A new dataset of Mesoscale Convective Complexes (MCC) derived from FY-2G satellite data

Ke Xu[1], Shuyun Zhao[1,2,*], Xinyu Ma[1], Jianchuan Shu[3], Wuke Wang[1], Qimin Deng[1], Zaheer Ahmad Babar[4]

[1]Department of Atmospheric Sciences, School of Environmental Studies, China University of Geosciences, Wuhan, China
[2]Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing, China
[3]Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, Institute of Tibetan Plateau Meteorology, China Meteorological Administration, Chengdu, China
[4]Flood Forecasting Division, Pakistan Meteorological Department, Lahore, Pakistan

[*]*Correspondence to*: Shuyun Zhao (zhaosy@cug.edu.cn)

Abstract. Mesoscale Convective Complexes (MCCs) are major convective weather systems occurring in midlatitude regions, typically associated with significant weather phenomena such as heavy rainfall, thunderstorms, strong winds, and hail. Based on the cloud-top temperature (CTT) data of the FY-2G satellite, and through multi-threshold screening combined with morphological analysis, an automated algorithm for MCC identification and tracking was developed. The algorithm is then applied to generate an hourly dataset of MCC variables over mainland China from June 2015 to December 2024. The dataset encompasses variables describing the spatial extent of the cold-core region (CTT < -52℃) of MCCs, the minimum cloud-top temperature within the cold cloud shields, and the geographic coordinates (longitude and latitude) of the centroids of the cold cloud shields. This work also conducts a preliminary analysis of the spatial and temporal distribution characteristics of MCCs over mainland China based on the dataset. Results indicate that MCCs occur more frequently in Southwest China than in other regions of the country, and over 70% of MCC events occur in summer both in Southwest China and mainland China as a whole. Moreover, MCC frequency in Southwest China exhibits significant interannual variability.

1 Introduction

MCCs represent an important class of organized deep convection characterized by large (meso-α scale), long-lived, and circular cloud shields. MCCs form through the merger of multiple thunderstorms, often persist for 6-12 hours, and typically produce widespread heavy precipitation and severe weather. MCCs play a significant role in the global hydrological cycle and frequently cause damaging floods through prolonged rainfall (Maddox, 1980). It was found that 91% of major floods in East Asia are linked to meso-scale convective systems (Zhou et al., 2023).

Previous studies have conducted extensive and systematic research on MCCs, based on manual census. Augustine and Howard (1989) documented MCC activity during 1985-1987 in North America using digitized satellite imagery, significantly advancing understanding of MCC structure and evolution in the region. Prior researches have also investigated

MCCs in South America, the western Pacific, the Indian subcontinent and Africa, respectively, thoroughly documenting the climatological characteristics of MCCs in these regions (Velasco, 1987; Fritsch, 1991; Miller, 1991; Laing, 1993; Febrizky, 2023). It is indicated in these research that MCCs exhibit distinct nocturnal characteristics, primarily forming over continental regions and preferentially occurring on the leeward or downwind sides of large-scale elevated terrain, such as plateaus and major mountain ranges,~~areas, and often occurring on the leeward side or downwind regions of major topographic features~~ with lifetimes of 8-14 hours. Abisusmita (2023) investigated three MCC events that occurred in South Sulawesi during 2018–2020, all located over the Makassar Strait, and revealed that their ~~formation~~variation was jointly influenced by the Southern Oscillation, the Madden-Julian Oscillation, and sea surface temperature anomalies.

Southwest China—encompassing the provinces of Sichuan, Guizhou, and Yunnan, along with Chongqing Municipality—exhibits complex topography and a pronounced monsoon climate. Frequent summertime MCCs in this region often produce heavy rainfall that triggers severe mountain hazards, underscoring the critical need to quantify their hydrometeorological characteristics for improved disaster risk reduction (Zhang et al., 2025). While numerous studies have investigated MCCs over Southwest China, the majority rely on case analysis or labor-intensive manual surveys, which are often constrained by limited temporal coverage and spatial extent. Jing et al. (2013) manually identified MCCs over China from 2005 to 2011 using FY-2 satellite imagery and MICAPS data, and performed a comprehensive analysis of their spatiotemporal distributions and meteorological characteristics. Their results show that MCCs are predominantly concentrated in the Yunnan–Guizhou–Guangxi–Guangdong region, where they exhibit a pronounced noctural frequancy maximum in June and July. These convective systems are closely associated with heavy rainfall, often leading to flooding~~ and occur most frequently at night during June and July. They are often associated with heavy rainfall that leads to flooding, originate from relatively consistent genesis locations, and can be classified into five distinct synoptic circulation patterns at 200 hPa.The results show that MCCs are predominantly concentrated in the Yunnan-Guizhou-Guangxi-Guangdong region, occur most frequently at night during June and July, are often associated with heavy rainfall that leads to flooding, originate from relatively consistent genesis locations, and can be classified into five distinct synoptic circulation patterns at 200 hPa.~~ .

The aforementioned studies on MCCs have predominantly relied on manual identification methods, typically involving the analysis of enhanced infrared imagery. While this approach is considered reliable, it is inherently labor-intensive and time-consuming, rendering it impractical for generating long-term, large-scale MCC datasets. Laing and Fritsch (1993) employed a computerized MCC program, with pixeled cloud shield temperature, latitude, and longitude values as input, to document the characteristics of MCCs over the Indian subcontinent region. Lakshmanan (2003) pioneered the use of the K-Means clustering algorithm for thunderstorm identification. Matthews et al. (2010) introduced the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm for thunderstorm detection. Compared with the K-Means clustering algorithm, the DBSCAN algorithm does not require the number of clusters to be predefined, is insensitive to initial centroid selection, and can effectively identify thunderstorms with diverse and irregular morphologies. Yan et al. (2020) used an improved DBSCAN method for identifying the three-dimensional structure of thunderstorms based on radar reflectivity. Shah et al. (2015) developed a fully automated thunderstorm detection algorithm that adaptively determines optimal multi-

parameter thresholds by combining optimized radar reflectivity thresholding with the SALdEdA (structure, amplitude, location, eccentricity difference and areal difference) tracking technique. Guo (2016) employed digital image processing and fuzzy pattern recognition techniques for the automated identification and tracking of meso-scale convective system cloud clusters. Nevertheless, the method still necessitated manual intervention to handle newly formed cloud clusters, and its identification and tracking accuracy could be influenced by satellite projection geometry and coordinate transformation errors.

Although numerous automated detection methods have been developed, most are not specifically designed for MCC identification or have not been applied to regions over China. As a result, a long-term MCC dataset suitable for studying the evolution and variability of these meso-α-scale weather systems remains scarce. To address this gap, we develop an adaptive MCC detection algorithm based on FY-2G satellite data, which integrates cloud-top temperature thresholds, area criteria, and morphological characteristics of cloud shields. The open-source algorithm and the resulting MCC dataset are expected to support operational MCC recognition and tracking, and research on the variability of convective systems under global warming. By applying this method, we construct a comprehensive MCC event dataset spanning 10 years (June 2015 to December 2024) over mainland China, which will be regularly updated in the future.

2 Data

The FY-2G satellite, a member of China's Fengyun-2 geostationary meteorological satellite series operated by the China Meteorological Administration (CMA), was launched on 31 December 2014 and has since maintained a geostationary orbit at an altitude of approximately 35,800 km. It carries the second-generation Visible and Infrared Spin-Scan Radiometer (VISSR-II, also referred to as S-VISSR) as primary imaging payload. Through multispectral sensing in the visible, infrared, and water-vapor bands, cloud-top temperatures ~~which allows~~are monitored ~~–~~comprehensively ~~monitoring of cloud-top temperatures~~. The satellite's longitude has ~~been relocated twice~~undergone two relocations: it was positioned ~~initially~~ at 99.5° E ~~before~~prior to June 1, 2015~~;~~, then a~~t~~djusted to 105°E from that date until April 16, 2018~~;~~, and ~~finally moved back to 99.2°E, where it remains currently~~has since been stationed at 99.2°E. VISSR-II~~FY-2G~~ operates at a temporal resolution of one hour during the non-flood season and 30 minutes during the flood season, providing~~es~~ ~~meteorological observations at~~ hourly observational data to the public~~intervals, (with operational scan cycles ranging from as frequent as 30 to 60 minutes)~~, supporting applications in weather forecasting, climate research, and environmental monitoring.

