We appreciate all the reviewers for their valuable comments and suggestions, which helped us to improve our manuscript greatly.

Reviewer 1's Comments to Author:

Main comments:

This work developed an automated algorithm for mesoscale convective complex (MCC) identification and tracking, and then composed a long-term dataset of MCC variables over mainland China, based on the FY-2G cloud-top temperature data. MCC is an important type of convective system, often bringing about persistent heavy rain and secondary geologic disasters. Thus, this work has the potential in promoting the monitoring and research on the occurrence of MCC, and helping in disaster alleviation over the region. In general, the algorithm is described clearly, the manuscript is well-organized, and the dataset can be actually downloaded. However, there are some places (see specific comments) that need to be revised before it can be accepted for publication.

Specific comments:

1. At line 37, I think the authors mean "their variation" instead of "their formation", as the Southern Oscillation, the Madden-Julian Oscillation are not short-term signals.

Reply: Thanks for the comment. We have revised "their formation" to "their variation" at line 37.

2. In lines 46-49, the sentence should be re-arranged to make its means more clear.

Reply: Thanks for the comment. Yes, the original sentence is too long and contains an excess of information, which compromises its clarity. We have restructured the sentence by breaking it into shorter sentences and removing unclear information, please see the revised manuscript at lines 46-48.

3. At line 80, the sentence begins with "Through multispectral sensing in the visible, infrared, and water-vapor bands, " is not completed. It seems the latter part of the sentence is missed out.

Reply: Thank you for pointing this out. Yes, the latter half of the sentence "cloud-top temperatures are monitored comprehensively" was inadvertently omitted during the revision and saving of the draft. Now the sentence has been reinstated and is complete. Please see the revised manuscript at line 79.

4. In lines 80-83: "The satellite's longitude has been relocated twice: initially at 99.5° E prior to June 1, 2015, ..." can be modified to be more concise. And I don't understand whether the operational scan cycle is 30 or 60 minutes, or somewhere in the between.

Reply: Thanks for the comment. Now, the sentence is revised as "The satellite's longitude has undergone two relocations: it was positioned at 99.5°E before June 1,

2015; then at 105°E from that date until April 16, 2018; and has since been stationed at 99.2°E.", which is more concise than the original one. The operational scan cycle is one hour during non-flood season and 30 minutes during flood season. We revised the sentence to make this information clearer. Please see the revised manuscript at lines 79-82.

5. Remove the "measurements" in line 85.

Reply: Done, thanks.

6. At line 89, what does the "nominal" mean?

Reply: Thanks for the comment. To avoid ambiguity, we have changed the term "a nominal 5km spatial resolution" to "a spatial resolution of approximately 5 km × 5 km". Please see the revised manuscript at line 89.

7. At line 96, "reveal" -> "reveals". And in line 104, "where" -> "with"

Reply: Done, thanks.

8. The sentence in lines 105-108 needs to be re-arranged. For example, what is the link between "ecc criterion is relaxed from ⩾0.7 to ⩾0.6 when the cold-cloud shield reaches its maximum extent" and "enabling capturing the full MCC lifecycle" ?

Reply: Thanks for the comment. The original sentence is a little ambiguous. We have revised the sentence to clarify that the ecc criterion is relaxed from ≥0.7 at maximum extent to ≥0.6 throughout the MCC lifecycle, to accommodate morphological variability. By this, the logical link of the sentence is improved. Please see the revised text at Lines 109-110.

9. At line 110, it is described that "all above morphological and area criteria are satisfied continuously for at least 6 hours". However, in line 107, is seems that ecc ⩾0.6 is only required when the cold-cloud shield reaches maximum extent. Can the authors re-check and clarify these two places?

Reply: Thanks for the comment. Like comment #8, this conflict is also because that the original sentence in lines 105-108 did not make clear ecc ≥0.6 is required throughout the MCC lifecycle. After the revision according to comment #8, this conflict can be eliminated. Please see the revised text at Lines 109-110.

10. The caption of Table 1, remove the "the". And in the table, when a criterion is not used in either Maddox (1980) or this study, probably you can just use "~" instead of "null" or "Not specified".

Reply: Done, thanks.

11. At line 124, "0.1°grid" -> "0.1°×0.1°grid", and "optimizing" -> "to optimize".

Reply: Done, thanks.

