

Author response to Referees 1 and 2:

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.

General Comments of Referee 1:

- 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 seasurface 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.

and of Referee 2:

This study aims to study convective organization and dynamics of upper tropospheric clouds over the global tropics using satellite datasets and ERA reanalysis products. The authors apply novel machine learning method to merge these datasets in order to fill in the gap in these dataset to obtain a fully-connected understanding of tropical convective systems. Overall, I commend the authors for writing such a comprehensive manuscript and I really liked the scientific analyses presented here. I specifically liked the authors efforts to connect dots between convection, cloud systems, and organization using multiple metrics so as to provide a better idea of tropical convection. However, I think that manuscript needs some restructuring as currently it looks like a lot of ingredients are mixed together but it didn't result in an edible dish. I got lost and a bit confused in section 3 as the authors jump between multiple thoughts and I couldn't connect the dots well. Therefore, I recommend **major revision** to the manuscript.

The authors thank both reviewers for their very constructive comments. They both agree that the description of the results and conclusions and abstract miss clarity. Indeed, since this dataset was also new for the authors, they wanted to show as many results as possible, but this made the manuscript difficult to follow and did not give clear key messages. Therefore the authors have completely revised section 3 (Results), have added a clear goal of this article into the introduction and have shortened and clarified the abstract and section 4 (conclusions and Outlook), so that the key messages come out clearer. In the following we copied the changes made to the manuscript into this document.

- We have added in the introduction just after the first paragraph the goal:

The goal of this article is to present a coherent long-term 3D dataset which describes tropical UT cloud systems and which can be used on one hand for a process-oriented evaluation of convective parameterizations in climate models and on the other hand for the study of convective organization.

- We have also reformulated the last paragraph of the introduction:

Apart from the conclusions and outlook given in Section 4, the article is divided into two main sections: Section 2 describes the data, methods and evaluation and Section 3 highlights scientific results which show the applicability of these newly derived variables.

Section 2 first describes the collocated data, the neural network development as well as an evaluation of the predictions on the collocated data. In addition, it presents the creation of the 3D dataset containing the additional variables (2.3) and the cloud system reconstruction (2.4). The last subsection (2.5) gives a short overview of existing convective organization indices and proxies for defining the convective objects. Section 3 first shows the coherence of these ML-derived properties, in particular the rain intensity classification, using the complete 3D dataset (section 3.1). Then, in combination with a cloud system analysis, section 3.2 presents the MCS properties with respect to their life cycle stage and their convective depth. The last subsection (3.3) explores tropical convective organization: we compare different proxies for convection and resulting indices of convective organization, by investigating annual cycle and inter-annual variability. The latter is small over the considered time period (2008 – 2018), but we find interesting geographical patterns in changes of radiative heating rate fields in relation to the tropical convective organization.

- We have also shortened the last part of the abstract:

... This rain intensity classification is more efficient to detect large latent heating than cold cloud temperature. In combination with a cloud system analysis we found that deeper convection leads to larger heavy rain areas and a larger detrainment, with a slightly smaller thick anvil emissivity. This kind of analysis can be used for a process-oriented evaluation of convective precipitation parameterizations in climate models. Furthermore we have shown the usefulness of our data to investigate tropical convective organization metrics. A comparison of different tropical convective organization indices and proxies to define convective areas has revealed that all indices show a similar annual cycle in convective organization, in phase with convective core height and anvil detrainment. The geographical patterns and magnitudes in radiative heating rate inter-annual changes with respect to one specific convective organization index (I_{org}) for the period 2008 to 2018 are similar to the ones related to the El Niño Southern Oscillation. However, since the inter-annual anomalies of the convective organization indices are very small and noisy, it was impossible to find a coherent relationship with those of other tropical mean variables such as surface temperature, thin cirrus area or subsidence area.

- We have rewritten section 3 (results), with a short introduction about the structure of this section:

3. Results

As application examples we highlight results from analyses using this long-term 3D dataset. We particularly concentrate our interest on the ML-derived rain rate indicator. Section 3.1 shows the coherence of this newly derived variable. The cloud system approach enables us to study the behaviour of the MCSs with respect to their life cycle stage and convective depth. This process-oriented analysis presented in section 3.2 can be used to evaluate parameterizations in climate models (Stubenrauch et al., 2019). In section 3.3, we show results concerning mesoscale convective organization. Mesoscale convective organization has been identified by larger and higher systems, which also live longer than unorganized systems (e. g. Rossow and Pearl, 2007; Takahashi et al., 2021), and they also lead to increases in tropical rainfall (e. g. Tan et al., 2015). We first compare convective organization indices derived from objects defined by strong rain and by cold cloud temperature and then investigate changes in geographical patterns of radiative heating with respect to one of these indices (I_{org}).