In this study, we utilize the FY-2G CTT product (https://satellite.nsmc.org.cn/). In contrast to Black Body Temperature (TBB)~~ measurements~~, CTT is retrieved through multi-band infrared data inversion integrated with atmospheric modeling, yielding a more accurate representation of actual cloud-top thermal conditions. Because the MCC signatures—extensive cold-cloud shields, near-circular morphology, and prolonged lifetimes—are readily discerned in CTT imagery derived from infrared satellite observations, physically retrieved CTTs are particularly advantageous for MCC identification and tracking. The FY-2G CTT product has a ~~nominal 5km~~ spatial resolution of approximately ~~(each pixel~~≈ 5km × 5km~~)~~ and an hourly

temporal resolution; it is available from June 2015 to the present, of which the subset spanning from June 2015 to December 2024 are used for analysis in this study.

3 Algorithm and implementation

3.1 Parameter Optimization

This study develops an automated detection algorithm for MCCs using CTT data retrieved from the infrared channels of the FY-2G satellite. The initial detection criteria are anchored in the morphological and lifecycle criteria for MCCs proposed by Maddox (1980) (Table 1). However, manual census of MCC occurrence over China reveals that the area, morphology and lifecycle parameters—originally derived from MCC characteristics over North America—are not fully applicable to China, and therefore warrant region-specific modification.

It was demonstrated by Augustine (1991) that ~~The corresponding parameter adjustments are introduced below.~~the original cold-cloud shield requirement (CTT ≤ −32°C with area ≥10^5 km^2) in Maddox (1980) is overly permissive; although all genuine MCCs satisfy this condition, it does little to exclude other convective systems. By contrast, the deep-convection core requirement (CTT ≤ -52°C with an area ≥50000 km^2), has been shown to be overly restrictive in some contexts (Augustine, 1991; Fei, 2011). We carefully examined several MCC cases over China manually and also found that well-organized MCCs can have cloud shield with CTT ≤ -52°C slightly smaller than 50000 km^2. Based on these considerations, the threshold for the area with CTT ≤ -52°C is set to ≥ 40000 km^2.

~~At each time step, candidate grid cells meeting the prescribed CTT thresholds are first identified. The original cold-cloud shield requirement (CTT ≤ −32°C with area ≥10^5 km^2) is overly permissive; although all genuine MCCs satisfy this condition, it does little to exclude other convective systems and is therefore omitted (Augustine, 19912). By contrast, the deep-convection core requirement (CTT ≤ -52°C with an area ≥50000 km^2) proves too restrictive for Chinese MCC cases; accordingly, the area threshold is revised to ≥40000 km^2 to better match observed MCC events previously., originally proposed by Maddox (1980), has been shown to be overly restrictive in some contexts (Augustine, 1991; Fei, 2011). Consistent with this assessment, our manual census of MCC occurrences over China indicates that a considerable fraction of well-documented MCC events exhibit deep-convective cores with areas slightly below this threshold while retaining canonical MCC morphology and lifecycles. Accordingly, the area criterion is empirically relaxed to ≥ 40,000 km², jointly supported by previous findings and the results of this study.~~ To represent the quasi-circular morphology typical of MCCs, the cold cloud region is extracted and best-fitting ellipse is computed, ~~where~~with the ellipse eccentricity (ecc) serves as a quantitative measure of roundness. Following Xiang (1995), —who reported smaller eccentricity values for MCCs in China than in North America, —the ~~conventional~~ ecc criterion is relaxed from ecc ≥ 0.7 at maximum extent to ecc ≥ 0.6 throughout the MCC lifecycle to accommodate morphological variability. This adjustment enables the algorithm to capture the complete MCC lifecycle from initiation to dissipation.~~when the cold-cloud shield reaches maximum extent, enabling capturing the full MCC lifecycle from initiation to dissipation.~~ For identification accuracy, additional morphological quality-control constraint

is incorporated. The cloud cluster must cover at least 80% of the fitted ellipse's area (area ratio ≥ 0.8), thereby filtering fragmented or spuriously connected patches. Finally, a system is classified as an MCC only if all above morphological and area criteria are satisfied continuously for at least 6 hours, ensuring consistency in both structural integrity and lifecycle evolution (Table 1).

Table 1: Comparison of MCC parameters/criteria between Maddox (1980) and ~~the~~ this study

Criterion/Parameter	Original Standard in Maddox (1980)	Optimized Standard in This Study
Cold Cloud Area Threshold	TBB ≤ -32°C with area ≥ 10^5 km^2	~~null~~
Deep Convection Core Threshold	TBB ≤ -52°C with area ≥ 50000 km^2	CTT ≤ −52°C with area ≥ 40000 km^2
Ellipse Eccentricity	Ecc ≥ 0.7 at maximum extent	Ecc ≥ 0.6 throughout lifetime
Area Ratio Quality Control	~~Not specified~~	Cloud cluster area / ellipse area ≥ 0.8
Lifetime	≥ 6h	≥ 6h

3.2 Automated Identification Algorithm

Based on the aforementioned criteria, the steps of the automated identification algorithm were correspondingly designed and tailored for the specific analysis of the China region (~~Fig.~~Figure 1).

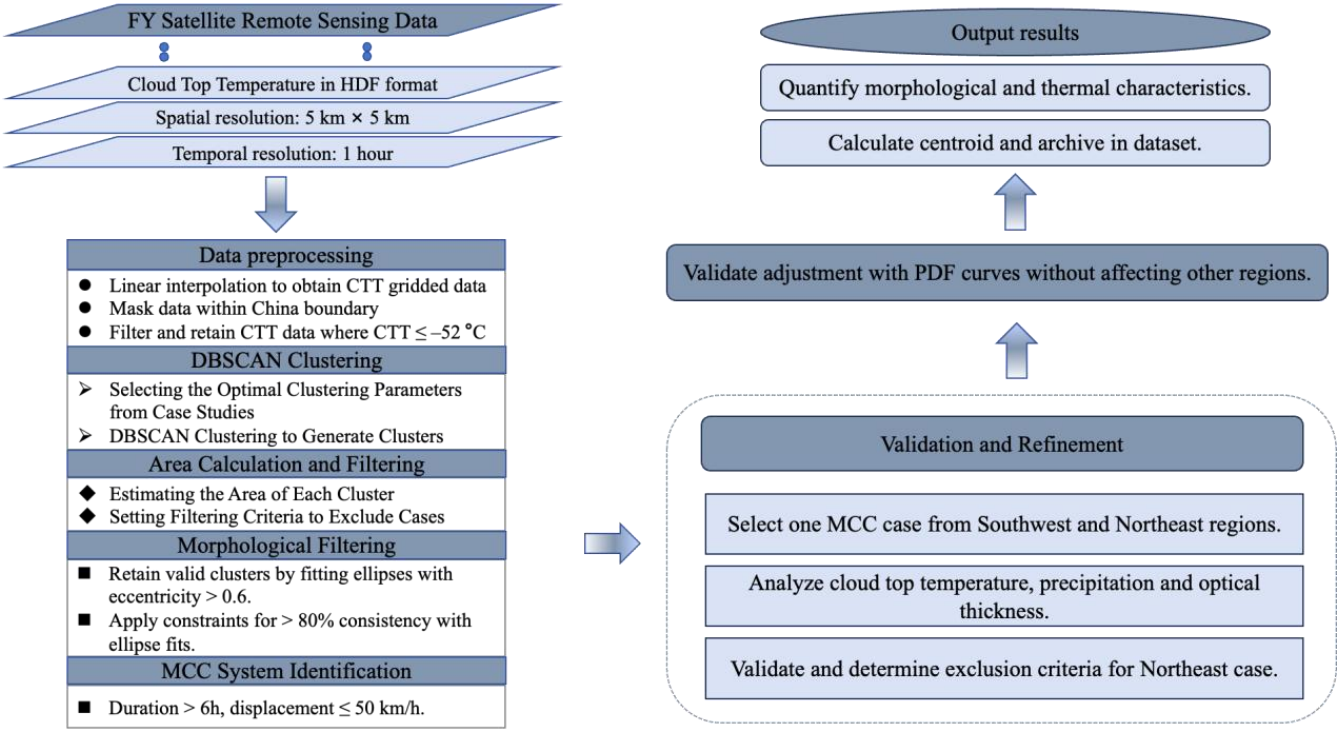


Figure. 1: Flowchart of the automated MCC identification and tracking process. The left panel delineates the steps for algorithmic identification, while the right panel outlines the procedures for data validation and output.

a. The first procedure involved geographically masking the original FY-2G satellite data to isolate the China region. The masked data were then resampled to a 0.1 ° ×0.1 ° grid, to optimize~~ing~~ computational efficiency for subsequent analysis. Finally, the initial extraction of potential cold cloud pixels was performed by applying a Cloud Top Temperature (CTT) threshold of ≤ -52°C (~~Fig.~~Figure 2a).

b. The second phase comprised the application of the density-based DBSCAN algorithm to cluster the discrete cold cloud pixels, a technique chosen for its proficiency in identifying dense clusters amidst noise (~~Fig.~~Figure 2b). Following the clustering application, the critical parametrization of DBSCAN was performed. The core parameters—neighborhood radius (eps) and minimum points (min_samples)—were optimized through analysis of 50 historical MCC event samples (Table 2), with the optimal parameter combination ultimately determined as eps = 0.5 and min_samples = 36.