12. At line 150, “lenient parameter” -> “permissive threshold”, “formative stages” -> “early stages”

Reply: Done, thanks.

13. The caption of Figure 2 (line 162), I think area $\geqslant 4\times10^4$ km^2 is not applied in the subplot of b, or it will be repetitive with the subplot of c. By the way, Figure 2 has not been cited anywhere in the text.

Reply: Thank you to the reviewer for such a thorough and insightful review. The “area $\geqslant 4\times10^4$ km^2” is now removed from the caption of Figure 2b. Please see the revised caption of Figure 2 at Line 165. We have also added citations of Figure 2 and its subplots in the revised main text, at lines 129, 131, 153, 155, to integrate the figure properly into the discussion.

14. At line 182, “the” -> “The”. And the citation of Fig. 4 should be put here instead of line 188.

Reply: Done, thanks.

15. The discussions in the two paragraphs of page 9 is a little puzzle for readers. For example, does the seasonally-filtered data mean winter-excluded data; and the unfiltered national data mean data of all seasons? Probably the authors can modify these two paragraphs to make them more clearly. And by the way, please don’t use “green and pink curves” or “black curve” to describe a scientific story.

Reply: Thanks for the comment. The two paragraphs in page 9 have been re-organized to improve their accuracy and standardization. For example, we clarified that the seasonal filtering refers to excluding winter months (December–February), and modified “national data” to MCC cases across China. And all non-standard expressions, such as “green and pink curves”, “black curve”, in the original paragraphs in page 9, have been replaced with descriptions based on physical variables. Please see the revised text at lines 203-221.

16. The caption of Figure 5, “Area Probability Distribution Curve” -> “Probability distribution function of MCC areal extent”. And the description of the three lines in the figure can be more concise.

Reply: Thanks for the comments. During the re-organization of the two paragraphs in page 9 according to comment #15, the full name of PDF is given in the main text, thus the caption of Figure 5 has been revised to “PDF of MCC areal extent ...”, and the description of the three lines is modified to “the full set of cases across China (green line), the validated subset over a selected region (20° – 30° N, 100° – 120° E) (purple line), and the subset across China after excluding winter months”, much more concise than before. Please see the revised caption at lines 224-225.

17. At line 216, what does the “legitimate climatic signals” mean here. I think here is still about MCC, right? Climatic signals are too ambiguous.

Reply: Thanks for the comment. We replaced the “legitimate climatic signals” with

"true MCC cases". Please see the revised text at Line 216.

18. At line 217, the "atmospheric divergence" is not precise here to express seasonal variation in weather pattern.

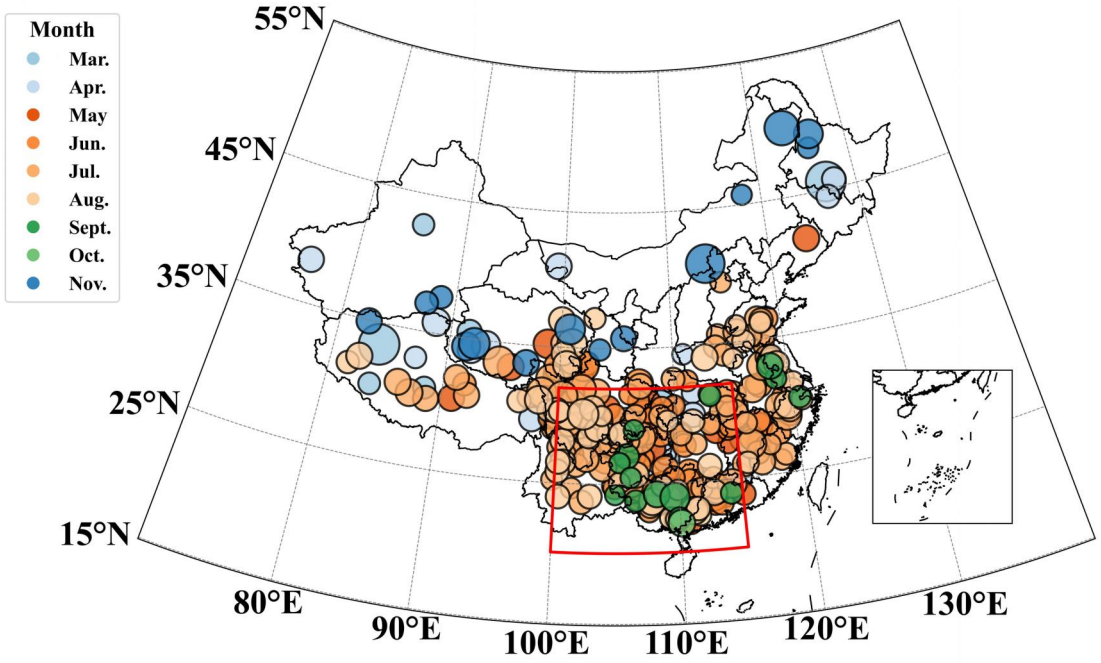
Reply: Thanks for the comment. The sentence with "atmospheric divergence" has been deleted when re-organizing the two paragraphs in page 9 according to comment #15. Please see the revised text at Lines 218-221.

