I appreciate the authors' efforts in addressing my previous comments and revising the manuscript based on the initial feedback. While most of the authors' responses satisfactorily address my concerns, some aspects regarding the robustness of the methodology remain questionable, despite the additional sensitivity analysis provided.
In this analysis, the authors examine how much the classification changes when noise is applied to four parameters — the convective fraction (F), convective area (A), characteristic length (L), and percentile (P). This is a valuable addition, as the manuscript presents a methodological approach. However, the analysis only assesses the method's sensitivity to errors in these parameters, without evaluating the typical magnitude of such errors in practice.
Based on the authors' response regarding the definition of the scene area, I am concerned that typical errors in F (and possibly P) could exceed the 10% threshold mentioned in the sensitivity analysis. I encourage the authors to either provide evidence that this is not the case or revise their methodology for defining the scene area. I also have follow-up (more minor) comments regarding the interpretation of P and the sensitivity to the tracking algorithm. Please see my detailed comments below.

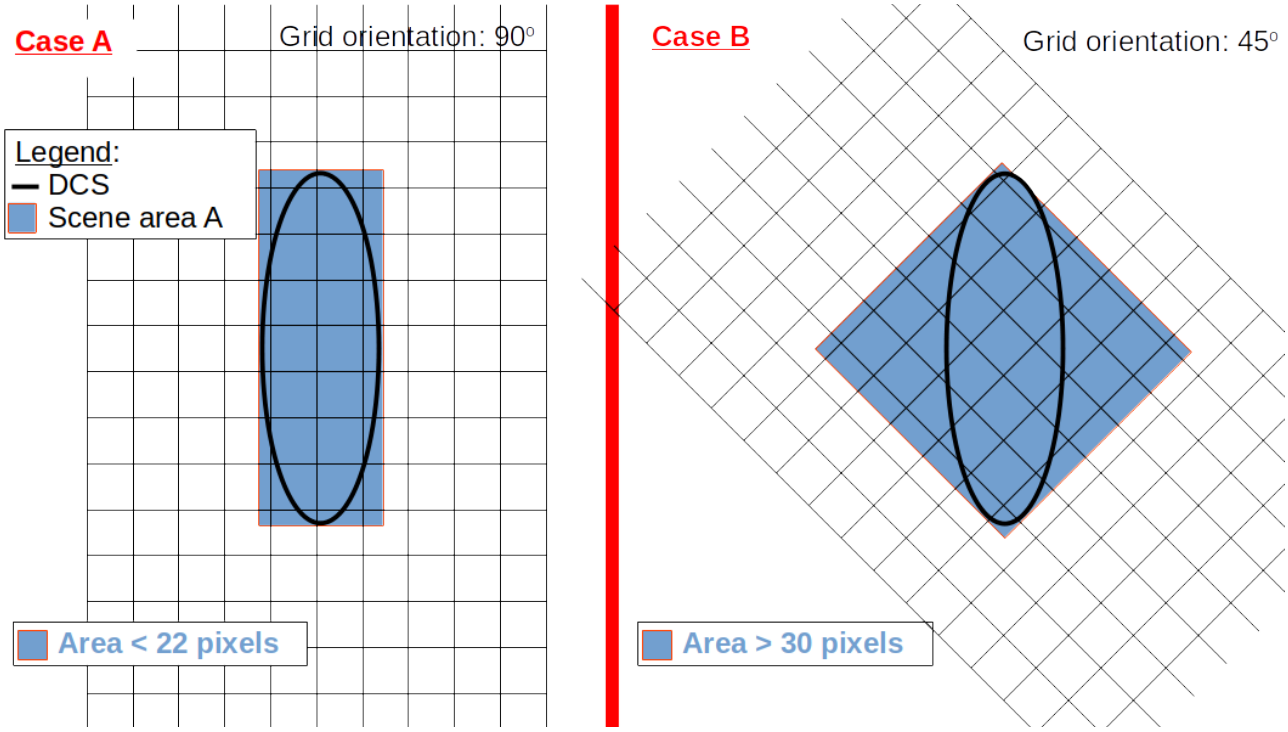
Scene area

I thank the authors for their detailed explanation of how the scene area is defined for each case. To my understanding, the scene area corresponds to the smallest rectangle that fully encloses the DCS (with one pixel of padding) and is aligned with the radar swath grid. I see two potential sources of bias in this definition: (1) the shape of the DCS, particularly its deviation from a rectangular form, and (2) its orientation relative to the grid.