Table 2: Algorithm performance evaluation across parameter configurations using 50 manually verified MCC cases

Parameter selection	Event success rate	Moment individual recall	Moment individual precision
eps = 0.5, min_samples = 25	46/50	317/370	317/397
eps = 0.5, min_samples = 30	48/50	328/370	328/403
eps = 0.5, min_samples = 35	50/50	332/370	332/422
eps = 0.5, min_samples = 36	50/50	342/370	342/419
eps = 0.5, min_samples = 40	50/50	325/370	325/404
eps = 0.5, min_samples = 45	45/50	306/370	306/368
eps = 0.5, min_samples = 50	37/50	247/370	247/309
eps = 0.5, min_samples = 55	32/50	212/370	212/267
eps = 0.6, min_samples = 85	21/50	138/370	138/182
eps = 1.0, min_samples = 220	7/50	39/370	39/41
eps = 1.0, min_samples = 240	3/50	12/370	12/12

c. The third procedure involved the estimation of the spatial extent for each cluster generated in the previous step. The area was derived by computing the cumulative area of all pixels belonging to a given cluster. The area of each cluster~~individual pixel area~~, which incorporates a geometric correction for the Earth's sphericity, is defined as Eq. (1):

$$S_i = R^2 \cdot \Delta\varphi_i \cdot \Delta\lambda_i \cdot |\cos(\varphi_i)| \times n_i \,, \quad (1)$$

where R denotes the Earth's radius ($\approx$ 6,371 km), $\Delta\varphi_i$ and $\Delta\lambda_i$ represent the latitudinal and longitudinal spans in radians, $|\cos(\varphi_i)|$ is the correction factor for the latitudinal variation in longitudinal arc length. The total cluster area S_i was subsequently obtained by multiplying the unit pixel area by n_i, the number of grid points in the cluster.

d. The fourth phase consisted of a multi-tiered filtering process. It commenced with a preliminary screening that eliminated cloud clusters with an area below 40,000 km². For clusters satisfying this area threshold, a two-stage morphological filtering approach was then executed based on minimum bounding ellipse fitting. The first filtering criterion mandated an ellipse-fitted eccentricity of ≥ 0.6, a deliberately permissive threshold~~lenient parameter~~ to encompass systems in their early~~formative~~ stages (~~Fig.~~Figure 2c). The second filtering criterion required clusters to maintain at least 80% areal coverage relative to their fitted ellipses, thereby enforcing morphological regularity and physical plausibility (~~Fig.~~Figure 2d).

e. The final processing stage entailed the definitive classification of the screened MCC candidates. These candidates first underwent standardized parameter extraction. Subsequently, their validity as MCC events was determined by evaluating spatiotemporal continuity against strict criteria: a maximum centroid displacement of 50 km and a minimum duration of continuous convective activity of 6 hours. This concluding phase of the algorithm ensured that only systems demonstrating the requisite persistence and spatial coherence were formally classified as MCC events and archived for subsequent analysis.

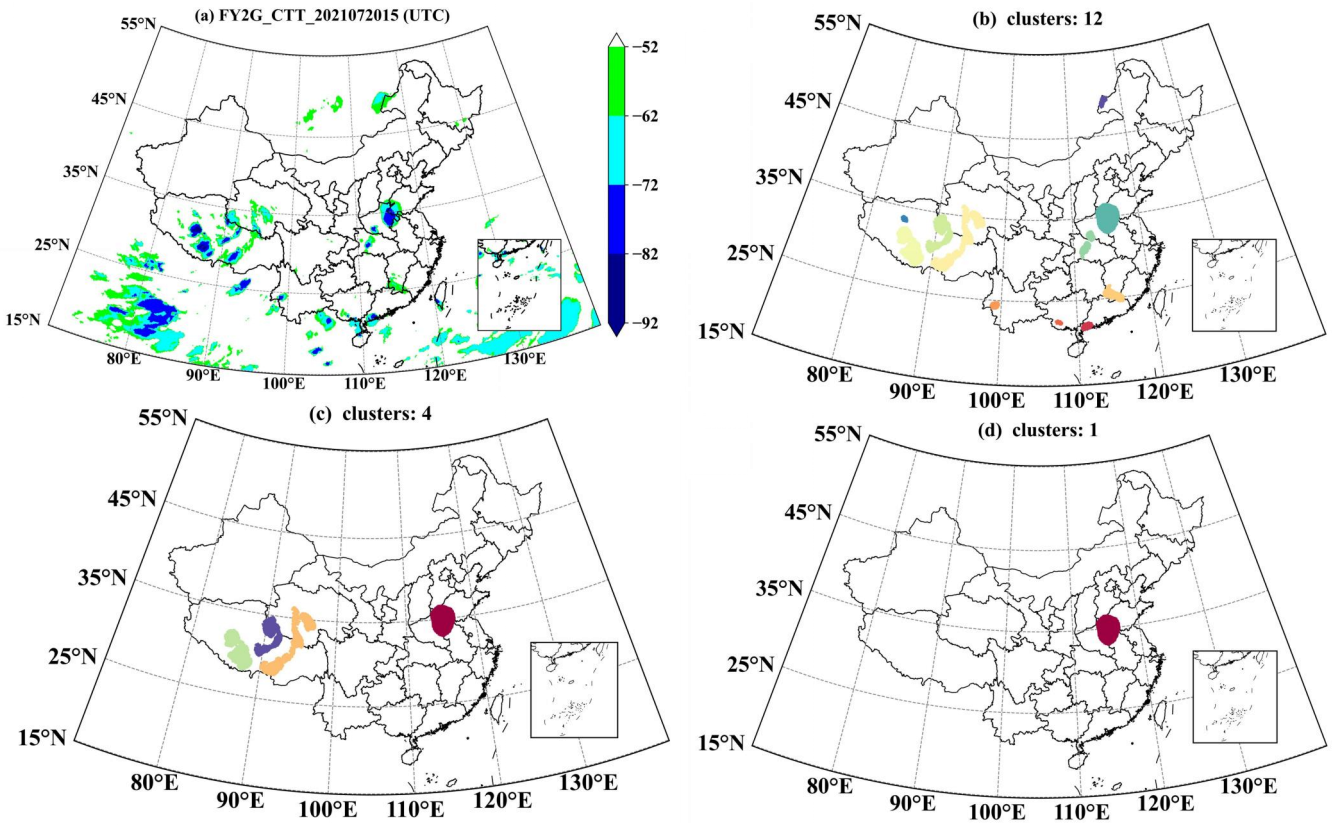


Figure 2: a. Spatial distribution of cloud-top temperature (CTT) below -52°C across China and adjacent regions at [2021-7-20-15:00 UTC], derived from FY-2G satellite observations. b. Spatial distribution of clustered cold-cloud systems identified through DBSCAN algorithm, where distinct colors represent individual clusters meeting MCC candidate criteria (CTT ≤ -52°C~~, area ≥ 4×10⁴ km²~~). c. Spatial distribution of clustered cold-cloud systems after area threshold filtering (≥4×10⁴ km²), where distinct colors represent individual valid MCC candidates. d. Final MCC candidates after morphological filtering, where color-filled areas represent systems meeting all criteria (elliptical eccentricity ≥ 0.6, area ratio ≥ 0.8).

3.3 Validation and Refinement

The initial analysis of the MCC dataset (June 2015–December 2024) produced by our algorithm reveals several key spatial patterns (~~Fig.~~Figure 3). The warm-season concentration in southern China is consistent with the canonical pattern described by Chang et al. (2015) and serves to benchmark the algorithm's performance. Notably, the occurrence frequency in ~~Northeast China~~northeastern China is significantly higher than in other regions, with this pattern being particularly pronounced in winter (Dec–Feb), which contrasts with the findings of Chang et al. (2015). It is essential to consider whether this pattern points to a genuine climatological feature or highlights a potential limitation of the identification algorithm in capturing the morphological characteristics of MCCs in ~~Northeast China~~northeastern China.

