Review

The authors investigate the robustness of 9 single number indices representing the degree of organization of convection (DOC) in a 10x10 degree domain. The convective regions are identified from thresholds on brightness temperature from geostationary satellite data (TOOCAN). The robustness of an organization metric is assessed from a sensitivity test on three categories of criteria; sensitivity to noise, sensitivity to changes in position and size of contiguous convective regions (8-connected convective gridboxes), and sensitivity to specifics of the dataset (mainly temporal and spatial resolution of sampling). The study addresses central questions surrounding the quality and versatility of different organization metrics and presents a new metric with high degree of robustness from these criteria on the specified domain. Furthermore, the study serves as a summary of current methods in assessing DOC and a great foundation for further improvements of organization metrics.

The sections of the paper are well structured, with informative methodology, illustrative examples, and concise explanation of the results.

The reviewer has a few minor concerns / questions that need to be addressed before approval of publication:

Conceptual

1. I_org assessment

It is interesting that the I_org metric is so sensitive in all categories of robustness criteria. I wonder to what extent the sensitivity relates to the number of objects considered. Is I_org robust when considering a large number of objects? How large fraction of the scenes considered in this study have less than 35 objects (where the I_org metric is no longer reliable)?

The I_org metric has been identified as unreliable for a small number of convective centroids (<20) before:
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019GL086927
Not in a systematic way as in the present study, but perhaps it should still be mentioned.

In several studies, I_org is used to assess aggregation from local minima in brightness temperature, which includes multiple convective centroids (convective cores) in a large convective object. With this approach the area of convective objects is implicitly included (as a large convective object introduce several closely connected convective cores), and the issue with a small number of convective centroids is addressed. Perhaps assessing I_org from the convective cores approach can be insightful to better understand the utility of this metric. Otherwise, perhaps just presenting the alternative approach and clarifying that the statement in the present study relates to the method applied to convective objects.
In this study:
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019AV000155
the authors mention that a similar result is obtain from using the object-based approach and the convective cores approach. Albeit, in that study the domain is the whole tropics, so the number of convective centroids is likely sufficiently large regardless.

2. ROME assessment

From working with the ROME metric to assess the tropical domain with DOC, I know that the metric is highly correlated with mean area of the domain. It was interesting to see that the impact of changes to the proximity of convective regions was so small. I suspect changes to the proximity of convective objects has a greater relative effect on the metric if the scenes are sub-sampled to scenes with similar energetic constraints (similar mean convective area, similar vertical velocity, similar mean precipitation etc.). Perhaps the distance component of ROME is also more significant when very large distances are considered (where the squared edge distance is a larger number of multiples of the smaller pair object). Further, considering that the proximity scaling applies to every object pair, a larger number of objects all moving together may also highlight the proximity scaling. I reserve the possibility, that the metric simply is unable to factor in the proximity of objects as it occurs in realistic settings, past the change in proximity which results in joining two objects. However, it would be interesting to test some of these considerations to highlight the limitations / utility of this metric.

3. ABCOP assessment

In the conclusion, it sounds like it is recommended to use this metric. While the metric captures changes in proximity, and is robust in most criteria, from what I understand, the metric does not correctly capture fundamental changes in aggregation; adding a random single convective gridbox increases aggregation, and merging objects decreases aggregation, which are the opposite signs of change from the conceptual interpretation. Perhaps it can be highlighted that these features make the metric unsound in this regard.

Technical

When testing condition 4 (changing the size of one object) - from the schematic in figure 6, it appears that the edge of the test object effectively move closer to the other objects when the area of the test object increases. Consequently, there will be a proximity component affecting metrics that depend on the edge distance between objects. To avoid a proximity influence when testing condition 4 on ABCOP and ROME, the test object could be uniformly extended Eastward in these cases.

In the methodology section it was mentioned that scenes with one or no objects were removed. What fraction of scenes contain only one object, and are they significant for describing degree of organization? Further, are they small objects or very large objects spanning most of the domain. Is it important for a metric to be able to handle a singular large object for the 10x10 degree domain?

In the introduction in line 27 there is a statement: 'Such studies have been so far performed only for example cases'. It could be nice with a reference to this. In the ROME metric paper, for example, there is an evaluation of different organization metrics for example cases, so maybe that paper can be referenced here:

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019JD031801

Language

There are some instances of grammatical errors, and some typos. Here are a few I noticed: [line: 'instance' - suggestion]

Line 16: 'Convection is the main responsible' - mainly responsible

Line 26: 'organization has not a rigorous' - organization does not currently have a rigorous

Line: 33: 'Moreover, it briefly recalls the convective organization indices' - Sect. 2.2 (it was hard to interpret what 'it' was referring in this line)

Line 212: 'more peaked' - exhibit a narrower peak around zero

Line 435-436: 'ROME is the index that depends less on' - is the least sensitive to

Line 454: 'OIDRA is also the index that is less sensitive' - the least sensitive

Line 459: 'It increases when an object is increasing' - 'when an object is increasing in size' (I think there are a couple more instances similar to this)

Line: 460: 'larger than the average mean size' - repetition, so keep one

Line 200: 'Move from 0.91 and 0.78' - Move from 0.91 to 0.78'

Line 303: figure 8 description: 'with a 3 times worst resolution' - coarse-grained by a factor of 3

Line 435: 'ROME depends very much more on the size' - is highly dependent on the size