- We have changed the subtitle of section 3.1 from 'Properties of UT clouds' to:
3.1 Coherence of ML-derived rain intensity classification

- We have taken out the titles of the two subsections ('Behaviour of precipitating areas' and 'Behaviour of convective core and anvil properties') of section 3.2 and have really focused on the 'process-oriented behaviour of mesoscale convective systems' for the evaluation of climate models, with merged figures, so that this subsection only includes two figures instead of four. We hope that section 3.2 is much clearer now:

3.2 Process-oriented behaviour of mesoscale convective systems

The cloud system concept described in section 2.4 permits to link the convective core and anvil properties: The fraction of the convective core area within a cloud system indicates the life cycle stage (e. g. Machado et al., 1998), with a large fraction indicating the developing stage and a decreasing fraction during dissipation. Once the systems have reached maturity, the minimum temperature within a convective core is a proxy for the convective depth.

According to Takahashi et al. (2021), using a convection-tracking analysis on data from Intergrated Multisatellite Retrievals for GPM (IMERG), the fraction of precipitating cores (adjacent grid cells with a rain rate > 5 mm/hr) within precipitation systems (adjacent grid cells with rain rate > 0.5 mm/hr) first increases and then decreases during the evolution of these systems. The maximum of the strong rain area relative to the whole precipitating area as well as the maximum and average intensity of the precipitation increase with the life time of the systems. This behaviour was also found by Roca et al. (2017).

Our data do not provide the absolute system life time, but the convective core fraction within a system indicates the maturity stage in a normalized life cycle. Figure 7 presents the statistical evolution during the life cycle of a) the precipitating area relative to the whole MCS area and b) the strong rain area relative to the precipitating area, for single core MCSs. As the rain rate classification was obtained per CIRS footprint, a grid cell of $0.5^\circ \times 0.5^\circ$ can be declared as precipitating by using different thresholds on the fraction of footprints with rain rate > 0 mm/hr. The same applies for grid cells including strong rain. Results using three different thresholds to define the precipitating and strongly precipitating areas are compared. For all thresholds, the precipitating area is very large in the beginning of the life cycle, when the anvil is just developing and then decreases, while the fraction of strong rain stays constant until the anvil reaches 40% of the system size and only then decreases. With our coarse spatial resolution we did not see the increase in strong rain after the developing stage, which has been observed by Fiolleau and Roca (2013) and Takahashi et al. (2021), using data with better time and space resolution. This means that we miss the very first development of the convective tower itself, as can also be seen in Figure 7c, which presents the evolution of the convective core size and the convective core top height. The latter varies much less than the convective core size, with an average of already 12.8 km for a convective core fraction close to 1. So due to the coarse spatial resolution and considering only high-level clouds, we start to identify the systems when they are already near to their maximum height, which is attained just before the decrease of the heavy rain portion.

The core size increases rapidly and then stays stable until dissipation of the system. We identify MCS maturity by a core fraction between 0.2 and 0.4, because by then the core size has attained its maximum.

Once the convective systems are mature, we can study their properties with respect to their convective depth: Figure 8 presents, as a function of the minimum temperature within the convective cores, a) the strong rain area relative to the precipitating area (again considering three thresholds), b) the volume of the thick anvil ($\tau_{\text{thick}} > 0.5$) relative to the volume of the convective core, which is a proxy for detrainment, and the emissivity of the thick anvil and c) the size of the surrounding thin cirrus relative to the total anvil size as well as the 50% warmest surface temperature underneath the MCSs. We deduce that deeper convection clearly leads to 1) larger areas of heavy rain within the precipitating

areas, in agreement with earlier studies, 2) to a larger volume detrainment but with a slightly smaller emissivity and 3) to more surrounding thin cirrus. From Figure 8 we also conclude that deeper convection occurs in general in the warmer regions of the tropics, as expected.

