Author response to Referee #1:

Dear Referee,

We sincerely thank you for your thoughtful and constructive comments, which have greatly helped us improve the manuscript. Below, we provide our point-by-point responses to each comment. All changes made to the manuscript according to the comments of the three reviewers have been carefully highlighted in the revised version.

General comments:

Latent heating and radiative heating are two major diabatic heating sources of the atmosphere. However, the observations of global latent and radiative heating profiles are difficult to obtain. Active sensors onboard satellites provide a way to measure these profiles, such as the products from TRMM and GPM. However, TRMM and GPM satellites are low-orbit satellites thus the spatiotemporal coverage of the heating profile data are limited. The previous work from the author(s) have applied a ML technique to extend the radiative heating profile to greater coverage. This paper does similar thing but for latent heating (LH), with further analysis of LH profiles for different surface conditions, different environmental conditions and MCS characteristics. One of the main conclusions is that although the mean values show good agreement, the ML-expanded latent heating profiles show much smaller variability than the target data. This is different from the ML-expanded radiative heating profiles. The authors made a further analysis and suggested that the data are only robust for coarse-resolution larger than 2.5deg.

Although some results of this paper are interesting and this expanded dataset would be useful for scientific community, I feel that the values and limitations need to be carefully elaborated and better demonstrated. One major concern I have is that since the authors explicitly suggest that "our ML-expanded LH dataset is suitable at scales larger than about 2.5◦", how robust can we trust the results for MCSs in the size of around 1x1deg, or 100kmx100km, as the authors show in Fig. 13 to 15?

We thank the reviewer for his very constructive and thoughtful comments.

Indeed, it is very difficult to evaluate very precisely this new dataset, since we had only the collocated data for a direct evaluation, with a sparse sampling (which was the reason why we used ANN to expand the data).

Most ML datasets are only evaluated using a part of the collocated data (as in our section 2.5.2). The results in Table S1 show relatively small mean absolute errors, of about 0.5 K/day in the case of heavy rain. However, instantaneous predictions do have indeed much larger uncertainties as we have shown that the variability of the predicted values are smaller than the original ones.

Therefore, we used the complete TRMM data for a further comparison (section 3.2), but the TRMM data themselves have a very difficult diurnal sampling. We partly eliminated the uneven time sampling problem of TRMM by concentrating on the latitudinal band near the equator. Still, we had the problem to compare specific times with the complete sampling, but since we could show that on average the AIRS observations at 1:30 AM/PM agree well on average with those of the complete TRMM data over ocean, we could build scatter plots of monthly mean values, but still these are monthly means.

Further inspired by the reviewer's comment, we have re-collocated the two datasets again, by including variables which were not used for the training (like latitude, longitude, date, fraction of certain rain) as well as the produced CIRS-ML LH, and we used this complete dataset for further investigations.

We analyzed the collocated data further, though these consist of these relatively narrow orbit pieces (shown in red in Fig. 1), so that averaging instantaneous values over larger regions do not fill out these

regions for daily statistics. By considering the latitudinal means of the collocated TRMM data and the collocated ML data (the data displayed in Fig. 3, but this time for the whole statistics of the collocated data and not only the 20% test data), we show that in regions of large rainfall the ML data underestimate LP and in regions of low rainfall the ML data very slightly overestimate LP. This could already be indirectly derived from the scatter plots in Fig. 6, but this time we compare data taken at the same local time.

In order to further understand the noise and biases, we tried to link discrepancies in LP with those in rain fraction over the 0.5° grid cells. Therefore, we compared CIRS-ML certain rain fraction as function of TRMM rain fraction and TRMM-SLH LP (new Fig. S10), and they look very coherent, with an average increase of the certain rain fraction with both, as one would expect. When separating cases with small and large fraction of certain rain within the grid cells, the CIRS-ML LP is much larger for a large fraction, as one would again expect. However, one would not expect cases with a large certain rain fraction for very small TRMM-SLH LP. Indeed, this is true as the average behavior shows a mean close to 0.

