaly in East Asia. *J. Meteor. Res.*, **6**, 25–37.

Huang, Y., G. Feng, and W. Dong, 2011: Temporal changes in the patterns of extremes air temperature and precipitation in the various regions of China in recent 50 years. *Acta Meteor.*, **69**, 125–136, https://doi.org/10.11676/qxxb2011.011. (in Chinese)

Huo, R., L. Li, H. Chen, C.-Y. Xu, J. Chen, and S. Guo, 2021: Extreme precipitation changes in europe from the last millennium to the end of the twenty-first century. *J. Climate*, **34**, 567–588, https://doi.org/10.1175/JCLI-D-19-0879.1.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the kobe collection. *Int*. *J. Climatol*., **25**, 865–879, https://doi.org/10.1002/joc.1169.

Japan Meteorological Agency, 2006: Characteristics of global sea surface temperature analysis data (COBE-SST) for climate use. *Monthly Report on Climate System*, **12**, 116pp.

Kalnay et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Renew. Energy*, B. Sorensen, 146–194, https://doi.org/10.4324/9781315793245-16.

Kirchmeier-Young, M. C., and X. Zhang, 2020: Human influence has intensified extreme precipitation in North America. *Proc. Natl. Acade. Sci.*, **117**, 13308–13313, https://doi.org/10.1073/pnas.1921628117.

Li, C., Z. Xiao, and X. Zhang., 2012: Climatic characteristics of regions of China for precipitation in various the past 60 years. *Meteor. Mon.*, **38**, 419–424, https://doi.org/10.7519/j.issn.1000-0526.2012.04.005. (in Chinese)

Li, D., P. Shao, and H. Wang, 2013: The position variations of the north boundary of east asian subtropical summer monsoon in 1951–2009. *J. Desert Res.*, **33**, 1511–1519. (in Chinese)

Li, M., Y. Zhao, Y. Li, X. Zhou, and S. Zhang, 2021: Changes in slight and extreme precipitation events from 1963 to 2017 over northern xinjiang, China. *Front. Earth Sci.*, **9**, 622889 https://doi.org/10.3389/feart.2021.622889.

Li, W., Y. Feng, D. Li, and Z. Yu, 2016: Micro-blog topic detection method based on BTM topic model and K-means clustering algorithm. *Autom. Control Comput. Sci.*, **50**, 271–277, https://doi.org/10.3103/S0146411616040040.

Liang, L., L. Li, and Q. Liu, 2011: Precipitation variability in Northeast China from 1961 to 2008. *J. Hydrol.* , **404**, 67–76, https://doi.org/10.1016/j.jhydrol.2011.04.020.

Liu, T., X. Zhu, R. Guo, K. Xu, and S. Zhang, 2022: Applicability of ERA5 reanalysis of precipitation data in China, *Arid Land Geogr.*, **45**, 66–79. (in Chinese)

Lu, X., Q. Li, W. Zhao, A. Xiao, G. Li, and Z. Yu, 2021: Spatiotemporal characteristics of rainfall in south China from 1967 to 2018. *J. Appl. Meteor. Climatol.*, **60**, 1333–1345, https://doi.org/10.1175/JAMC-D-20-0191.1.

Lyon, B., A. G. Barnston, and D. G. DeWitt, 2014: Tropical pacific forcing of a 1998–1999 climate shift: observational analysis and climate model results for the boreal spring season. *Climate Dyn.*, **43**, 893–909, https://doi.org/10.1007/s00382–013–1891–9.

Ma, Y., L. Zhao, J.-S. Wang, and T. Yu, 2021: Increasing difference of china summer precipitation statistics between percentage anomaly and probability distribution methods due to tropical warming. *Earth Space Sci.*, **8**, https://doi.org/10.1029/2021EA001777.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observation, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, LA, University of California, 281–297.

Visalakshi, N. K. and K. Thangavel, 2009: Impact of normalization in distributed k-means clustering. *Int. J. Soft Comput*., **4**, 168–172.

O'Gorman, P. A., and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-

century climate change. *Proc. Natl. Acad. Sci.*, **106**, 14773–14777, https://doi.org/10.1073/pnas.0907610106.

Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson, 2017: Precipitation variability increases in a warmer climate. *Sci. Rep.*, **7**, 17966, https://doi.org/10.1038/s41598-017-17966-y.

Philandras, C. M., P. T. Nastos, J. Kapsomenakis, K. C. Douvis, G. Tselioudis, and C. S. Zerefos, 2011: Long term precipitation trends and variability within the Mediterranean region. *Nat. Hazards Earth Syst. Sci.*, **11**, 3235–3250, https://doi.org/10.5194/nhess-11-3235-2011.