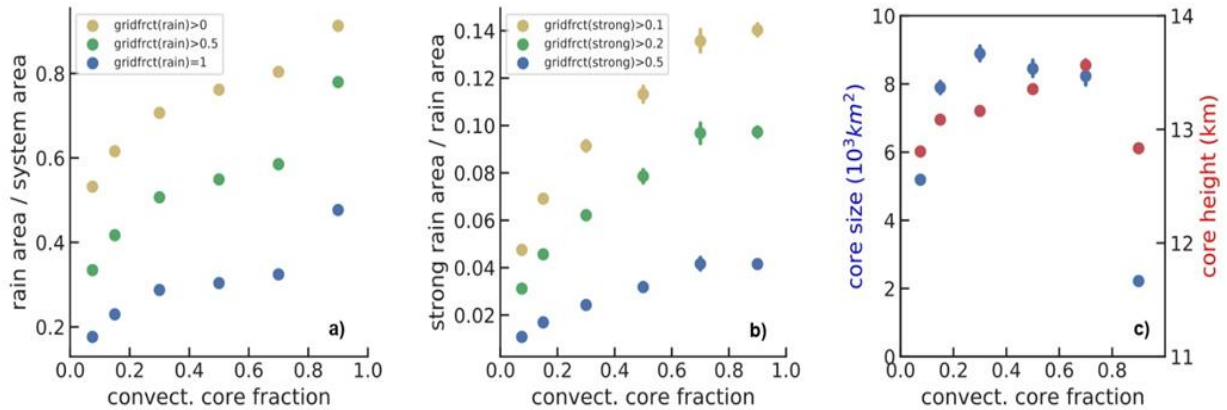


Figure 7: MCS properties as function of their life cycle stage, given by fraction of convective core area within the system (1 corresponds to developing phase, with no anvil and 0.1 to dissipating stage). Only cloud systems with a single convective core are considered: a) ratio of precipitating area over MCS size, b) ratio of strong rain area over precipitating area and c) size and top height of the convective cores. For a) and b) different thresholds on the rain fraction per grid cell are compared. The condition on strong rain also includes the condition that at least 50% of the grid cells are covered by any rain. Statistics combines observations at 1:30AM and 1:30PM, for 2008-2018.

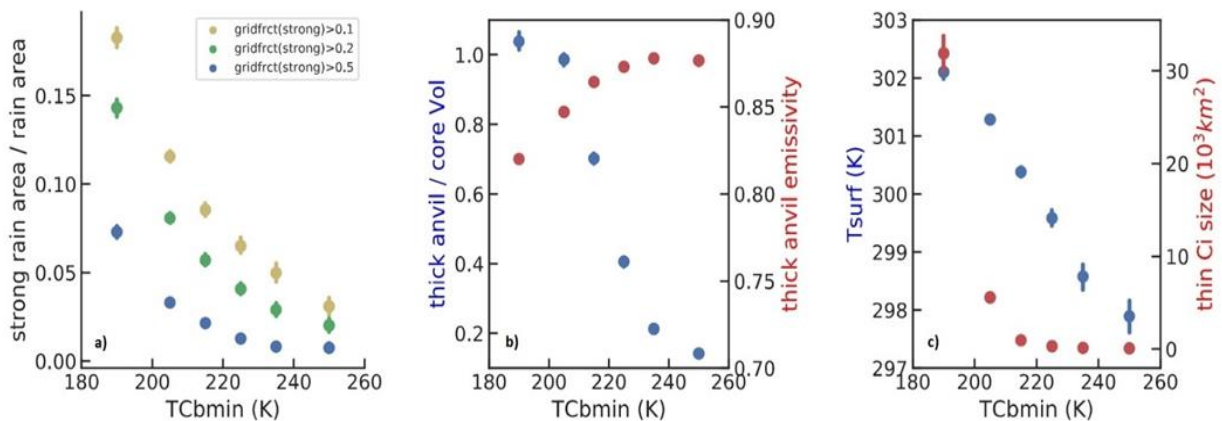


Figure 8: Mature MCS system properties as function of their convective depth, given by decreasing minimum temperature within the convective cores: a) ratio of strong rain area over precipitating area, b) volume detrainment and thick anvil emissivity and c) 50% warmest surface temperature underneath the MCSs and ratio of thin cirrus over total anvil area. For a) different thresholds on the fraction of strong rain per grid cell are compared and at least 50% of the grid cells are covered by any rain. Statistics combines observations at 1:30AM and 1:30PM, for 2008-2018.

- Section 3.3 is indeed not linked to section 3.2, but shows just another example of application of our new 3D dataset. We have stated this in the introduction and in the paragraph following the title of section 3 (see above).

- We have also reformulated the first part of section 3.3 (comparison of convective organization indices by means of their annual cycle) to make it clearer for key messages:

3.3 Tropical convective organization

In this section we demonstrate the usefulness of this new dataset by analyzing the convective organization in the tropics. We compare results using various metrics of convective organization and proxies to define convective objects, as described in section 2.5. A spatial resolution of 0.5° relates to an organization of MCSs at a scale which is more linked to the large-scale environment and circulation.

We first consider the annual cycle of convective organization and then highlight an application on inter-annual variability.