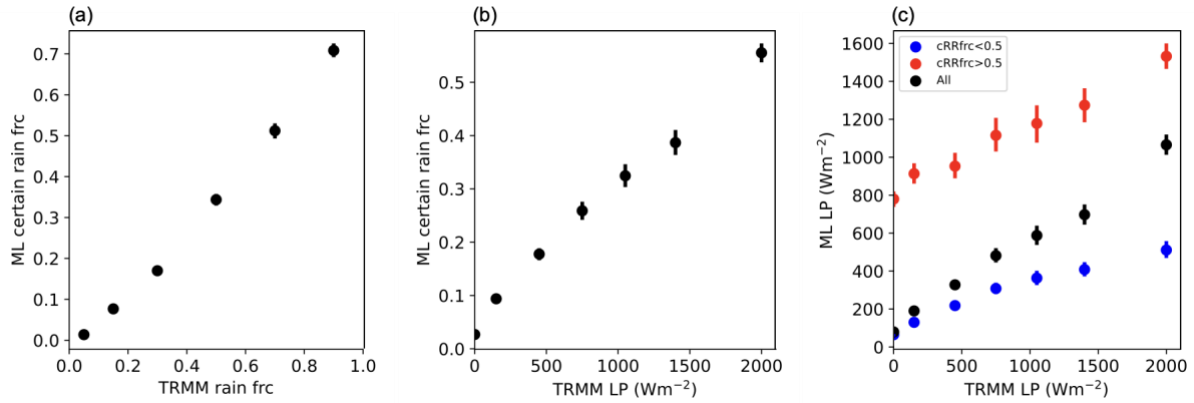

Figure S10: Average relationships for high-level clouds between (a) CIRS-ML certain rain fraction and TRMM rain fraction, (b) CIRS-ML certain rain fraction and TRMM-SLH LP and (c) CIRS-ML LP and TRMM-SLH LP for all and for cases with CIRS-ML certain rain fraction < 0.5 and > 0.5. Statistics of collocated data over ocean 30°N–30°S, over the time period from 2004–2013.

The distributions of CIRS-ML certain rain fraction in the CIRS-ML LP – TRMM-SLH LP space, shown in the new Fig. S11, explain then the noise, in particular for small TRMM LP. So, this noise can be mostly explained by a few individual cases which show a mismatch between the certain rain coverage obtained from ANN classification of certain rain identified over CloudSat samples (1.25 km × 2.5 km), and the TRMM radar samples (5 km × 5 km). Nevertheless, what is important to note is that the CIRS-ML LP seems to be coherent with the certain rain coverage, even though this variable was not used in the training! Considering the relationship of the averages, this leads to a slightly larger CIRS-ML LP compared to TRMM LP for small TRMM LP and a slightly smaller CIRS-ML for large TRMM LP, in addition to a large underestimation of extremely larger values, as expected from the ML technique. We needed to use the rain information from CloudSat, which itself was already expanded by ML, because this is available in the CIRS production.

We have improved the text in section 2.3 and added these findings in section 3.2.

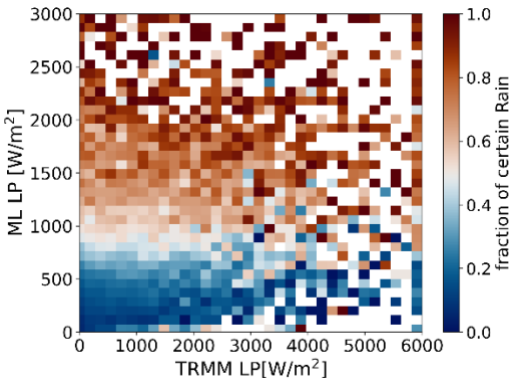

Figure S11: Averages of CIRS-ML certain rain fraction as function of TRMM-SLH LP and CIRS-ML LP of high-level clouds over the ocean, from collocated data for the period 2004-2013, spanning 30°N to 30°S, at a spatial resolution of 0.5°. Each square corresponds to an interval of 75 W/m^2 in ML LP and 150 W/m^2 in TRMM-SLH LP.