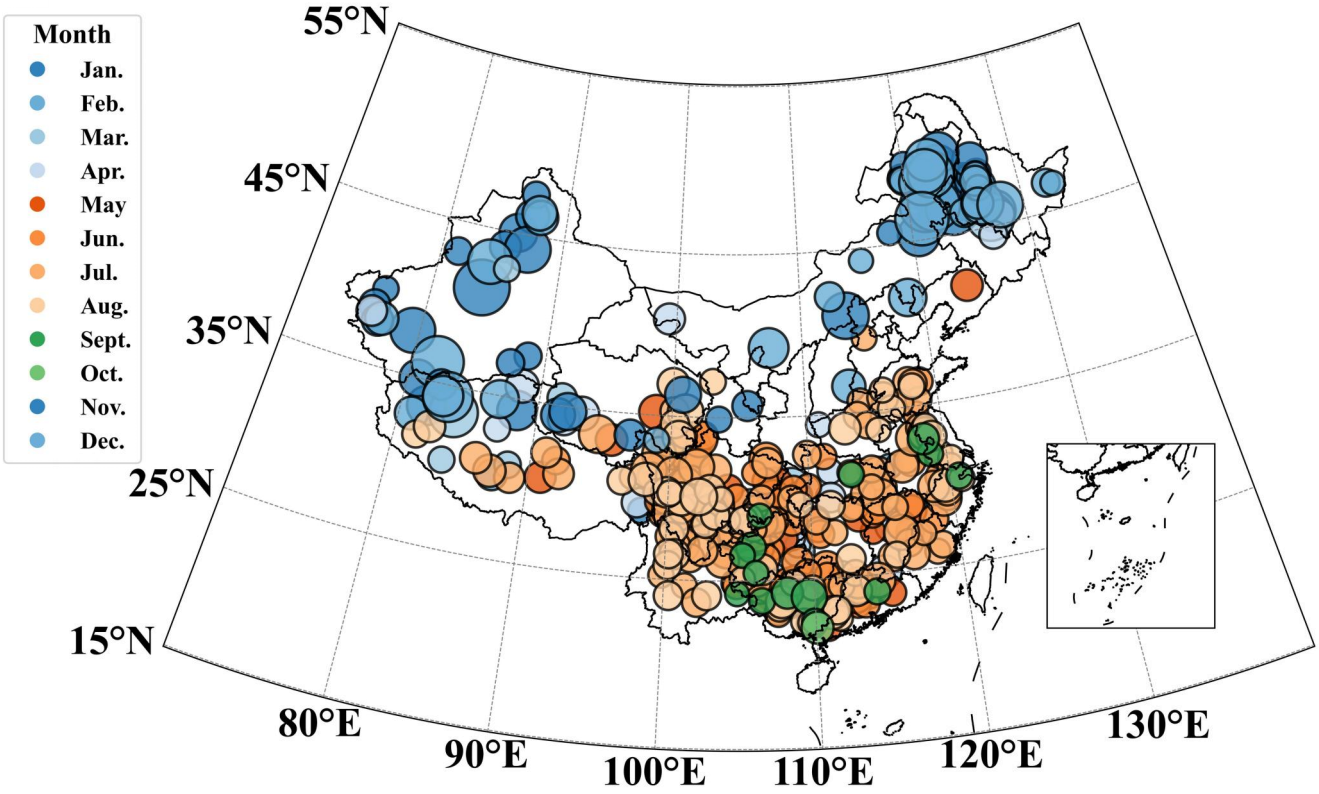


Figure 3: Spatial distribution of MCCs identified by the algorithm across China (2015 – 2024). Each marker represents an individual MCC, with its size proportional to the areal extent (CTT ≤ -52°C) at the onset of its lifetime. Colors represent different months.~~National MCC distribution patterns (2015-2024) with: Color coding: Monthly variation (see colorbar); Marker size: Cluster area with CTT ≤ -52°C (log-scaled, 10^4–10^5 km^2 range); Geographic base: China political boundaries with terrain shading.~~

To evaluate the accuracy of the elevated frequency of MCCs identified in northeastern China, comparative analysis was conducted between two ~~typical algorithm-identified~~randomly selected –cases identified by the algorithm, one from northeastern China~~this region~~ and the other from southern China, with the latter ~~manually confirmed~~verified ~~.~~through manual inspection ~~a comparative analysis was conducted between a typical algorithm-identified case from this region and a manually confirmed case from southern China~~(~~Fig.~~Figure 4). ~~t~~The CTT distribution of the northeastern~~Northeast~~ China case lacks a coherent structure, exhibiting dispersed cold regions without a defined core (~~Fig.~~Figure 4a). In contrast, the s~~S~~outhern China case displays a compact deep convective system featuring a pronounced cold center below −92 ℃, indicative of high cloud tops (~~Fig.~~Figure 4d). This contrast is further corroborated by the optical thickness distribution: the ~~Northeast China~~northeastern China case appears nearly cloud-free or covered by thin clouds (~~Fig.~~Figure 4b), while the ~~Southern China~~southern China case shows extremely high values ($\tau \approx 100$), suggestive of deep cumulonimbus (~~Fig.~~Figure 4e). Similarly, the rainfall distribution for the ~~Northeast China~~northeastern China case shows negligible precipitation (Figure 4c), whereas the ~~Southern China~~southern China case features a strong rainfall maximum co-located with the cloud system

(~~Fig.~~Figure 4f) ~~(Fig. 4)~~. Validation of additional algorithm-identified cases in ~~Northeast China~~northeastern China consistently yielded similar results, leading to the conclusion that the ~~Northeast China~~northeastern China case does not represent a genuine MCC event. This demonstrates that conventional MCC identification algorithms based solely on satellite CTT thresholds~~criteria carry a inherent risk of~~are prone to mis~~classification~~identification in ~~Northeast China~~northeastern China, underscoring the necessity for region-specific ~~ating algorithm~~ adaptations that reflect the ~~to account for the~~ region's unique~~distinctive~~ climatic characteristics.

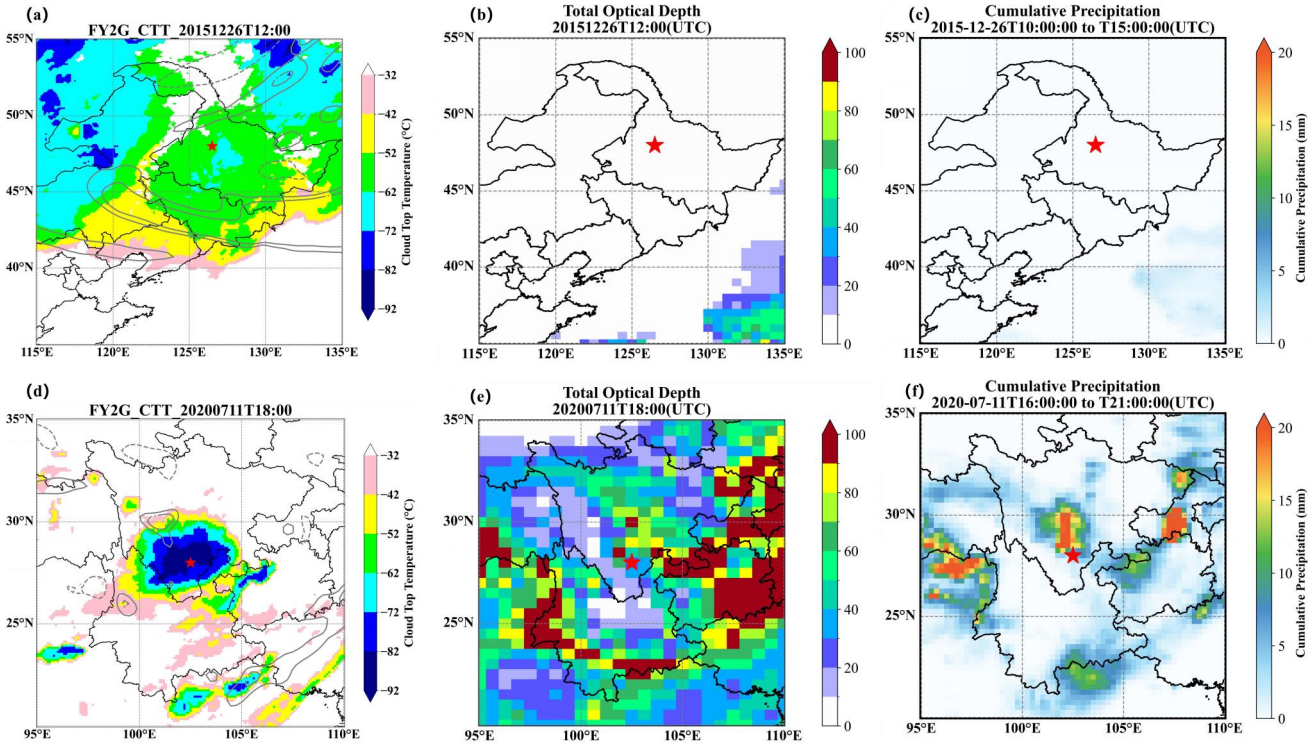


Figure 4: (~~Left~~a-b) Cloud-top temperature (shaded, ℃) with 500-hPa relative vorticity (contours, 10^{-5} s^{-1}); (~~Middle~~c-d) Total optical thickness (unitless); (~~Right~~e-f) Six-hour accumulated precipitation (mm) centered on peak -52 ℃ cloud area, for ~~in~~ a northeastern China case (upper panel) and ~~in~~ a southern China case (lower panel)

