

## **Review of “Convective Organization and 3D Structure of Tropical Cloud Systems deduced from Synergistic A-Train Observations and Machine Learning” by C. Stubenrauch et al.**

Atmospheric Chemistry and Physics

December 8, 2022

Stubenrauch et al. apply a machine-learning method to various satellite datasets to build a complete 3D description of upper tropospheric clouds. Their method fills gaps in the observational record due to sampling limitations. This allows the authors to examine metrics that are more closely related to physical processes than one can obtain with the raw satellite data alone. The technical analysis is very well done. I applaud the authors for carefully implementing the machine learning method and documenting how different necessary but subjective choices, such as thresholds and variable definitions, affect the results. However, the section that presents scientific results lacks coherence, and at times it reads like a list of findings that do not always connect to one another in an obvious way. This makes the writing unclear. I think the authors could improve the presentation of the results if they decide on a clear purpose for the paper and then write a more coherent presentation that fits that purpose. This could require some restructuring of the paper, so I recommend *major revision* for the paper. I would be happy to recommend this work for publication if my comments are addressed.

### **General Comments**

- Section 3 presents a list of findings, first about relationships between anvil-cloud properties and deep convection, then proxies for the life cycle of mesoscale convective systems (MCS), then differences between MCS over warm sea-surface temperature and MCS over cool sea-surface temperature, then convective aggregation across the entire tropical belt. I was expecting a coherent link from one set of findings to another, and I didn't see this in the paper. I think the paper would be much clearer if the results were presented in a logical order in which each piece of analysis builds on the one before it, or if the authors explicitly state that they are investigating aspects of tropical convection that are related but do not necessarily build on one another as they are presented in the text. To accomplish this, it would help if the authors first decide the purpose of the paper. Is the purpose to document the machine-learning dataset and present some preliminary findings that can potentially be used for process-oriented evaluations of numerical models? (This is emphasized in the introduction as a key motivating factor.) If so, then please state in each subsection of section 3 that you are documenting distinct aspects of tropical convective clouds for potential use in model evaluations, and that the findings from one subsection do not directly build on the subsection before. Alternatively, is the purpose of the

paper to show new evidence about convective aggregation and climate? About half of the abstract text is devoted to this topic, but no clear conclusions are reached about this because the results are very sensitive to the particular aggregation index that is used. Furthermore, if the authors want to emphasize the convective aggregation results, then why use so much discussion on process-oriented behavior in Section 3.2 without explaining the link between these results and convective aggregation in Section 3.3? Overall, I could not identify the key messages that the authors want their readers to take away from the paper, other than the fact that they have produced a new dataset. This lack of a message is confusing. I recommend that the authors decide the purpose of the paper and the most important things that they want readers to learn, then revise the manuscript accordingly.

- It would help to provide a table that states the meaning of all variables and acronyms. It was hard for me to keep track of everything as I was reading, so I had to frequently stop and look up the meaning of different variables. I think it would help the reader if the authors provide a table with all of the definitions in one place.

### Specific Comments and Typos

- Line 14: change “allows” to “allows us”
- Line 39: change “Pendergrast et al.” to “Pendergrass” (only one author)
- Line 57: perhaps change “consolidating the hypothesis” to “supporting the hypothesis”
- Line 71: change “become” to “have become”
- Line 122: change “skewedness” to “skewness”
- Line 132: perhaps change “we use randomly chosen 80% of the dataset” to “we use 80% of the dataset chosen at random”
- Line 153: it would help to define DZ in the text here. DZ is defined in Table 1, but by the time I read this sentence I did not remember what DZ stands for, so I had to pause and look it up. It would be easier for the reader if you define DZ here because this is the first place where it is referenced in the text.
- Line 200: remove commas after “both” and “classification”
- Line 238: change “as convective core” to “as a convective core”
- Line 240: remove comma after “UT clouds”
- Line 314: change “area of the larger area” to “area of the larger object”?
- Line 318: “we highlight results on the ML derived variables by investigation relationships”. Relationships between what? Please specify.
- Line 328: change “grid cells which include heavy precipitation” to “grid cells that include heavy precipitation”
- Line 416: change “the less sensitive” to “the least sensitive”
- Line 436: change “lorg is the metric that is less related to these variables” to “lorg is the metric that is least related to these variables”

- Line 547: the README page for the AIRS data from the CIRS L2 webpage is not working. This makes it difficult for others to obtain and use the data. Are the authors able to fix this?
- Figure 1: the legend of the bottom left panel partly covers the curves in the plot. This is confusing because the text mentions that the curves have a bimodal distribution (line 173), but one of the modes of the distribution is covered by the legend in Fig. 1. Please revise Fig. 1 so that all curves are not obscured by the legends.
- Figure 9: Only three variables are listed in the caption (“Mature MCS system properties as function of their convective depth, given by the minimum temperature within the 735 convective cores: ratio of areas with thin cirrus over total anvil, thick anvil emissivity and ratio of strong rain area over rainy area (right) within mature convective systems as a function of convective depth, given by decreasing minimum temperature within the convective cores.”) However, four subplots are shown, so some description is missing. Please consider adding labels (a), (b), (c), etc. to the subplots and including these labels in the caption. This will clarify the figure because the variable names on the vertical axes are not stated in the caption.
- Figure 13: Change “high troposphere” to “upper troposphere” on the figure for consistency with the caption. Alternatively, you could state the pressure ranges on the figure so the reader can see the levels of each subplot without looking back and forth at the caption (e.g. replace “high troposphere” with “100-200 hPa” on the figure, and so on).