Figure 9 presents the annual cycle of a) I_{org} , b) ROME, c) COP, d) total area and e) mean area of the convective objects and f) their number. We also investigate different variables to define these objects, in particular precipitation intensity (given by ML-derived rain rate indicator) of UT clouds ($p_{cid} < 350$ hPa) and cold cloud temperature ($T_{cid} < 230$ K) of opaque clouds ($\tau_{cid} > 0.95$). The latter definition is similar to $TB < 230$ K, but without any contamination of colder thinner Ci. Since ROME is strongly related to the size of the convective objects, we have also computed the annual anomalies of the three indices by using only 2% of the largest precipitation intensities for constructing the convective objects. This corresponds to comparing the indices for a constant total area of convection. A similar approach was undertaken in a study by Blaeckberg and Singh (2022), using the precipitation intensity and ROME as proxies for convection and convective organization, respectively.

All three indices reveal a clear annual cycle of convective organization, with a minimum in April and November and a maximum in July / August and in January / February, though with differences in magnitude and width of the oscillations due to the choice of proxy for convective area. While the amplitude of the annual cycle is the smallest for I_{org} (0.04), the seasonal anomalies of COP are the less sensitive and those of ROME the most sensitive to the choice of proxy. The latter look very similar to those of the mean area of the convective objects (with a correlation coefficient larger than 0.9). Thus the seasonal anomalies of ROME primarily reflect the ones of the mean areas of the convective objects. We also observe in Figure 9e, that the minima and maxima in the annual cycle of the mean area of objects with intensive rain are shifted compared to those with cold opaque cloud. When using a fixed total area of intense precipitation, the magnitude of the seasonal anomalies is much smaller and the shift in behaviour compared to using the proxy of cold opaque clouds disappears.

The annual cycle of the total area of convective objects can be reconstructed by the one of the mean area times the one of the number of the convective objects (Figures 9 d-f): The relative flatness of the seasonal cycle of the total area of cold cloud objects can be explained by a nearly opposite seasonal cycle in their mean size and number, whereas for intense precipitation their cycles are in phase which then leads to a pronounced cycle in their total area.

The absolute values of the convective organization indices, presented in Figure S3 of the supplement, depend more strongly on the proxy used to define the convective objects: The absolute maximum of COP is the same for all proxies during boreal summer, while for other seasons COP, like I_{org} , is larger when considering precipitation intensity. While ROME primarily reflects the mean area, ABCOP reflects the total area of the convective objects (both with a correlation coefficient larger than 0.8). Whereas the mean area of the convective objects is clearly an indication of meso-scale convective organization, the total area may not be directly linked to convective organization, only perhaps if one considers local regions. I_{org} , only considering the distance between convective objects, seems to add another aspect. I_{org} indicates a more organized convection when precipitation intensity is used instead of cold cloud temperature to identify the convective objects, though the total area of intense precipitation areas is smaller than the one of the cold cloud objects. The difference in the absolute peak values of the I_{org} and COP anomalies between boreal summer and boreal winter may be explained by regional shifts in intense precipitation occurrence, as shown in Figure S4 and in agreement with earlier results (e. g. Berry and Reeder, 2014).

Figure 10 presents the annual cycle of MCS properties discussed in the earlier sections: The core height of the MCSs, the anvil horizontal detrainment estimated by the ratio of anvil over convective core size and the vertical extent of the anvil (Figure S5) are in phase with the annual cycle of I_{org} and COP. This shows that in seasons with larger tropical convective organization the MCSs have in general higher cloud tops and larger anvils. The fraction of single core MCSs and the ratio of thin cirrus over total anvil size are in opposite phase. The first confirms that convective organization corresponds to multi-core MCSs. The latter may be not directly expected as we have seen that this ratio increases with the height of the MCSs (convective depth). On second thought it may be explained by the fact that clustering of convective systems leaves less space for the thin cirrus between them. It is interesting to note that the annual cycle of the relative subsidence area (clear sky and low-level clouds) is also in phase with the

one of I_{org} and COP. This means that larger heating of the upper and middle tropospheric heating by more organized MCS leads to more cooling of the lower troposphere in the subsidence regions, which has recently been found by Stubenrauch et al. (2021).