Based on the preceding analysis of individual MCC cases, we found that ~~some~~ winter ~~event~~cases automatically identified in ~~Northeast China~~northeastern China were likely false positives. To mitigate these spurious detections, a seasonal filtering excluding winter observations was applied. A~~Following th~~fter~~is~~ filtering, the spatial distribution of MCC occurrences across China shows significantly improved agreement with the patterns documented in~~by~~ Chang et al. (2015). ~~Our analysis demonstrates that the spatial distribution of MCC occurrences across China shows significantly enhanced agreement with the patterns documented by Chang et al. (2015) when winter data are excluded.~~ To quantitatively evaluate the performance of the seasonal ~~algorithm~~filtering~~'s performance~~, we defined a validation reference zone (20°-30°N, 100°-120°E) within which

all algorithm-detected ~~event~~cases were manually verified against key MCC characteristics, including cloud-top structure, lifetime duration, and associated precipitation. This rigorously validated region subsequently served as a benchmark for assessing the detection accuracy in other geographic domains.

A comparative analysis of the probability distribution function (PDF~~s~~) of MCC areal extent between the full set of cases across China and the subset obtained after seasonal filtering shows that the latter exhibits greater consistency with ~~indicates a pronounced discrepancy between the national dataset including all seasons and~~ the validated reference ~~dataset~~region (Figure 5). The areal extent PDF of all MCCs without seasonal filtering exhibits a long tail to the right, attributable to the false and large winter cases over northeastern China shown in Figure 3~~The unfiltered national dataset exhibits an anomalously enhanced contribution from large MCCs, particularly for areas exceeding 40×10^4 km^2, suggesting the presence of unrealistically extensive systems~~. The comparative analysis demonstrates that the exclusion of winter cases effectively suppresses false MCC identifications in northeastern China while preserving true MCC cases in other regions.

~~After excluding winter months (December–February), the resulting seasonally filtered national dataset shows a substantially reduced large-area tail and a much closer agreement with the reference distribution. This result indicates that the excess occurrence of extremely large MCCs in the unfiltered dataset is primarily attributable to false-positive detections during the winter season, especially over Northeast China.~~ The seasonal filtering~~ing~~ therefore effectively ~~suppresses spurious large-area MCC identifications and~~ improves the overall reliability of the ~~national-scale~~identified -MCC dataset.

The above results highlight that integrating seasonal and regional specificity is critical for advancing MCC detection methodology. Our empirical results demonstrate the limited applicability of the standard MCC identification criteria to northeastern China during the winter season. This finding underscores the necessity for future algorithm optimization to account for regionally-specific parameters.

~~Comparative analysis of the probability density functions (PDFs) for MCC areas shows a clear divergence in distribution shape between the unfiltered national data and the ground-truth benchmark, contrasting the green and pink curves in Figure 5. Conversely, the seasonally-filtered national dataset (black curve) exhibits striking concordance with the validated reference distribution. This finding substantiates that the conspicuous divergence in PDF shapes can be primarily ascribed to false-positive MCC identifications, with a significant concentration in Northeast China's winter season.~~

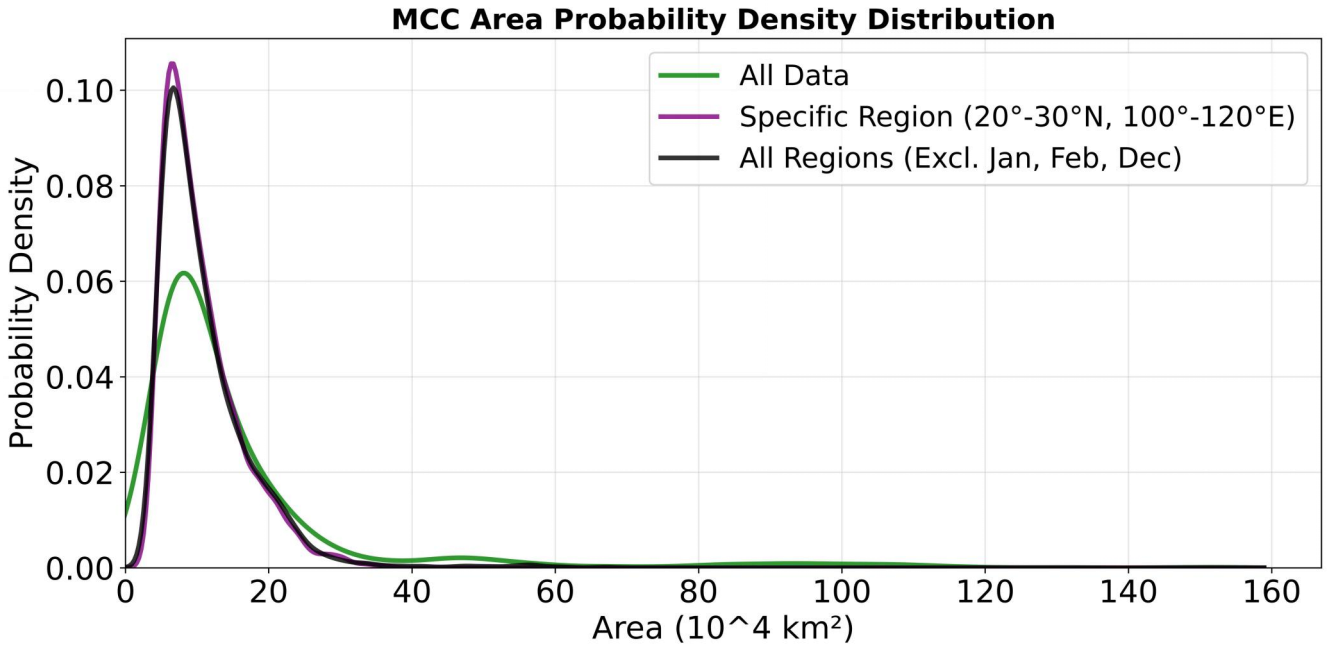


Figure 5. ~~Probability distribution function (~~PDF~~)~~ of MCC areal extent for the full set of cases across China~~national dataset including all seasons~~, the validated subset ~~reference dataset~~ over a selected region (20°–30°N, 100°–120°E), and the ~~national dataset~~subset across China after excluding winter months (December–February).~~Area Probability Distribution Curve. The green line represents the area probability distribution for the China region, the purple line shows the area probability distribution for a specific ground-truth region, and the black line indicates the area probability distribution for the China region after removing winter data.~~

~~This study demonstrates that the exclusion of winter observations effectively suppresses false MCC identifications in Northeast China while preserving legitimate climatic signals~~true MCC events in other regions. ~~This finding provides quantitative support for the limited applicability of conventional MCC criteria in areas with pronounced seasonal variation~~atmospheric divergence. ~~The results highlight that incorporating seasonal and regional specificity is critical for advancing MCC detection methodology. Our empirical results demonstrate the limited applicability of the standard MCC identification criteria to Northeast China during the winter season. This finding underscores the necessity for future algorithm optimization to incorporate regionally-specific parameters.~~

3.4 ~~Results~~Data Output

The algorithmic outputs were consolidated into a structured dataset representing the spatiotemporal and physical characteristics of the identified MCC events. Each record includes quantified metrics such as cloud shield area, minimum CTT, and centroid location, stored alongside necessary metadata in standardized text formats (Table 3). This final dataset is readily accessible for climatological analysis, model validation, and comparative studies.