Non-rectangular DCSs that are oriented at approximately 45° relative to the satellite grid will result in significantly larger scene areas than those that are rectangular and aligned with the grid. For the method to yield a meaningful classification of convective element arrangements within DCSs, such orientation- and shape-related biases are undesirable. While it is acceptable for F to be treated as an algorithmic variable representing the convective fraction, this remains valid only if there is a clear and consistent relationship between F and the actual convective fraction (i.e., larger F values should correspond to higher convective fractions). Based on my understanding of the current methodology, I am concerned that this relationship may not hold given the current definition of scene area.
To illustrate this point, consider a simplified example of an elliptical DCS (aspect ratio ≈ 1/3) observed by two radars with different grid orientations (cases A and B in the figure below). As it is the same DCS, the method should ideally yield similar convective fractions and classifications in both cases. However, as shown, case B produces a significantly larger scene area than case A — a difference of approximately (30–22)/22 = 36%. This increase in scene area could lead to a decrease in the derived convective fraction by about 26%, potentially resulting in different classifications for the same physical system depending solely on its orientation relative to the radar grid (as suggested by the additional sensitivity analysis).

Differences in DCS geometry would add on this orientation bias, making the convective fraction F less representative of the actual physical convective fraction. Large variations in scene area would also influence the P metric, which is likely to be larger for case B than case A, possibly exceeding the 15% threshold reported in the sensitivity analysis.
I encourage the authors to either (1) provide evidence that typical errors in F and P do not exceed the thresholds discussed in Figure B3, or (2) revise their method to reduce these biases. In addition, I recommend that the authors include a clear description of how the scene area is defined in the manuscript, as this aspect is critical to the methodology and

may substantially influence the classification results. Finally, the limitations associated with the definition of scene area should be discussed in the "Limits of applicability" section.



Figures:

Some figures still lack panel labels (letters) for reference. This is particularly the case for Figure 14, where letters are mentioned in the caption but do not appear in the figure itself. For clarity and consistency, I recommend adding panel letters (e.g., a, b, c, …) to all relevant figures, as these are preferable to references such as "left" and "right."

Interpretation of the percentile P:

- L. 465-466: "computing a percentile P, which quantifies the deviation of the real scene's spatial arrangement from a random arrangement (Figure 9, Red arrow in Figure 6)."

That wording implicitly assumes that higher percentiles mean stronger deviations from random.
But that is not necessarily true, low percentiles can also represent strong but opposite-direction deviations (more regular patterns instead of random). Please reformulate.

- L. 470-473:

For clarity, I copy the Authors' response to my comment on the previous round of revision.

"We don't use the clustering notion to illustrate the meaning of the value of P. Besides, rarity is somehow a vague notion that can lead to ambiguous interpretations. We don't mention it directly. Instead, we express the notion of how many times an event occurs within the total number of attempts in the context of the generated scenes, as the percentile computation. Besides, as described in the paper, our objective is broader than simply distinguishing between aggregated and clustered, as with metrics such as Iorg/Lorg, for example. Rather,we aim to quantify the deviation from randomness and the structuring of convective cores. Therefore, we believe that the original phrase is appropriate in this context."