Ren, G., H. Wu, and Z. Chen, 2000: Spatial patterns of change trend in rainfall of China. *Quart. J. Appl. Meteor.*, **11**, 322–330. (in Chinese)

Ren, G., Y. Zhan, Y. Ren, Y. Chen, T. Wang, Y. Liu, and X. Sun, 2015: Spatial and temporal patterns of precipitation variability over mainland China: I: Climatology. *Adv. Water Sci.* , **26**, 299–310. (in Chinese)

Ren, Y., H. Yu, J. Huang, J. Zhou, H. Luo, and S. Cheng, 2025: Impact of mid-high latitude circulation and surface thermal forcing on drought events in Central Asia. *Climate Dyn.*, **63**, 90, https://doi.org/10.1007/s00382-024-07577-5.

Sardar, T. H., and Z. Ansari, 2018: An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm. *Future Comput. Inf. J.*, **3**, 200–209, https://doi.org/10.1016/j.fcij.2018.03.003.

Shafi, J., and A. Waheed, 2020: K-means clustering analysing abrupt changes in air quality. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, 26–30, https://doi.org/10.1109/ICECA49313.2020.9297493.

Tabari, H., 2020: Climate change impact on flood and extreme precipitation increases with water availability. *Sci. Rep.*, **10**, 13768, https://doi.org/10.1038/s41598-020-70816-2.

Tang, J., D. Wang, Z. Zhang, L. He, J. Xin, and Y. Xu, 2017: Weed identification based on K-means feature learning combined with convolutional neural network. *Comput. Electron. Agric.*, **135**, 63–70, https://doi.org/10.1016/j.compag.2017.01.001.

Tleis, M., R. Callieris, and R. Roma, 2017: Segmenting the organic food market in Lebanon: an application of k-means cluster analysis. *British Food J.*, **119**, 1423–1441, https://doi.org/10.1108/BFJ-08-2016-0354.

Wang, H., R. Fu, A. Kumar, and W. Li, 2010: Variability and predictability of southeastern united states summer precipitation. *J. Hydroclim.*, **11**, 1007–1018, https://www.cpc.ncep.noaa.gov/products/people/hwang/pdf/2009_Wang_JCLI_-sub1.pdf.

Wang, Y., X. Chen, and F. Yan, 2015: Spatial and temporal variations of annual precipitation during 1960–2010 in China. *Quat. Int.*, **380–381**, 5–13, https://doi.org/10.1016/j.quaint.2014.12.047.

Wen, L., K. Zhao, Z. Yang, H. Chen, H. Huang, G. Chen, and Z. Yang, 2020: Microphysics of stratiform and convective precipitation during meiyu season in eastern China. *J. Geophys. Res. Atmos.*, **125**, https://doi.org/10.1029/2020JD032677.

Yao, S., D. Jiang, and G. Fan, 2017: Seasonality of precipitation over China. *Chinese J. Atmos. Sci.* , **41**, 1191–1203, https://doi.org/10.3878/j.issn.1006-9895.1703.16233. (in Chinese)

Yokoi, S., and Coauthors, 2011: Application of cluster analysis to climate model performance metrics. *J. Appl. Meteor. Climatol.*, **50**, 1666–1675, https://doi.org/10.1175/2011JAMC2643.1.

You, Q., S. Kang, G. Ren, K. Fraedrich, N. Pepin, Y. Yan, and L. Ma, 2011: Observed changes in snow depth and number of snow days in the eastern and central Tibetan Plateau. *Climate Res.*, **46**, 171–183, https://doi.org/10.3354/cr00985.

Zhai, P., A. Sun, F. Ren, X. Liu, B. Gao, and Q. Zhang, 1999: Changes of climate extremes in China. *Climate Change*, **42**, 203–218, https://doi.org/10.1023/A:1005428602279.

Zhai, P., X. Zhang, H. Wan, and X. Pan, 2005: Trends in total precipitation and frequency of daily precipitation extremes over China. *J. Climate*, **18**, 1096–1108, https://doi.org/10.1175/JCLI-3318.1.

Zhan, Y., X. Ju, S. Fan, Z. Zhou, G. Tang, and G. Ren, 2021: An analysis of minute summer precipitation in China during 1965—2019. *Acta Meteor. Sin.*, **79**, 598–611, https://doi.org/10.11676/qxxb2021.036. (in Chinese)

Zhang, W., and T. Zhou, 2020: Increasing impacts from extreme precipitation on population over China with global warming. *Sci. Bull.*, **65**, 243–252, https://doi.org/10.1016/j.scib.2019.12.002.

Zheng, J., Y. Yin, and B. Li, 2010: A new scheme for climate regionalization in China. *Acta Geogr. Sin.*, **6**, 3–12. (in Chinese)

Zheng, J., S. Sun, J. Wu, and A. Xu, 2014: Analysis on multi-scale ambien field for ahort-time server torrential rain on meiyu front. *Meteor. Mon.*, **40**, 570–579, https://doi.org/10.7519/j.issn.1000-0526.2014.05.007. (in Chinese)

Citation: Ma, Y. Y., J. S. Wang, L. Zhao, W. J. Li, H. Yang, W. Wen, H. Zhou, and M. Chen, 2025: A Precipitation Classification Scheme for China with Special Consideration of Extreme Events. *J. Trop. Meteor.*, **31**, 497–510, https://doi.org/10.3724/j.1006-8775.2025.041.