Table 3. Summary of variables in the MCC dataset

Variable name	Description	Units
Time	~~Observation~~ Ttime (yyyymmddhh) of each MCC record	~~YYYYMMDDHH~~
Area	Areal extent of ~~the~~ MCC ~~deep convective core~~cold cloud shield (CTT≤ −52℃)	km^2
Maj.Axis	Major axis length of ~~best fitting ellipse~~ MCC	km
Ecc	Eccentricity of ~~best fitting ellipse~~MCC	-1
Lat	Latitude of ~~cluster~~MCC centroid	°N
Lon	Longitude of ~~cluster~~MCC centroid	°E
CTT_Min	Minimum cloud-top temperature within ~~cluster~~MCC	K

4 Results

We perform a preliminary statistical analysis of the comprehensive MCC dataset to delineate the spatial and seasonal distribution of MCC occurrences across China (~~Fig.~~Figure 6). The results reveal pronounced regional heterogeneity. Over the past decade, the total number of MCC events has been highest in Southwest China, far exceeding that in other regions.~~cumulative MCC activity has been the most frequent in Southwest China, far exceeding that in other regions.~~ The Yangtze River Basin, the South China coast, and the Sichuan Basin are also identified as hotspots for MCC occurrence. A distinct north-south seasonal contrast is evident, with MCCs in northern China relatively rare and occurring during spring and autumn, whereas those in southern China frequent and predominantly concentrated in summer. Furthermore, the spatial scales of the MCCs, represented by the circle sizes in Figure 6, indicate that summer systems rarely attain an areal extent exceeding 20×10^4 km^2~~extremely large areal extents~~. An exception is Southwest China, where MCCs from June to August are both larger in scale and more frequent than their autumn (Sep - Nov) counterparts.

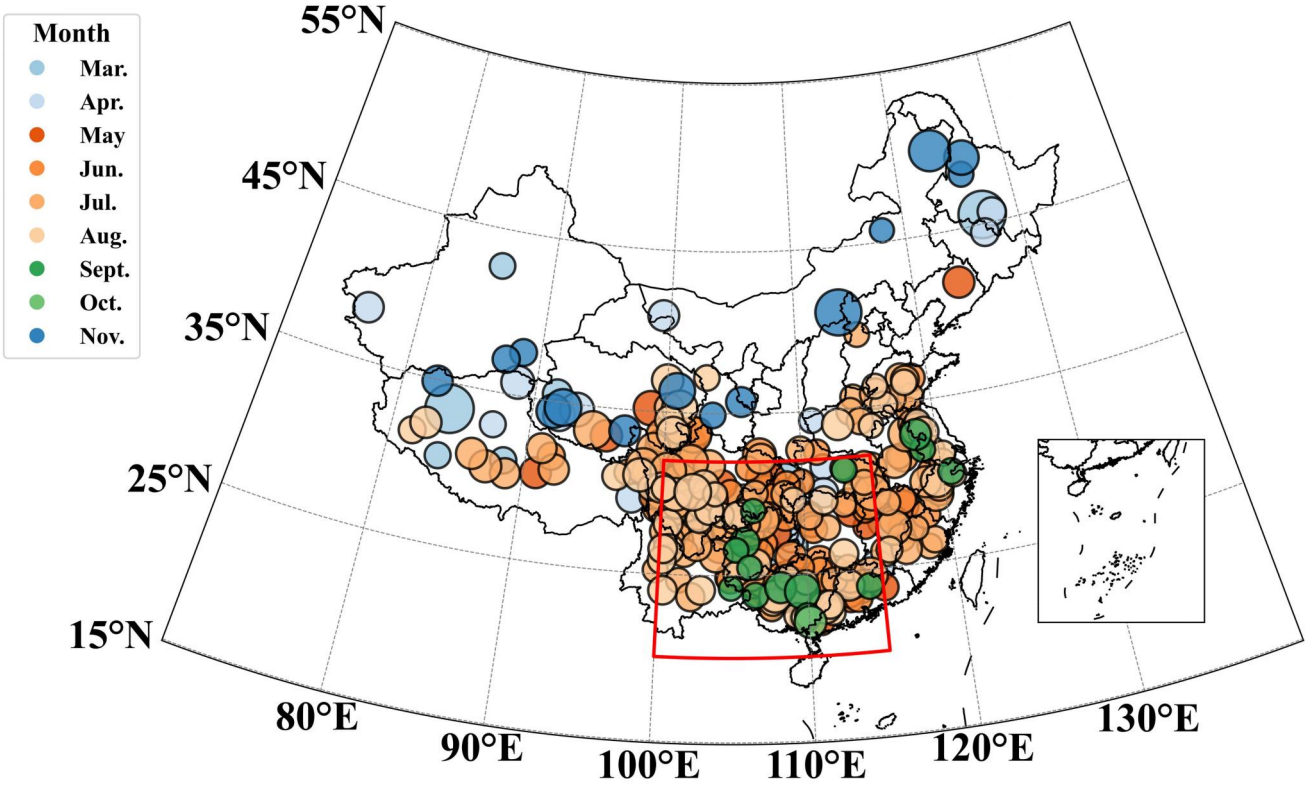
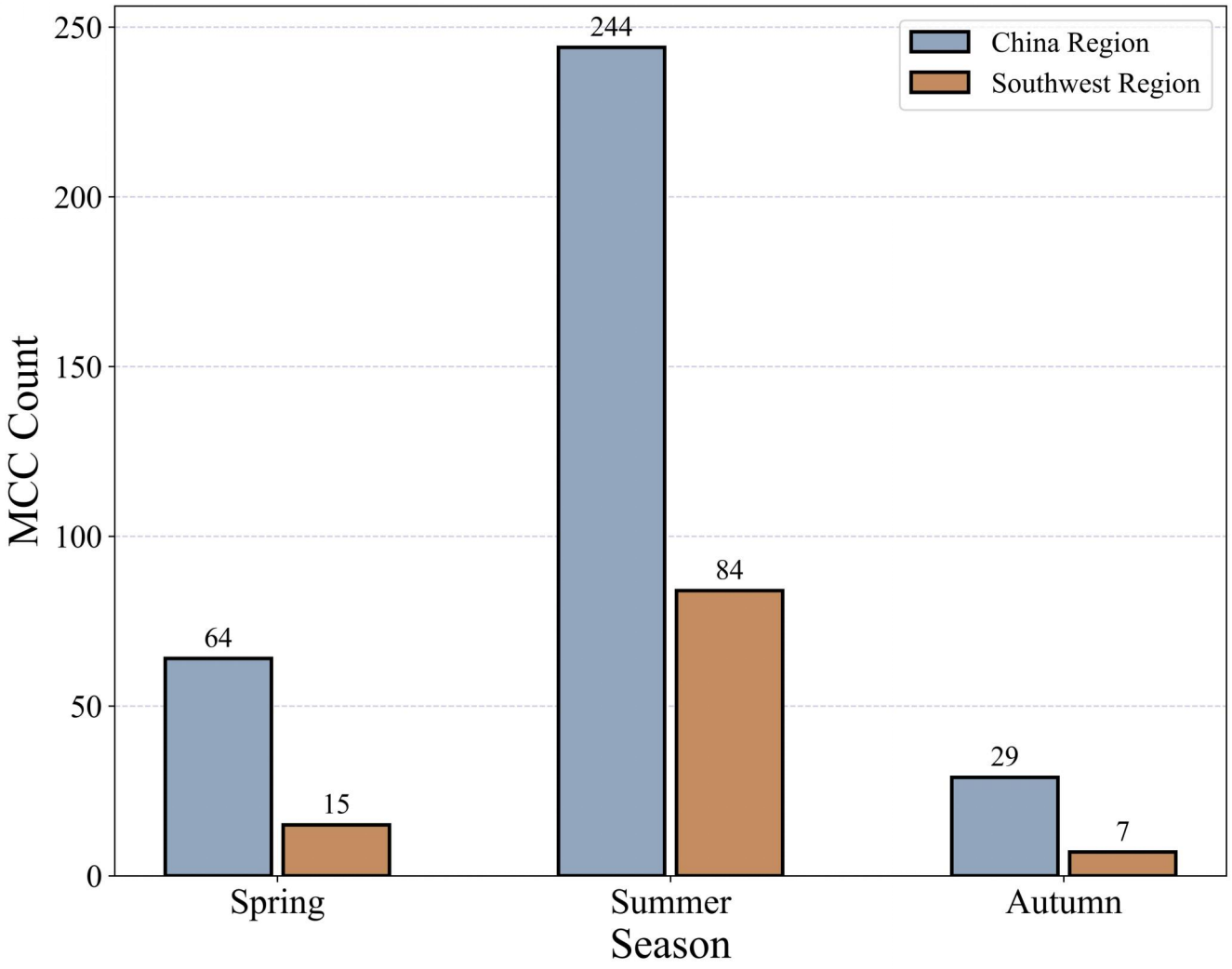


Figure 6: ~~Spatial distribution of MCC events across China (2015 - 2024, winter-excluded). , where color-coding indicates monthly variation (see colorbar) and marker size represents the cluster area with cloud-top temperature ≤ -52°C.Each marker represents an individual MCC event, with marker size proportional to the areal extent of the MCC cluster (cloud-top temperature ≤ -52°C). Color-coding indicates monthly variation (see colorbar)~~Same as Figure3, but with winter months excluded. Red box denotes the southwestern China in this study.