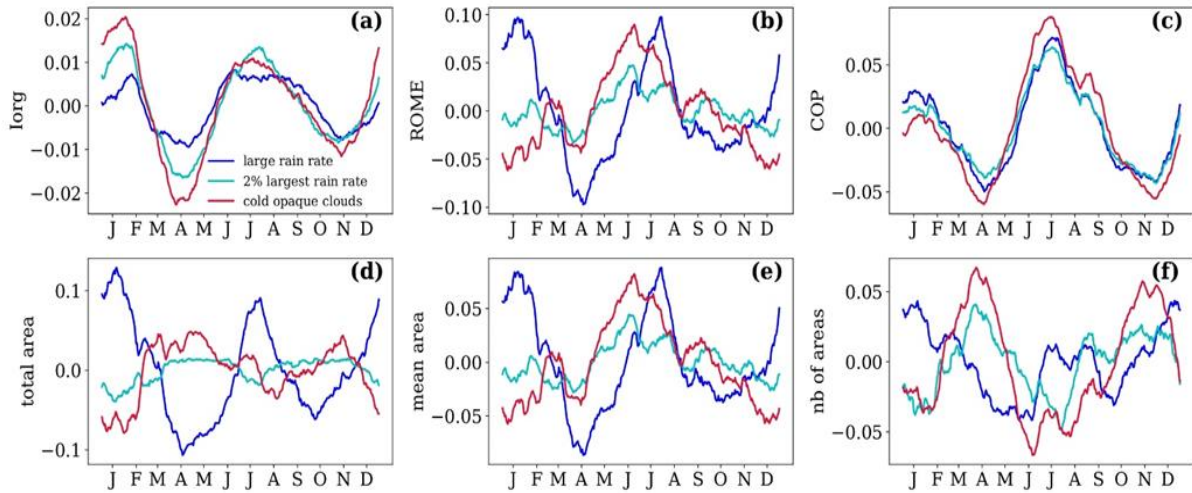


Figure 9: Annual cycle of a) I_{org} , b) ROME, c) COP, d) total area of convective objects, e) mean convective area and f) number of convective areas. These areas are built from grid cells covered by at least 90% UT clouds, with rain rate indicator > 2 (dark blue), $T_{cld} < 230$ K and $\epsilon_{cld} > 0.95$ (red), or using the 2% largest rain rate indicator values (cyan). The latter leads to a constant total area of convection. Monthly statistics of UT clouds averaged over four observation times from 2008 to 2018.

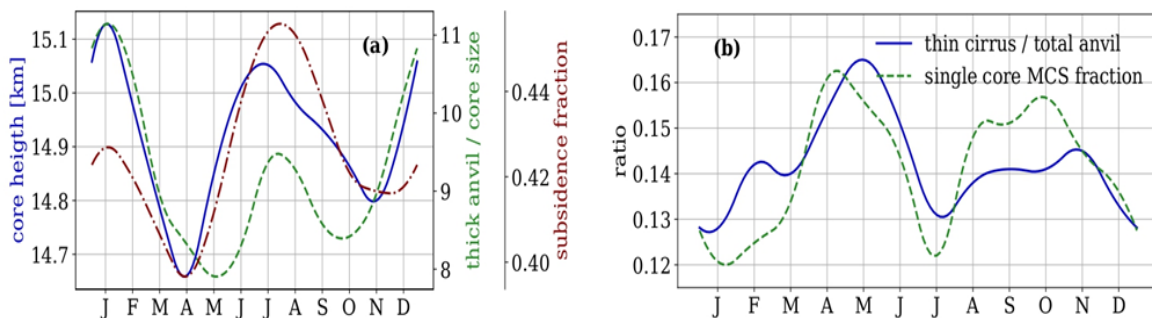


Figure 10: right: Annual cycle of a) MCS core top height (blue), horizontal detrainment (green) and fraction of subsidence area (given by clear sky and lowlevel clouds) over the tropics (red) and b) ratio of thin cirrus over total anvil size (blue) and fraction of single core MCSs (green). Monthly statistics averaged over four observation times from 2008 to 2018.

- We have also added a figure with histograms at the end of section 3.3 (analysis of inter-annual variability), again for a clearer key message:

...

Whereas the geographical patterns of the derivatives of heating / cooling with respect to I_{org} show a coherent picture, we did not find any correlation between the very small inter-annual anomalies of I_{org} (shown in Figure 12) and the ones of the tropical means of different variables like surface temperature, thin cirrus area and subsidence area. The correlations depend on the proxies for the definition of the convective areas and in particular on the metrics for convective organization. Already the time series of the inter-annual anomalies of the different indices have a different behaviour as can be seen in Figure S8 in the supplement. We have also investigated tighter thresholds on the variables which define deep convection (like rain rate indicator > 2.5 or $T_{cld} < 210$ K), however we are left with only about 0.5% total area, which increases the noise level. In addition, we found that the results also change when we exclude objects with the size of only one grid cell (not shown), as already pointed out by Jin et al. (2022). Therefore we do not consider it meaningful to use the discussed convective organization indices for an estimation of tropical mean changes with respect to changes in convective organization.