19. The caption of Figure 6, "cluster area" is not precise. Here each marker represents an MCC event, right? Then the marker size must represent the areal extent of the MCC event at some time point, mustn't it?

Reply: Thanks for the comment. It reminds us that this comment is also applied to Figure 3. Thus, we first revised the caption of Figure 3 and give the clear information of each marker and its size as "Each marker represents an individual MCC, with its size proportional to the areal extent (CTT ≤ -52° C) at the onset of its lifetime." Then, we revised the caption of Figure 6 as "Same as Figure3, but with winter months excluded." Please see the revised captions at lines 180-183, and line 248.

20. At line 243, "demarcated in Fig. 6" can not been found in the Figure 6.

Reply: Thanks for the comment. A red box has been added to Fig.6 as the following. Please see the revised Figure 6 in the manuscript in page 12.



Reviewer 2' Comments to Author:

This paper developed an automated algorithm for MCC identification and tracking based on the FY-2G satellite data. Such a dataset is valuable for various applications, such as climate change research and disaster risk reduction. The algorithm is proved to be physically sound and can effectively capture MCC occurrence and movement over mainland China. In general, the paper is well written: the method is clearly described, the analysis and discussions are thorough and comprehensible. However, there are some places that either need more clarification or should be revised before it is published. These are listed in the following specific comments and questions.

Question about the method:

In the north China, cases that are "nealy cloud-free or covered by thin cloud" could be miss-identified as MCC (around line 185) . Can this happen for other seasons and in other regions? If so, how can we justify the correctness of the identified MCCs by simply eliminate the winter season?

Reply: Thank you for the valuable comment. MCC was first defined by Maddox (1980) based on cases observed over America during the warm season (March to September). The definition relies on cloud top temperature thresholds (CTT≤ -32℃ and/or CTT≤ -52℃), requiring that the cold cloud shield be circular and clustered in shape, and exceed a minimum areal extent (achieving meso-α scale). In winter, however, the application of CTT-based criteria to northeastern China presents challenges. The land surface in this region can become extremely cold, allowing clouds with limited vertical development to exhibit CTT values low enough to meet the MCC thresholds. Moreover, this region is frequently influenced by the Northeast China cold vortex (NCCV) — a synoptic-scale system whose cloud shield often appears round, clustered, and expansive, closely resembling the signature of warm-season MCCs. These factors contribute to a high rate of false positives when the CTT-based algorithm is applied during winter. By contrast, during other seasons, without convection, it is not easy for clouds to attain the necessary vertical development to achieve such low CTT values, or exhibit the organized, meso-α scale structure characteristic of true MCCs. Therefore, in the present work we opted to simply exclude winter months from the analysis to ensure data reliability. Nevertheless, in future versions of the automated identification algorithm, we plan to incorporate additional physical variables (e.g., vertical velocity, stability indices) to improve detection accuracy and reduce dependence on manual intervention, such as the seasonal filtering applied here.

Specific comments and corrections:

1. Line 23: "both in South China and mainland China". People may misunderstand these as two separate regions of China. Maybe it's better to say "both in South

China and mainland China as a whole".

Reply: Done, thanks.

2. Line 35: "region of major topographic features" is an ambiguous expression. Do you mean region of large-scale plateau?

Reply: Thanks for the comment. Here "regions of major topographic features" means not only large-scale plateau, but also large mountain range such as the Rockies. To avoid ambiguity, we revised the phrase as "large-scale elevated terrain, such as plateaus and major mountain ranges". Please see the revised text at Lines 35-36.

3. Line 39: The hyphens here should be em-dashes.

Reply: Done, thanks.