Given the significantly higher frequency of MCC occurrences and associated extreme rainfall and secondary disasters in Southwest China compared to other regions, this area (demarcated in ~~Fig.~~Figure 6) is analyzed separately in addition to the analysis for mainland China as a whole.

As a primary distribution area for MCCs in China, Southwest China accounts for approximately 32% of the national total frequency, underscoring its prominence in the regional climatology (~~Fig.~~Figure 7). The seasonal distribution of MCC occurrences in this region aligns with the broader pattern observed across mainland China, characterized by a dominant peak in boreal summer (June–August), a secondary frequency in spring (March–May), and notably subdued activity in autumn (September – November). This temporal regime is principally governed by monsoon-modulated dynamics and the thermodynamic environment, in which the availability of low-level moisture and the degree of convective instability during the warm season serve as critical controlling factors (Zhang et al., 2025). As evidenced in Fig~~s. 7 and 8~~ure 7, the summer months alone account for a substantial majority (75–80%) of all annual MCC events identified in the region.



315

Figure 7: Seasonal distribution of MCC frequency over China (blue bars) and the southwestern region (brown bars).

Figure 8 depicts the annual frequency of MCC occurrences over mainland China and Southwest China from 2015 to 2024. Both time series exhibit substantial interannual variability. The mainland China series fluctuates between a minimum of 17 (2015) and a maximum of 59 (2020), with notably higher frequencies also recorded in 2016 and 2024. Similarly, Southwest China witnesses MCC counts ranging from 4 (2022) to 21 (2020). The coherence of the peak in 2020, as well as the periodicity, across mainland China and Southwest China implicates a common set of large-scale climatic drivers that are particularly conducive to MCC development during that year.

320

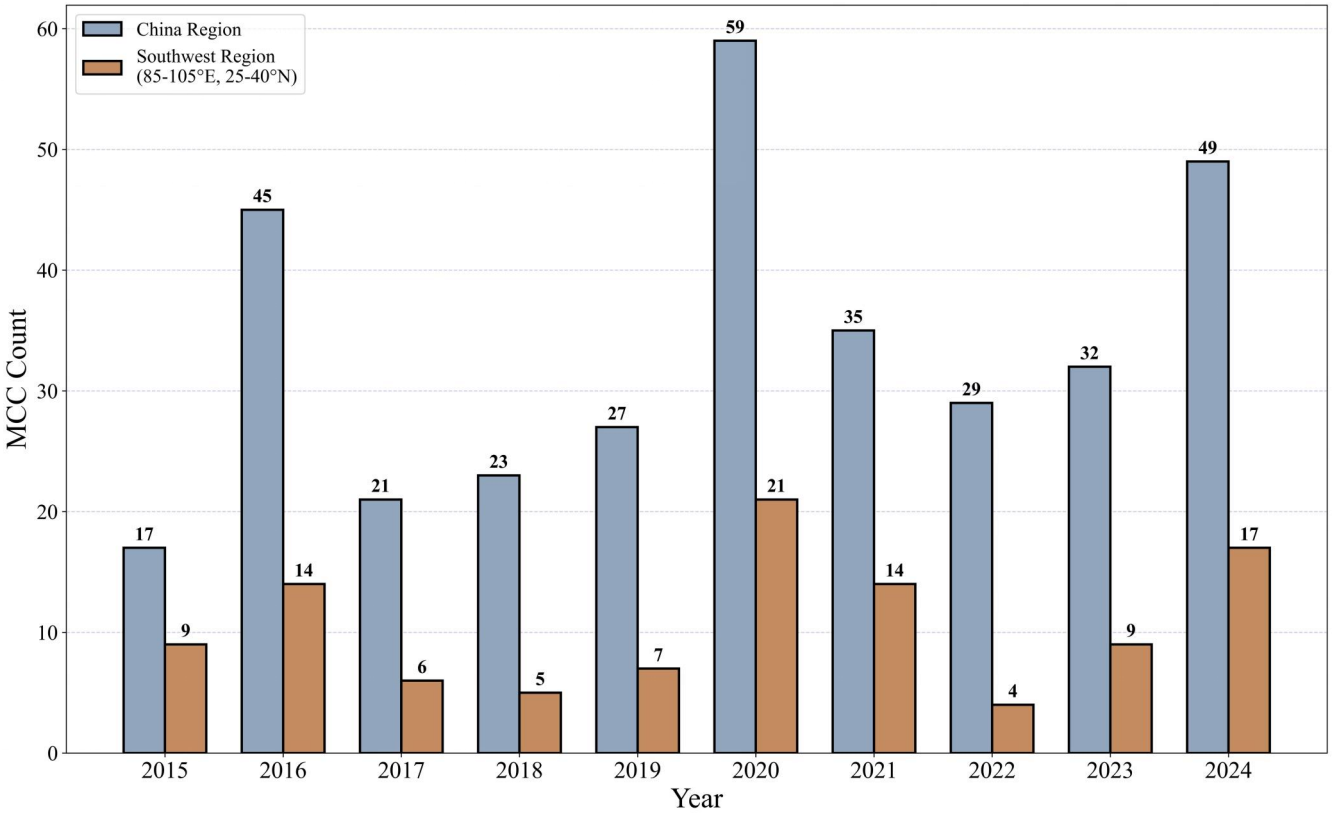


Figure 8: Interannual variation of MCC frequency over China (blue bars) and the southwestern region (brown bars).

The interannual variability of MCC occurrence in Southwest China appears to be influenced by the El Niño-Southern Oscillation (ENSO). For instance, the peak years of MCC occurrence in Southwest China (2016, 2020, 2024), consistently coincided with the decaying phase of El Niño and/or the developing phase of La Niña. Conversely, the annual trough of MCC occurrence in 2018 corresponded to the decaying year of La Niña and the developing year of El Niño. Statistical fitting of the 2015－2024 time series further corroborates this relationship, revealing a significant periodicity of 4.27 years (p-value < 0.0001), which is strikingly close to the mean period of the ENSO. This result strongly suggests that the statistical characteristics of annual MCC occurrences in Southwest China are modulated by large-scale air-sea interactions, although the underlying mechanisms require further investigation.

5 Data availability

The database is available via Zenodo at https://doi.org/10.5281/zenodo.17349888 (Xu, 2025).

6 Conclusions

This study develops an automatic MCC detection and tracking algorithm based on the cloud top temperature from the FY-2G satellite, by employing a combination of DBSCAN clustering, multi-threshold screening, and morphological filtering. And a long-term (2015－2014) MCC dataset across mainland China is generated. The algorithm effectively captures MCC events. The Exclusion of winter data substantially enhances detection accuracy in ~~Northeast China~~northeastern China, where cold surface conditions and thin and ~~very~~ high clouds can lead to false identifications. This necessity of region-specific adaptations to the MCC indentification protocol.

Analysis of the spatial distribution of MCCs across mainland China reveals pronounced regional heterogeneity. Southwestern China stands out as a primary hotspot for MCC occurrence, exhibiting a notably higher frequency than other regions that accounts for approximately 32% of the national total. MCC activity demonstrates clear seasonal variability, with a primary peak during boreal summer and a secondary maximum in spring. Furthermore, both mainland China and Southwest China exhibit marked interannual variability in annual MCC frequency. The interannual fluctuations in MCC occurrence over Southwest China are likely modulated by the El Niño-Southern Oscillation.