While we have seen that the convective organization indices vary much more seasonally than inter-annually, Figure 13 suggests that the difference of the density distributions of convective core height

and strong rain area within the MCSs between April and July or between cool years (2008/11) and warm years (2015/16) is of the same order, with a shift towards higher core height and a longer tail in strong rain area. However, the size distributions of the MCSs are similar. The tail in the mean area of strong precipitation within the MCSs is clearly larger in the case of warmer years. This indicates that a shift in tropical surface temperature changes only a small part of the MCSs, with more extreme values. Such a behavior cannot be identified using a convective organization index computed over the whole tropics.

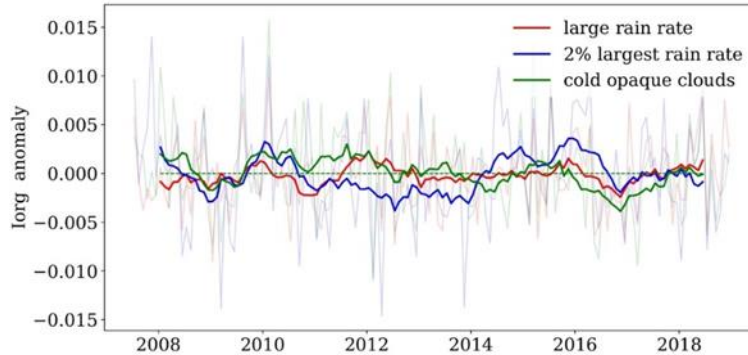


Figure 12: Time series of deseasonalized monthly anomalies of I_{org} , using different proxies to define the convective areas. The deseasonalization was done by computing 12-month running means. The monthly anomalies are shown in light grey.

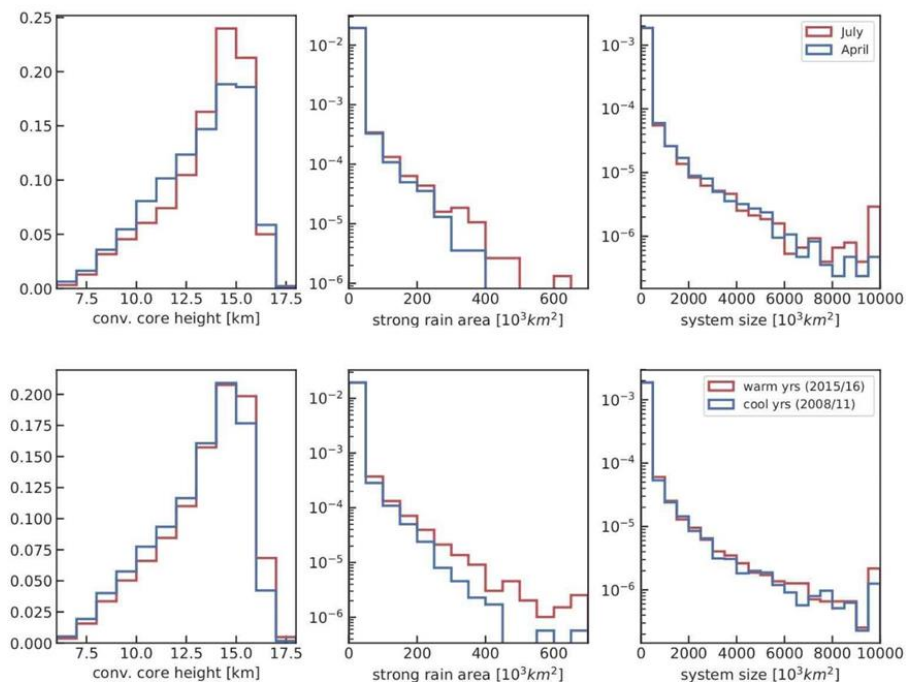


Figure 13: Density distributions of MCS system properties, comparing April and July (top) and cooler and warmer years (bottom): height of convective cores (left), size of areas with strong precipitation (middle) and system size (right). Statistics combines observations at 1:30AM and 1:30PM, for 2008-2018, for MCSs with core fraction > 0.1 .

- We have shortened and rewritten parts of section 4, so that key messages come out clearer, which are pointed out in *italic*:

4. Conclusions and Outlook

We have presented a methodology to extend spatially and temporally information on the cloud vertical structure and precipitation derived from active lidar and radar measurements of CALIPSO and CloudSat missions. This new approach made use of CIRS data obtained from advanced IR sounder measurements of AIRS and IASI combined with ERA-Interim reanalyses and machine learning technologies using ANN. *The resulting 3D dataset of UT cloud systems, covering 2008 to 2018, together with a similarly produced dataset of radiative heating rates (Stubenrauch et al., 2021), can be used to improve our understanding*

of the relationship between tropical convection and resulting anvils and how they are impacted by and feed back to climate change.