4. Lines 96-97 and lines 100-102: These sentences described how manual census reveals the problem of using the original parameters from Maddox (1980) in China. It's better to conjunct these sentences properly to increase the logic fluency and coherence.

Reply: Thanks for this helpful suggestion. We have re-arranged these sentences within one paragraph, by explicitly linking the limitations of the original Maddox (1980) and how to modified the deep-convection core area threshold in order to better apply it in China. Please see the revised text at Lines 98-103.

5. Line 102: please give proper references to the criteria stated here (i.e., CTT<=-52℃with an area >=50000 km^2). And does the 40000 km^2 criteria brought from previous studies or randomly set in this study?

Reply: Thanks for the comment. The original criteria (i.e., CTT $\leqslant -32℃$ with area $\geqslant 10^5$ km^2, CTT $\leq$ -52℃ with an area $\geq$50000 km^2) are from Maddox (1980), and we have added this reference. Please see the revised text at line 99.

The threshold of 40000 km^2 is set by comprehensive consideration of previous studies (i.e., Augustine, 1991; Fei, 2011) and our examination of several MCC cases in China. We have added this illustration, please see the revised text at lines 100-103.

6. Table 1: Both TBB and CTT are used in this table. Are they different variables? Or just two different expressions of the same variable? If the latter, it's better to use the same expression.

Reply: Thank you for the comment. Physically speaking, TBB (Black Body Temperature) and CTT (Cloud-Top Temperature) are not strictly identical. TBB is the brightness temperature directly retrieved from satellite infrared measurements, whereas CTT is a more general concept that can be derived through various methods, including radiative transfer modeling or multi-channel observations.

In our study, the cloud-top temperature data used are from the FY-2G satellite, and its official product is named "CTT" by the data provider (National Satellite Meteorological Center, China). In the seminal work of Maddox (1980), the equivalent variable was referred to as "TBB" based on the satellite data available at that time. To

ensure data traceability and consistency with the official product nomenclature, and to respect the original terminology used in the referenced literature, we prefer to retain the respective terms in Table 1.

7. Lines 129-130: the hyphen here should be em-dash.

Reply: Done, thanks.

8. Equation 1: The equation and text are inconsistent. The equation actually calculated the total clutter area Si, rather than "the individual pixel area" as stated in lines 141-142.

Reply: Thanks for the comment. We have corrected this place by changing "the individual pixel area" to "area of each cluster". Please see the revised text at line 142.

9. Lines 180-181: this sentence is confusing. a better expression may be "comparative analysis was conducted between two typical algorithm-identified cases, one from this region and the other from southern China, with the latter manually confirmed."

Reply: Done, thanks for this valuable suggestion.

10. Line 182: "the CTT" should be "The CTT". Additionally, please refer to the explicit figure on basis of which the discussion is made.

Reply: Thanks for such an insightful review. "the CTT" is now changed to "The CTT". And we have added figure citation in the proper places to clearly indicate the figure on which the analysis is based. Please see the revised text at lines 184-191.

11. Lines 198-199: Are this statement addressed from figure 5? If so, it's better to move it below to appropriate places, perhaps after line 205.

Reply: Thanks for the comment. The statement at original lines 198-199 is not addressed from figure 5, but still based on preceding analysis of individual MCC cases. To make this background explicit, we revised this place by adding sentences before this statement: "Based on the preceding analysis of individual MCC cases, we found that winter cases automatically identified in northeastern China were likely false positives. To mitigate these spurious detections, a seasonal filtering excluding winter observations was applied." Please see the revised text at lines 202-204.

12. Line 202: The title could be "Data Output", to avoid confusion with the title of section 4.

Reply: Done, thanks.

13. Section 3.4: It is recommended to provide a table here to summarize the key information of the dataset, for instance the variables, dimensions, attributes.