We understand fully the concern of the reviewer that the MCS results around small MCSs may be biased. Therefore, we have excluded MCSs with a size less than 1°, and instead of a fixed size threshold we use the 40% largest and smallest MCSs. The results become even clearer (new Fig. 14). This shows that the results are robust. In addition, we know that when averaging over larger areas, the LP is less biased and that for small LP there is a slight overestimation and for very large LP there is a large underestimation. This means that the points in Fig. 14 move very slightly to the left for small LP and to the right for large LP, and this more for the small MCSs than for the large MCSs, which means that the effect we show here is even slightly underestimated. So, the conclusion that the ACRE is slightly larger for the same LP (or rain intensity) in the case of larger MCSs is certainly robust!

However, we took out the old Fig. 14a which presented a very similar relationship between ACRE and LP for developing, mature and dissipating stages, because since in general the size of the MCSs increases from developing towards dissipating (which means dissipating convection but not dissipating cloud shield) as we know from Fig.12 of (Stubenrauch et al. 2019) or Fig. 9 from (Protopapadaki et al. 2017), and therefore the LP averages over developing MCSs may be stronger negatively biased than the dissipating ones. This would lead then to the conclusion that ACRE for MCSs with similar LP may be slightly larger for dissipating MCSs than for developing ones, but we can't give any quantitative estimate.

All in all, we can give qualitative statements about relationships, but it is difficult to quantify the results. Nevertheless, since we have an idea about the direction of the biases, it helps in the interpretation. We included this kind of interpretation into the manuscript.

Specific comments:

- Line 24: as RH is usually referred as relative humidity, maybe change the abbreviation (e.g. Qrad)

  Thank you for this valuable suggestion. Indeed, this will help clarity. We have incorporated this change and replaced "RH" with "Qrad" throughout the revised manuscript.

- Line 61: maybe need a little bit more introduction on CRE and ACRE, for example, from which data source did you obtain the CRE/ACRE data (clear-sky and full-sky radiation)? Sometimes the authors state CRE and radiative heating profiles together (even the units in Fig.5 have K/day and W/m2 on the same panel, this confuses me a bit), do they represent similar effect?

The cloud radiative effect (CRE) is calculated as the difference between all-sky radiative heating rates and clear-sky radiative heating rates, with the unit K/day. The atmospheric cloud radiative effect (ACRE), already defined for example by Li et al. (2015) and Harrop and Hartmann (2016) as the difference in cloud radiative effects between the TOA and the surface, and corresponds to the vertically integrated CRE, with unit W/m².

To clarify this point, we have revised the sentence to:

"In our analyses, we use the following definitions: LH refers to the latent heating profile; LP denotes the vertically integrated latent heating; Qrad represents the radiative heating profile; CRE (cloud radiative effect) refers to the difference between all-sky and clear-sky radiative heating rates, expressed in units of K/day; and ACRE (atmospheric cloud radiative effect) represents the vertically integrated CRE, with units of W/m²."

- Fig.1: more details of the figure are needed: what do the different blue colors mean? what are the thin and thick swaths?

  Taken into account. We have added one sentence in caption of Fig. 1 to:

  "The narrow swaths represent TRMM–PR orbits, while the broader swaths represent Aqua–AIRS orbits. Shades of blue indicate variations in sampling time difference."

- Line 160: maximum should be minimum

  We changed this to 'with a large peak at…' in the text.

- Line 162: what are the thresholds in your rain intensity categorization?

  The thresholds for rain intensity categorization were mentioned in section 2.3, but indeed, only the thresholds used in the rain intensity classification per AIRS footprint were mentioned. We have now completed the description including the propagation to the 0.5° grid cells.