The primary objective of this work is the development of a novel MCC detection algorithm; consequently, the analysis of the spatiotemporal distribution of MCCs across mainland China remains preliminary. Our findings, however, clearly identify Southwestern China as a region of exceptionally high frequency of MCCs and associated disaster risk. This insight paves the way for our future research, which will leverage the generated dataset to focus on this hotspot region. We will investigate the region's key meteorological characteristics, MCC triggering conditions, and maintenance mechanisms, ultimately elucidating their relationship with ENSO. Furthermore, the rapid advancement of satellite technology and increasing availablility of high-resolution data present a critical avenue for future research: the adaptation of this algorithm for practical application in operational convective systems monitoring and nowcasting.

Author contributions

SZ supervised the entire study. KX and SZ jointly conceived the study and designed the methodology. They were also responsible for data curation and formal analysis. KX uploaded the dataset to the repository and drafted the original manuscript. SZ and XM provided critial feedback and assisted in revision. All authors participated in discussions throughout the research process and approved the final manuscript.

Acknowledgement

This work was financially supported by the National Key Research and DevelopmentProgram of China (2024YFF0807904), the Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province (SZKT202306).

References

Abisusmita, R. W., Arsyad, M., and Subaer, S.: Karakteristik Mesoscale Convective Complex (MCC) di Wilayah Sulawesi Selatan dan Sekitarnya, Jurnal Fisika Unand, 12, 283-290, doi: 10.25077/jfu.12.2.282-289.2023, 2023.

Augustine, J. A. and Howard, K. W.: Mesoscale Convective Complexes over the United States during 1986 and 1987, Mon. Wea. Rev., 119, 1575-1589, doi: 10.1175/1520-0493(1991)119<1575:MCCOTU>2.0.CO;2, 1991.

Bai J, W. H., Tao, Z.: Retrieval of cloud motion winds from GMS satellite infrared imagery. (in Chinese), Acta Scientiarum Naturalium Universitatis Pekinensis, 33(1), 85-92, doi: 10.13209/j.0479-8023.1997.011, 1997.

Blamey, R. C. and Reason, C. J. C.: Mesoscale Convective Complexes over Southern Africa, J. Climate, 25, 753-766, doi: 10.1175/JCLI-D-10-05013.1, 2012.

Chang C, D. Z.: Distribution characteristics of MCCs east of the Tibetan Plateau in China during summer 2007-2012. (in Chinese), Journal of the Meteorological Sciences, 35(4), 445-453, doi: 10.3969/2014jms.0008, 2015.

Cotton, W. and Anthes, R.: Storm and Cloud Dynamics Academic Press, San Diego, Calif. 1989.

Duan L, G. G.: An automatic MCS identification and tracking method using Fengyun satellite imagery. (in Chinese), Electronic Science and Technology, 29(4), 116-119+126, doi: 10.16180/j.cnki.issn1007-7820.2016.04.030, 2016.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, kdd, 96, 226-231, 1996.

Fei Z, W. H., Zhang Y, Song S, Liu J: Automatic identification and tracking of MCS based on geostationary satellite infrared imagery. (in Chinese), Journal of Applied Meteorological Science, 22(1), 115-122, doi: CNKI:SUN:YYQX.0.2011-01-016, 2011.

Houze Jr, R. A.: Orographic effects on precipitating clouds, Reviews of Geophysics, 50, doi: 10.1029/2011RG000365, 2012.

Houze, R. A.: Structure and Dynamics of a Tropical Squall – Line System, Mon. Wea. Rev., 105, 1540-1567, doi: 10.1175/1520-0493(1977)105<1540:SADOAT>2.0.CO;2, 1977.

Hua, S., Xu, X., and Chen, B.: Influence of Multiscale Orography on the Initiation and Maintenance of a Precipitating Convective System in North China: A Case Study, Journal of Geophysical Research: Atmospheres, 125, e2019JD031731, doi: 10.1029/2019JD031731, 2020.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. D. W., Stumpf, G. J., Eilts, M. D., and Thomas, K. W.: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm, Wea. Forecasting, 13, 263-276, doi: 10.1175/1520-0434(1998)013<0263:TSCIAT>2.0.CO;2, 1998.

Lakshmanan, V., Rabin, R., and DeBrunner, V.: Multiscale storm identification and forecast, Atmospheric Research, 67-68, 367-380, doi: 10.1016/S0169-8095(03)00068-1, 2003.

Le Barbé, L., Lebel, T., and Tapsoba, D.: Rainfall Variability in West Africa during the Years 1950–90, J. Climate, 15, 187-202, doi: 10.1175/1520-0442(2002)015<0187:RVIWAD>2.0.CO;2, 2002.

Li Y, W. J., Zheng X, Guo W, Huang W: A study of Mesoscale Convective Complexes (MCC) in Southwest-South China. (in Chinese), Atmospheric Sciences, 13(4), 415-422, doi: CNKI:SUN:DQXK.0.1989-04-004, 1989.

Machado, L. A. T. and Rossow, W. B.: Structural Characteristics and Radiative Properties of Tropical Cloud Clusters, Mon. Wea. Rev., 121, 3234-3260, doi: 10.1175/1520-0493(1993)121<3234:SCARPO>2.0.CO;2, 1993.

Machado, L. A. T., Desbois, M., and Duvel, J.-P.: Structural Characteristics of Deep Convective Systems over Tropical Africa and the Atlantic Ocean, Mon. Wea. Rev., 120, 392-406, doi: 10.1175/1520-0493(1992)120<0392:SCODCS>2.0.CO;2, 1992.

Machado, L. A. T., Rossow, W. B., Guedes, R. L., and Walker, A. W.: Life Cycle Variations of Mesoscale Convective Systems over the Americas, Mon. Wea. Rev., 126, 1630-1654, doi: 10.1175/1520-0493(1998)126<1630:LCVOMC>2.0.CO;2, 1998.

Maddox, R. A.: Mesoscale convective complexes, Bulletin of the American Meteorological Society, 1374-1387, 1980.

Mathon, V. and Laurent, H.: Life cycle of Sahelian mesoscale convective cloud systems, Q.J.R. Meteorol. Soc., 127, 377-406, doi: 10.1002/qj.49712757208, 2001.

Matthews, J. and Trostel, J.: An improved storm cell identification and tracking (SCIT) algorithm based on DBSCAN clustering and JPDA tracking methods, American Meteorological Society, Atlanta, GA. Available from: http://ams. confex. com/ams/90annual/techprogram/paper_164442. htm, 2010.

Miller, D. and Fritsch, J. M.: Mesoscale Convective Complexes in the Western Pacific Region, Mon. Wea. Rev., 119, 2978-2992, doi: 10.1175/1520-0493(1991)119<2978:MCCITW>2.0.CO;2, 1991.

Schumacher, R. S. and Rasmussen, K. L.: The formation, character and changing nature of mesoscale convective systems, Nature Reviews Earth & Environment, 1, 300-314, doi: 10.1038/s43017-020-0057-7, 2020.

Shah, S., Notarpietro, R., and Branca, M.: Storm Identification, Tracking and Forecasting Using High-Resolution Images of Short-Range X-Band Radar, Atmosphere, 6, 579-606, doi: 10.3390/atmos6050579, 2015.

Silva Dias, M. A. F. and Ferreira, R. N.: Application of a linear spectral model to the study of Amazonian squall lines during GTE/ABLE 2B, J. Geophys. Res., 97, 20405-20419, doi: 10.1029/92JD01333, 1992.

Velasco, I. and Fritsch, J. M.: Mesoscale convective complexes in the Americas, J. Geophys. Res., 92, 9591-9613, doi:10.1029/JD092iD08p09591, 1987.

Vila, D. A., Machado, L. A. T., Laurent, H., and Velasco, I.: Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) Using Satellite Infrared Imagery: Methodology and Validation, Wea. Forecasting, 23, 233-245, doi: 10.1175/2007WAF2006121.1, 2008.

Xiang X, J. J.: Mesoscale Convective Complexes in ~~Southern China~~southern China. (in Chinese), Journal of Applied Meteorological Science, 6(1), 9-17, doi: CNKI:SUN:YYQX.0.1995-01-001, 1995.

Xu K.: A new dataset of MCC derived from FY-2G satellite data, Zenodo [data set], doi: 10.5281/zenodo.17349888 , 2025.

Yan W, H. X., Zhao Y, Yang T, Ni H: Introduction to a 3D structure identification technique for thunderstorm cells based on an improved DBSCAN clustering algorithm. (in Chinese), Journal of Tropical Meteorology, 36(4), 542-551, doi: 10.16032/j.issn.1004-4965.2020.050, 2020.

Zhang, Y., Chen, H., Li, P., and Li, J.: The characteristics of summer mesoscale convective systems with different moving paths over Southwest China, Climate Dynamics, 63, 77, doi: 10.1007/s00382-025-07586-y, 2025.