I understand that the objective of the paper is broader than simply distinguishing between aggregated and clustered patterns, and the proposed method indeed achieves this, as indicated by the combination of P with the three other variables. However, my comment focuses specifically on the variable P.
The variable P measures the fraction of randomly generated scenes that have a lower L than the actual scene. Therefore to my understanding, the variable P, as defined in the current version of the manuscript, is more of a measure of the degree of clustering expressed in a statistical sense (on a probabilistic scale) than a measure of the deviation from randomness. This is because both very low and very high values of P indicate strong deviation from randomness, as noted above. Conversely, as P increases from 0 to 1, the spatial arrangement transitions progressively from more regular to more clustered scenes. If the Authors would like to put more emphasis on the deviation from randomness than the degree of aggregation, I would suggest to use a modified metric, for example 1-2.min(P,1-P), that will more closely quantify this aspect.

- <u>L.539-540</u> "On the other hand, the P distribution, indicating the probability of the characteristic length to differ from a random arrangement (Figure 12, d) is very much skewed to the highest values (around 1) with a very long tail. This indicates that only very few scenes can be associated with a randomly generated pattern."

The P distribution does not directly indicate a probability of differing from a random arrangement. First, very low P values are also indicative of deviation from randomness, as noted above. Second, such P-values (percentiles) only quantify the conditional probability of observing a value less extreme than the observed statistic (here L) under the null hypothesis (here, H0 = randomness (H1 = non-randomness)). Formally, they measure $\text{Prob}(L <= L_{obs} | H0)$, which has no direct relation with $P(H1|L<=L_{obs}) = 1 – \text{Prob}(H0|L<=L_{obs})$.

To illustrate, consider a bag containing many dice — some fair, some modified. You randomly draw one die, roll it, and obtain a six. Under the null hypothesis of a fair die, the probability of rolling a six is 1/6, and thus the P-value is relatively high (not indicative of deviation). However, this outcome alone tells us nothing about whether the die was fair or not. The P-value only quantifies how likely the observation is if the die were fair; it does not give the probability that the die is unfair given the observation.
Similarly, in your analysis, the P metric does not directly represent the probability of non-randomness. While the mapping between P and the probability of deviation from randomness may appear more meaningful when the reference (random) distribution is approximately bell-shaped, there is no one-to-one correspondence. The interpretation of P thus depends both on the shape of the null distribution and, in a Bayesian sense, on the prior probability of non-randomness, which can vary across cases.

- <u>L. 574:</u> "very unlikely to be randomly spatially distributed with high L values"

This statement remains somewhat ambiguous. It could be misinterpreted as referring to the unlikelihood of observing L relative to all high values in a random distribution, rather than the intended meaning: that the observed spatial pattern is specifically unlikely under the null hypothesis and falls on the side of clustered rather than regular patterns.

Sensitivity to the TOOCAN seed definition:

In the previous round of review, I shared my concern that the method used for detecting DCSs in the TOOCAN algorithm may restrain the diversity of spatial arrangements of DCCs, even when these DCCs are identified through radar measurements.

While the response provides a useful description of the TOOCAN procedure, it does not demonstrate that the IR-based identification of convective systems is fully independent of radar-based DCC detection (microwave measurements). Given that cloud-top temperature and convective precipitation intensity are often correlated (e.g., Arkin and Meissner, 1986), it seems surprising to me that the TOOCAN definition of DCS, which emerge from single convective seeds, would have no impact on the spatial arrangement of DCCs identified in radar.

Since convective seeds in the TOOCAN algorithm are defined as very cold brightness temperature minima, and only one seed is retained per DCS, one would expect the radar-identified DCCs to be spatially close to these IR-based seeds. This makes it difficult to conceive how the TOOCAN DCS definition could have no influence on the spatial distribution of radar-detected DCCs. A brief quantitative check (e.g., distance statistics between TOOCAN seeds and radar DCC centroids), or reference to an earlier validation study would help clarify this point.

Reference:
Arkin, P. A., and B. N. Meisner, 1987: The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western Hemisphere during 1982-84. *Mon. Wea. Rev.*, **115**, 51–74, https://doi.org/10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2.