  "A rain intensity classification (0 for no rain, 1 for light rain, 2 for heavy rain) was also obtained by ANN classification per footprint, but trained with precipitation rate data from CloudSat (2C-PRECIP-COLUMN, Haynes et al., 2009). The rain intensity classification considers light rain to be =< 5 mm h^{-1} and heavy rain > 5 mm h^{-1}. In combination with a CloudSat 2C-PRECIP-COLUMN quality flag, which indicates certain rain, also expanded via a binary ANN classification, a 'rain rate indicator' was constructed and then averaged per 0.5° grid cell (Stubenrauch et al., 2023). The rain intensity classification used for the scene identification for the ANN training and production is based on this averaged rain indicator, with heavy rain starting probably already around 2.5 mm/h. We could show that this category corresponds to an average certain rain fraction of at least 0.8 (not shown), while the light rain category corresponds only to 0.3 over ocean (0.5 over land). The advantage of using the CIRS-ML rain intensity classification, is that the CIRS-ML data are available together with the whole CIRS-AIRS and CIRS-IASI data records, so we can use them for a scene identification for the training as well as for the application of the ANN models developed for these different scenes discussed in section 2.5.1."

- Line 164: How do you define UT clouds in this paper?

  High-level clouds are defined in line 118 (previous version):

  "CIRS cloud types are defined according to p_{cld} and ε_{cld} as high clouds (P_{cld} < 440 hPa)…"

UT clouds are defined in line 331 (previous version):

"We consider UT clouds with P_{cld} < 350 hPa…"

Since the definition of UT clouds is introduced much later, it may cause confusion with high-level clouds. To address this, we added the sentence "UT clouds with P_{cld} < 350 hPa are part of the high cloud category." in revised manuscript (line 149).

Additionally, there was a confusion in the manuscript. The training of the ANNs was separately done for high-level clouds and mid- and low-level clouds. We carefully reviewed the use of "UT clouds" and "high-level clouds" in both the figures and the text, and made necessary revisions, mostly in section 2.

- Line 192: I am not an expert of machine learning, but 20 iterations look a lot to me if you don't see any improvement in the loss function. Is it true that it is still not overfitting after 20 iterations without improvement?

  The appropriate early stopping iteration count depends on the model complexity, dataset size, and training configuration. Since our model is relatively simple, the dataset is large, and the learning rate is small (1e-4), it may take more epochs to see improvement in the validation loss, as for heavy rain shown in Fig. S4. However, also for the scenes for which after 10 iterations no further improvement was seen the validation loss and training loss are similar, indicating no overfitting. So the 20 iterations used here seem to be appropriate.

- Line 247: Is there any explanation of the two peaks of LH profiles over ocean?

  The two peaks in the LH profiles over ocean (Fig. 4) at 450 hPa and 850 hPa correspond to the contributions from high-level clouds and low-level clouds, respectively. We have updated line 247 to:

  "Over ocean, the mean LH profiles have two peaks, at 450 hPa and at 850 hPa, corresponding to the contributions from high-level clouds and low-level clouds (shige et al., 2004), while over land the LH is mostly produced by high-level clouds."

- Fig. 5: which variables are the left and right y axes corresponding to?

  Thank you for pointing this out. The dual y-axes in Fig. 5 both correspond to integrated latent/radiative heating. The values are originally in W/m², which is the correct unit for the integrated quantities. The K/day axis represents the converted heating rate. To avoid confusion, we have removed the K/day axis and kept only the W/m² axis.

- Fig. 5: why is the SW radiative effect is almost zero? My direct intuition is that when cloud exists, shortwave radiative effect should be pretty negative.

  In Fig. 5, the SW radiative effect is not the TOA effect but the effect within the atmospheric column, so it should be slightly positive, as the SW heats within the clouds, while indeed the heating is less underneath an opaque cloud than for clear sky (see for example Fig. 3 of Stubenrauch et al. 2021). It is also very small, because the SW ACRE shown is the 24h mean (SW ACRE at 1:30PM weighted by $1/(\pi \times \cos\theta)$).