Reply: Thanks for this valuable suggestion. We have added a table in Section 3.4 that summarizes the key information of the MCC dataset, including the variables, their physical meaning, units, and typical dimensions. Please see the revised manuscript at

lines 231-233. The table is also presented bellow:

Table 3. Summary of variables in the MCC dataset

Variable name	Description	Units
Time	Time (yyyymmddhh) of each MCC record	~
Area	Areal extent of MCC cold cloud shield (CTT≤ −52°C)	km^2
Maj.Axis	Major axis length of MCC	km
Ecc	Eccentricity of MCC	1
Lat	Latitude of MCC centroid	°N
Lon	Longitude of MCC centroid	°E
CTT_Min	Minimum cloud-top temperature within MCC	K

14. Line 230: The phrase “cumulative MCC” is confusing. Are there other types of MCC?

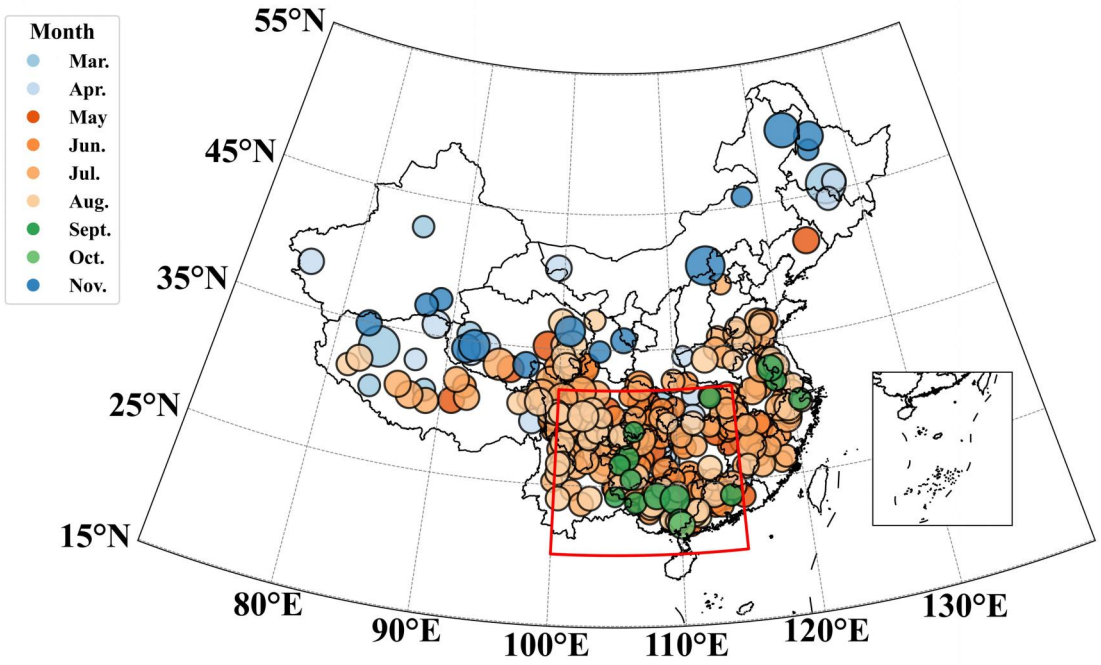
Reply: Thanks for the comment. The phrase “cumulative MCC” was intended to indicate the total number of MCC events over the past decade. We have changed this term to “total number of MCC events” to avoid ambiguity. Please see the revised text at line 237.

15. Line 235: The phrase “extremely large” is ambiguous. Could give a quantitative number?

Reply: Thanks for this suggestion. We have revised the text by replacing the ambiguous phrase “extremely large” with a quantitative description. Based on our dataset, summer MCCs rarely exceed an areal extent of 20×10^4 km^2, which is now explicitly stated in the manuscript. Please see the revised text at line 242.

16. Line 243: It’s unclear where the southwest China here refers to. Please consider illustrating the area in the figure, or providing an exact depiction of its extension in the text.

Reply: Thanks for this comment. To clarify the spatial extent of Southwest China, we have added a clear outline of this region directly on the figure 6. Please see the revised manuscript in page 12. The figure is also presented bellow.



17. Lines 245-248: Please mention figure 7 here when doing description based on the figure.

Reply: Done, thanks.

18. Line 251: "Zhang, 2025" should be "Zhang et al., 2025". "in Figs. 7 and 8" should be "in Figure 7".

Reply: Done, thanks.

19. Line 260: Perhaps "The coherence of the peak in 2020 across..." can be "The coherence of the peak in 2020, as well as the periodicity, across..."

Reply: Done, thanks.

20. Line 281: "2015-2014" should be "2015-2024".

Reply: Done, thanks.

21. Line 282-283: "where cold surface conditions can lead to" perhaps is better to be "where cold surface conditions and thin very high clouds can lead to"?

Reply: Done, thanks.