Though the uncertainties in the predicted variables and classifications are relative large (with an accuracy of about 65 to 70% for the rain intensity classification), this new dataset allows to study their *horizontal structures on specific snapshots in time*. For a complete instantaneous coverage, necessary to compute indices of tropical convective organization, the gaps between the orbits have been filled iteratively with the four observations per day of AIRS and IASI data, starting with those closest in time (already leading to 90% coverage). We have demonstrated that *the newly developed precipitation intensity classification is slightly more efficient to detect large latent heating and therefore deep convection compared to the cold cloud temperature*.

The cloud system approach developed by Protopapadaki et al. (2017) has been slightly modified, and the normalized vertical extent obtained from the ML approach has been employed to slightly improve the identification of the convective cores, in particular in the cooler tropical regions. *The cloud system concept allows a process-oriented evaluation of parameterizations in climate models. In agreement with earlier studies (e. g. Schumacher and Houze, 2003; Roca et al., 2014, Takahashi et al., 2021), we found that deeper convection leads to larger areas of heavy rain. These results also confirm the quality of the ML-derived precipitation rate classification. With increasing convective depth mature MCSs also show an increase in volume detrainment, while the anvil emissivity slightly decreases.*

Moreover we have shown the *usefulness of our new dataset by investigating convective organization metrics*. By comparing different organization indices (I_{org} , COP, ABCOP and ROME) and proxies to define convective objects, we have shown that the indices indicate a similar annual cycle of convective organization. *However, ABCOP and ROME are strongly correlated to the total and mean area of the objects, respectively. While the mean area of the objects is certainly an indication of convective organization, their total area at tropical scale seems to be less linked to organization.* The index I_{org} , which only considers the distance between convective objects, seems to add another information. The core height of the MCSs and their anvil detrainment are in phase with the annual cycle of I_{org} and COP, as well as the relative subsidence area. This shows also a link between the MCSs and subsidence areas. It is interesting to note that the annual cycles of the total area of cold cloud objects and of intense precipitation objects are very different. This can be related to a nearly opposite cycle in their mean size and number for the first, and to a cycle in phase for the latter.

Changes in gradients of tropospheric radiative heating relate to changes in atmospheric circulation. The geographical patterns and magnitudes in radiative heating rate changes with respect to I_{org} are similar for both proxies, but slightly larger for strong precipitation areas. This may be expected as *intense precipitation should be a more direct proxy for convection than cold cloud top. Furthermore the HR pattern changes are similar to the ones related to ENSO during this period.*

However, *the time series of the inter-annual anomalies of convective organization strongly depend on the convective organization metrics, and correlations between these anomalies and those of tropical means of different atmospheric variables do not show consistent results.* The tail of the distribution of strong rain areas seems to be more related to warmer tropics than the indices themselves. Therefore *one has to be careful using only one of these organization indices and proxies to study climate change.* More detailed studies are necessary to show the behaviour of these indices with spatial resolution and domain size.

This data base of UT cloud systems, their vertical structure and precipitation areas is being constructed within the framework of the GEWEX (Global Energy and Water Exchanges) Process Evaluation Study on Upper Tropospheric Clouds and Convection (GEWEX UTCC PROES) to advance our knowledge on the climate feedbacks of UT clouds. It will be made available within this year via <https://gewex-utcc-proes.aeris-data.fr/>. For the future it will also be interesting to use this dataset for the study of cold pools, using data of Garg et al. (2020).

In order to continue this dataset beyond 2018, we are now preparing a new version of CIRS data, using ERA5 (Hersbach et al., 2020) instead of ERA-Interim ancillary data, and newly calibrated AIRS L1C radiances (Manning et al., 2019) as input.

Specific comments of Referee 1:

- 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.

- **As some variables (in particular ‘cloud fuzziness’) have been only explained in section 3, which made it more difficult to follow the discussion of the results, we made sure that all variables are explained and named in section 2.1; the variables derived from the target variables are now also listed, together with their computation from, in Table 1. In the discussion of the results we now use the variables name instead of the formulas.**

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”

All taken into account

- 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.

Done, and we also write in section 3 ‘vertical extent’, so that the discussion is easier to follow

- 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”?

All taken into account

- Line 318: “we highlight results on the ML derived variables by investigation relationships”. Relationships between what? Please specify.

We changed this sentence to:

As application examples we highlight results from analyses using this long-term 3D dataset.

- 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”

All taken into account

- 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?

Thank you very much for pointing this out! The first author has contacted the French Data Centre AERIS, and they have repaired this (this is unfortunately not the first time that this happened...), and we have also put a readme for the available variables directly at <https://cirs.aeris-data.fr/>)

- 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.