  However, since our main focus is the coherence between TRMM–SLH LP and ML-predicted LP, and not the radiative effect, we have removed the parts on SW and LW radiative effects and the corresponding text to avoid unnecessary complexity.

- Fig. 5: The ‘pink solid line’ mentioned in the caption is not shown in the figure.

We have revised Fig. 5, keeping the separation ocean – land, and one part went into the supplement to keep the flow of the paper. So the new Fig. 5 shows original TRMM with diurnal sampling and the CIRS-ML production separately for the 4 observation times, and the supplemental figure (Fig. S9) shows the comparison between the CIRS-ML data and the collocated results (for this we needed to rerun the collocation in order to add latitude and longitude to the dataset; this took a little effort, because the machine on which the collocation was run changed in the meantime), separately for all 4 observation times

So, Fig. 5 shows the effect of diurnal sampling, and the Fig.S9 the biases in regions with strong rain (tropical peak region) and in regions with not much rain (subtropics), with a slight underestimation in the first and a very slight overestimation in the latter. The underestimation is largest when the rain is the heaviest, as seen in Fig. 6.

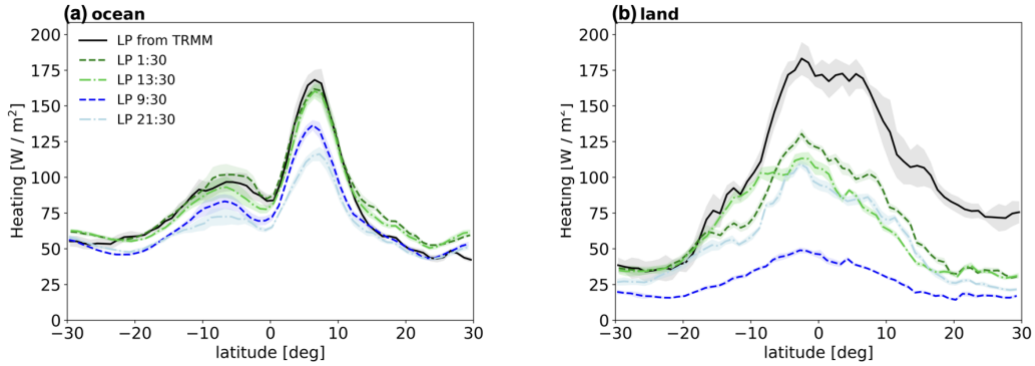

Figure 5: Zonal averages of vertically integrated LH (LP) at 4 specific observation times (1h30, 9h30 AM/PM) from the CIRS-ML production and for the original TRMM, including diurnal sampling, (a) over ocean and (b) over land. Latitude intervals are 1°. LP from TRMM data for the period 2008–2013 is represented by a solid black line. The green dashed line and bright green dash-dotted line represent LP from ML regression using TRMM–AIRS (2008–2013) as inputs, at 1:30 AM and 1:30 PM, respectively. The blue dashed line and light blue dash-dotted line represent LP from ML regression using TRMM–IASI (2008–2013) as inputs, at 9:30 AM and 9:30 PM, respectively. Shaded areas correspond indicate inter-annual variabilities.

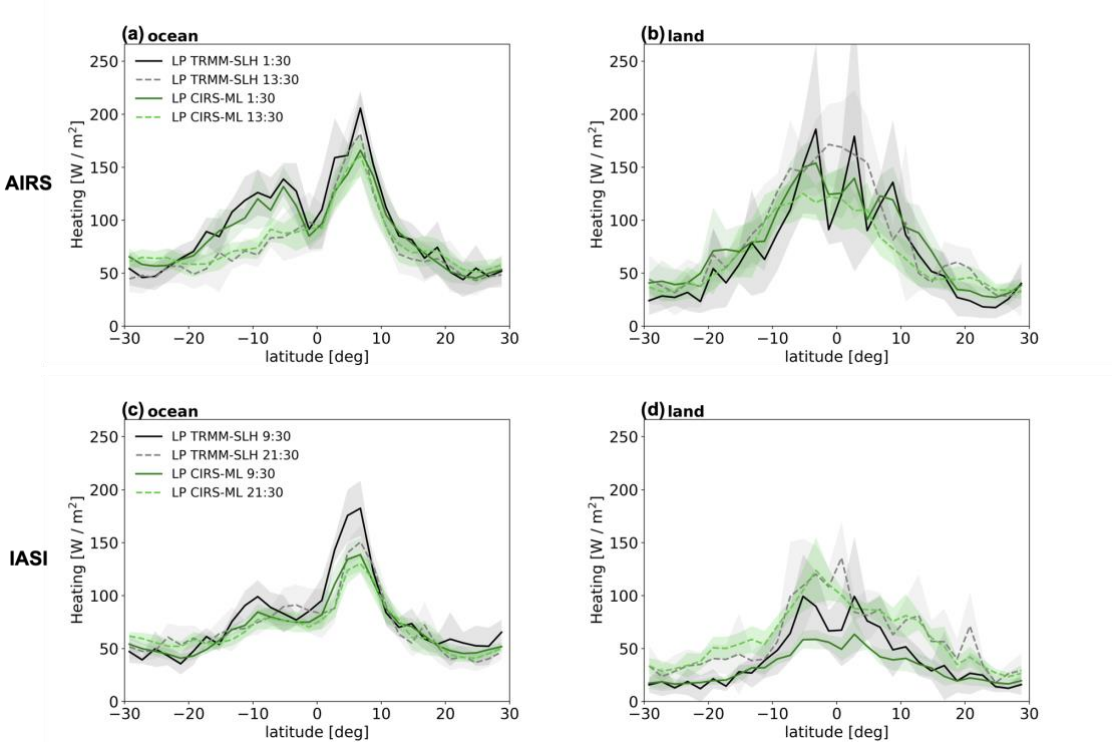

Figure S9: Zonal averages of vertically integrated LH (LP) of collocated data (a) at 1:30 AM and PM over ocean, (b) at 1:30 AM and PM over land, (c) at 9:30 AM and PM over ocean, (b) at 9:30 AM and PM over land. Black solid lines: TRMM LP for AM observations. Gray dashed line: TRMM LP for PM observations. Dark green solid line: LP from ML regression using AIRS or IASI as inputs for AM observations. Light green dashed line: LP from ML regression using AIRS or IASI as inputs for PM observations. Shaded areas correspond to inter-annual variabilities. The latitude intervals are 2°, and the time period is 2008-2013.

- Line 334: do they need to be neighboring grid cells or only by cloud height to be merged into the same system?

  As the cloud pressure of UT clouds retrieved by CIRS has an average uncertainty of about 25 hPa, we only merged adjacent grid cells with similar cloud pressure (slightly above the mean error). The UT clouds systems are constructed from neighboring grid cells, containing at least 90% UT clouds (P_{cld} < 350 hPa) and having similar cloud top heights, within a range of 8 hPa $\times$ In(P_{cld} [hPa]).

  We have added one word to the sentence of line 420:

  "First, adjacent grid cells with UT clouds of similar height…"

- Line 335: I don't quite understand this sentence "the size ... is computed... by the number...". Is there anything missing?

  To clarify this point, we have revised the sentence on line 421 to:

  "Then the size of convective cores is determined first by counting the number of grid cells with cloud emissivity > 0.98 within regions where the cloud emissivity exceeds 0.93 and then by multiplying this number by the grid cell size of 0.5° (approximately 3000 km^2).".

We needed to group first regions with slightly smaller emissivity in order to reduce the noise in the determination of the number of cores. If one groups only grid cells with emissivity > 0.98, then one obtains many more multi-cores (see Protopapdaki et al. 2017).

- Fig. 9: this is another place that I don't quite understand what this paper use, is it CRE or radiative heating?

  We apologize for the confusion caused by the unclear definitions of CRE and radiative heating earlier. In Fig. 9, it is the CRE being shown, as it represents the net effect of clouds after subtracting the clear-sky radiative heating.