Taken into account

- 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 7 has been removed as a separate figure and Figures 8 and 9 (now Figures 7 and 8) are new (see above)

- 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).

Taken into account

Specific comments of Referee 2:

1) First, some figures are highly pixelated (Figures 3, 7, 10, 11) and therefore I recommend providing high resolution version of all the figures.

Taken into account; the author have had problems to include the figures in pdf format into the word text, which reduced the initial resolution. All figures will be separately provided in good resolution (This will be checked also by the editors before publishing)

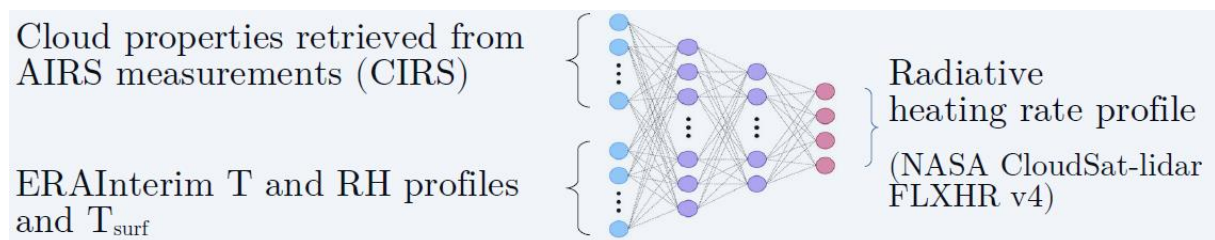
2) I was wondering that whether the authors performed any tests regarding the variables of interest? For e.g., did the authors test other atmospheric state variables to predict the cloud properties? If not, can the authors comment on how they came up with these input/output variables. It would also be useful if the authors can comment on relative importance of each input variables, if possible. This will aid the readers to choose which variables are more important and significant for cloud properties and rain classification.

The authors have done this kind of test for the radiative heating rates (published in Stubenrauch et al., 2021, <https://doi.org/10.5194/acp-21-1015-2021>): one surprising result was that once basic surface, atmospheric and cloud properties were included, the results were already very good, and by adding other variables, the results improved only by 5 to 10% (section 3.1 of the article from 2021). We have done similar tests for the variables published in this article, with a similar results. For the choice of the input variables, we started with the CIRS cloud properties, and added then atmospheric and surface properties from ERA-Interim. The latter have been used in the CIRS cloud property retrieval, therefore the whole input variables are coherent. We found out that the scenes for which

the models were developed are much more important than the detailed choice of the input variables (once the basic variables were included). We also tested the resolution of the atmospheric profiles, by using once 10 and once 20 layers, but again there was not much difference in the results.

3) I suggest adding a schematic of the designed ANN as it would be easier to visualize the connected network and their hidden layers.

This ANN network approach is actually quite standard, so we do not want to add an additional figure on it. Just for your information, it would look a bit like this:



4) Previous studies (de Szoeké et al. 2017, Garg et al., 2020) have shown that cold pools have a strong relationship with tropical convection and specially with precipitation and convective areas. I highly recommend connecting these results with global tropical oceanic cold pool properties shown in these studies as it will help the authors connect dots between clouds, precipitation and convection over the global tropics.

Thank you very much for these articles! Indeed, in the future we want to combine our 3D dataset with others, in order to understand better the different interactions. We have added a sentence in section 4, and the first author also shared this information within the GEWEX UTCC PROES group.

For the future it will also be interesting to use this dataset for the study of cold pools, using data of Garg et al. (2020).

5) Line 390: Correct "MSCs" to "MCSs".

6) I suggest correcting TB to T_b throughout the manuscript.

Both taken into account

7) I recommend summarizing the conclusions section in some take-away points so that it better concludes and summarize your major points. Currently the conclusions section is also not linked at all.

We have shortened and rewritten parts of section 4, so that key messages come out clearer, which are pointed out in italic (see above).

8) In the Code/Data availability, the authors have provided links to the relevant datasets but I am just wondering if the authors intend to provide open source code of their ANN as well?

We will provide the 3D dataset (HR's, vertical structure and rain rate indicator) via the GEWEX UTCC PROES website. The first author has already contacted the French data centre AERIS and will start as soon as possible with the documentation, and this distribution should be within this year. In the meantime we continue to test the data. We do not foresee to distribute the ANN code, because they a several programs (training, data production and evaluation), they have not been written in such a way that they will be easy to use, and we do not have any funding for rewriting and document them.

We prefer to concentrate our effort on providing and document the dataset (which needs extensive testing).