  We have already revised the corresponding text in the manuscript to clarify this point.

- Line 364: not sure I understand this statement. Are you comparing MCS intensity relations with MCS size and opacity? or comparing radiative heating profile with LH profiles?

  Section 4 was brought in to describe the construction of the mesoscale convective systems. We then thought to illustrate the effect of different proxies for precipitation intensity. Some of these proxies are very indirect, like the size of the systems or the minimum cooling above the precipitating part of the convective system (the LW cooling above the cloud stands for the opacity). For the latter, we select the grid cell within the precipitating part of the convective system with the minimum value of the cooling above the cloud. We compare the latent heating rate profiles with the one for heavy rain MCSs, which is the most direct proxy. Figure 9 was meant to test the coherence of the data. The LW cooling above the core comes from the CIRS-ML data and the core identification from CIRS and using CIRS-ML rain classification. And we can show that the CIRS-ML latent heating profiles change in coherence with the different proxies.

  Furthermore, we show that on average the minimum LW cooling above the heavily raining MCS's is the largest. This shows again that the minimum cooling above the participating part of the convective system seems to be an interesting proxy for rain intensity.

  We have rewritten the paragraph on the description of these results, and hope that it is clearer now.

- Line 372-373: from where did you obtain this conclusion that LP is more reliable over ocean?

  Indeed, the wording was not well chosen. We have rewritten this phrase as:

  "Since we found the CIRS-ML LP at 1:30 AM/PM more similar to the LP of the diurnally sampled TRMM-SLH over ocean, we consider in the following only precipitating clouds over ocean."

  From Fig. 5, we found that the CIRS-ML LP is more similar to the LP of the diurnally sampled TRMM-SLH over the ocean, particularly at 1:30 AM/PM. Therefore, these values could be taken as the average over the whole diurnal cycle.

- Fig. 14: this is an interesting result that LP-ACRE relation shows no distinction between developing, mature and dissipating stages, but is dependent on MCS sizes. Not sure how to understand/explain this.

  Indeed, we found this result also interesting, and so far we have no explanation for this, but we thought this result is important to show, so that the community may go further on this research.

It may be important for convective parameterizations. However, to avoid confusion, we have removed this panel from Fig. 14 in the revised manuscript.

- Line 457: What are 'the larger ones' refer to?

  Sorry, this was a typo: What we meant to say is that larger systems have, on average, a smaller core fraction than smaller ones. This error has been fixed, and "larger" is now changed to "smaller."

References:

Harrop, B. E., and D. L. Hartmann: The role of cloud radiative heating within the atmosphere on the high cloud amount and top-of-atmosphere cloud radiative effect, J. Adv. Model. Earth Syst., 8, 1391–1410, doi:10.1002/2016MS000670, 2016.

Haynes, J. M., L'Ecuyer, T. S., Stephens, G. L., Miller, S. D., Mitrescu, C., Wood, N. B., and Tanelli, S.: Rainfall retrieval over the ocean with spaceborne W-band radar, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD009973, 2009.

Li, Y., D. W. J. Thompson, and S. Bony: The Influence of Atmospheric Cloud Radiative Effects on the Large-Scale Atmospheric Circulation. J. Climate, 28, 7263–7278, https://doi.org/10.1175/JCLI-D-14-00825.1, 2015.

Stubenrauch, C. J., Caria, G., Protopapadaki, S. E., and Hemmer, F.: 3D radiative heating of tropical upper tropospheric cloud systems derived from synergistic A-Train observations and machine learning, Atmospheric Chemistry and Physics, 21, 1015–1034, https://doi.org/10.5194/acp-21-1015-2021, 2021.

Stubenrauch, C. J., Mandorli, G., and Lemaitre, E.: Convective organization and 3D structure of tropical cloud systems deduced from synergistic A-Train observations and machine learning, Atmospheric Chemistry and Physics, 23, 5867–5884, https://doi.org/10.5194/acp-23-5867-2023, 2023.